MADHYA PRADESH BHOJ(OPEN) UNIVERSITY BHOPAL

Numerical Methods & Statistical Analysis

BSDS-304

NUMERICAL METHODS AND STATISTICAL ANALYSIS

मध्यप्रदेश भोज (मुक्त) विश्वविद्यालय – भोपाल
MADHYA PRADESH BHOJ (OPEN) UNIVERSITY - BHOPAL

COURSE WRITERS

Dr. N. Dutta, Professor (Mathematics), Head, Department of Basic Sciences & Humanities, Heritage Institute of Technology, Kolkata

Units (1, 2.0-2.3, 2.6-2.10, 3)

Manisha Pant, Former Lecturer, Department of Mathematics, H.N.B. Garhwal University, Srinagar (A Central University), U.K.

Units (2.4, 2.5.2, 5.5.3, 4.2, 4.3, 4.6.5, 5.2, 5.6)

C. R. Kothari, Ex-Associate Professor, Department of Economic Administration & Financial Management, University of Rajasthan

Units (2.5-2.5.1, 4.7-4.7.5)

J. S. Chandan, Retd. Professor, Medgar Evers College, City University of New York

Units (4.0-4.1, 4.3.1-4.6.4, 4.8-4.12, 5.0-5.1, 5.3-5.5, 5.7-5.11)

Published by Registrar, MP Bhoj (open) University, Bhopal in 2020

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

SYLLABI-BOOK MAPPING TABLE

Numerical Methods and Statistical Analysis

Syllabi	Mapping in Book
UNIT - I Introduction, Limitation of Number Representation, Arithmetic rules for Floating Point Numbers, Errors in Numbers, Measurement of Errors, Solving Equations, Introduction, Bisection Method, Regula Falsi Method, Secant Method, Convergence of the iterative methods.	**Unit-1:** Representation of Numbers **(Pages 3-38)**
UNIT - II Interpolation, Introduction, Lagrange Interpolation, Finite Differences, Truncation Error in Interpolation, Curve Fitting, Introduction, Linear Regression, Polynomial Regression, Fitting Exponential and Trigonometric Functions	**Unit-2:** Interpolation and Curve Fitting **(Pages 39-113)**
UNIT - III Numerical Differentiation and Integration, Introduction, Numerical Differentiation Formulae, Numerical Integration Formulae, Simpson's Rule, Errors in Integration Formulae, Gaussian Quadrature Formulae, Comparison of Integration Formulae, Solving Numerical Differential Equations, Introduction, Euler's Method, Taylor Series Method, Runge-Kutta Method, Higher Order Differential Equations.	**Unit-3:** Numerical Differentiation and Integration **(Pages 115-184)**
UNIT - IV Introduction to Statistical Computation, History of Statistics, Meaning and scope of Statistics, Various measures of Average, Median, Mode, Geometric Mean, Harmonic Mean, Measures of Dispersion, Range, Standard Deviation, Probability Distributions, Introduction, Counting Techniques, Probability, Axiomatic or Modern Approach to Probability, Theorems on Probability, Probability Distribution of a Random Variable, Mean and Variance of a Random Variable, Standard Probability Distributions, Binomial Distribution, Hyper geometric Distribution Geometrical Distribution, Uniform Distribution (Discrete Random Variable), Poisson Distribution, Exponential Distribution, Uniform Distribution (Continuous Variable), Normal Distribution	**Unit-4:** Statistical Computation and Probability Distributiona **(Pages 185-290)**
UNIT - V Estimation, Sampling Theory, Parameter and Statistic, Sampling Distribution of Sample Mean, Sampling Distribution of the Number of Successes, The Student's Distribution, Theory of Estimation, Point Estimation, Interval Estimation, Hypothesis Testing, Test of Hypothesis, Test of Hypothesis Concerning Mean, Test of Hypothesis Concerning Proportion, Test of Hypothesis Concerning Standard Deviation	**Unit-5:** Estimation and Hypothesis Testing **(Pages 291-328)**

CONTENTS

INTRODUCTION

NOTES

Numerical method and Statistical analysis is the study of algorithms to find solutions for problems of continuous mathematics or considered as a mathematical science pertaining to the collection, analysis, interpretation or explanation, and presentation of data and can be categorized as Inferential Statistics and Descriptive Statistics.

Numerical method helps in obtaining approximate solutions while maintaining reasonable bounds on errors. Although numerical analysis has applications in all fields of engineering and the physical sciences, yet in the 21st century life sciences and both the arts have adopted elements of scientific computations. Ordinary differential equations are used for calculating the movement of heavenly bodies, i.e., planets, stars and galaxies. Besides, it evaluates optimization occurring in portfolio management and also computes stochastic differential equations to solve problems related to medicine and biology. Airlines use sophisticated optimization algorithms to finalize ticket prices, airplane and crew assignments and fuel needs. Insurance companies too use numerical programs for actuarial analysis. The basic aim of numerical analysis is to design and analyse techniques to compute approximate and accurate solutions to unique problems. In numerical analysis, two methods are involved, namely direct and iterative methods. Direct methods compute the solution to a problem in a finite number of steps whereas iterative methods start from an initial guess to form successive approximations that converge to the exact solution only in the limit. Iterative methods are more common than direct methods in numerical analysis. The study of errors is an important part of numerical analysis. There are different methods to detect and fix errors that occur in the solution of any problem. Round-off errors occur because it is not possible to represent all real numbers exactly on a machine with finite memory. Truncation errors are assigned when an iterative method is terminated or a mathematical procedure is approximated and the approximate solution differs from the exact solution.

Statistical analysis is very important for taking decisions and is widely used by academic institutions, natural and social sciences departments, governments 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 that pertained to state wealth, collection of taxes, study of population, and so on. In the beginning of the Indian, Greek and Egyptian civilizations, data was collected for the purpose of planning and organizing civilian and military projects. Proper records of such vital events as births and deaths have been kept since the Middle Ages. By the end of the 19th century, the field of statistics extended from simple data collection and record keeping to interpretation of data and drawing useful conclusions from it. Statistics can be called a science that deals with numbers or figures describing the state of affairs of various situations with which we are generally and specifically concerned. To a layman, it often means columns of figures, or perhaps tables, graphs and charts relating to population, national income, expenditures, production, consumption, supply, demand, sales, imports, exports, births, deaths and accidents. Similarly, statistical records kept at universities may reflect the number of students,

NOTES

the percentage of female and male students, the number of divisions and courses in each division, the number of professors, the tuition received, the expenditures incurred, and so on. Hence, the subject of statistics deals primarily with numerical data gathered from surveys or collected using various statistical methods.

This book is divided into five units. The topics discussed is designed to be a comprehensive and easily accessible book covering the limitations of number representation, measurement of errors, solving equations, Regula Falsi method, secant method, interpolation, Lagrange interpolation, curve fitting, regression, numerical differentiation, Simpson's rule, Gaussian quadrature formulae, solving numerical differential equations, Euler's method, Taylor series method, Runge-Kutta method, history of statistics, various measures of statistical computation, probability distribution, standard probability distribution, sampling theory, point estimation and test of hypothesis.

The book follows the Self-Instructional Mode (SIM) wherein each unit begins with an 'Introduction' to the topic. The 'Objectives' are then outlined before going on to the presentation of the detailed content in a simple and structured format. 'Check Your Progress' questions are provided at regular intervals to test the student's understanding of the subject. 'Answers to Check Your Progress Questions', a 'Summary', a list of 'Key Terms', and a set of 'Self-Assessment Questions and Exercises' are provided at the end of each unit for effective recapitulation.

UNIT 1 REPRESENTATION OF NUMBERS

NOTES

Structure

1.0 INTRODUCTION

The use of computers to solve problems involving real numbers is referred to as 'Numerical Calculations.' A finite string of digits can represent a large number of real numbers. Most scientific computers limit the amount of digits that can be used to represent a single number to a set number. Numerical error is the combined effect of two kinds of error in a calculation. The first is caused by the finite precision of computations involving floating point or integer values. The second usually called truncation error is the difference between the exact mathematical solution and the approximate solution obtained when simplifications are made to the mathematical equations to make them more amenable to calculation. The number of significant figures in a measurement, such as 2.531, is equal to the number of digits that are known with some degree of confidence (2, 5 and 3) plus the last digit (1), which is an estimate or approximation. Zeros within a number are always significant. Zeros that do nothing but set the decimal point are not significant. Trailing zeros that are not needed to hold the decimal point are significant. A round-off error, also called rounding error, is the difference between the calculated approximation of a number and its exact mathematical value. Numerical analysis specifically tries to estimate this error when using approximation equations and/or algorithms, especially when using finitely many digits to represent real numbers.

NOTES

In root finding and curve fitting, a root-finding algorithm is a numerical method, or algorithm, for finding a value x such that $f(x) = 0$, for a given function f. Such an x is called a root of the function f. Generally speaking, algorithms for solving problems numerically can be divided into two main groups: direct methods and iterative methods. Direct methods are those which can be completed in a predetermined unite number of steps. Iterative methods are methods which converge to the solution over time. These algorithms run until some convergence criterion is met. When choosing which method to use one important consideration is how quickly the algorithm converges to the solution or the method's convergence rate.

In this unit, you will learn about the limitations of number representation, arithmetic rules for floating point numbers, errors in numbers and measurement of errors, solving equation, bisection method and convergence of the iterative method, secant method and Regula-Falsi method.

1.1 OBJECTIVES

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

- Understand the basic concept of limitations of number representation
- Explain about the arithmetic rules for floating point numbers
- Analysis the errors in numbers and measurement of errors
- Define solving equation
- Discuss about the bisection method and convergence of the iterative method
- Elaborate on the secant method
- Explain Regula-Falsi method

1.2 INTRODUCTION TO NUMERICAL COMPUTING

Numerical methods are useful in almost all fields of Science and Engineering, especially when analytical solutions are either not available or are very complicated. There are several problems for which numerical solutions are simpler than analytical solutions. The development of computers and the advancement in software engineering have spurred further research in numerical analysis. Well-defined algorithms lead to faster computation, improved storage capacity better accuracy and stability.

The methods employed in numerical analysis are at times approximate and the data used in computation are of finite decimal representation. Thus in most of the cases, the results obtained by numerical methods have some errors. Before defining numerical computing we must be aware of sources of errors in a numerical solution and accordingly handle the case. Numerical analysis basically deals with the development of suitable methods for obtaining applicable numerical solutions for mathematical problems along with an indication of the accuracy of the solution.

Numerical methods have been specifically developed for finding accurate numerical solutions using a computer. While performing arithmetic operations on real numbers, using a computer, we use fixed number of decimal digits. Most numbers usually have infinite decimal representation, but for machine computation the numbers are given and stored with a finite number of digits.

1.3 LIMITATIONS OF NUMBER REPRESENTATIONS

Numerical methods are methods used for solving problems through numerical calculations providing a table of numbers and/or graphical representations or figures. Numerical methods emphasize that how the algorithms are implemented. Thus, the objective of numerical methods is to provide systematic methods for solving problems in a numerical form. Often the numerical data and the methods used are approximate ones. Hence, the error in a computed result may be caused by the errors in the data or the errors in the method or both. Generally, the numbers are represented in **decimal** (base 10) form, while in computers the numbers are represented using the **binary** (base 2) and also the **hexadecimal** (base 16) forms. To perform a numerical calculation, approximate them first by a representation involving a finite number of **significant digits**. If the numbers to be represented are very large or very small, then they are written in **floating point notation**. The Institute of Electrical and Electronics Engineers (IEEE) has published a standard for binary floating point arithmetic. This standard, known as the IEEE Standard 754, has been widely adopted. The standard specifies formats for **single precision** and **double precision** numbers. The simplest way of reducing the number of significant digits in the representation of a number is simply to ignore the unwanted digits known as chopping. All these topics are discussed in the following section.

Significant Figures

In approximate representation of numbers, the number is represented with a finite number of digits. All the digits in the usual decimal representation may not be significant while considering the accuracy of the number. Consider the following numbers:

$$1514,\ 15.14,\ 1.324,\ 1524$$

Each of them has four significant digits and all the digits in them are significant. Now consider the following numbers,

$$0.00215,\ 0.0215,\ 0.000215,\ 0.0000125$$

The leading zeroes after the decimal point in each of the above numbers are not significant. Each number has only three significant digits, even though they have different number of digits after the decimal point.

1.3.1 Arithmetic Rules for Floating Point Numbers

Every real number is usually represented by a finite or infinite sequence of decimal digits. This is called decimal system representation. For example, we can represent

$\frac{1}{4}$ as 0.25, but $\frac{1}{3}$ as 0.333... Thus $\frac{1}{4}$ is represented by two significant digits only, while $\frac{1}{3}$ is represented by an infinite number of digits. Most computers have two forms of storing numbers for performing computations. They are fixed-point and floating point. In a fixed-point system, all numbers are given with a fixed number of decimal places. For example, 35.123, 0.014, 2.001. However, fixed-point representation is not of practical importance in scientific computation, since it cannot deal with very large or very small numbers.

NOTES

In a floating-point representation, a number is represented with a finite number of significant digits having a floating decimal point. We can express the floating decimal number as follows:

$$623.8 \text{ as } 0.6238 \times 10^3,\ 0.0001714 \text{ as } 0.1714 \times 10^{-3}$$

A very large number can also be representated with floating-point representation, keeping the first few significant digits such as $0.14263218 \times 10^{39}$. Similarly, a very small number can be written with only the significant digits, leaving the leading zeros such as $0.32192516 \times 10^{-19}$.

In the decimal system, very large and very small numbers are expressed in scientific notation as follows: $4 \cdot 69 \times 10^{23}$ and $1 \cdot 601 \times 10^{-19}$. Binary numbers can also be expressed by the floating point representation. The floating point representation of a number consists of two parts: the first part represents a signed, fixed point number called the *mantissa* **(***m***)**; the second part designates the position of the decimal (or binary) point and is called the *exponent* **(***e***)**. The fixed point mantissa may be a fraction or an integer. The number of bits required to express the exponent and mantissa is determined by the accuracy desired from the computing system as well as its capability to handle such numbers. For example, the decimal number + 6132.789 is represented in floating point as follows:

$$\overset{\text{sign}}{\underbrace{0 \quad 0 \cdot 6132789}_{\text{mantissa}}} \quad \overset{\text{sign}}{\underbrace{0 \quad 04}_{\text{exponent}}}$$

The mantissa has a 0 in the leftmost position to denote a plus. Here, the mantissa is considered to be a fixed point fraction. This representation is equivalent to the number expressed as a fraction 10 times by an exponent, that is $0.6132789 \times 10^{+04}$. Because of this analogy, the mantissa is sometimes called the *fraction part*.

Consider, for example, a computer that assumes integer representation for the mantissa and radix 8 for the numbers. The octal number $+ 36.754 = 36754 \times 8^{-3}$ in its floating point representation will look like this:

$$\overset{\text{sign}}{\underbrace{0 \quad 36754}_{\text{mantissa}}} \quad \overset{\text{sign}}{\underbrace{1 \quad 03}_{\text{exponent}}}$$

When this number is represented in a register in its binary-coded form, the actual value of the register becomes 0 011 110 111 101 100 and 1 000 011.

Most computers and all electronic calculators have a built-in capacity to perform floating-point arithmetic operations.

Example 1.1: Determine the number of bits required to represent in floating point notation the exponent for decimal numbers in the range of $10^{\pm 86}$.

Solution: Let n be the required number of bits to represent the number $10^{\pm 86}$.

$$2^n = 10^{86}$$

$$n \log 2 = 86$$

$$n = \frac{86}{\log 2} = \frac{86}{0.3010} = 285.7$$

$$\text{Therefore,} \quad 10^{\pm 86} = 2^{\pm 285.7}$$

The exponent ±285 can be represented by a 10-bit binary word. It has a range of exponents (+511 to –512).

1.4 ERRORS IN NUMBERS AND MEASUREMENT OF ERRORS

The errors in a numerical solution are basically of two types. They are *truncation error* and *computational error*. The error which is inherent in the numerical method employed for finding numerical solution is called the truncation error. The computational error arises while doing arithmetic computation due to representation of numbers with a finite number of decimal digits.

The truncation error arises due to the replacement of an infinite process such as summation or integration by a finite one. For example, in computation of a transcendental function we use Taylor series/Maclaurin series expansion but retain only a finite number of terms. Similarly, a definite integral is numerically evaluated using a finite sum with a few function values of the integral. Thus, we express the error in the solution obtained by numerical method.

Inherent errors are errors in the data which are obtained by physical measurement and are due to limitations of the measuring instrument. The analysis of errors in the computed result due to the inherent errors in data is similar to that of round-off errors.

1.4.1 Generation and Propagation of Round-Off Error

During numerical computation on a computer, a round-off error is generated by taking an infinite decimal representation of a real, rational number such as 1/3, 4/7, etc., by a finite size decimal form. In each arithmetic operation with such approximate rounded-off numbers there arises a round-off error. Also round-off errors present in the data will propagate in the result. Consider two approximate floating point numbers rounded-off to four significant digits.

$$x = 0.2234 \times 10^3 \text{ and } y = 0.1112 \times 10^2$$

The sum $x + y = 0.23452 \times 10^3$ is rounded-off to 0.23456×10^3 with an absolute error, 2×10^{-2}. This is the new round-off error generated in the result. Besides this error, the result will have an error propagated from the round-off errors in x and y.

1.4.2 Round-Off Errors in Arithmetic Operations

To get an insight into the propagation of round-off errors, let us consider them for the four basic operations of addition, subtraction, multiplication and division. Let x_T and y_T be two real numbers whose round-off errors in their approximate representations x and y are ε_1 and ε_2 respectively, so that

$$x_T = x + \varepsilon_1 \quad \text{and} \quad y_T = y + \varepsilon_2$$

Their addition gives, $(x_T + y_T) = (x + y) + \varepsilon_1 + \varepsilon_2$

Hence, the propagated round-off error is given by,

$$(x_T + y_T) - (x + y) = \varepsilon_1 + \varepsilon_2$$

Thus the propagated round-off error is the sum of two approximate numbers (having round-off errors) equal to the sum of the round-off errors in the individual numbers.

The multiplication of two approximate numbers has the propagated round-off error given by,

$$x_T \times y_T - xy = \varepsilon_1 y + \varepsilon_2 x + \varepsilon_1 \varepsilon_2$$

Since the product $\varepsilon_1\varepsilon_2$ is a small quantity of higher order, then ε_1 or ε_2 may take the propagated round-off error as $\varepsilon_1 x_1 + \varepsilon_2 y_1$ and the relative propagated error is given by,

$$\frac{\varepsilon_1 x + \varepsilon_2 y}{xy} = \frac{\varepsilon_1}{x} + \frac{\varepsilon_2}{y}$$

This is equal to the sum of the relative errors in the numbers x and y.

Similarly, for division we get the relative propagated error as,

$$\frac{\dfrac{x_T}{y_T} - \dfrac{x}{y}}{\dfrac{x}{y}} = \frac{\varepsilon_1}{x} - \frac{\varepsilon_2}{y}$$

Thus, the relative error in division is equal to the difference of the relative errors in the numbers.

1.4.3 Errors in Evaluation of Functions

The propagated error in the evaluation of a function $f(x)$ of a single variable x having a round-off error ε is given by,

$$f(x + \varepsilon) - f(x) \approx \varepsilon \quad f'(x)$$

In the evaluation of a function of several variables $x_1, x_2, \ldots, x_n$, the propagated round-off error is given by $\sum_{i=1}^{n} \varepsilon_1 \frac{\partial f}{\partial x_i}$, where $\varepsilon_1, \varepsilon_2, ..., \varepsilon_n$ are the round-off errors in $x_1, x_2, ..., x_n$, respectively.

NOTES

Significance Errors

During arithmetic computations of approximate numbers having fixed precision, there may be loss of significant digits in some cases. The error due to loss of significant digits is termed as significance error. Significance error is more serious than round-off errors, since it affects the accuracy of the result.

There are two situations when loss of significant digits occur. These are,

(*i*) Subtraction of two nearly equal numbers.

(*ii*) Division by a very small divisor compared to the dividend.

For example, consider the subtraction of the nearly equal numbers $x = 0.12454657$ and $y = 0.12452413$, each having eight significant digits. The result $x - y = 0.22440000 \times 10^{-4}$, is correct to four significant figures only. This result when used in further computations leads to serious error in the result.

Consider the problem of computing the roots of the quadratic equation,

$$ax^2 + bx + c = 0$$

The roots of this equation are,

$$\frac{-b+\sqrt{b^2-4ac}}{2a} \quad \text{and} \quad \frac{-b-\sqrt{b^2-4ac}}{2a}$$

If $b^2 >> 4ac$, then the evaluation of $-b+\sqrt{b^2-4ac}$ leads to subtraction of nearly equal numbers. One can avoid this by rewriting the expression,

$$\frac{-b+\sqrt{b^2-4ac}}{2a}$$

It can be written as,

$$\frac{\left(-b+\sqrt{b^2-4ac}\right)\left(-b-\sqrt{b^2-4ac}\right)}{2a \times \left(b-\sqrt{b^2-4ac}\right)} = \frac{-2c}{b+\sqrt{b^2-4ac}}$$

Let the quadratic equation be,

$$x^2 + 100.0001x + 0.01 = 0$$

Using the first formula, we get the smaller root $= 0.10050000 \times 10^{-3}$, whereas exact root is $0.10000\,0000 \times 10^{-3}$. But using the last expression we get the smaller root as $0.10000000 \times 10^{-3}$ which does not has the effect of significance error.

Consider an example where loss of significant digits occur due to division by a small number.

Computation of $f(x) = \frac{1-\cos x}{x^2}$, for small values of x would have loss of significant digits.

The Table 1.1 shows the computed values of $f(x)$ upto six decimal places along with the correct values and error.

NOTES

Table 1.1 *Computed Value of f(x) upto Six Decimal Places*

x	Computed $f(x)$	Correct $f(x)$	Error
0.1	0.499584	0.499583	– 0.000001
0.01	0.50008	0.499996	– 0.000012
0.001	0.506639	0.500000	– 0.006639
0.0001	0.500000	0.745058	0.245058

Table 1.1 shows that the error in the computed value becomes more serious for smaller value of x. It may be noted that the correct values of $f(x)$ can be computed by avoiding the divisions by small number by rewriting $f(x)$ as given below.

$$f(x) = \frac{1-\cos x}{x^2} \times \frac{1+\cos x}{1+\cos x}$$

i.e.,

$$f(x) = \frac{\sin^2 x}{x^2(1+\cos x)}$$

1.4.4 Characteristics of Numerical Computation

A numerical solution can never be exact but attempts are made to know the accuracy of the approximate solution. Thus one attempts to get an approximate solution which differs from the exact solution by less than a specified tolerance limit.

Some numerical methods find the solution by a direct method but many others are of repetitive nature. The first step in the solution procedure is to take an approximate solution. Then the numerical method is applied repeatedly to get better results till the solution is obtained up to a desired accuracy. This process is known as iteration.

To get a numerical solution on a computer, one has to write an algorithm. An algorithm is a sequence of unambiguous steps used to solve a given problem. In the design of such computer programs one considers the input data required to implement the numerical method and writes the computer program in a suitable programming language. The output of the program should give the solution with the desired accuracy.

It may be noted that the iterative method gives rise to a sequence of results. The convergence of this sequence to get the output upto a desired accuracy is dependent on the initial data. Hence, one has to suitably choose the input data. Thus, if for some input data the sequence is not convergent for certain pre-assigned number of iterations then the input data is changed. It is for this reason that one has to limit the number of iterations to be employed while designing the computer program.

While computing a solution with the help of an algorithm, one has to check the correctness of the solution obtained. To do so, one has to have some test data whose solution is known.

Example 1.2: The numbers 28.483 and 27.984 are both approximate and are correct up to the last digits shown. Compute their difference. Indicate how many significant digits are present in the result and comment.

NOTES

Solution: We have $28.483 - 27.984 = 00.499$. The result has only three significant digits. This is due to the loss of significant digits during subtraction of nearly equal numbers.

Example 1.3: Round the number $x = 2.2554$ to three significant figures. Find the absolute error and the relative error.

Solution: The rounded-off number is 2.25.

The absolute error is 0.0054.

The relative error is $\simeq \dfrac{0.0054}{2.25} = 0.0024$

The percentage error is 0.24 per cent.

Example 1.4: If $\pi = 3.14$ instead of $\dfrac{22}{7}$, find the relative error.

Solution: Relative error $= \left(\dfrac{22}{7} - 3.14\right) \Big/ \dfrac{22}{7} = 0.00090.$

Example 1.5: Determine the number of correct digits in $x = 0.2217$, if it has a relative error $\varepsilon_r = 0.2 \times 10^{-1}$.

Solution: Absolute error $= 0.2 \times 10^{-1} \times 0.2217 = 0.004434$

Hence, x has only one correct digit $x \simeq 0.2$.

Example 1.6: Round-off the number 4.5126 to four significant figures and find the relative percentage error.

Solution: The number 4.5126 rounded-off to four significant figures is 4.153.

Relative error $= \dfrac{-\ 0.0004}{4.5126} \times 100 = -\ \ 0.0088$ per cent

Example 1.7: Given $f(x, y, z) = \dfrac{5xy^2}{z^2}$, find the relative maximum error in the evaluation of $f\,(x,\ y,\ z)$ at $x = y = z = 1$, if x, y, z have absolute errors $\Delta x = \Delta y = \Delta z = 0.1$

Solution: The value of $f(x, y, z)$ at $x = y = z = 1$ is 5. The maximum absolute error in the evaluation of $f(x, y, z)$ is,

$$|(\Delta f)_{\max}| = \left|\frac{\partial f}{\partial x}\Delta x\right| + \left|\frac{\partial f}{\partial y}\Delta y\right| + \left|\frac{\partial f}{\partial z}\Delta z\right|$$

$$= \left|\frac{5y^2}{z^2}\Delta x\right| + \left|\frac{10xy}{z^2}\Delta y\right| + \left|\frac{-10xy^2}{z^3}\Delta z\right|$$

At, $x = y = z = 1$, the maximum relative error is,

$$(E_R)_{\max} = \frac{25 \times 0.1}{5} = 0.5$$

Example 1.8: Find the relative propagated error in the evaluation of $x + y$ where $x = 13.24$ and $y = 14.32$ have round-off errors $\varepsilon_1 = 0.004$ and $\varepsilon_2 = 0.002$ respectively.

Solution: Here, $x + y = 27.56$ and $\varepsilon_1 + \varepsilon_2 = 0.006$.

Thus, the required relative error $= \dfrac{0.006}{27.56} = 0.0002177$.

NOTES

Example 1.9: Find the relative percentage error in the evaluation of $u = xy$ with $x = 5.43$, $y = 3.82$ having round-off errors 0.01 in both x and y.

Solution: Now, $xy = 5.43 \times 3.82 \simeq 20.74$

The relative error in x is $\dfrac{0.01}{5.43} \simeq 0.0018$.

The relative error in y is $\dfrac{0.01}{3.82} \simeq 0.0026$.

Thus, the relative propagated error in x and $y = 0.0044$.

The percentage relative error = 0.44 per cent.

Example 1.10: Given $u = xy + yz + zx$, find the estimate of relative percentage error in the evaluation of u for $x = 2.104$, $y = 1.935$ and $z = 0.845$. What are the approximate values correct to the last digit?

Solution: Here, $u = x(y + z) + yz = 2.104\,(1.935 + 0.845) + 1.935 \times 0.845$

$$= 5.849 + 1.635 = 7.484$$

$$\text{Error, } \Delta u = (y+z)\Delta x + (z+x)\Delta y + (x+y)\Delta z$$
$$= 0.0005 \times 2(x+y+z) \qquad (\because\ \Delta x = \Delta y = \Delta z = 0.0005)$$
$$\simeq 2 \times 4.884 \times 0.0005 \simeq 0.0049$$

Hence, the relative percentage error $= \dfrac{0.0049}{7.884} \times 100 = 0.062$ per cent.

Example 1.11: The diameter of a circle measured to within 1 mm is $d = 0.842$ m. Compute the area of the circle and give the estimated relative error in the computed result.

Solution: The area of the circle A is given by the formula, $A = \dfrac{\pi d^2}{4}$.

Thus, $A = \dfrac{3.1416}{4} \times (0.842)^2\ \text{m}^2 = 0.5568\ \text{m}^2$.

Here the value of π is taken upto 4th decimal palce since the data of d has accuracy upto the 3rd decimal place. Now the relative percentage error in the above computation is,

$$E_p = \frac{2\pi d}{4} \times \frac{4\Delta d}{\pi d^2} \times 100 = \frac{2\Delta d}{d} \times 100 = \frac{2 \times 0.01}{0.842} = 0.24 \text{ per cent}$$

Example 1.12: The length a and the width b of a plate is measured accurate up to 1cm as $a = 5.43$ m and $b = 3.82$ m. Compute the area of the plate and indicate its error.

Solution: The area of the plate is given by,

$$A = ab = 3.82 \times 5.43 \text{ sq. m.} = 20.74\ \text{m}^2.$$

The estimate of error in the computed value of A is given by,

$$\begin{aligned}\Delta A &= \Delta a.b + \Delta b.a \\ &= 0.01\times 3.82 + 0.01\times 5.43, \quad \text{since} \quad \Delta a = \Delta b = 0.01 \\ &= 0.0925 \simeq 10 \text{ m}^2\end{aligned}$$

1.4.5 Computational Algorithms

For solving problems with the help of a computer, one should first analyse the mathematical formulation of the problem and consider a suitable numerical method for solving it. The next step is to write an algorithm for implementing the method. An algorithm is defined as a finite sequence of unambiguous steps to be followed for solving a given problem. Finally, one has to write a computer program in a suitable programming language. A computer program is a sequence of computer instructions for solving a problem.

It is possible to write more than one algorithm to solve a specific problem. But one should analyse them before writing a computer program. The analysis involves checking their correctness, robustness, efficiency and other characteristics. The analysis is helpful for solving the problem on a computer. The analysis of correctness of an algorithm ensures that the algorithm gives a correct solution of the problem. The analysis of robustness is required to ascertain if the algorithm is capable of tackling the problem for possible cases or for all possible variations of the parameters of the problem. The efficiency is concerned with the computational complexities and the total time required to solve the problem.

Computer oriented numerical methods must deal with algorithms for implementation of numerical methods on a computer. The following algorithms of some simple problems will make the concept clear.

Consider the problem of solving a pair of linear equations in two unknowns given by,

$$\begin{aligned}a_1x + b_1y &= c_1 \\ a_2x + b_2y &= c_2\end{aligned}$$

where a_1, b_1, c_1, a_2, b_2, c_2 are real constants. The solution of the equations are given by cross multiplication as,

$$x = \frac{b_2c_1 - b_1c_2}{a_1b_2 - a_2b_1}, \qquad y = \frac{c_2a_1 - c_1a_2}{a_1b_2 - a_2b_1}$$

It may be noted that if $a_1 b_2 - a_2 b_1 = 0$, then the solution does not exist. This aspect has to be kept in mind while writing the algorithm as given below.

Algorithm: Solution of a pair of equations $a_1x + b_1y = c_1$, $a_2\ x + b_2\ y = c_2$

Step 1: Read a_1, b_1, c_1, a_2, b_2, c_2

Step 2: Compute $d = a_1 b_2 - a_2 b_1$

Step 3: Check if $d = 0$, then go to Step 8 else
go to next step

Step 4: Compute x

Step 5: Compute y

Step 6: Write 'x =', x, 'y =', y

Step 7: Go to Step 9

NOTES

Step 8: Write 'No Solution'

Step 9: Stop

Example 1.13: Write an algorithm to compute the roots of a quadratic equation, $ax^2 + bx + c = 0$.

Solution: We know that the roots of the quadratic equation are given by,

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

Further, if $b^2 \geq 4ac$, the roots are real, otherwise they are complex conjugates. This aspect is to be considered while writing an algorithm.

Algorithm: Computation of roots of a quadratic equation.

Step 1: Read a, b, c

Step 2: Compute $d = b^2 - 4ac$

Step 3: Check if $d \geq 0$, go to Step 4 else go to Step 8

Step 4: Compute $x_1 = (-b + \sqrt{d})/(2a)$

Step 5: Compute $x_2 = (-b - \sqrt{d})/(2a)$

Step 6: Write 'Roots are real', x_1, x_2

Step 7: Go to Step 11

Step 8: Compute $x_i = \sqrt{-d}/(2a)$

Step 9: Compute $x_r = -b/(2a)$

Step 10: Write 'Roots are complex', 'Real part =' x_r 'Imaginary part =', x_i

Step 11: Stop

Check Your Progress

1. What are the two parts of floating point representation?
2. Define truncation and computational errors.
3. How will you define the inherent errors?
4. What is propagated round-off error?
5. What are significance errors?
6. Write the situations when loss of significant digits occur.
7. Why we write an algorithm?
8. Define features and purpose of computational algorithms.

1.5 SOLVING EQUATION

NOTES

In this section, we consider numerical methods for computing the roots of an equation of the form,

$$f(x) = 0 \tag{1.1}$$

where $f(x)$ is a reasonably well-behaved function of a real variable x. The function may be in algebraic form or polynomial form given by,

$$f(x) = a_n x^n + a_{n-1} x^{n-1} + ... + a_1 x + a_0 \tag{1.2}$$

It may also be an expression containing transcendental functions such as cos x, sin x, e^x, etc. First, we would discuss methods to find the isolated real roots of a single equation. Later, we would discuss methods to find the isolated roots of a system of equations, particularly of two real variables x and y, given by

$$f(x, y) = 0 \ , \ g(x, y) = 0 \tag{1.3}$$

A root of an equation is usually computed in two stages. First, we find the location of a root in the form of a crude approximation of the root. Next we use an iterative technique for computing a better value of the root to a desired accuracy in successive approximations/computations. This is done by using an iterative function.

Methods for Finding Location of Real Roots

The location or crude approximation of a real root is determined by the use of any one of the two methods, (a) Graphical and (b) Tabulation.

Graphical Method: In the graphical method, we draw the graph of the function $y = f(x)$, for a certain range of values of x. The abscissae of the points where the graph intersects the x-axis are crude approximations for the roots of the Equation (1.1). For example, consider the equation,

$$f(x) = x^2 + 2x - 1 = 0$$

From the graph of the function $y = f(x)$ shown in Figure 1.1, we find that it cuts the x-axis between 0 and 1. We may take any point in [0, 1] as the crude approximation for one root. Thus, we may take 0.5 as the location of a root. The other root lies between – 2 and – 3. We can take – 2.5 as the crude approximation of the other root.

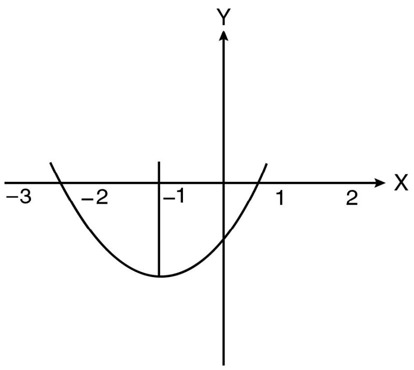


Fig. 1.1 *Graph of* $y = x^2 + 2x - 1$

NOTES

In some cases, where it is complicated to draw the graph of $y = f(x)$, we may rewrite the equation $f(x) = 0$, as $f_1(x) = f_2(x)$, where the graphs of $y = f_1(x)$ and $y = f_2(x)$ are standard curves. Then we find the x-coordinate(s) of the point(s) of intersection of the curves $y = f_1(x)$ and $y = f_2(x)$, which is the crude approximation of the root (s).

For example, consider the equation

$$x^3 - 15.2x - 13.2 = 0$$

This can be rewritten as,

$$x^3 = 15.2x + 13.2$$

where it is easy to draw the graphs of $y = x^3$ and $y = 15.2x + 13.2$. Then, the abscissa of the point(s) of intersection can be taken as the crude approximation(s) of the root(s).

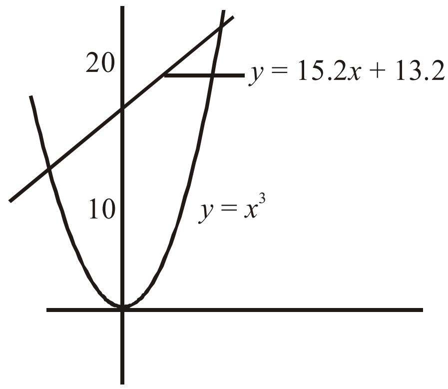
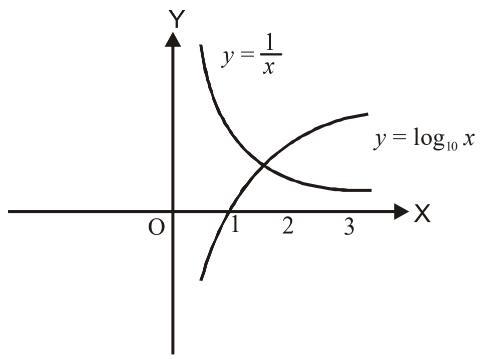


***Fig. 1.2** Graph of $y = x^3$ and $y = 15.2x + 13.2$*

Example 1.14: Find the location of the root of the equation $x \log_{10} x = 1$.

Solution: The equation can be rewritten as $\log_{10} x = \frac{1}{x}$.

Now the curves $y = \log_{10} x$ and $y = \frac{1}{x}$, can be easily drawn and are shown in Figure below.



Graph of $y = \frac{1}{x}$ *and* $y = \log_{10} x$

The point of intersection of the curves has its x-coordinates value 2.5 approximately. Thus, the location of the root is 2.5.

Tabulation Method: In the tabulation method, a table of values of $f(x)$ is made for values of x in a particular range. Then, we look for the change in sign in the values of $f(x)$ for two consecutive values of x. We conclude that a real root lies between these values of x. This is true if we make use of the following theorem on continuous functions.

Theorem 1.1: If $f(x)$ is continuous in an interval (a, b), and $f(a)$ and $f(b)$ are of opposite signs, then there exists at least one real root of $f(x) = 0$, between a and b.

Consider for example, the equation $f(x) = x^3 - 8x + 5 = 0$.

Constructing the following table of x and $f(x)$,

x	−4	−3	−2	−1	0	1	2	3
$f(x)$	−27	2	13	12	5	−2	−3	8

we observe that there is a change in sign of $f(x)$ in each of the sub-intervals (–3, –4), (0, 1) and (2, 3). Thus we can take the crude approximation for the three real roots as – 3.2, 0.2 and 2.2.

1.5.1 Bisection Method and Convergence of the Iterative Method

Bisection Method: The bisection method is a root finding method which repeatedly bisects an interval and then selects a subinterval in which a root must lie for further processing. It is an extremely simple and robust method, but it is relatively slow. It is normally used for obtaining a rough approximation to a solution which is then used as a starting point for more rapidly converging methods. When an interval contains more than one root, the bisection method can find one of them. When an interval contains a singularity, the bisection method converges to that singularity. The notion of the bisection method is based on the fact that a function will change sign when it passes through zero. By evaluating the function at the middle of an interval and replacing whichever limit has the same sign, the bisection method can halve the size of the interval in each iteration to find the root.

Thus, the bisection method is the simplest method for finding a root to an equation. It needs two initial estimates x_a and x_b which bracket the root. Let $f_a = f(x_a)$ and $f_b = f(x_b)$ such that $f_a f_b \le 0$. Evidently, if $f_a f_b = 0$ then one or both of x_a and x_b must be a root of $f(x) = 0$. As shown in Figure 1.3 is a graphical representation of the bisection method showing two initial guesses x_a and x_b bracketing the root.

NOTES

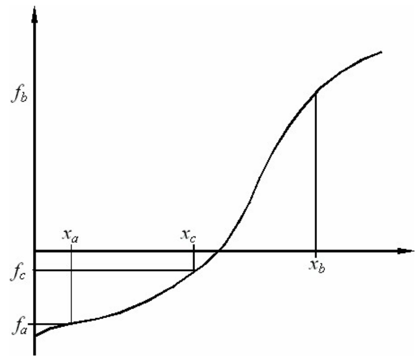


***Fig. 1.3** Graph of the Bisection Method showing Two Initial Estimates x_a and x_b Bracketing the Root*

The method is applicable when we wish to solve the equation $f(x) = 0$ for the real variable x, where f is a continuous function defined on an interval $[a, b]$ and $f(a)$ and $f(b)$ have opposite signs.

The bisection method involves successive reduction of the interval in which an isolated root of an equation lies. This method is based upon an important theorem on continuous functions as stated below.

Theorem 1.2: If a function $f(x)$ is continuous in the closed interval $[a, b]$, and $f(a)$ and $f(b)$ are of opposite signs, i.e., $f(a)f(b) < 0$, then there exists at least one real root of $f(x) = 0$ between a and b.

The bisection method starts with two guess values x_0 and x_1. Then this interval $[x_0, x_1]$ is bisected by a point $x_2 = \frac{1}{2}(x_0 + x_1)$, where $f(x_0) \cdot f(x_1) < 0$. We compute $f(x_2)$. If $f(x_2) = 0$, then x_2 is a root. Otherwise, we check whether $f(x_0) \cdot f(x_2) < 0$ or $f(x_1) \cdot f(x_2) < 0$. If $f(x_2)/f(x_0) < 0$, then the root lies in the interval (x_2, x_0). Otherwise, if $f(x_0) \cdot f(x_1) < 0$, then the root lies in the interval (x_2, x_1).

The sub-interval in which the root lies is again bisected and the above process is repeated until the length of the sub-interval is less than the desired accuracy.

The bisection method is also termed as bracketing method, since the method successively reduces the gap between the two ends of an interval surrounding the real root, i.e., brackets the real root.

The algorithm given below clearly shows the steps to be followed in finding a real root of an equation, by bisection method to the desired accuracy.

Algorithm: Finding root using bisection method.

Step 1: Define the equation, $f(x) = 0$

Step 2: Read epsilon, the desired accuracy

Setp 3: Read two initial values x_0 and x_1 which bracket the desired root

Step 4: Compute $y_0 = f(x_0)$

Step 5: Compute $y_1 = f(x_1)$

Step 6: Check if $y_0 y_1 < 0$, then go to Step 6
else go to Step 2

Step 7: Compute $x_2 = (x_0 + x_1)/2$

Step 8: Compute $y_2 = f(x_2)$

Step 9: Check if $y_0 y_2 > 0$, then set $x_0 = x_2$
else set $x_1 = x_2$

Step 10: Check if $|(x_1 - x_0)/x_1| >$ epsilon, then go to Step 3

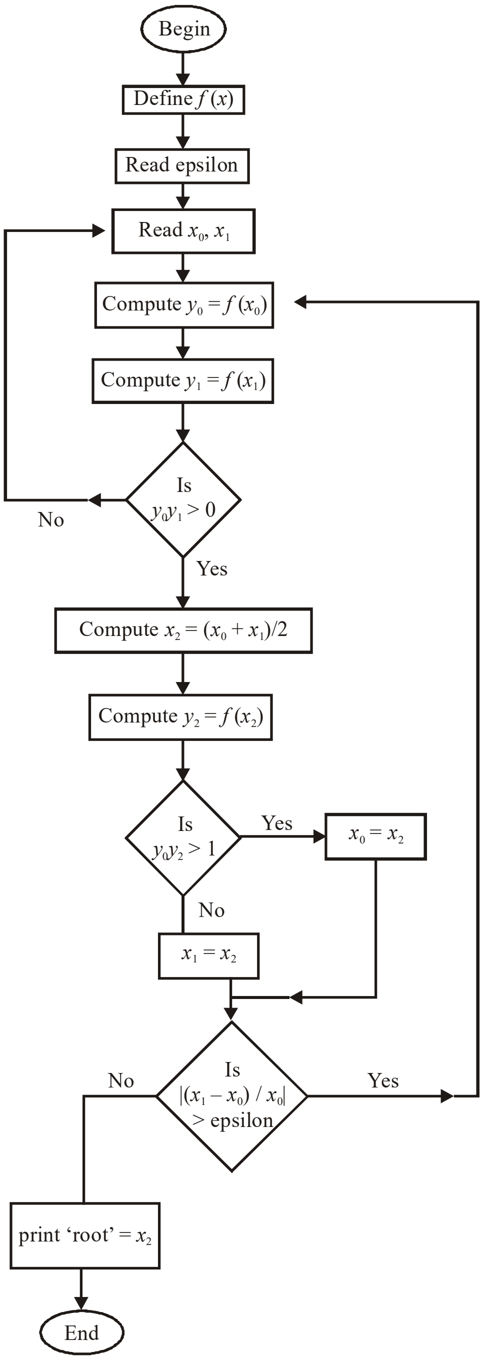
Step 11: Write x_2, y_2

Step 12: End

NOTES

Next, we give the flowchart representation of the above algorithm to get a better understanding of the method. The flowchart also helps in easy implementation of the method in a computer program.

Flowchart for Bisection Algorithm



Example 1.15: Find the location of the smallest positive root of the equation $x^3 - 9x + 1 = 0$ and compute it by bisection method, correct to two decimal places.

NOTES

Solution: To find the location of the smallest positive root we tabulate the function $f(x) = x^3 - 9x + 1$ below.

x	0	1	2	3
$f(x)$	1	−2	−9	1

We observe that the smallest positive root lies in the interval [0, 1]. The computed values for the successive steps of the bisection method are given in the table.

n	x_0	x_1	x_2	$f(x_2)$
1	0	1	0.5	−3.37
2	0	0.5	0.25	−1.23
3	0	0.25	0.125	−0.123
4	0	0.125	0.0625	0.437
5	0.0625	0.125	0.09375	0.155
6	0.09375	0.125	0.109375	0.016933
7	0.109375	0.125	0.11718	−0.053

From the above results, we conclude that the smallest root correct to two decimal places is 0.11.

Simple Iteration Method: A root of an equation $f(x) = 0$, is determined using the method of simple iteration by successively computing better and better approximation of the root, by first rewriting the equation in the form,

$$x = g(x) \tag{1.4}$$

Then, we form the sequence $\{x_n\}$ starting from the guess value x_0 of the root and computing successively,

$$x_1 = g(x_0), \quad x_2 = g(x_1), \ldots, \quad x_n = g(x_{n-1})$$

In general, the above sequence may converge to the root ξ as $n \to \infty$ or it may diverge. If the sequence diverges, we shall discard it and consider another form $x = h(x)$, by rewriting $f(x) = 0$. It is always possible to get a convergent sequence since there are different ways of rewriting $f(x) = 0$ in the form $x = g(x)$. However, instead of starting computation of the sequence, we shall first test whether the form of $g(x)$ can give a convergent sequence or not. We give below a theorem which can be used to test for convergence.

Theorem 1.3: If the function $g(x)$ is continuous in the interval $[a, b]$ which contains a root ξ of the equation $f(x) = 0$, and is rewritten as $x = g(x)$, and $|g'(x)| \le l \le 1$ in this interval, then for any choice of $x_0 \in [a, b]$, the sequence $\{x_n\}$ determined by the iterations,

$$x_{k+1} = g(x_k), \quad \text{for} \quad k = 0, 1, 2, \ldots \tag{1.5}$$

This converges to the root of $f(x) = 0$.

Proof: Since $x = \xi$, is a root of the equation $x = g(x)$, we have

$$\xi = g(\xi) \tag{1.6}$$

The first iteration gives $x_1 = g(x_0)$ (1.7)

Subtracting Equation (1.7) from Equation (1.6), we get

$$\xi - x_1 = g(\xi) - g(x_0)$$

Applying mean value theorem, we can write

$$\xi - x_1 = (\xi - x_0)g'(s_0), \quad x_0 < s_0 < \xi \tag{1.8}$$

Similarly, we can derive

$$\xi - x_2 = (\xi - x_1)g'(s_1), \quad x_1 < s_1 < \xi \tag{1.9}$$

....

$$\xi - x_{n+1} = (\xi - x_n)g'(s_n), \quad x_n < s_n < \xi \tag{1.10}$$

From Equations (1.8), (1.9) and (1.10), we get

$$\xi - x_{n+1} = (\xi - x_0)g'(s_0)g'(s_1)... \quad g'(s_n) \tag{1.11}$$

Since $|g'(x_i)| < l$ for each x_i, the above Equation (1.11) becomes,

$$|\xi - x_{n+1}| < l^{n+1}|\xi - x_0| \tag{1.12}$$

Evidently, since $l < 1$, $l^{n+1} \to 0$, as $n \to \infty$, the right hand side tends to zero and thus it follows that the sequence $\{x_n\}$ converges to the root ξ if $|\varphi'(\xi)| < 1$. This completes the proof.

Order of Convergence: The order of convergence of an iterative process is determined in terms of the errors e_n and e_{n+1} in successive iterations. An iterative process is said to have kth order of convergence if $\lim_{n\to\infty} \frac{e_{n+1}}{e_n^k} < M$, where M is a finite number.

Roughly speaking, the error in any iteration is proportional to the kth power of the error in the previous iteration.

Evidently, the simple iteration discussed in this section has its order of convergence 1.

The above iteration is also termed as **fixed point iteration** since it determines the root as the fixed point of the mapping defined by $x = g(x)$.

Algorithm: Computation of a root of $f(x) = 0$ by linear iteration.

Step 1: Define $g(x)$, where $f(x) = 0$ is rewritten as $x = g(x)$

Step 2: Input x_0, epsilon, maxit, where x_0 is the initial guess of root, epsilon is accuracy desired, maxit is the maximum number of iterations allowed.

Step 3: Set $i = 0$

Step 4: Set $x_1 = g(x_0)$

Step 5: Set $i = i + 1$

Step 6: Check, if $|(x_1 - x_0)/x_1| <$ epsilon, then print 'Root is', x_1
else go to Step 6

Step 7: Check, if $i < n$, then set $x_0 = x_1$ and go to Step 3

Step 8: Write 'No convergence after', n, 'Iterations'

Step 9: End

NOTES

Example 1.16: In order to compute a real root of the equation $x^3 - x - 1 = 0$, near $x = 1$, by iteration, determine which of the following iterative functions can be used to give a convergent sequence.

$$(i)\ x = x^3 - 1 \qquad (ii)\ x = \frac{x+1}{x^2} \qquad (iii)\ x = \sqrt{\frac{x+1}{x}}$$

Solution:

(*i*) For the form $x = x^3 - 1$, $g(x) = x^3 - 1$ and $g'(x) = 3x^2$. Hence, $|g'(x)| > 1$, for x near 1. So, this form would not give a convergent sequence of iterations.

(*ii*) For the form $x = \frac{x+1}{x^2}$, $g(x) = \frac{x+1}{x^2}$. Thus, $g'(x) = -\frac{1}{x^2} - \frac{2}{x^3}$ and $|g'(1)| = 3 > 1$. Hence, this form also would not give a convergent sequence of iterations.

(*iii*) For the form, $g(x) = \sqrt{\frac{x+1}{x}}$, $g'(x) = \frac{1}{2}\left(\frac{x+1}{x}\right)^{-\frac{1}{2}} \cdot \left(-\frac{1}{x^2}\right)$.

$\therefore\ |g'(1)| = \frac{1}{2\sqrt{2}} < 1$. Hence, the form $x = \sqrt{\frac{x+1}{x}}$ would give a convergent sequence of iterations.

Example 1.17: Compute the real root of the equation $x^3 + x^2 - 1 = 0$, correct to five significant digits, by iteration method.

Solution: The equation has a real root between 0 and 1 since $f(x) = x^3 + x^2 - 1$ has opposite signs at 0 and 1. For using iteration, we first rewrite the equation in the following different forms:

$$(i)\ x = \frac{1}{x^2} - 1 \qquad (ii)\ x = \sqrt{\frac{1}{x} - 1} \qquad (iii)\ x = \frac{1}{\sqrt{x+1}}$$

For the form (*i*), $g(x) = -1 + \frac{1}{x^2}$, $g'(x) = -\frac{2}{x^3}$ and for x in (0, 1), $|g'(x)| > 1$. So, this form is not suitable.

For the form (*ii*) $g'(x) = \frac{1}{2} \cdot \frac{1}{\sqrt{\frac{1}{x} - 1}}\left(-\frac{1}{x^2} - 1\right)$ and $|g'(x)| > 1$ for all x in (0, 1).

Finally, for the form (*iii*) $g'(x) = -\frac{1}{2} \cdot \frac{1}{(x+1)^{\frac{3}{2}}}$ and $|g'(x)| < 1$ for x in (0, 1).

Thus this form can be used to form a convergent sequence for finding the root.

We start the iteration $x = \frac{1}{\sqrt{1+x}}$ with $x_0 = 1$. The results of sucecessive iterations are,

$x_1 = 0.70711 \quad x_2 = 0.76537 \quad x_3 = 0.75236 \quad x_4 = 0.75541$
$x_5 = 0.75476 \quad x_6 = 0.75490 \quad x_7 = 0.75488 \quad x_8 = 0.75488$

Thus, the root is 0.75488, correct to five significant digits.

Example 1.18: Compute the root of the equation $x^2 - x - 0.1 = 0$, which lies in (1, 2), correct to five significant figures.

Solution: The equation is rewritten in the following form for computing the root by iteration:

$x = \sqrt{x+0.1}$. Here, $g'(x) = \frac{1}{2\sqrt{x+0.1}}$ and $|g'(x)| < 1$, for x in (1, 2).

The results for successive iterations, taking $x_0 = 1$, are

$x_1 = 1.0488 \qquad x_2 = 1.0718 \qquad x_3 = 1.0825$
$x_4 = 1.0874 \qquad x_5 = 1.0897.$

Thus, the root is 1.09, correct to three significant figures.

Example 1.19: Solve the following equation for the root lying in (2, 4) by using the method of linear iteration: $x^3 - 9x + 1 = 0$. Show that there are various ways of rewriting the equation in the form, $x = g(x)$ and choose the one which gives a convergent sequence for the root.

Solution: We can rewrite the equation in the following different forms:

$(i) \quad x = \frac{1}{9}(x^3+1) \qquad (ii) \quad x = 9/x - \frac{1}{x^2} \qquad (iii) \quad x = \sqrt{9 - \frac{1}{x}}$

In case of (*i*), $g'(x) = \frac{1}{3}x^2$ and for x in [2, 4], $|g'(x)| > 1$. Hence it will not give rise to a convergent sequence.

In case of (*ii*) $g'(x) = 2x - \frac{9}{x^2} + \frac{2}{x^3}$ and for x in [2, 4], $|g'(x)| > 1$

In case of (*iii*) $g'(x) = \left(9 - \frac{1}{x}\right)^{-\frac{1}{2}} \frac{1}{2x^2}$ and $|g'(x)| < 1$

Thus, the forms (*ii*) and (*iii*) would give convergent sequences for finding the root in [2, 3].

We start the iterations taking $x_0 = 2$ in the iteration scheme (*iii*). The result for successive iterations are,

$x_0 = 2.0 \qquad x_1 = 2.91548 \quad x_4 = 2.94282$
$x_2 = 2.94228 \qquad x_3 = 2.94281$

Thus, the root can be taken as 2.94281, correct to four decimal places.

NOTES

1.5.2 Newton-Raphson Method

Newton-Raphson method is a widely used numerical method for finding a root of an equation $f(x) = 0$, to the desired accuracy. It is an iterative method which has a faster rate of convergence and is very useful when the expression for the derivative $f(x)$ is not complicated. Newton-Raphson method, also called the Newton's method, is a root finding algorithm that uses the first few terms of the Taylor series of a function $f(x)$ in the neighborhood of a suspected root. In the Newton-Raphson method, to find the root start with an initial guess x_1 at the root, the next guess x_2 is the intersection of the tangent from the point $[x_1, f(x_1)]$ to the x-axis. The next guess x_3 is the intersection of the tangent from the point $[x_2, f(x_2)]$ to the x-axis as shown in Figure 1.4.

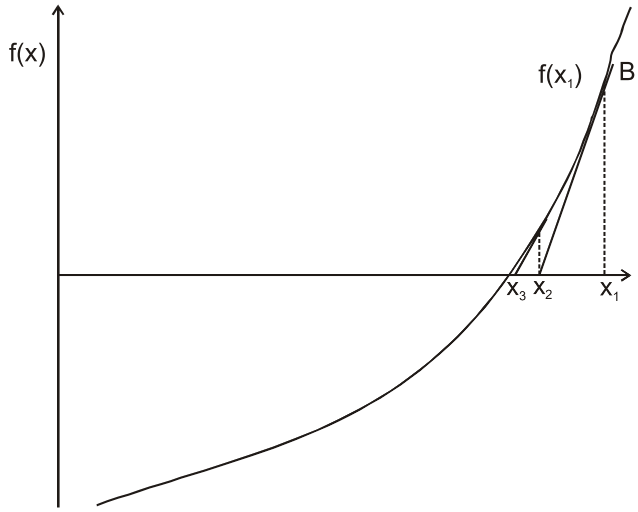


Fig. 1.4 *Graph of the Newton-Raphson Method*

The Newton-Raphson method can be derived from the definition of a slope as follows:

$$f'(x_1) = \frac{f(x_1) - 0}{x_1 - x_2} \Rightarrow x_2 = x_1 - \frac{f(x_1)}{f'(x_1)}$$

As a general rule, from the point $[x_n, f(x_n)]$, the next guess is calculated as follows:

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$

The derivative or slope $f(x_n)$ can be approximated numerically as follows:

$$f'(x_n) = \frac{f(x_n + \Delta x) - f(x_n)}{\Delta x}$$

To derive the formula for this method, we consider a Taylor's series expansion of $f(x_0 + h)$, x_0 being an initial guess of a root of $f(x) = 0$ and h a small correction to the root.

$$f(x_0 + h) = f(x_0) + h \quad f'(x_0) + \frac{h^2}{2\ !} \quad f''(x_0) + \ldots$$

NOTES

Assuming h to be small, we equate $f(x_0 + h)$ to 0 by neglecting square and higher powers of h.

$$\therefore \quad f(x_0) + h\, f'(x_0) = 0\sqrt{2}$$

or,

$$h = -\frac{f(x_0)}{f'(x_0)}$$

Thus, we can write an improved value of the root as,

$$x_1 = x_0 + h$$

i.e.,

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}$$

Successive approximations $x_2, x_3, ..., x_{n+1}$ can thus be written as,

$$x_2 = x_1 - \frac{f(x_1)}{f'(x_1)}$$

$$x_3 = x_2 - \frac{f(x_2)}{f'(x_2)}$$

... ...

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \qquad (1.13)$$

If the sequence $\{x_n\}$ converges, we get the root.

Algorithm: Computation of a root of $f(x) = 0$ by Newton-Raphson method.

Step 1: Define $f(x)$, $f'(x)$

Step 2: Input x_0, epsilon, maxit

[x_0 is the initial guess of root, epsilon is the desired accuracy of the root and maxit is the maximum number of iterations allowed]

Step 3: Set $i = 0$

Step 4: Set $f_0 = f(x_0)$

Step 5: Compute $df_0 = f'(x_0)$

Step 6: Set $x_1 = x_0 - f_0/df_0$

Step 7: Set $i = i + 1$

Step 8: Check if $|(x_1 - x_0)\ |x_1| <$ epsilon, then print 'Root is', x_1 and stop

else if $i < n$, then set $x_0 = x_1$ and go to Step 3

Step 9: Write 'Iterations do not converge'

Step 10: End

Example 1.20: Use Newton-Raphson method to compute the positive root of the equation $x^3 - 8x - 4 = 0$, correct to five significant digits.

Solution: Newton-Raphson iterative scheme is given by,

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}, \text{ for } n = 0, 1, 2, ...$$

For the given equation, $f(x) = x^3 - 8x - 4$.

First we find the location of the root by the method of tabulation. The table for $f(x)$ is,

NOTES

x	0	1	2	3	4
$f(x)$	-4	-13	-12	-1	28

Evidently, the positive root is near $x = 3$. We take $x_0 = 3$ in Newton-Raphson iterative scheme.

$$x_{n+1} = x_n - \frac{x_n^3 - 8x_n - 4}{3x_n^2 - 8}$$

We get, $x_1 = 3 - \dfrac{27 - 24 - 4}{27 - 8} = 3.0526$

Similarly, $x_2 = 3.05138$ and $x_3 = 3.05138$.

Thus, the positive root is 3.0514, correct to five significant digits.

Example 1.21: Find a real root of the equation $x^3 + 7x^2 + 9 = 0$, correct to five significant digits.

Solution: First we find the location of the real root by tabulation. We observe that the real root is negative and since $f(-7) = 9 > 0$ and $f(-8) = -55 < 0$, a root lies between -7 and -8.

For computing the root to the desired accuracy, we take $x_0 = -8$ and use Newton-Raphson iterative formula,

$$x_{n+1} = x_n - \frac{x_n^3 + 7x_n^2 + 9}{3x_n^2 + 14x_n}, \quad \text{for} \quad n = 0, 1, 2, ...$$

The successive iterations give,

$$x_1 = -7.3125$$
$$x_2 = -7.17966$$
$$x_3 = -7.17484$$
$$x_4 = -7.17483$$

Hence, the desired root is -7.1748, correct to five significant digits.

Example 1.22: For evaluating $\sqrt{a}$, deduce the iterative formula $x_{n+1} = \dfrac{1}{2}\left(x_n + \dfrac{a}{x_n}\right)$, by using Newton-Raphson scheme of iteration. Hence, evaluate $\sqrt{2}$ using this, correct to four significant digits.

Solution: We observe that, $\sqrt{a}$ is the solution of the equation $x^2 - a = 0$.

Now, using $f(x) = x^2 - a$ in the Newton-Raphson iterative scheme,

$$x_{n+1} = x_n - \frac{x_n^2 - a}{2x_n}$$

We have, $$x_{n+1} = x_n - \frac{x_n^2 - a}{2x_n}$$

$$x_{n+1} = x_n - \frac{x_n^2 - a}{2x_n}$$

i.e.,
$$x_{n+1} = \frac{1}{2}\left(x_n + \frac{a}{x_n}\right), \quad \text{for} \quad n = 0, 1, 2, \ldots$$

Now, for computing $\sqrt{2}$, we assume $x_0 = 1.4$. The successive iterations give,

$$x_1 = \frac{1}{2}\left(1.4 + \frac{2}{1.4}\right) = \frac{3.96}{2.8} = 1.414$$
$$x_2 = \frac{1}{2}\left(1.414 + \frac{2}{1.414}\right) = 1.41421$$

Hence, the value of $\sqrt{2}$ is 1.414 correct to four significant digits.

Example 1.23: Prove that $\sqrt[k]{a}$ can be computed by the iterative scheme,

$x_{n+1} = \frac{1}{k}\left[(k-1)x_n + \frac{a}{x_n^{k-1}}\right]$. Hence evaluate $\sqrt[3]{2}$, correct to five significant digits.

Solution: The value $\sqrt[k]{a}$ is the positive root of $x^k - a = 0$. Thus, the iterative scheme for evaluating $\sqrt[k]{a}$ is,

$$x_{n+1} = x_n - \frac{x_n^k - a}{kx_n^{k-1}}$$

or, $$x_{n+1} = \frac{1}{k}\left[(k-1)x_n + \frac{a}{x_n^{k-1}}\right], \quad \text{for} \quad n = 0, 1, 2, \ldots$$

Now, for evaluating $\sqrt[3]{2}$, we take $x_0 = 1.25$ and use the iterative formula,

$$x_{n+1} = \frac{1}{3}\left[2x_n + \frac{2}{x_n^2}\right].$$

We have, $$x_1 = \frac{1}{3}\left[1.25 \times 2 + \frac{2}{(1.25)^2}\right] = 1.26$$

$$x_2 = 1.259921, \quad x_3 = 1.259921$$

Hence, $\sqrt[3]{2} = 1.2599$, correct to five significant digits.

Example 1.24: Find by Newton-Raphson method, the real root of $3x - \cos x - 1 = 0$, correct to three significant figures.

Solution: The location of the real root of $f(x) = 3x - \cos x - 1 = 0$, is [0, 1] since $f(0) = -2$ and $f(1) > 0$.

We choose $x_0 = 0$ and use Newton-Raphson scheme of iteration.

$$x_{n+1} = x_n - \frac{3x_n - \cos x_n - 1}{3 + \sin x_n}, \quad n = 0, 1, 2, \ldots$$

The results for successive iterations are,

$$x_1 = 0.667,\ x_2 = 0.6075,\ x_3 = 0.6071$$

Thus, the root is 0.607 correct to three significant figures.

Example 1.25: Find a real root of the equation $x^x + 2x - 6 = 0$, correct to four significant digits.

Solution: Taking $f(x) = x^x + 2x - 6$, we have $f(1) = -3 < 0$ and $f(2) = 2 > 0$. Thus, a root lies in [1, 2]. Choosing $x_0 = 2$, we use Newton-Raphson iterative scheme given by,

NOTES

$$x_{n+1} = x_n - \frac{x_n^{x_n} + 2x_n - 6}{x_n^{x_n}(\log_e x_n + 1) + 2}, \quad \text{for} \quad n = 0, 1, 2, ...$$

The computed results for successive iterations are,

$$x_1 = 2 - \frac{4 + 4 - 6}{4 \times (\log_e 2x^2 + 1) + 2} = 1.72238$$

$$x_2 = 1.72321$$

$$x_3 = 1.72308$$

Hence, the root is 1.723 correct to four significant figures.

Order of Convergence: We consider the order of convergence of the Newton-Raphson method given by the formula,

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$

Let us assume that the sequence of iterations $\{x_n\}$ converge to the root ξ. Then, expanding by Taylor's series about x_n, the relation $f(\xi) = 0$, gives

$$f(x_n) + (\xi - x_n) f'(x_n) + \frac{1}{2}(\xi - x_n)^2 f''(x_n) + ... = 0$$

$$\therefore \quad -\frac{f(x_n)}{f'(x_n)} = \xi - x_n + \frac{1}{2}(\xi - x_n)^2 \quad . \quad \frac{f''(x_n)}{f'(x_n)} + ...$$

$$\therefore \quad x_{n+1} - \xi \approx \frac{1}{2}(\xi - x_n)^2 \quad . \quad \frac{f''(x_n)}{f'(x_n)}$$

Taking ε_n as the error in the nth iteration and writing $\varepsilon_n = \varepsilon_n - \xi$, we have,

$$\varepsilon_{n+1} \approx \frac{1}{2}\varepsilon_n^{\,2} \quad \frac{f''(\xi)}{f'(\xi)} \tag{1.14}$$

Thus, $\varepsilon_{n+1} = k\varepsilon^2 n$, where k is a constant.

This shows that the order of convergence of Newton-Raphson method is 2. In other words, the Newton-Raphson method has a quadratic rate of convergence.

The condition for convergence of Newton-Raphson method can easily be derived by rewriting the Newton-Raphson iterative scheme as $x_{n+1} = \varphi(x_n)$ with $\varphi(x) = x - \dfrac{f(x)}{f'(x)}$.

Hence, using the condition for convergence of the linear iteration method, we can write $\varphi'(x) = \dfrac{f(x) \quad f''(x)}{[f'(x)]^2}$.

NOTES

Thus, the sufficient condition for the convergence of Newton-Raphson method is,

$$\left|\frac{f(x) \quad f''(x)}{[f'(x)]^2}\right| < 1, \text{ in the interval near the root.}$$

i.e., $$|f(x) \quad f''(x)| < |f'(x)|^2 \qquad (1.15)$$

1.5.3 Secant Method

Secant method can be considered as a discretized form of Newton-Raphson method. The iterative formula for this method is obtained from formula of Newton-Raphson method on replacing the derivative $f'(x_0)$ by the gradient of the chord joining two neighbouring points x_0 and x_1 on the curve $y = f(x)$.

Thus, we have

$$f'(x_0) \approx \frac{f(x_1) - f(x_0)}{x_1 - x_0}$$

The iterative formula is given by,

$$x_2 = x_1 - \frac{f(x_1)}{f(x_1) - f(x_0)}(x_1 - x_0)$$

This can be rewritten as,

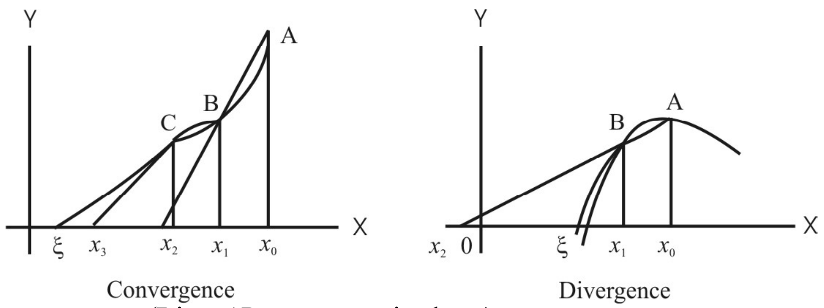
$$x_2 = \frac{x_0 f(x_1) - x_1 f(x_0)}{f(x_1) - f(x_0)}$$

Secant formula in general form is,

$$x_n = x_{n-1} - f(x_{n-1})$$

$$\frac{x_{n-1} - x_{n-1}}{f(x_{n-1}) - f(x_{n-2})}$$

The iterative formula is equivalent to the one for Regula-Falsi method. The distinction between secant method and Regula-Falsi method lies in the fact that unlike in Regula-Falsi method, the two initial guess values do not bracket a root and the bracketing of the root is not checked during successive iterations, in secant method. Thus, secant method may not always give rise to a convergent sequence to find the root. The geometrical interpretation of the method is shown in Figure 1.5.



(Line *AB* meets *x*-axis alone)

Fig. 1.5 *Secant Method*

Algorithm: To find a root of $f(x) = 0$, by secant method.

Step 1: Define $f(x)$.

Step 2: Input x_0, x_1, error, maxit [x_0, x_1, are initial guess values, error is the prescribed precision and maxit is the maximum number of iterations allowed].

Step 3: Set $i = 1$

Step 4: Compute $f_0 = f(x_0)$

Step 5: Compute $f_1 = f(x_1)$

Step 6: Compute $x_2 = (x_0 f_1 - x_1 f_0)/(f_1 - f_0)$

Step 7: Set $i = i + 1$

Step 8: Compute accy $= |x_2 - x_1| / |x_1|$

Step 9: Check if accy < error, then go to Step 14

Step 10: Check if $i \geq$ maxit then go to Step 16

Step 11: Set $x_0 = x_1$

Step 12: Set $x_1 = x_2$

Step 13: Go to Step 6

Step 14: Print 'Root =', x_2

Step 15: Go to Step 17

Step 16: Print 'Iterations do not converge'

Step 17: Stop

NOTES

1.5.4 Regula-Falsi Method

Regula-Falsi method is also a bracketing method. As in bisection method, we start the computation by first finding an interval (a, b) within which a real root lies. Writing $a = x_0$ and $b = x_1$, we compute $f(x_0)$ and $f(x_1)$, and check if $f(x_0)$ and $f(x_1)$ are of opposite signs. For determining the approximate root x_2, we find the point of intersection of the chord joining the points $(x_0, f(x_0))$ and $(x_1, f(x_1))$ with the x-axis, i.e., the curve $y = f(x_0)$ is replaced by the chord given by,

$$y - f(x_0) = \frac{f(x_1) - f(x_0)}{x_1 - x_0}(x - x_0) \quad (1.16)$$

Thus, by putting $y = 0$ and $x = x_2$ in Equation (1.16), we get

$$x_2 = x_0 - \frac{f(x_0)}{f(x_1) - f(x_0)}(x_1 - x_0) \quad (1.17)$$

Next, we compute $f(x_2)$ and determine the interval in which the root lies in the following manner. If (*a*) $f(x_2)$ and $f(x_1)$ are of opposite signs, then the root lies in (x_2, x_1). Otherwise if (*b*) $f(x_0)$ and $f(x_2)$ are of opposite signs, then the root lies in (x_0, x_2). The next approximate root is determined by changing x_0 by x_2 in the first case and x_1 by x_2 in the second case.

The aforesaid process is repeated until the root is computed to the desired accuracy ε, i.e., the condition

$|(x_{k+1} - x_k)/x_k| < \varepsilon,$ should be satisfied.

Regula-Falsi method can be geometrically interpreted by the following Figure 1.6.

NOTES

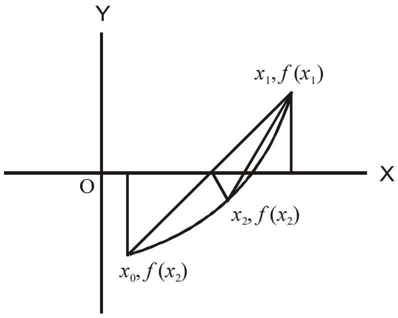


Fig. 1.6 *Regula-Falsi Method*

Algorithm: Computing root of an equation by Regula-Falsi method.

Step 1: Define $f(x)$

Step 2: Read epsilon, the desired accuracy

Step 3: Read maxit, the maximum number of iterations

Step 4: Read x_0, x_1, two initial guess values of root

Step 5: Compute $f_0 = f(x_0)$

Step 6: Compute $f_1 = f(x_1)$

Step 7: Check if $f_0 f_1 < 0$, then go to the next step
else go to Step 4

Step 8: Compute $x_2 = (x_0 f_1 - x_1 f_0) / (f_1 - f_0)$

Step 9: Compute $f_2 = f(x_2)$

Step 10: Check if $|f_2| <$ epsilon, then go to Step 18

Step 11: Check if $f_2 f_0 < 0$ then go to the next step
else go to Step 15

Step 12: Set $x_1 = x_2$

Step 13: Set $f_1 = f_2$

Step 14: Go to Step 7

Step 15: Set $x_0 = x_2$

Step 16: Set $f_0 = f_2$

Step 17: Go to Step 7

Step 18: Write 'Root =' , x_2, f_3

Step 19: End

Example 1.26: Use Regula-Falsi method to compute the positive root of $x^3 - 3x - 5 = 0$, correct to four significant figures.

Solution: First we find the interval in which the root lies. We observe that $f(2) = -3$ and $f(3) = 13$. Thus, the root lies in [2, 3]. For using the Regula-Falsi method, we use the formula,

$$x_2 = x_0 - \frac{f(x_0)}{f(x_1) - f(x_0)}(x_1 - x_0)$$

with $x_0 = 2$, and $x_1 = 3$, we have

NOTES

$$x_2 = 2 + \frac{3}{13+3}(3-2) = 2.1875$$

Again, since $f(x_2) = f(2.1875) = -1.095$, we consider the interval [2.1875, 3]. The next approximation is $x_3 = 2.2461$. Also, $f(x_3) = -0.4128$. Hence, the root lies in [2.2461, 3].

Repeating the iterations, we get

$$x_4 = 2.2684, f(x_4) = -\ 0.1328$$
$$x_5 = 2.2748, f(x_5) = -\ 0.0529$$
$$x_6 = 2.2773, f(x_6) = -\ 0.0316$$
$$x_7 = 2.2788, f(x_7) = -\ 0.0028$$
$$x_8 = 2.2792, f(x_8) = -\ 0.0022$$

The root correct to four significant figures is 2.279.

Roots of Polynomial Equations

Polynomial equations with real coefficients have some important characteristics regarding their roots. A polynomial equation of degree n is of the form $p_n(x) = a_n x^n + a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + ... + a_2x^2 + a_1x + a_0 = 0$.

(*i*) A polynomial equation of degree n has exactly n roots.

(*ii*) Complex roots occur in pairs, i.e., if $\alpha + i\beta$ is a root of $p_n(x) = 0$, then $\alpha - i\beta$ is also a root.

(*iii*) Descarte's rule of signs can be used to determine the number of possible real roots (positive or negative).

(*iv*) If $x_1, x_2, ..., x_n$ are all real roots of the polynomial equation, then we can express $p_n(x)$ uniquely as,

$$p_n(x) = a_n(x - x_1)(x - x_2)...(x - x_n)$$

(*v*) $p_n(x)$ has a quadratic factor for each pair of complex conjugate roots. Let, $\alpha + i\beta$ and $\alpha - i\beta$ be the roots, then $\{x^2 - 2\alpha\, x + (\alpha^2 + \beta^2)\}$ is the quadratic factor.

(*vi*) There is a special method, known as Horner's method of synthetic substitution, for evaluating the values of a polynomial and its derivatives for a given x.

1.5.5 Descarte's Rule

The number of positive real roots of a polynomial equation is equal to the number of changes of sign in $p_n(x)$, written with descending powers of x, or less by an even number.

Consider for example, the polynomial equation,

$$3x^5 + 2x^4 + x^3 - 2x^2 + x - 2 = 0$$

Clearly there are three changes of sign and hence the number of positive real roots is three or one. Thus, it must have a real root. In fact, every polynomial equation of odd degree has a real root.

We can also use Descarte's rule to determine the number of negative roots by finding the number of changes of signs in $p_n(-x)$. For the above equation, $p_n(-x) = -3x^5 + 2x^4 - x^3 - 2x^2 - x - 2 = 0$ and it has two changes of sign. Thus, it has either two negative real roots or none.

Check Your Progress

9. The roots of an equation are computed in how many stages?
10. Define tabulation method.
11. State the procedure of bisection method.
12. How is the order of convergence of an iterative process determined?
13. State a property of Newton-Raphson method.
14. Define secant method.
15. Give the procedure of Regula-Falsi method.

1.6 ANSWERS TO 'CHECK YOUR PROGRESS'

1. The floating point representation of a number consists of mantissa and exponent.
2. The errors in a numerical solution are basically of two types. They are truncation error and computational error. The error which is inherent in the numerical method employed for finding numerical solution is called the truncation error. The computational error arises while doing arithmetic computation due to representation of numbers with a finite number of decimal digits.
3. Inherent errors are errors in the data which are obtained by physical measurement and are due to limitations of the measuring instrument. The analysis of errors in the computed result due to the inherent errors in data is similar to that of round-off errors.
4. Propagated round-off error is the sum of two approximate numbers (having round-off errors) equal to the sum of the round-off errors in the individual numbers.
5. During arithmetic computations of approximate numbers having fixed precision, there may be loss of significant digits in some cases. The error due to loss of significant digits is termed as significance error.
6. There are two situations when loss of significant digits occur. These are,
   (*i*) Subtraction of two nearly equal numbers
   (*ii*) Division by a very small divisor compared to the dividend
7. To get a numerical solution on a computer, one has to write an algorithm.

8. For solving problems with the help of a computer, one should first analyse the mathematical formulation of the problem and consider a suitable numerical method for solving it. The next step is to write an algorithm for implementing the method.
9. A root of an equation is usually computed in two stages. First, we find the location of a root in the form of a crude approximation of the root. Next we use an iterative technique for computing a better value of the root to a desired accuracy in successive approximations/computations.
10. In the tabulation method, a table of values of $f(x)$ is made for values of x in a particular range. Then, we look for the change in sign in the values of $f(x)$ for two consecutive values of x. We conclude that a real root lies between these values of x.
11. The bisection method involves successive reduction of the interval in which an isolated root of an equation lies. The sub-interval in which the root lies is again bisected and the above process is repeated until the length of the sub-interval is less than the desired accuracy.
12. The order of convergence of an iterative process is determined in terms of the errors e_n and e_{n+1} in successive iterations.
13. Newton-Raphson method is a widely used numerical method for finding a root of an equation $f(x) = 0$, to the desired accuracy. It is an iterative method which has a faster rate of convergence and is very useful when the expression for the derivative $f(x)$ is not complicated.
14. Secant method can be considered as a discretized form of Newton-Raphson method. The iterative formula for this method is obtained from formula of Newton-Raphson method on replacing the derivative by the gradient of the chord joining two neighbouring points x_0 and x_1 on the curve $y = f(x)$.
15. Regula-Falsi method is also a bracketing method. As in bisection method, we start the computation by first finding an interval (a, b) within which a real root lies. Writing $a = x_0$ and $b = x_1$, we compute $f(x_0)$ and $f(x_1)$ and check if $f(x_0)$ and $f(x_1)$ are of opposite signs. For determining the approximate root x_2, we find the point of intersection of the chord joining the points $(x_0, f(x_0))$ and $(x_1, f(x_1))$ with the x-axis, i.e., the curve $y = f(x_0)$ is replaced by the chord given by,

$$y - f(x_0) = \frac{f(x_1) - f(x_0)}{x_1 - x_0}(x - x_0)$$

1.7 SUMMARY

- Numerical methods are methods used for solving problems through numerical calculations providing a table of numbers and/or graphical representations or figures. Numerical methods emphasize that how the algorithms are implemented.
- To perform a numerical calculation, approximate them first by a representation involving a finite number of significant digits. If the numbers to be represented are very large or very small, then they are written in floating point notation.

- The Institute of Electrical and Electronics Engineers (IEEE) has published a standard for binary floating point arithmetic.
- In approximate representation of numbers, the number is represented with a finite number of digits. All the digits in the usual decimal representation may not be significant while considering the accuracy of the number.
- In a floating representation, a number is represented with a finite number of significant digits having a floating decimal point.
- Floating point representation of a number consists of mantissa and exponent.
- The errors in a numerical solution are basically of two types termed as truncation error and computational error.
- The error which is inherent in the numerical method employed for finding numerical solution is called the truncation error.
- The truncation error arises due to the replacement of an infinite process such as summation or integration by a finite one.
- Inherent errors are errors in the data which are obtained by physical measurement and are due to limitations of the measuring instrument.
- Numerical methods can be employed for computing the roots of an equation of the form, $f(x) = 0$, where $f(x)$ is a reasonably well-behaved function of a real variable x.
- The location or crude approximation of a real root is determined by the use of any one of the following methods, (a) Graphical and (b) Tabulation.
- In general, the roots of an equation can be computed using bisection and simple iteration methods.
- The bisection method is also termed as bracketing method.
- Newton-Raphson method is a widely used numerical method for finding a root of an equation $f(x) = 0$, to the desired accuracy.
- Secant method can be considered as a discretized form of Newton-Raphson method. The iterative formula for this method is obtained from formula of Newton-Raphson method on replacing the derivative by the gradient of the chord joining two neighbouring points x_0 and x_1 on the curve $y = f(x)$.
- Regula-Falsi method is also a bracketing method.

1.8 KEY TERMS

- **Truncation error:** This error is inherent in the numerical method employed for finding numerical solution. It occurs due to the replacement of an infinite process such as summation or integration by a finite one.
- **Computational error:** This error occurs during arithmetic computation due to representation of numbers having a finite number of decimal digits.
- **Inherent error:** This error occurs in the data type which is obtained using physical measurement and also due to limitations of the measuring instruments.
- **Significance error:** This error occurs due to loss of significant digits.

NOTES

- **Tabulation method:** In the tabulation method, a table of values of $f(x)$ is made for values of x in a particular range.
- **Bisection method:** It involves successive reduction of the interval in which an isolated root of an equation lies.

1.9 SELF-ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. What are floating point numbers?
2. Find the percentage error in approximating $\frac{5}{6}$ by 0.8333 correct upto four significant figures.
3. Write the characteristics of numerical computation.
4. Find the relative error in the computation of $x - y$ for $x = 12.05$ and $y = 8.02$ having absolute errors $\Delta x = 0.005$ and $\Delta y = 0.001$.
5. Find the percentage error in computing $y = 3x^2 - 6x$ at $x = 1$, if the error in x is 0.05.
6. Given $a = 1.135$ and $b = 1.075$ having absolute errors $\Delta a = 0.011$ and $\Delta b = 0.12$. Estimate the relative percentage error in the computation of $a - b$.
7. What are isolated roots?
8. What is crude approximation in graphical method?
9. Why is bisection method also termed as bracketing method?
10. What is the order of convergence of the Newton-Raphson method?
11. State the similarity between secant method and Regula-Falsi method.

Long-Answer Questions

1. Round-off the following numbers to three decimal places:

   (*i*) 0.230582 (*ii*) 0.00221118 (*iii*) 2.3645 (*iv*) 1.3455
2. Round-off the following numbers to four significant figures:

   (*i*) 49.3628 (*ii*) 0.80022 (*iii*) 8.9325 (*iv*) 0.032588

   (*v*) 0.0029417 (*vi*) 0.00010211 (*vii*) 410.99
3. Round-off each of the following numbers to three significant figures and indicate the absolute error in each.

   (*i*) 49.3628 (*ii*) 0.9002 (*iii*) 8.325 (*iv*) 0.0039417
4. Find the sum of the following approximate numbers, correct to the last digits.

   0.348, 0.1834, 345.4, 235.2, 11.75, 0.0849, 0.0214, 0.0002435
5. Find the number of correct significant digits in the approximate number 11.2461. Given is its absolute error = 0.25×10^{-2}.

6. Given are the following approximate numbers with their relative errors. Determine the absolute errors.

   (*i*) $x_A = 12165, \quad \varepsilon_R = 0.1\%$ (*ii*) $x_A = 3.23, \quad \varepsilon_R = 0.6\%$

   (*iii*) $x_A = 0.798, \quad \varepsilon_R = 10\%$ (*iv*) $x_A = 67.84, \quad \varepsilon_R = 1\%$

7. Round-off the following numbers to four significant digits.

   (*i*) 450.92 (*ii*) 48.3668 (*iii*) 9.3265 (*iv*) 8.4155

   (*v*) 0.80012 (*vi*) 0.042514 (*vii*) 0.0049125 (*viii*) 0.00020215

8. Write the following numbers in floating-point form rounded to four significant digits.

   (*i*) 100000 (*ii*) – 0.0022136 (*iii*) – 35.666

9. Determine the number of correct digits in the number *x* in each of the following (the relative errors are given).

   (*i*) $x = 0.2217, \quad \varepsilon_R = 0.2 \times 10^{-1}$ (*ii*) $x = 32.541, \quad \varepsilon_R = 0.1$

   (*iii*) $x = 0.12432, \quad \varepsilon_R = 10\%$ (*iv*) $x = 0.58632, \quad \varepsilon_R = 1\%$

10. Find the percentage error in computing $z = \sqrt{x}$ for $x = 4.44$, if *x* is correct to its last digit only.

11. Let $u = 4x^6 + 3x - 9$. Find the relative percentage error in computing *u* at $x = 1.1$, if the error in *x* is 0.05.

12. Use graphical method to find the location of a real root of the equation $x^3 + 10_x - 15 = 0$.

13. Draw the graphs of the function $f(x) = \cos x - x$, in the range $[0, \pi/2)$ and find the location of the root of the equation $f(x) = 0$.

14. Compute the root of the equation $x^3 - 9x + 1 = 0$ which lies between 2 and 3 correct upto three significant digits using bisection method.

15. Compute the root of the equation $x^3 + x^2 - 1 = 0$, near 1, by the iterative method correct upto two significant digits.

16. Compute using Newton-Raphson method the root of the equation $e^x = 4^x$, near 2, correct upto four significant digits.

17. Find the real root of $x \log_{10} x - 1.2 = 0$ correct upto four decimal places using Regula-Falsi method.

1.10 FURTHER READING

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

Chandan, J.S., Jagjit Singh and K.K. Khanna. 1995. *Business Statistics*. New Delhi: Vikas Publishing House.

Elhance, D.N. 2006. *Fundamental of Statistics*. Allahabad: Kitab Mahal.

NOTES

Freud, J.E., and F.J. William. 1997. *Elementary Business Statistics – The Modern Approach*. New Jersey: Prentice-Hall International.

Goon, A.M., M.K. Gupta, and B. Das Gupta. 1983. *Fundamentals of Statistics*. Vols. I & II, Kolkata: The World Press Pvt. Ltd.

Gupta, S.C. 2008. *Fundamentals of Business Statistics*. Mumbai: Himalaya Publishing House.

Kothari, C.R. 1984. *Quantitative Techniques*. New Delhi: Vikas Publishing House.

Levin, Richard. I., and David. S. Rubin. 1997. *Statistics for Management.* New Jersey: Prentice-Hall International.

Meyer, Paul L. 1970. *Introductory Probability and Statistical Applications*. Massachusetts: Addison-Wesley.

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

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

Anderson, David R., Dennis J. Sweeney and Thomas A. Williams. *Essentials of Statistics for Business and Economics*. Mumbai: Thomson Learning, 2007.

S.P. Gupta. 2021. *Statistical Methods*. Delhi: Sultan Chand and Sons.

UNIT 2 INTERPOLATION AND CURVE FITTING

NOTES

Structure

2.0 INTRODUCTION

Interpolation is the process of defining a function that takes on specified values at specified points. Polynomial interpolation is the most known one-dimensional interpolation method. Its advantages lies in its simplicity of realization and the good quality of interpolants obtained from it. You will learn about the various interpolation methods, namely Lagrange's interpolation, Newton's forward and backward difference interpolation formulae, iterative linear interpolation and inverse interpolation.

Curve fitting is the process of constructing a curve, or mathematical function, which has the best fit to a series of data points, possibly subject to constraints.

In mathematics, the trigonometric functions, also called the circular functions, are functions of an angle. They relate the angles of a triangle to the lengths of its sides. The most familiar trigonometric functions are the sine, cosine and tangent. In the context of the standard unit circle (a circle with radius 1 unit), where a triangle is formed by a ray originating at the origin and making some angle with the x-axis, the sine of the angle gives the length of the y-component (the opposite to the angle or the rise) of the triangle, the cosine gives the length of the x-component (the adjacent of the angle or the run), and the tangent function gives the slope (y-component divided by the x-component). Trigonometric functions are commonly defined as ratios of two sides of a right triangle containing the angle, and can equivalently be defined as the lengths of various line segments from a unit circle.

Regression analysis, is the mathematical process of using observations to find the line of best fit through the data in order to make estimates and predictions about the behaviour of variables. This technique is used to determine the statistical relationship between two or more variables and to make prediction of one variable on the basis of one or more other variables.

In this unit, you will learn about the interpolation, curve fitting, trigonometric function and regression.

2.1 OBJECTIVES

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

- Describe the method of iterative linear interpolation
- Understand polynomial interpolation
- Explain the importance of Lagrange's interpolation
- Perform interpolation of equally spaced tabular values
- Explain finite, forward and backward differences
- Evaluate interpolation using symbolic, shift and central difference operators
- Know differences of polynomials
- Define Newton's forward and backward interpolation formulae
- Explain extrapolation and inverse interpolation
- Understand the concept of curve fitting
- Explain the various trigonometric functions
- Discuss regression analysis in detail

2.2 INTERPOLATION

The problem of interpolation is very fundamental problem in numerical analysis. The term interpolation literally means reading between the lines. In numerical analysis, interpolation means computing the value of a function $f(x)$ in between values of x in a table of values. It can be stated explicitly as 'given a set of $(n + 1)$ values y_0, $y_1, y_2,..., y_n$ for $x = x_0, x_1, x_2, ..., x_n$ respectively. The problem of interpolation is to compute the value of the function $y = f(x)$ for some non-tabular value of x.'

The computation is often made by finding a polynomial called interpolating polynomial of degree less than or equal to n such that the value of the polynomial is equal to the value of the function at each of the tabulated points. Thus if,

$$\varphi(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n \tag{2.1}$$

NOTES

is the interpolating polynomial of degree $\leq n$, then

$$\varphi(x_i) = y_i, \quad \text{for} \quad i = 0, 1, 2, ..., \quad n \qquad (2.2)$$

It is true that, in general, it is difficult to guess the type of function to approximate $f(x)$. In case of periodic functions, the approximation can be made by a finite series of trigonometric functions. Polynomial interpolation is a very useful method for functional approximation. The interpolating polynomial is also useful as a basis to develop methods for other problems such as numerical differentiation, numerical integration and solution of initial and boundary value problems associated with differential equations.

The following theorem, developed by Weierstrass, gives the justification for approximation of the unknown function by a polynomial.

Theorem 2.1: Every function which is continuous in an interval (a, b) can be represented in that interval by a polynomial to any desired accuracy. In other words, it is possible to determine a polynomial $P(x)$ such that $|f(x) - P(x)| < \varepsilon$, for every x in the interval (a, b) where ε is any prescribed small quantity. Geometrically, it may be interpreted that the graph of the polynomial $y = P(x)$ is confined to the region bounded by the curves $y = f(x) - \varepsilon$ and $y = f(x) + \varepsilon$ for all values of x within (a, b), however small ε may be.

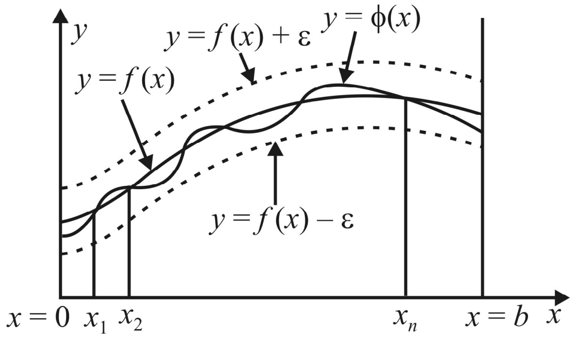


Fig. 2.1 *Interpolation*

The following theorem is regarding the uniqueness of the interpolating polynomial.

Theorem 2.2: For a real-valued function $f(x)$ defined at $(n + 1)$ distinct points $x_0, x_1, ..., x_n$, there exists exactly one polynomial of degree $\leq n$ which interpolates $f(x)$ at $x_0, x_1, ..., x_n$.

We know that a polynomial $P(x)$ which has $(n + 1)$ distinct roots $x_0, x_1, ..., x_n$ can be written as,

$$P(x) = (x - x_0)(x - x_1) .....(x - x_n)\, q(x)$$

where $q(x)$ is a polynomial whose degree is either 0 or $(n + 1)$ which is less than the degree of $P(x)$.

Suppose that two polynomials $\varphi(x)$ and $\psi(x)$ are of degree $\leq n$ and that both interpolate $f(x)$. Here $P(x) = \varphi(x) - \psi(x)$ at $x = x_0, x_1, ..., x_n$. Then $P(x)$ vanishes at the $n + 1$ points $x_0, x_1, ..., x_n$. Thus $P(x) = 0$ and $\phi(x) = \psi(x)$.

NOTES

2.2.1 Iterative Linear Interpolation

In this method, we successively generate interpolating polynomials, of any degree, by iteratively using linear interpolating functions.

Let $p_{01}(x)$ denote the linear interpolating polynomial for the tabulated values at x_0 and x_1. Thus, we can write as,

$$p_{01}(x) = \frac{(x_1 - x)f_0 - (x_0 - x)f_1}{x_1 - x_0}$$

This can be written with determinant notation as,

$$p_{01}(x) = \frac{\begin{vmatrix} f_0 & x_0 - x \\ f_1 & x_1 - x \end{vmatrix}}{x_1 - x_0} \qquad (2.3)$$

This form of $p_{01}(x)$ is easy to visualize and is convenient for desk computation. Thus, the linear interpolating polynomial through the pair of points (x_0, f_0) and (x_j, f_j) can be easily written as,

$$p_{0j}(x) = \frac{1}{x_j - x_0}\begin{vmatrix} f_0 & x_0 - x \\ f_j & x_j - x \end{vmatrix}, \text{ for } \quad j = 1, 2, ..., n \qquad (2.4)$$

Now, consider the polynomial denoted by $p_{01j}(x)$ and defined by,

$$p_{01j}(x) = \frac{1}{x_j - x_1}\begin{vmatrix} p_{01}(x) & x_1 - x \\ p_{0j}(x) & x_j - x \end{vmatrix}, \text{ for } \quad j = 2, 3, ..., n \qquad (2.5)$$

The polynomial $p_{01j}(x)$ interpolates $f(x)$ at the points x_0, x_1, x_j $(j > 1)$ and is a polynomial of degree 2, which can be easily verified that,

$p_{0ij}(x_0) = f_0$, $p_{0ij}(x_i) = f_i$ and $p_{0ij}(x_j) = f_j$ because $p_{01}(x_0) = f_0 = p_{0ij}(x_0)$, etc.

Similarly, the polynomial $p_{012j}(x)$ can be constructed by replacing $p_{01}(x)$ by $p_{012}(x)$ and $p_{0j}(x)$ by $p_{01j}(x)$.

Thus,

$$p_{012j}(x) = \frac{1}{x_j - x_2}\begin{vmatrix} p_{012}(x) & x_2 - x \\ p_{01j}(x) & x_j - x \end{vmatrix}, \text{ for } \quad j = 3, 4, ..., n \qquad (2.6)$$

Evidently, $p_{012j}(x)$ is a polynomial of degree 3 and it interpolates the function at x_0, x_1, x_2 and x_j.

i.e., $\quad p_{012j}(x_0) = f_0$; $p_{012j}(x_i) = f_1$; $p_{012j}(x_2) = f_2$ and $p_{012j}(x_j) = f_j$

This process can be continued to generate higher and higher degree interpolating polynomials.

The results of the iterated linear interpolation can be conveniently represented as given in the following table.

x_k	f_k	p_{0j}	p_{01j}	...	$x_j - x$
x_0	f_0				$x_0 - x$
x_1	f_1	p_{01}			$x_1 - x$
x_2	f_2	p_{02}	p_{012}		$x_2 - x$
x_3	f_3	p_{03}	p_{013}		$x_3 - x$
...	...	...	...	...	...
x_j	f_j	p_{0j}	p_{01j}		$x_j - x$
...	...	...	...	...	...
x_n	f_n	p_{0n}	p_{01n}		$x_n - x$

The successive columns of interpolation results can be conveniently filled by computing the values of the determinants written using the previous column and the corresponding entries in the last column $x_j - x$. Thus, for computing p_{01j}'s for $j = 2, 3, ..., n$, we evaluate the determinant whose elements are the boldface quantities and divide the determinant's value by the difference $(x_j - x) - (x_1 - x)$.

Example 2.1: Find $s(2.12)$ using the following table by iterative linear interpolation:

x	2.0	2.1	2.2	2.3
$s(x)$	0.7909	0.7875	0.7796	0.7673

Solution: Here, $x = 2.12$. The following table gives the successive iterative linear interpolation results. The details of the calculations are shown below in the table.

x_j	$s(x_j)$	p_{0j}	p_{01j}	p_{012j}	$x_j - x$
2.0	0.7909				−0.12
2.1	0.7875	**0.78682**			−0.02
2.2	0.7796	0.78412	**0.78628**		0.08
2.3	0.7673	0.78146	0.78628	**0.78628**	0.18

$$p_{01} = \frac{1}{2.1 - 2.0}\begin{vmatrix} 0.7909 & -0.12 \\ 0.7875 & -0.02 \end{vmatrix} = 0.78682$$

$$p_{02} = \frac{1}{2.2 - 2.0}\begin{vmatrix} 0.7909 & -0.12 \\ 0.7796 & -0.08 \end{vmatrix} = 0.78412$$

$$p_{03} = \frac{1}{2.3 - 2.0}\begin{vmatrix} 0.7909 & -0.12 \\ 0.7673 & 0.18 \end{vmatrix} = 0.78146$$

$$p_{012} = \frac{1}{2.2 - 2.1}\begin{vmatrix} 0.78682 & -0.02 \\ 0.78412 & 0.08 \end{vmatrix} = 0.78628$$

$$p_{013} = \frac{1}{2.3 - 2.1}\begin{vmatrix} 0.78682 & -0.02 \\ 0.78146 & 0.18 \end{vmatrix} = 0.78628$$

$$p_{012} = \frac{1}{2.3 - 2.2}\begin{vmatrix} 0.78628 & 0.08 \\ 0.78628 & 0.18 \end{vmatrix} = 0.78628$$

The boldfaced results in the table give the value of the interpolation at x = 2.12. The result 0.78682 is the value obtained by linear interpolation. The result 0.78628 is obtained by quadratic as well as by cubic interpolation. We conclude that there is no improvement in the third degree polynomial over that of the second degree.

Notes **1.** Unlike Lagrange's methods, it is not necessary to find the degree of the interpolating polynomial to be used.

2. The approximation by a higher degree interpolating polynomial may not always lead to a better result. In fact it may be even worse in some cases.

Consider, the function $f(x) = 4$.

We form the finite difference table with values for $x = 0$ to 4.

x	$f(x)$	$\Delta f(x)$	$\Delta^2 f(x)$	$\Delta^3 f(x)$	$\Delta^4 f(x)$
0	1				
		3			
1	4		9		
		12		27	
2	16		36		81
		48		108	
3	64		144		
		192			
4	256				

Newton's forward difference interpolating polynomial is given below by taking $x_0 = 0$,

$$u = \frac{x - x_0}{h} = x, \quad \varphi(x) = 1 + 3x + \frac{9}{2}x(x-1) + \frac{27}{6}x(x-1)(x-2) + \frac{81}{24}x(x-1)(x-2)(x-3)$$

Now, consider values of $\varphi(x)$ at $x = 0.5$ by taking successively higher and higher degree polynomials.

Thus,

$$\varphi_1(0.5) = 1 + 0.5 \times 3 = 2.5, \quad \text{by linear interpolation}$$

$$\varphi_2(0.5) = 2.5 + \frac{0.5 \times (-0.5)}{2} \times 9 = 1.375, \quad \text{by quadratic interpolation}$$

$$\varphi_3(0.5) = 1.375 + \frac{0.5 \times (-0.5) \times (-1.5)}{6} \times 27 = 3.0625, \quad \text{by cubic interpolation}$$

$$\varphi_4(0.5) = 3.0625 + \frac{(0.5)(-0.5)(-1.5)(-2.5)}{24} \times 81 = -0.10156, \quad \text{by quartic interpolation}$$

We note that the actual value $4^{0.5} = 2$ is not obtainable by interpolation. The results for higher degree interpolating polynomials become worse.

Note: Lagrange's interpolation formula and iterative linear interpolation can easily be implemented for computations by a digital computer.

Example 2.2: Determine the interpolating polynomial for the following table of data:

x	1	2	3	4
y	-1	-1	1	5

NOTES

Solution: The data is equally spaced. We thus form the finite difference table.

x	y	Δy	$\Delta^2 y$
1	-1		
		0	
2	-1		2
		2	
3	1		2
		4	
4	5		

Since the differences of second order are constant, the interpolating polynomial is of degree two. Using Newton's forward difference interpolation, we get

$$y = y_0 + u\Delta y_0 + \frac{u(u-1)}{2!}\Delta^2 y_0,$$

Here, $x_0 = 1, \quad u = x - 1.$

Thus, $y = -1 + (x-1) \times 0 + \dfrac{(x-1)(x-2)}{2} \times 2 = x^2 - 3x + 1.$

Example 2.3: Compute the value of $f(7.5)$ by using suitable interpolation on the following table of data.

x	3	4	5	6	7	8
$f(x)$	28	65	126	217	344	513

Solution: The data is equally spaced. Thus for computing $f(7.5)$, we use Newton's backward difference interpolation. For this, we first form the finite difference table as shown below.

x	$f(x)$	$\Delta f(x)$	$\Delta^2 f(x)$	$\Delta^3 f(x)$
3	28			
		37		
4	65		24	
		61		6
5	126		30	
		91		6
6	217		36	
		127		6
7	344		42	
		169		
8	513			

The differences of order three are constant and hence we use Newton's backward difference interpolating polynomial of degree three.

$$f(x) = y_n + v\ \nabla y_n + \frac{v(v+1)}{2\ !}\nabla^2 y_n + \frac{v(v+1)(v+2)}{3\ !}\nabla^3 y_n,$$

$$v = \frac{x - x_n}{h},\ \text{for}\ \ x = 7.5, \quad x_n = 8$$

NOTES

$$\therefore \qquad v = \frac{7.5-8}{1} = -0.5$$

$$f(7.5) = 513 - 0.5\times 169 + \frac{(-0.5)\ \ (-0.5+1)}{2}\times 42 + \frac{-0.5\times 0.5\times 1.5}{6}\times 6$$

$$= 513 - 84.5 - 5.25 - 0.375$$

$$= 422.875$$

Example 2.4: Determine the interpolating polynomial for the following data:

x	2	4	6	8	10
$f(x)$	5	10	17	29	50

Solution: The data is equally spaced. We construct the Newton's forward difference interpolating polynomial. The finite difference table is,

x	$f(x)$	$\Delta f(x)$	$\Delta^2 f(x)$	$\Delta^3 f(x)$	$\Delta^4 f(x)$
2	5				
		5			
4	10		2		
		7		3	
6	17		5		1
		12		4	
8	29		9		
		21			
10	50				

Here, $x_0 = 2$, $u = (x-x_0)/h = (x-2)/2$.

The interpolating polynomial is,

$$f(x) = f(x_0) + u\,\Delta f(x_0) + \frac{u(u-1)}{2!}\Delta^2 f(x_0) + \ldots$$

$$= 5 + \frac{x-2}{2}\times 5 + \frac{x-2}{2}\left(\frac{x-2}{2}-1\right)\frac{2}{2!} + \frac{x-2}{2}\left(\frac{x-2}{2}-1\right)\left(\frac{x-2}{2}-2\right)\frac{3}{3!}$$

$$+ \frac{x-2}{2}\left(\frac{x-2}{2}-1\right)\left(\frac{x-2}{2}-2\right)\left(\frac{x-2}{2}-3\right)\frac{1}{4!}$$

$$= \frac{1}{384}(x^4 + 4x^3 - 52x^2 + 1040x)$$

Example 2.5: Find the interpolating polynomial which takes the following values:

$y(0) = 1, y(0.1) = 0.9975,\ y(0.2) = 0.9900, y(0.3) = 0.9980$. Hence compute $y\,(0.05)$.

Solution: The data values of x are equally spaced we form the finite difference table,

x	y	Δy	$\Delta^2 y$	$\Delta^3 y$
0.0	1.0000			
		− 25		
0.1	0.9975		− 50	
		− 75		25
0.2	0.9900		− 25	
		−100		
0.3	0.9800			

Here, $h = 0.1$. Choosing $x_0 = 0.0$, we have $s = \frac{x}{0.1} = 10x.$ Newton's forward difference interpolation formula is,

$$y = y_0 + s\ \Delta y_0 + \frac{s(s-1)}{2!}\Delta^2 y_0 + \frac{s(s-1)(s-2)}{3!}\Delta^3 y_0$$

$$= 1 + 10x(-0.0025) + \frac{10x(10x-1)}{2!}(-0.0050) + \frac{10x(10x-1)(10x-2)}{6}\times 0.0025$$

$$= 1.0 - 0.25x - 0.25x^2 + 0.25x + \frac{2.5}{6}x^3 - \frac{300}{4}\times 0.0025x^2 + \frac{0.025}{6}x$$

$$= 1.0 + 0.004x - 0.375x^2 + 0.421x^3$$

$$y(0.05) = 1.0002$$

Example 2.6: Compute $f(0.23)$ and $f(0.29)$ by using suitable interpolation formula with the table of data given below.

x	0.20	0.22	0.24	0.26	0.28	0.30
$f(x)$	1.6596	1.6698	1.6804	1.6912	1.7024	1.7139

Solution: The data being equally spaced, we use Newton's forward difference interpolation for computing $f(0.23)$, and for computing $f(0.29)$, we use Newton's backward difference interpolation. We first form the finite difference table,

x	$f(x)$	$\Delta f(x)$	$\Delta^2 f(x)$
0.20	1.6596		
		102	
0.22	1.6698		4
		106	
0.24	1.6804		2
		108	
0.26	1.6912		4
		112	
0.28	1.7024		3
		115	
0.30	1.7139		

We observe that differences of order higher than two would be irregular. Hence, we use second degree interpolating polynomial. For computing $f(0.23)$, we take $x_0 = 0.22$ so that $u = \frac{x - x_0}{h} = \frac{0.23 - 0.22}{0.02} = 0.5.$

Using Newton's forward difference interpolation, we compute

$$f(0.23) = 1.6698 + 0.5\times 0.0106 + \frac{(0.5)(0.5-1.0)}{2}\times 0.0002$$

$$= 1.6698 + 0.0053 - 0.000025$$

$$= 1.675075$$

$$\approx 1.6751$$

Again for computing $f(0.29)$, we take $x_n = 0.30$,

so that $v = \frac{x - x_n}{n} = \frac{0.29 - 0.30}{0.02} = -0.5$

NOTES

Using Newton's backward difference interpolation we evaluate,

$$f(0.29) = 1.7139 - 0.5 \times .0115 + \frac{(-0.5)(-0.5+1.0)}{2} \times 0.0003$$

$$= 1.7139 - 0.00575 - 0.00004$$

$$= 1.70811$$

$$\simeq 1.7081$$

Example 2.7: Compute values of e^x at $x = 0.02$ and at $x = 0.38$ using suitable interpolation formula on the table of data given below.

x	0.0	0.1	0.2	0.3	0.4
e^x	1.0000	1.1052	1.2214	1.3499	1.4918

Solution: The data is equally spaced. We have to use Newton's forward difference interpolation formula for computing e^x at $x = 0.02$, and for computing e^x at $x = 0.38$, we have to use Newton's backward difference interpolation formula. We first form the finite difference table.

x	$y = e^x$	Δy	$\Delta^2 y$	$\Delta^3 y$	$\Delta^4 y$
0.0	1.0000				
		1052			
0.1	1.1052		110		
		1162		13	
0.2	1.2214		123		−2
		1285		11	
0.3	1.3499		134		
		1419			
0.4	1.4918				

For computing $e^{0.02}$, we take $x_0 = 0$

$$\therefore \quad u = \frac{x - x_0}{h} = \frac{0.02 - 0.0}{0.1} = 0.2$$

By Newton's forward difference interpolation formula, we have

$$e^{0.02} = 1.0 + 0.2 \times 0.1052 + \frac{0.2(0.2-1)}{2} \times 0.0110 + \frac{0.2(0.2-1)(0.2-2)}{6} \times 0.0013$$

$$+ \frac{0.2 \quad (0.2-1)(0.2-2)(0.2-3)}{24} \times -0.0002$$

$$= 1.0 + .02104 - 0.00088 + 0.00006 + 0.00001$$

$$= 1.02023 \simeq 1.0202$$

For computing $e^{0.38}$ we take $x_n = 0.4$. Thus, $v = \frac{0.38 - 0.4}{0.1} = -0.2$

By Newton's backward difference interpolation formula, we have

$$e^{0.38} = 1.4918 + (-0.2) \times 0.1419 + \frac{(-0.2)(-0.2+1)}{2} \times 0.0134$$

$$+ \frac{(-0.2)(-0.2+1)(-0.2+2)}{6} \times 0.0011 + \frac{-0.2(-0.2+1)(-0.2+2)(-0.2+3)}{24} \times (-0.0002)$$

$$= 1.4918 - 0.02838 - 0.00107 - 0.00005 - 0.00001$$

$$= 1.49287 - 0.02844$$

$$= 1.46443 \approx 1.4644$$

2.2.2 Lagrange's Interpolation

Lagrange's interpolation is useful for unequally spaced tabulated values. Let $y = f(x)$ be a real valued function defined in an interval (a, b) and let $y_0, y_1, ..., y_n$ be the $(n + 1)$ known values of y at $x_0, x_1, ..., x_n$, respectively. The polynomial $\varphi(x)$, which interpolates $f(x)$, is of degree less than or equal to n. Thus,

$$\varphi(x_i) = y_i, \quad \text{for} \quad i = 0, 1, 2, ..., n \tag{2.7}$$

The polynomial $\varphi(x)$ is assumed to be of the form,

$$\varphi(x) = \sum_{i=0}^{n} l_i(x) y_i \tag{2.8}$$

where each $l_i(x)$ is a polynomial of degree $\leq n$ in x and is called Lagrangian function.

Now, $\varphi(x)$ satisfies Equation (2.7) if each $l_i(x)$ satisfies,

$$\begin{aligned} l_i(x_j) &= 0 \quad \text{when} \quad i \neq j \\ &= 1 \quad \text{when} \quad i = j \end{aligned} \tag{2.9}$$

Equation (2.9) suggests that $l_i(x)$ vanishes at the $(n+1)$ points $x_0, x_1, ... x_{i-1}, x_{i+1}, ..., x_n$. Thus, we can write,

$$l_i(x) = c_i (x - x_0)(x - x_1) ... (x - x_{i-1})(x - x_{i+1})...(x - x_n)$$

where c_i is a constant given by $l_i(x_i) = 1$,

i.e., $c_i \ (x_i - x_0) \ (x_i - x_1)...(x_i - x_{i-1}) \ (x_i - x_{i+1})... \ (x_i - x_n) = 1$

Thus, $$l_i(x) = \frac{(x - x_0)(x - x_1)...(x - x_{i-1})(x - x_{i+1})...(x - x_n)}{(x_i - x_0)(x_i - x_1)...(x_i - x_{i-1})(x_i - x_{i+1})...(x_i - x_n)} \text{ for } i = 0, 1, 2, ..., n \tag{2.10}$$

Equations (2.8) and (2.10) together give Lagrange's interpolating polynomial.

Algorithm: To compute $f(x)$ by Lagrange's interpolation.

Step 1: Read n [n being the number of values]

Step 2: Read values of x_i, f_i for $i = 1, 2, ..., n$.

Step 3: Set sum = 0, $i = 1$

Step 4: Read x [x being the interpolating point]

Step 5: Set $j = 1$, product = 1

Step 6: Check if $j \neq i$, product = product $\times (x - x_j)/(x_i - x_j)$ else go to Step 7

Step 7: Set $j = j + 1$

Step 8: Check if $j > n$, then go to Step 9 else go to Step 6

Step 9: Compute sum = sum + product $\times f_i$

Step 10: Set $i = i + 1$

Step 11: Check if $i > n$, then go to Step 12

else go to Step 5

Step 12: Write x, sum

NOTES

Example 2.8: Compute $f(0.4)$ for the table below by Lagrange's interpolation.

x	0.3	0.5	0.6
$f(x)$	0.61	0.69	0.72

Solution: The Lagrange's interpolation formula gives,

$$f(0.4) = \frac{(0.4-0.5)(0.4-0.6)}{(0.3-0.5)(0.3-0.6)} \times 0.61 + \frac{(0.4-0.3)(0.4-0.6)}{(0.5-0.3)(0.5-0.6)} \times 0.69 + \frac{(0.4-0.3)(0.4-0.5)}{(0.6-0.3)(0.6-0.5)} \times 0.72$$

$$= 0.203 + 0.69 - 0.24 = 0.653 \simeq 0.65$$

Thus, $f(0.4) = 0.65$.

Example 2.9: Using Lagrange's formula, find the value of $f(0)$ from the table given below.

x	-1	-2	2	4
$f(x)$	-1	-9	11	69

Solution: Using Lagrange's interpolation formula, we find

$$f(0) = \left[\frac{(0+2)(0-2)(0-4)}{(-1+2)(-1-2)(-1-4)} \times (-1)\right] + \left[\frac{(0+1)(0-2)(0-4)}{(-2+1)(-2-2)(-2-4)} \times (-9)\right]$$

$$+ \left[\frac{(0+1)(0+2)(0-4)}{(2+1)(2+2)(2-4)} \times 11\right] + \left[\frac{(0+1)(0+2)(0-2)}{(4+1)(4+2)(4-2)} \times 69\right]$$

$$= -\frac{16}{15} + \frac{9}{3} + \frac{11}{3} - \frac{69}{15} = \frac{20}{3} - \frac{85}{15}$$

$$= \frac{20}{3} - \frac{17}{3} = 1$$

Example 2.10: Determine the interpolating polynomial of degree three for the table given below.

x	-1	0	1	2
$f(x)$	1	1	1	-3

Solution: We have Lagrange's third degree interpolating polynomial as,

$$f(x) = \sum_{i=0}^{3} l_i(x) f(x_i)$$

where

$$l_0(x) = \frac{(x-0)(x-1)(x-2)}{(-1-0)(-1-1)(-1-2)} = -\frac{1}{6}x(x-1)(x-2)$$

$$l_1(x) = \frac{(x+1)(x-1)(x-2)}{(0+1)(0-1)(0-2)} = \frac{1}{2}(x+1)(x-1)(x-2)$$

$$l_2(x) = \frac{(x+1)(x-0)(x-2)}{(1+1)(1-0)(1-2)} = -\frac{1}{2}(x+1)x(x-2)$$

$$l_3(x) = \frac{(x+1)(x-0)(x-1)}{(2+1)(2-0)(2-1)} = \frac{1}{6}(x+1)\,x\,(x-2)$$

$$f(x) = -\frac{1}{6}x(x-1)(x-2)\times 1 + \frac{1}{2}(x+1)(x-1)(x-2)\times 1 - \frac{1}{2}(x+1)x(x-2)\times 1 + \frac{1}{6}(x+1)x(x-2)\times(-3)$$

$$= -\frac{1}{6}(4x^3 - 4x - 6)$$

$$= \frac{-1}{3}(2x^3 - 2x - 3)$$

NOTES

Example 2.11: Evaluate the values of $f(2)$ and $f(6.3)$ using Lagrange's interpolation formula for the table of values given below.

x	1.2	2.5	4	5.1	6	6.5
$f(x)$	6.84	14.25	27	39.21	51	58.25

Solution: It is not advisable to use a higher degree interpolating polynomial. For evaluation of $f(2)$ we take a second degree polynomial using the values of $f(x)$ at the points $x_0 = 1.2$, $x_1 = 2.5$ and $x_2 = 4$.

Thus,

$$f(2) = l_0(2) \times 6.84 + l_1(2) \times 14.25 + l_2(2) \times 27$$

where

$$l_0(2) = \frac{(2-2.5)(2-4)}{(1.2-2.5)(1.2-4)} = 0.275$$

$$l_1(2) = \frac{(2-1.2)(2-4)}{(2.5-1.2)(2.5-4)} = 0.821$$

$$l_2(2) = \frac{(2-1.2)(2-2.5)}{(4-1.2)(4-2.5)} = -0.095$$

$$\therefore f(2) = 0.275 \times 6.84 + 0.821 \times 14.25 - 0.095 \times 27 = 11.015 \simeq 11.02$$

For evaluation of $f(6.3)$, we consider the values of $f(x)$ at $x_0 = 5.1$, $x_1 = 6.0$, $x_2 = 6.5$.

Thus, $f(6.3) = l_0(6.3) \times 39.21 + l_1(6.3) \times 51 + l_2(6.3) \times 58.25$

where

$$l_0(6.3) = \frac{(6.3-6.0)(6.3-6.5)}{(5.1-6.0)(5.1-6.5)} = -0.048$$

$$l_1(6.3) = \frac{(6.3-5.1)(6.3-6.5)}{(6-5.1)(6.0-6.5)} = 0.533$$

$$l_2(6.3) = \frac{(6.3-5.1)(6.3-6.0)}{(6.5-5.1)(6.5-6.0)} = 0.514$$

$$\therefore f(6.3) = -0.048 \times 39.21 + 0.533 \times 51 + 0.514 \times 58.25$$

$$= 55.241 \cong 55.24$$

Since, the computed result cannot be more accurate than the data, the final result is rounded-off to the same number of decimals as the data. In some cases, a higher degree interpolating polynomial may not lead to better results.

2.2.3 Finite Difference for Interpolation

For interpolation of an unknown function when the tabular values of the argument x are equally spaced, we have two important interpolation formulae, viz., (*i*) Newton's forward difference interpolation formula

(*ii*) Newton's backward difference interpolation formula

NOTES

We will first discuss the finite differences which are used in evaluating the above two formulae.

Finite Differences

Let us assume that values of a function $y = f(x)$ are known for a set of equally spaced values of x given by $\{x_0, x_1, ..., x_n\}$, such that the spacing between any two consecutive values is equal. Thus, $x_1 = x_0 + h$, $x_2 = x_1 + h, ..., x_n = x_{n-1} + h$, so that $x_i = x_0 + ih$ for $i = 1, 2, ..., n$. We consider two types of differences known as forward differences and backward differences of various orders. These differences can be tabulated in a finite difference table as explained in the subsequent sections.

Forward Differences

Let $y_0, y_1, ..., y_n$ be the values of a function $y = f(x)$ at the equally spaced values of $x = x_0, x_1, ..., x_n$. The differences between two consecutive y given by $y_1 - y_0, y_2 - y_1, ..., y_n - y_{n-1}$ are called the first order forward differences of the function $y = f(x)$ at the points $x_0, x_1, ..., x_{n-1}$. These differences are denoted by,

$$\Delta y_0 = y_1 - y_0,\quad \Delta y_1 = y_2 - y_1,\ \ ...,\ \Delta y_{n-1} = y_n - y_{n-1} \tag{2.11}$$

where Δ is termed as the forward difference operator defined by,

$$\Delta f(x) = f(x+h) - f(x) \tag{2.12}$$

Thus, $\Delta y_i = y_{i+1} - y_i$, for $i = 0, 1, 2, ..., n - 1$, are the first order forward differences at x_i.

The differences of these first order forward differences are called the second order forward differences.

Thus,
$$\begin{aligned}\Delta^2 y_i &= \Delta(\Delta y_i)\\ &= \Delta y_{i+1} - \Delta y_i,\ \text{for } i = 0, 1, 2, ..., n-2\end{aligned} \tag{2.13}$$

Evidently,

$$\Delta^2 y_0 = \Delta y_1 - \Delta y_0 = y_2 - y_1 - (y_1 - y_0) = y_2 - 2y_1 + y_0$$

And,
$$\Delta^2 y_i = y_{i+2} - y_{i+1} - (y_{i+1} - y_i)$$

i.e.,
$$\Delta^2 y_i = y_{i+2} - 2y_{i+1} + y_i,\quad \text{for}\quad i = 0, 1, 2, ..., n-2 \tag{2.14}$$

Similarly, the third order forward differences are given by,

$$\Delta^3 y_i = \Delta^2 y_{i+1} - \Delta^2 y_i,\ \ \text{for}\ \ i = 0, 1, 2, ..., n-3$$

i.e.,
$$\Delta^3 y_i = y_{i+3} - 3y_{i+2} + 3y_{i+1} - y_i \tag{2.15}$$

NOTES

Finally, we can define the nth order forward difference by,

$$\Delta^n y_0 = y_n - ny_{n-1} + \frac{n(n-1)}{2!} y_{n-2} + ... + (-1)^n y_0 \quad (2.16)$$

The coefficients in above equations are the coefficients of the binomial expansion $(1 - x)^n$.

The forward differences of various orders for a table of values of a function $y = f(x)$, are usually computed and represented in a diagonal difference table. A diagonal difference table for a table of values of $y = f(x)$, for six points $x_0, x_1, x_2, x_3, x_4, x_5$ is shown here.

Diagonal difference Table for $y = f(x)$:

i	x_i	y_i	Δy_i	$\Delta^2 y_i$	$\Delta^3 y_i$	$\Delta^4 y_i$	$\Delta^5 y_i$
0	x_0	y_0					
			Δy_0				
1	x_1	y_1		$\Delta^2 y_0$			
			Δy_1		$\Delta^3 y_0$		
2	x_2	y_2		$\Delta^2 y_1$		$\Delta^4 y_0$	
			Δy_2		$\Delta^3 y_1$		$\Delta^5 y_0$
3	x_3	y_3		$\Delta^2 y_2$		$\Delta^4 y_1$	
			Δy_3		$\Delta^3 y_2$		
4	x_4	y_4		$\Delta^2 y^3$			
			Δy_4				
5	x_5	y_5					

The entries in any column of the differences are computed as the differences of the entries of the previous column and one placed in between them. The upper data in a column is subtracted from the lower data to compute the forward differences. We notice that the forward differences of various orders with respect to y_i are along the forward diagonal through it. Thus Δy_0, $\Delta^2 y_0$, $\Delta^3 y_0$, $\Delta^4 y_0$ and $\Delta^5 y_0$ lie along the top forward diagonal through y_0. Consider the following example.

Example 2.12: Given the table of values of $y = f(x)$,

x	1	3	5	7	9
y	8	12	21	36	62

form the diagonal difference table and find the values of $\Delta f(5)$, $\Delta^2 f(3)$, $\Delta^3 f(1)$.

Solution: The diagonal difference table is,

i	x_i	y_i	Δy_i	$\Delta^2 y_i$	$\Delta^3 y_i$	$\Delta^4 y_i$
0	1	8				
			4			
1	3	12		5		
			9		1	
2	5	21		6		4
			15		5	
3	7	36		11		
			26			
4	9	62				

NOTES

From the table, we find that $\Delta f(5) = 15,$ the entry along the diagonal through the entry 21 of $f(5)$.

Similarly, $\Delta^2 f(3) = 6,$ the entry along the diagonal through $f(3)$. Finally, $\Delta^3 f(1) = 1.$

Backward Differences

The backward differences of various orders for a table of values of a function $y = f(x)$ are defined in a manner similar to the forward differences. The backward difference operator ∇ (inverted triangle) is defined by $\nabla f(x) = f(x) - f(x-h).$

Thus, $\nabla y_k = y_k - y_{k-1}, \text{ for } k = 1, 2, ..., n$

i.e., $\nabla y_1 = y_1 - y_0, \nabla y_2 = y_2 - y_1, ..., \nabla y_n = y_n - y_{n-1}$ (2.17)

The backward differences of second order are defined by,

$$\nabla^2 y_k = \nabla y_k - \nabla y_{k-1} = y_k - 2y_{k-1} + y_{k-2}$$

Hence,

$$\nabla^2 y_2 = y_2 - 2y_1 + y_0, \text{ and } \nabla^2 y_n = y_n - 2y_{n-1} + y_{n-2} \quad (2.18)$$

Higher order backward differences can be defined in a similar manner.

Thus, $\nabla^3 y_n = y_n - 3y_{n-1} + 3y_{n-2} - y_{n-3},$ etc. (2.19)

Finally,

$$\nabla^n y_n = y_n - ny_{n-1} + \frac{n(n-1)}{2\ i} y_{n-2} - ... + (-1)^n y_0 \quad (2.20)$$

The backward differences of various orders can be computed and placed in a diagonal difference table. The backward differences at a point are then found along the backward diagonal through the point. The following table shows the backward differences entries.

Diagonal difference Table of backward differences:

i	x_i	y_i	∇y_i	$\nabla^2 y_i$	$\nabla^3 y_i$	$\nabla^4 y_i$	$\nabla^5 y_i$
0	x_0	y_0					
			∇y_1				
1	x_1	y_1		$\nabla^2 y_2$			
			∇y_2		$\nabla^3 y_3$		
2	x_2	y_2		$\nabla^2 y_3$		$\nabla^4 y_4$	
			∇y_3		$\nabla^3 y_4$		
3	x_3	y_3		$\nabla^2 y_4$		$\nabla^4 y_5$	
			∇y_4		$\nabla^3 y_5$		
4	x_4	y_4		$\nabla^2 y_5$			
			∇y_5				
5	x_5	y_5					

The entries along a column in the table are computed (as discussed in previous example) as the differences of the entries in the previous column and are placed in between. We notice that the backward differences of various orders with respect to y_i are along the backward diagonal through it. Thus, $\nabla y_5, \nabla^2 y_5, \nabla^3 y_5, \nabla^4 y_5$ and $\nabla^5 y_5$ are along the lowest backward diagonal through y_5.

We may note that the data entries of the backward difference table in any column are the same as those of the forward difference table, but the differences are for different reference points.

Specifically, if we compare the columns of first order differences we can see that,

$$\Delta y_0 = \nabla y_1, \Delta y_1 = \nabla y_2, ..., \Delta y_{n-1} = \nabla y_n$$

Hence, $\Delta y_i = \nabla y_{i+1}$, for $i = 0, 1, 2, ..., n-1$

Similarly, $\Delta^2 y_0 = \nabla^2 y_2, \Delta^2 y_1 = \nabla^2 y_3, ..., \Delta^2 y_{n-2} = \nabla^2 y_n$

Thus, $\Delta^2 y_i = \nabla^2 y_{i+2}$, for $i = 1, 2, ..., n-2$

In general, $\Delta^k y_i = \nabla^k y_{i+k}$.

Conversely, $\nabla^k y_i = \Delta^k y_{i-k}$.

Example 2.13: Given the following table of values of $y = f(x)$:

x	1	3	5	7	9
y	8	12	21	36	62

Find the values of $\nabla y_{(7)}, \nabla^2 y_{(9)}, \nabla^3 y_{(9)}$.

Solution: We form the diagonal difference table,

x_i	y_i	∇y_i	$\nabla^2 y_i$	$\nabla^3 y_i$	$\nabla^4 y_i$
1	8				
		4			
3	12		5		
		9		1	
5	21		6		4
		15		5	
7	36		11		
		26			
9	62				

From the table, we can easily find $\nabla y_{(7)} = 15$, $\nabla^2 y_{(9)} = 11$, $\nabla^3 y_{(9)} = 5$.

2.2.4 Symbolic Operators

We consider the finite differences of an equally spaced tabular data for developing numerical methods. Let a function $y = f(x)$ has a set of values $y_0, y_1, y_2, ...,$

NOTES

corresponding to points $x_0, x_1, x_2,...,$ where $x_1 = x_0 + h$, $x_2 = x_0 + 2h,....,$ are equally spaced with spacing h. We define different types of finite differences such as forward differences, backward differences and central differences, and express them in terms of operators.

The forward difference of a function $f(x)$ is defined by the operator Δ, called the forward difference operator given by,

$$\Delta f(x) = f(x+h) - f(x) \tag{2.21}$$

At a tabulated point x_i, we have

$$\Delta f(x_i) = f(x_i + h) - f(x_i) \tag{2.22}$$

We also denote $\Delta f(x_i)$ by Δy_i, given by

$$\Delta y_i = y_{i+1} - y_i, \quad \text{for} \quad i = 0, 1, 2, ... \tag{2.23}$$

We also define an operator E, called the shift operator which is given by,

$$E\,f(x) = f(x + h) \tag{2.24}$$

$$\therefore \qquad \Delta f(x) = Ef(x) - f(x)$$

Thus, $\Delta = E - 1$ is an operator relation. (2.25)

While Equation (2.21) defines the first order forward difference, we can define second order forward difference by,

$$\Delta^2 y_i = \Delta(\Delta y_i) = \Delta(y_{i+1} - y_i)$$

$$\therefore \qquad \Delta^2 y_i = \Delta y_{i+1} - \Delta y_i \tag{2.26}$$

2.2.5 Shift Operator

The shift operator is denoted by E and is defined by $E\,f(x) = f(x + h)$. Thus,

$$Ey_k = y_{k+1}$$

Higher order shift operators can be defined by, $E^2 f(x) = Ef(x + h) = f(x + 2h)$.

$$E^2 y_k = E(Ey_k) = E(y_{k+1}) = y_{k+2}$$

In general, $$E^m f(x) = f(x + mh)$$

$$E^m yk = y_{k+m}$$

Relation between forward difference operator and shift operator

From the definition of forward difference operator, we have

$$\begin{aligned}\Delta y(x) &= y(x+h) - y(x)\\ &= Ey(x) - y(x)\\ &= (E-1)y(x)\end{aligned}$$

This leads to the operator relation,

$$\Delta = E - 1$$

or, $$E = 1 + \Delta \qquad (2.27)$$

Similarly, for the second order forward difference, we have

$$\begin{aligned}\Delta^2 y(x) &= \Delta y(x+h) - \Delta y(x)\\ &= y(x+2h) - 2y(x+h) + y(x)\\ &= E^2 y(x) - 2Ey(x) + y(x)\\ &= (E^2 - 2E + 1)y(x)\end{aligned}$$

This gives the operator relation, $\Delta^2 = (E-1)^2$.

Finally, we have $\Delta^m = (E-1)^m$, for $m = 1, 2, \ldots$ (2.28)

Relation between the backward difference operator with shift operator

From the definition of backward difference operator, we have

$$\begin{aligned}\nabla f(x) &= f(x) - f(x-h)\\ &= f(x) - E^{-1} f(x) = (1 - E^{-1})\, f(x)\end{aligned}$$

This leads to the operator relation, $\nabla \equiv 1 - E^{-1}$ (2.29)

Similarly, the second order backward difference is defined by,

$$\begin{aligned}\nabla^2 f(x) &= \nabla f(x) - \nabla f(x-h)\\ &= f(x) - f(x-h) - f(x-h) + f(x-2h)\\ &= f(x) - 2f(x-h) + f(x-2h)\\ &= f(x) - E^{-1} f(x) + E^{-2} f(x)\\ &= (1 - E^{-1} + E^{-2})\, f(x)\\ &= (1 - E^{-1})^2\, f(x)\end{aligned}$$

This gives the operator relation, $\nabla^2 \equiv (1 - E^{-1})^2$ and in general,

$$\nabla^m \equiv (1 - E^{-1})^m \qquad (2.30)$$

Relations between the operators E, D and Δ

We have by Taylor's theorem,

$$f(x+h) = f(x) + hf'(x) + \frac{h^2}{2!} f''(x) + \ldots$$

Thus, $Ef(x) = f(x) + hDf(x) + \frac{h^2 D^2}{2!} f(x) + \ldots$, where $D = \frac{d}{dx}$

or, $$\begin{aligned}(1+\Delta) f(x) &= \left(1 + hD + \frac{h^2 D^2}{2!} + \ldots\right) f(x)\\ &= e^{hD} f(x)\end{aligned}$$

NOTES

Thus, $$e^{hD} = 1 + \Delta = E \qquad (2.31)$$

Also, $$hD = \log \ (1 + \Delta)$$

or, $$hD = \Delta - \frac{\Delta^2}{2} + \frac{\Delta^3}{3} - \frac{\Delta^4}{4} + ...$$

$$\therefore \qquad D = \frac{1}{h} \left(\Delta - \frac{\Delta^2}{2} + \frac{\Delta^3}{3} - \frac{\Delta^4}{4} + ... \right)$$

2.2.6 Central Difference Operator

The central difference operator denoted by δ is defined by,

$$\delta y(x) = y\left(x + \frac{h}{2} \right) - y\left(x - \frac{h}{2} \right)$$

Thus,

$$\delta y(x) = (E^{1/2} - E^{-1/2}) y(x)$$

Giving the operator relation, $\delta = E^{1/2} - E^{-1/2}$ or $\delta E^{1/2} = E - 1$

Also,

$$\delta\, y_n = (E^{1/2} - E^{-1/2}) y(x_n) = E^{1/2} y_n - E^{-1/2} y_n$$

i.e., $$\delta\, y_n = y_{n+1/2} - y_{n-1/2}$$

Further,

$$\delta^2 y_n = \delta\,(\delta\, y_n) = \delta\, y_{n+1/2} - \delta\, y_{n-1/2}$$
$$= \left(E^{1/2} - E^{-1/2}\right) \ \left(y_{n+1/2}\right) - \left(E^{1/2} - E^{-1/2}\right) \ \left(y_{n-1/2}\right)$$
$$= E^{1/2}\left(y_{n+1/2} - y_{n-1/2}\right) - E^{-1/2}\left(y_{n+1/2} - y_{n-1/2}\right)$$
$$= y_{n+1} - y_n - y_n + y_{n-1}$$
$$= y_{n+1} - 2y_n + y_{n-1} = (E^{1/2} - E^{-1/2})^2 y_n \ = \left[\Delta^2 y_{n-1} \equiv \nabla^2 y_{n+1}\right]$$
$$= (E + E^{-1} - 2) y_n$$

$$\therefore \quad \delta^2 \equiv E + E^{-1} - 2 \qquad (2.32)$$

Even though the central difference operator uses fractional arguments, still it is widely used. This is related to the averaging operator and is defined by,

$$\mu = \frac{1}{2}(E^{1/2} + E^{-1/2}) \qquad (2.33)$$

Squaring, $\mu^2 = \frac{1}{4}(E + 2 + E^{-1}) = \frac{1}{4}(\delta^2 + 2 + 2) = 1 + \frac{1}{4}\delta^2$

$$\therefore \mu^2 = 1 + \frac{1}{4}\delta^2 \qquad (2.34)$$

It may be noted that, $\delta\ y_{1/2} = y_1 - y_0 = \nabla y_1$

Also, $\delta\, E^{1/2} y_1 = \delta\, y_{\frac{1}{2}+1} = y_2 - y_1 = \Delta y_1$

$$\therefore\ \delta E^{1/2} = \Delta = E - 1 \qquad (2.35)$$

Further,

$$\delta^3 y_n = \delta(\delta^2 y_n) = \delta\left(\delta y_{n+\frac{1}{2}} - \delta\, y_{n-\frac{1}{2}}\right)$$

$$= \delta^2 y_{n+\frac{1}{2}} - \delta^2 y_{n-\frac{1}{2}} = \delta(y_{n+1} - 2y_n + y_{n-1})$$

Example 2.14: Prove the following operator relations:

(*i*) $\Delta \equiv \nabla E$ (*ii*) $(1+\Delta)\,(1-\nabla) = 1$

Solution:

(*i*) Since, $\Delta f(x) = f(x+h) - f(x) = Ef(x) - f(x)$, $\Delta \equiv E - 1$ (1)

and since $\nabla f(x) = f(x) - f(x-h) = (1 - E^{-1}) f(x)$, $\nabla \equiv 1 - E^{-1}$ (2)

Thus, $\nabla \equiv \frac{E-1}{E}$ or $\nabla E \equiv E - 1 \equiv \Delta$

Hence proved.

(*ii*) From Equation (1), we have $E \equiv \Delta + 1$ (3)

and from Equation (2) we get $E^{-1} \equiv 1 - \nabla$ (4)

Combining Equations (3) and (4), we get $(1+\Delta)(1-\nabla) \equiv 1$.

Example 2.15: If f_i is the value of $f(x)$ at x_i where $x_i = x_0 + ih$, for $i = 1,2,...,$ prove that,

$$f_i = E^i f_o = \sum_{j=0}^{i} \binom{i}{j} \Delta^i f_0$$

Solution: We can write $Ef(x) = f(x+h)$

Using Taylor series expansion, we have

$$Ef(x) = f(x) + hf'(x) + \frac{h^2}{2!} f''(x) + ...$$

$$= f(x) + hDf(x) + \frac{h^2}{2!} D^2 f(x) + ..., \quad \text{where} \quad D = \frac{d}{dx}$$

$$\therefore \qquad (1+\Delta) f(x) = \left(1 + hD + \frac{h^2 D^2}{2!} + ...\right) f(x), \quad \text{since} \quad E = 1 + \Delta$$

$$= e^{hD}.f(x)$$

$$\therefore \qquad 1 + \Delta = e^{hD}$$

Hence, $e^{ihD} = (1+\Delta)^i$

NOTES

Now, $f_i = f(x_i) = f(x_0 + ih) = E^i f(x_0)$

$\therefore$ $f_i = (1+\Delta)^i f(x_0)$, since $E \equiv 1+\Delta$

$$f_i = \sum_{j=0}^{i} \binom{i}{j} \Delta^i f_0, \text{ using binomial expansion.}$$

Hence proved.

Example 2.16: Compute the following differences:

(i) $\Delta^n e^x$ (ii) $\Delta^n x^n$

Solution:

(i) We have, $\Delta e^x = e^{x+h} - e^x = e^x(e^h - 1)$

Again, $\Delta^2 e^x = \Delta(\Delta e^x) = (e^h - 1)\Delta e^x = (e^h-1)^2 e^x$

Thus by induction, $\Delta^n e^x = (e^h - 1)^n e^x$.

(ii) We have,

$$\Delta(x^n) = (x+h)^n - x^n$$
$$= n\,h\,x^{n-1} + \frac{n(n-1)}{2!} h^2 x^{n-2} + ... + h^n$$

Thus, $\Delta(x^n)$ is a polynomial of degree $(n-1)$

Also, $\Delta(h^n) = 0$. Hence, we can say that $\Delta^2(x^n)$ is a polynomial of degree $(n-2)$ with the leading term $n(n-1)h^2 x^{n-2}$.

Proceeding n times, we get

$$\Delta^n(x^n) = n(n-1)...\,1h^n = n!h^n$$

Example 2.17: Prove that,

(i) $\Delta\left\{\dfrac{f(x)}{g(x)}\right\} = \dfrac{g(x)\ \Delta f(x) - f(x)\ \Delta g(x)}{g(x)\, g(x+h)}$

(ii) $\Delta\{\log f(x)\} = \log\left\{1 + \dfrac{\Delta f(x)}{f(x)}\right\}$

Solution:

(i) We have,

$$\Delta\left\{\frac{f(x)}{g(x)}\right\} = \frac{f(x+h)}{g(x+h)} - \frac{f(x)}{g(x)}$$
$$= \frac{f(x+h)\,g(x) - f(x)g(x+h)}{g(x+h)\,g(x)}$$

NOTES

$$= \frac{f(x+h)\,g(x) - f(x)\,g(x) + f(x)\,g(x) - f(x)\,g(x+h)}{g(x+h)\,g(x)}$$

$$= \frac{g(x)\{f(x+h) - f(x)\} - f(x)\{g(x+h) - g(x)\}}{g(x)\,g(x+h)}$$

$$= \frac{g(x)\;\Delta f(x) - f(x)\,\Delta g(x)}{g(x)\,g(x+h)}$$

(*ii*) We have,

$$\Delta\{\log f(x)\} = \log\{f(x+h)\} - \log\{f(x)\}$$

$$= \log\frac{f(x+h)}{f(x)} = \log\left\{\frac{f(x+h) - f(x) + f(x)}{f(x)}\right\}$$

$$= \log\left\{\frac{\Delta f(x)}{f(x)} + 1\right\}$$

2.2.7 Differences of a Polynomial

We now look at the differences of various orders of a polynomial of degree n, given by

$$y = f(x) = a_n x^n + a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + ... + a_1 x + a_0$$

The first order forward difference is defined by,

$\Delta f(x) = f(x+h) - f(x)$ and is given by,

$$\Delta y = a_n\{(x+h)^n - x^n\} + a_{n-1}\{(x+h)^{n-1} - x^{n-1}\} + ... + a_1(x+h-x)$$

$$= a_n\{n\,h\,x^{n-1} + \frac{n(n-1)}{2\;!}h^2 x^{n-2} + ...\} + a_{n-1}\{(n-1)h\,x^{n-2} + ...\}$$

$$= b_{n-1}x^{n-1} + b_{n-2}x^{n-2} + ... + b_1 x + b_0$$

where the coefficients of various powers of x are collected separately.

Thus, the first order difference of a polynomial of degree n is a polynomial of degree $n-1$, with $b_{n-1} = a_n.nh$

Proceeding as above, we can state that the second order forward difference of a polynomial of degree n is a polynomial of degree $n-2$, with coefficient of x^{n-2} as $n(n-1)h^2 a_0$.

Continuing successively, we finally get $\Delta^n y = a_0 n!h^n$, a constant.

We can conclude that for polynomial of degree n, all other differences having order higher than n are zero.

It may be noted that the converse of the above result is partially true and suggests that if the tabulated values of a function are found to be such that the differences of the kth order are approximately constant, then the highest degree of the interpolating polynomial that should be used is k. Since the tabulated data may have round-off errors, the acutal function may not be a polynomial.

NOTES

Example 2.18: Compute the horizontal difference table for the following data and hence, write down the values of $\nabla f(4)$, $\nabla^2 f(3)$ and $\nabla^3 f(5)$.

x	1	2	3	4	5
$f(x)$	3	18	83	258	627

Solution: The horizontal difference table for the given data is as follows:

x	$f(x)$	$\nabla f(x)$	$\nabla^2 f(x)$	$\nabla^3 f(x)$	$\nabla^4 f(x)$
1	3	–	–	–	–
2	18	15	–	–	–
3	83	65	50	–	–
4	258	175	110	60	–
5	627	369	194	84	24

From the table we read the required values and get the following result:

$$\nabla f(4) = 175,\ \nabla^2 f(3) = 50,\ \nabla^3 f(5) = 84$$

Example 2.19: Form the difference table of $f(x)$ on the basis of the following table and show that the third differences are constant. Hence, conclude about the degree of the interpolating polynomial.

x	0	1	2	3	4
$f(x)$	5	6	13	32	69

Solution: The difference table is given below

x	$f(x)$	$\Delta f(x)$	$\Delta^2 f(x)$	$\Delta^3 f(x)$
0	5			
		1		
1	6		6	
		7		6
2	13		12	
		19		6
3	32		18	
		37		
4	69			

It is clear from the above table that the third differences are constant and hence, the degree of the interpolating polynomial is three.

2.2.8 Newton's Forward Difference Interpolation Formula

Newton's forward difference interpolation formula is a polynomial of degree less than or equal to n. This is used to find the value of the tabulated function at a non-tabular point. Consider a function $y = f(x)$ whose values $y_0, y_1, ..., y_n$ at a set of equidistant points $x_0, x_1, ..., x_n$ are known.

NOTES

Let $\varphi(x)$ be the interpolating polynomial, such that

$$\varphi(x_i) = f(x_i) = y_i$$
$$x_i = x_0 + ih, \text{ for } i = 0, 1, 2, ..., n \qquad (2.36)$$

We assume the polynomial $\varphi(x)$ to be of the form,

$$\varphi(x) = a_0 + a_1(x-x_0) + a_2(x-x_0)(x-x_1) + a_3(x-x_0)(x-x_1)(x-x_2) + \\ ... + a_n(x-x_0)(x-x_1)...(x-x_{n-1}) \qquad (2.37)$$

The coefficients a_i's in Equation (2.37) are determined by satisfying the conditions in Equation (2.36) successively for $i = 0, 1, 2,...,n$.

Thus, we get

$$y_0 = \varphi(x_0) = a_0, \quad \text{gives} \quad a_0 = y_0$$

$$y_1 = \varphi(x_1) = a_0 + a_1(x_1 - x_0), \quad \text{gives} \quad a_1 = \frac{y_1 - y_0}{h}$$

$$\therefore \quad a_1 = \frac{\Delta y_0}{h}$$

$$y_2 = \varphi(x_2) = a_0 + a_1(x_2 - x_0) + a_2(x_2 - x_0)(x_2 - x_1)$$

or,

$$y_2 = y_0 + \frac{\Delta y_0}{h} + a_2(2h) \quad h$$

$$a_2 = \frac{y_2 - 2y_1 + y_0}{2h^2} = \frac{\Delta^2 y_0}{2\,!\;h^2}$$

Proceeding further, we get successively,

$$a_3 = \frac{\Delta^3 y_0}{3\,!\,h^3}, ..., a_n = \frac{\Delta^n y_0}{n\,!\,h^n}$$

Using these values of the coefficients, we get Newton's forward difference interpolation in the form,

$$\varphi(x) = y_0 + \frac{x-x_0}{h}\Delta y_0 + \frac{(x-x_0)(x-x_1)}{2\,!\;h^2}\Delta^2 y_0 + \frac{(x-x_0)}{h}\frac{(x-x_1)}{h}\frac{(x-x_2)}{h}\frac{\Delta^3 y_0}{3!} + ...$$
$$+ ... + \frac{(x-x_0)}{h}\frac{(x-x_1)}{h}...\frac{(x-x_{n-1})}{h}\frac{\Delta^n y_0}{n!}$$

This formula can be expressed in a more convenient form by taking $u = \frac{x-x_0}{h}$ as shown here.

We have,

$$\frac{x-x_1}{h} = \frac{x-(x_0+h)}{h} = \frac{x-x_0}{h} - 1 = u - 1$$
$$\frac{x-x_2}{h} = \frac{x-(x_0+2h)}{h} = \frac{x-x_0}{h} - 2 = u - 2$$
$$\frac{x-x_{n-1}}{h} = \frac{x-\{x_0+(n-1)h\}}{h} = \frac{x-x_0}{h} - (n-1) = u - n + 1$$

NOTES

Thus, the interpolating polynomial reduces to:

$$\varphi(u) = y_0 + u \quad \Delta y_0 + \frac{u(u-1)}{2\ !}\Delta^2 y_0 + \frac{u(u-1)(u-2)}{3\ !}\Delta^3 y_0$$

$$+ \dots + \frac{u(u-1)(u-2)\dots(u-n+1)}{n\ !}\Delta^n y_0 \qquad (2.38)$$

This formula is generally used for interpolating near the beginning of the table. For a given x, we choose a tabulated point as x_0 for which the following condition is satisfied.

For better results, we should have

$$|u| = \left|\frac{x - x_0}{h}\right| \le 0.5$$

The degree of the interpolating polynomial to be used is less than or equal to n and is determined by the order of the differences when they are nearly same so that the differences of higher orders are irregular due to the propagated round-off error in the data.

2.2.9 Newton's Backward Difference Interpolation Formula

Newton's forward difference interpolation formula cannot be used for interpolating at a point near the end of a table, since we do not have the required forward differences for interpolating at such points. However, we can use a separate formula known as Newton's backward difference interpolation formula. Let a table of values $\{x_i, y_i\}$, for $i = 0, 1, 2, ..., n$ for equally spaced values of x_i be given. Thus, $x_i = x_0 + ih$, $y_i = f(x_i)$, for $i = 0, 1, 2, ..., n$ are known.

We construct an interpolating polynomial of degree n of the form,

$$y \approx \varphi(x) = b_0 + b_1(x - x_n) + b_2(x - x_n)(x - x_{n-1}) + \dots + b_n(x - x_n)(x - x_{n-1})\dots(x - x_1) \qquad (2.39)$$

We have to determine the coefficients $b_0, b_1, ..., b_n$ by satisfying the relations,

$$\varphi(x_i) = y_i, \quad \text{for} \quad i = n, n-1, n-2, ..., 1, 0 \qquad (2.40)$$

Thus, $\quad \varphi(x_n) = y_n, \quad \text{gives} \quad b_0 = y_n \qquad (2.41)$

Similarly, $\quad \varphi(x_{n-1}) = y_{n-1}, \quad \text{gives} \quad y_{n-1} = b_0 + b_1(x_{n-1} - x_n)$

or, $$b_1 = \frac{y_n - y_{n-1}}{h} = \frac{\nabla y_n}{h} \qquad (2.42)$$

Again

$$\varphi(x_{n-2}) = y_{n-2}, \quad \text{gives} \quad y_{n-2} = b_0 + b_1(x_{n-2} - x_n) + b_2(x_{n-2} - x_n)(x_{n-1} - x_n)$$

or, $$y_{n-2} = y_n + \frac{y_n - y_{n-1}}{h}(-2h) + b_2(-2h)(-h)$$

$$\therefore \quad b_2 = \frac{y_n - 2y_{n-1} + y_{n-2}}{2h^2} = \frac{\nabla^2 y_n}{2!\,h^2} \qquad (2.43)$$

By induction or by proceeding as mentioned earlier, we have

NOTES

$$b_3 = \frac{\nabla^3 y_n}{3\;!\;\;h^3},\;\; b_4 = \frac{\nabla^4 y_n}{4\;!\;\;h^4},\;\ldots,\; b_n = \frac{\nabla^n y_n}{n\;!\;\;h^n} \qquad (2.44)$$

Substituting the expressions for b_i in Equation (2.39), we get

$$\varphi(x) = y_n + \frac{\nabla y_n}{h}(x - x_n) + \frac{\nabla^2 y_n}{2\;!\;\;h^2}(x - x_n)\quad(x - x_{n-1}) + \ldots + \frac{\nabla^n y_n}{n\;!\;\;h^n}(x - x_n)(x - x_{n-1})\ldots(x - x_1)$$

(2.45)

This formula is known as Newton's backward difference interpolation formula. It uses the backward differences along the backward diagonal in the difference table.

Introducing a new variable $v = \dfrac{x - x_n}{h}$,

we have, $\dfrac{x - x_{n-1}}{h} = \dfrac{x - (x_n - h)}{h} = v + 1$.

Similarly, $\dfrac{x - x_{n-2}}{h} = v + 2, \ldots, \dfrac{x - x_1}{h} = v + \overline{n-1}$.

Thus, the interpolating polynomial in Equation (2.45) may be rewritten as,

$$\varphi(x) = y_n + v\nabla y_n + \frac{v(v+1)}{2\;!}\nabla^2 y_n + \frac{v(v+1)(v+2)}{3\;!}\nabla^3 y_n + \ldots + \frac{v(v+1)(v+2)\ldots(v+n-1)}{n\;!}\nabla^n y_n$$

(2.46)

This formula is generally used for interpolation at a point near the end of a table.

The error in the given interpolation formula may be written as,

$$E(x) = f(x) - \varphi(x)$$

$$= \frac{(x - x_n)(x - x_{n-1})\ldots(x - x_1)(x - x_0)\,f^{(n+1)}(\xi)}{(n+1)\;\;!}, \quad \text{where} \quad x_0 < \xi < x_n$$

$$= v(v+1)(v+2)\ldots(v+n)\frac{y^{(n+1)}(\xi)}{(n+1)\;\;!}h^{n+1}$$

2.2.10 Extrapolation

The interpolating polynomials are usually used for finding values of the tabulated function $y = f(x)$ for a value of x within the table. But, they can also be used in some cases for finding values of $f(x)$ for values of x near to the end points x_0 or x_n outside the interval $[x_0, x_n]$. This process of finding values of $f(x)$ at points beyond the interval is termed as extrapolation. We can use Newton's forward difference interpolation for points near the beginning value x_0. Similarly, for points near the end value x_n, we use Newton's backward difference interpolation formula.

Example 2.20: With the help of appropriate interpolation formula, find from the following data the weight of a baby at the age of one year and of ten years:

Age = x	3	5	7	9
Weight = y (kg)	5	8	12	17

NOTES

Solution: Since the values of x are equidistant, we form the finite difference table for using Newton's forward difference interpolation formula to compute weight of the baby at the age of required years.

x	y	Δy	$\Delta^2 y$
3	5		
		3	
5	8		1
		4	
7	12		1
		5	
9	17		

Taking $x = 2,\ u = \dfrac{x - x_0}{h} = -0.5.$

Newton's forward difference interpolation gives,

$$y \text{ at } x = 1,\ y(1) = 5 - 0.5 \times 3 + \frac{(-0.5)(-1.5)}{2} \times 1$$
$$= 5 - 1.5 + 0.38 = 3.88 \simeq 3.9 \text{ kg.}$$

Similarly, for computing weight of the baby at the age of ten years, we use Newton's backward difference interpolation given by,

$$v = \frac{x - x_n}{h} = \frac{10 - 9}{2} = 0.5$$
$$y \text{ at } x = 10,\ y(10) = 17 + 0.5 \times 5 + \frac{0.5 \times 1.5}{2} \times 1$$
$$= 17 + 2.5 + 0.38 \simeq 19.88$$

2.2.11 Inverse Interpolation

The problem of inverse interpolation in a table of values of $y = f(x)$ is to find the value of x for a given y. We know that the inverse function $x = g(y)$ exists and is unique, if $y = f(x)$ is a single valued function of x and $\dfrac{dy}{dx}$ exists and does not vanish in the neighbourhood of the point where inverse interpolation is desired.

When the values of x are unequally spaced, we can apply Lagrange's interpolation or iterative linear interpolation simply by interchanging the roles of x and y. Thus Lagrange's formula for inverse interpolation can be written as,

$$x = \sum_{i=0}^{n} l_i(y) x_i$$

where
$$l_i(y) = \prod_{\substack{j=0 \\ j \neq i}}^{n} [(y - y_j)/(y_i - y_j)]$$

When x values are equally spaced, we can apply the method of successive approximation as described below.

Consider Newton's formula for forward difference interpolation given by,

$$y = y_0 + u\,\Delta y_0 + \frac{u(u-1)}{2!}\Delta^2 y_0 + \frac{u(u-1)(u-2)}{3!}\Delta^3 y_0 + ...$$

Retaining only two terms on the RHS, we can write the first approximation,

$$u^{(1)} = \frac{1}{\Delta\, y_0}(y - y_0)$$

The second approximation can be written as,

$$u^{(2)} = \frac{1}{\Delta y_0}\left[(y - y_0) - \frac{u^{(1)}(u^{(1)}-1)}{2}\Delta^2 y_0\right]$$

on replacing u by $u^{(1)}$ in the coefficient of $\Delta^2 y_0$.

Similarly, the third approximation can be written as,

$$u^{(3)} = \frac{1}{\Delta y_o}\left[y - y_0 - \frac{u^{(2)}(u^{(2)}-1)}{2}\Delta^2 y_0 - \frac{u^{(2)}(u^{(2)}-1)(u^{(2)}-2)}{6}\Delta^3 y_0\right]$$

The process can be continued until two successive approximations have a reasonable accuracy. Then x is obtained by the relation,

$$x = x_0 + uh$$

Example 2.21: Using inverse interpolation, find the value of x for $y = 5$, from the given table.

x	1	3	4
y	3	12	19

Solution: Applying inverse interpolation,

$$x = \sum_{i=0}^{2} l_i(y).x_i$$

Thus, for $y = 5$, we have

$$x = \frac{(5-12)(5-19)}{(3-12)(3-19)}\times 1 + \frac{(5-3)(5-19)}{(12-3)(12-19)}\times 3 + \frac{(5-3)(5-12)}{(19-3)(19-12)}\times 4$$

$$= \frac{7\times 14}{9\times 16} + \frac{2\times -14}{9\times -7}\times 3 + \frac{2\times -7}{16\times 7}\times 4$$

$$= 0.6805 + 1.3333 - 0.5000$$

$$= 1.5138$$

$$= 1.514 \text{ correct upto four significant figures.}$$

Example 2.22: Given the following tabular values of cos*h* x, find x for which cos*h* x = 1.285.

x	0.738	0.739	0.740	0.741	0.742
cos*h* x	1.2849085	1.2857159	1.2865247	1.2873348	1.2881461

NOTES

Solution: Since finding x for an equally spaced table of $\cosh x$ is a problem of inverse interpolation, we employ the method of successive approximation using Newton's formula of inverse interpolation. We first form the finite difference table.

x	$f(x) = \cosh x$	$\Delta f(x)$	$\Delta^2 f(x)$	$\Delta^3 f(x)$
0.738	1.2849085			
		8074		
0.739	1.2857159		14	
		8088		−1
0.740	1.2865247		13	
		8101		−1
0.741	1.2873348		12	
		8113		
0.742	1.2881461			

Using Newton's forward difference interpolation formula for the first approximation $u = \frac{(x - x_0)}{h}$ we get,

$$u^{(1)} = \frac{1}{\Delta f(x_0)}(y - y_0)$$

For, $y = 1.285$, we take $x_0 = 0.739$.

$\therefore$

$$u^{(1)} = \frac{1}{0.0008088} \times (1.285 - 1.2857159) = -0.8851384, \text{ then } x \approx 0.739885$$

For a second approximation,

$$u^{(2)} = u^{(1)} - \frac{1}{\Delta f(x_0)} \frac{u^{(1)}(u^{(1)} - 1)}{2} \Delta^2 f(x_0)$$

$$= -0.8851384 - \frac{1}{0.0008088 \times 2} \times (-0.8851384) \times (-1.8851384) \times 0.0000013$$

$$= -0.8851384 - 0.0013409 = -0.8864793 \Rightarrow x = 0.7398864$$

Similarly,

$$u^{(3)} = u^{(1)} - \frac{u^{(2)}(u^{(2)} - 1)}{2} \frac{\Delta^2 f_0}{\Delta f_0} - \frac{1}{6} u^{(2)}(u^{(2)} - 1)(u^{(2)} - 2) \frac{\Delta^3 f_0}{\Delta f_0}$$

$$= -0.8851384 - 0.0013430 - 0.000073600$$

$$= -0.886555 \Rightarrow x = 0.7398865$$

Example 2.23: Find the divided difference interpolation for the following table of values:

x	4	7	9
$f(x)$	−43	83	327

Solution: We first form the Divided Difference (DD) table as given below.

x	$f(x)$	1st DD	2nd DD
4	−43		
		42	
7	83		16
		122	
9	327		

Newton's divided difference interpolation formula is,

$$f(x) \approx f(x_0) + (x - x_0)\, f(x_0, x_1) + (x - x_0)\,(x - x_1)\, f\left[x_0, \ \ x_1, \ \ x_2\right]$$

$$\therefore f(x) \approx -43 + (x - 4)\, 42 + (x - 4)\,(x - 7) \times 16$$

$$= 16x^2 - 134x + 237$$

Example 2.24: Given the following table of values of the function $y = \log_e x$, construct the Newton's forward difference interpolating polynomial. Comment on the degree of the polynomial and find $\log_e 1001$.

x	1000	1010	1020	1030	1040
$\log_e x$	3.00000	3.00432	3.00860	3.01284	3.01703

Solution: We form the difference table as given below:

x	y	Δy	$\Delta^2 y$
1000	3.00000		
		432	
1010	3.00432		−4
		428	
1020	3.00860		−4
		424	
1030	3.01284		−5
		419	
1040	3.01703		

We observe that, the differences of second order are nearly constant. Thus, the degree of the interpolating polynomial is 2 and is given by,

$$y = y_0 + u\Delta y_0 + \frac{u(u-1)}{2}\Delta^2 y_0, \text{ where } u = \frac{x - x_0}{h}$$

For $x = 1001$, we take $x_0 = 1000$.

$$\therefore u = \frac{1001 - 1000}{10} = 0.1$$

$$\log_e 1001 = 3.00000 + 0.1 \times 0.00432 + \frac{0.1 \times 0.9}{2} \times (-0.00004)$$

$$= 3.000430 \approx 3.00043$$

Example 2.25: Determine the interpolating polynomial for the following data table using both forward and backward difference interpolating formulae. Comment on the result.

x	0	1	2	3	4
$f(x)$	1.0	8.5	36.0	95.5	199.0

NOTES

Solution: Since the data points are equally spaced, we construct the Newton's forward difference interpolating polynomial for which we first form the finite difference table as given below:

x	$f(x)$	$\Delta f(x)$	$\Delta^2 f(x)$	$\Delta^3 f(x)$
0	1.0			
		7.5		
1.0	8.5		20.0	
		27.5		12.0
2.0	36.0		32.0	
		59.5		12.0
3.0	95.5		44.0	
		103.5		
4.0	199.0			

Since the differences of order 3 are constant, we construct the third degree Newton's forward difference interpolating polynomial given by,

$$f(x) \cong 1.0 + x \times 0.75 + \frac{x(x-1)}{2} \times 20 + \frac{x(x-1)(x-2)}{6} \times 12$$

Since $x_0 = 0, \quad h = 1.0$

$\therefore \quad u = \dfrac{x - x_0}{h} = x$

i.e., $f(x) = 1.0 + 1.5x + 4x^2 + 2x^3$, *on simplification.*

Taking $x_n = 4$, we also construct the backward difference interpolating polynomial given by,

$$f(x) = 199 + (x-4) \times 103.5 + \frac{(x-4)\ \ (x-3)}{2} \times 44$$
$$+ \frac{(x-4)\ \ (x-3)\ \ (x-3)}{6} \times 12$$
$$= 1.0 + 1.5x + 4x^2 + 2x^3, \text{ on simplification.}$$

This is the same as the forward difference interpolating polynomial, because the difference of third order is constant.

Example 2.26: Use Newton's divided difference interpolation to evaluate $f(18)$ and $f(12)$ for the following data:

x	4	5	7	10	11	13
$f(x)$	48	100	294	900	1210	2028

NOTES

Solution: We first form the divided difference table as given below.

x	$f(x)$	1st DD	2nd DD	3rd DD
4	48			
		52		
5	100		15	
		97		1
7	294		21	
		202		1
10	900		27	
		310		1
11	1210		33	
		409		
13	2028			

Since 3rd order divided differences are same, higher order divided differences vanish. We have the Newton's divided difference interpolation given by,

$$f(x) \approx f_0 + (x - x_0) f\left[x, x_1\right] + (x - x_0)(x - x_1) f\left[x_0, x_1, x_2\right]$$
$$+ (x - x_0)(x - x_1)(x - x_2) f\left[x_0, x_1, x_2, x_3\right]$$

For $x = 8$, we take $x_0 = 4$,

$$f(8) = 48 + (8-4)52 + (8-4)(8-5)\times 15 + (8-4)(8-5)(8-7)\times 1$$
$$= 48 + 208 + 180 + 12 = 448$$

For $x = 12$, we take $x_0 = 13$,

$$f(12) = 2028 + (12-13)\times 409 + (12-13)\ (12-11)\times 33$$
$$+ (12-13)\ (12-11)\ (12-10)\times 1$$
$$\therefore f(12) = 2028 - 409 - 33 - 2 = 1584$$

Example 2.27: Using inverse interpolation, find the zero of $f(x)$ given by the following tabular values.

x	0.3	0.4	0.6	0.7
$y = f(x)$	0.14	0.06	-0.04	-0.06

Solution: Using Lagrange's form of inverse interpolation, we calculate the formula using $y = 0.14, 0.06, -0.04$ and -0.06, as given below:

$$P_3(y) = \frac{(y-0.06)(y+0.04)(y+0.06)}{(0.14-0.06)(0.14+0.04)(0.14+0.06)}\times 0.3$$
$$+\frac{(y-0.14)(y+0.04)(y+0.06)}{(0.06-0.14)(0.06+0.04)(0.06+0.06)}\times 0.4$$
$$+\frac{(y-0.14)(y-0.06)(y+0.06)}{(0.04-0.14)(0.04-0.06)(-0.04+0.06)}\times 0.6$$
$$+\frac{(y-0.14)(y-0.06)(y+0.04)}{(-0.06-0.14)(-0.06-0.06)(-0.06+0.04)}\times 0.7$$

NOTES

Thus, $P_3(0) = \frac{-0.06 \times 0.04 \times 0.06 \times 0.3}{0.08 \times 0.18 \times 0.20} + \frac{0.14 \times 0.04 \times 0.06 \times 0.4}{0.08 \times 0.1 \times 0.12}$

$$+ \frac{0.14 \times 0.06 \times 0.06 \times 0.6}{0.18 \times 0.1 \times 0.02} - \frac{0.14 \times 0.06 \times 0.04 \times 0.7}{0.2 \times 0.12 \times 0.02}$$

$$= -0.015 + 0.14 + 0.84 - 0.49 = 0.475$$

Thus, the zero of $f(x)$ is 0.475 which is approximately equal to 0.48, since the accuracy depends on the accuracy of the data which is the significant digits.

2.2.12 Truncation Error in Interpolation

Refer Unit 1 (Level Head 1.4)

Check Your Progress

1. What do we generate in iterative linear interpolation?
2. Define interpolation.
3. How is Lagrange's interpolation useful?
4. Which interpolation will you use for equally spaced tabular values?
5. Define the shift operator.
6. What is the Newton forward difference interpolation formula used?
7. Define extrapolation.
8. Define the problem of inverse interpolation.

2.3 CURVE FITTING

In this section, we consider the problem of approximating an unknown function whose values, at a set of points, are generally known only empirically and are, thus subject to inherent errors, which may sometimes be appreciably larger in many engineering and scientific problems. In these cases, it is required to derive a functional relationship using certain experimentally observed data. Here the observed data may have inherent or round-off errors, which are serious, making polynomial interpolation for approximating the function inappropriate. In polynomial interpolation the truncation error in the approximation is considered to be important. But when the data contains round-off errors or inherent errors, interpolation is not appropriate.

The subject of this section is curve fitting by least square approximation. Here we consider a technique by which noisy function values are used to generate a smooth approximation. This smooth approximation can then be used to approximate the derivative more accurately than with exact polynomial interpolation.

There are situations where interpolation for approximating function may not be efficacious procedure. Errors will arise when the function values $f(x_i)$, $i = 1, 2, \ldots, n$ are observed data and not exact. In this case, if we use the polynomial interpolation, then it would reproduce all the errors of observation. In such situations one may take a large number of observed data, so that statistical laws in effect cancel the errors introduced by inaccuracies in the measuring equipment. The

NOTES

approximating function is then derived, such that the sum of the squared deviation between the observed values and the estimated values are made as small as possible.

Mathematically, the problem of curve fitting or function approximation may be stated as follows:

To find a functional relationship $y = g(x)$, that relates the set of observed data values $P_i(x_i, y_i)$, $i = 1, 2, ..., n$ as closely as possible, so that the graph of $y = g(x)$ goes near the data points P_i's though not necessarily through all of them.

The first task in curve fitting is to select a proper form of an approximating function $g(x)$, containing some parameters, which are then determined by minimizing the total squared deviation.

For example, $g(x)$ may be a polynomial of some degree or an exponential or logarithmic function. Thus $g(x)$ may be any of the following:

(*i*) $g(x) = \alpha + \beta x$ (*ii*) $g(x) = \alpha + \beta x + \gamma x^2$

(*iii*) $g(x) = \alpha e^{\beta x}$ (*iv*) $g(x) = \alpha e^{-\beta x}$

(*v*) $g(x) = \alpha \log(\beta x)$

Here α, β, γ are parameters which are to be evaluated so that the curve $y = g(x)$, fits the data well. A measure of how well the curve fits is called the goodness of fit.

In the case of least square fit, the parameters are evaluated by solving a system of normal equations, derived from the conditions to be satisfied so that the sum of the squared deviations of the estimated values from the observed values, is minimum.

2.3.1 Method of Least Squares

Let (x_1, f_1), (x_2, f_2),..., (x_n, f_n) be a set of observed values and $g(x)$ be the approximating function. We form the sums of the squares of the deviations of the observed values f_i from the estimated values $g(x_i)$,

i.e.,
$$S = \sum_{i=1}^{n} \{f_i - g(x_i)\}^2 \tag{2.47}$$

The function $g(x)$ may have some parameters, α, β, γ. In order to determine these parameters we have to form the necessary conditions for S to be minimum, which are

$$\frac{\partial S}{\partial \alpha} = 0, \frac{\partial S}{\partial \beta} = 0, \frac{\partial S}{\partial \gamma} = 0 \tag{2.48}$$

These equations are called normal equations, solving which we get the parameters for the best approximate function $g(x)$.

Curve Fitting by a Straight Line: Let $g(x) = \alpha + \beta x$, be the straight line which fits a set of observed data points (x_i, y_i), $i = 1, 2, ..., n$.

NOTES

Let S be the sum of the squares of the deviations $g(x_i) - y_i$, $i = 1, 2,..., n$, given by,

$$S = \sum_{i=1}^{n} (\alpha + \beta\, x_i - y_i)^2 \qquad (2.49)$$

We now employ the method of least squares to determine α and β so that S will be minimum. The normal equations are,

$$\frac{\partial S}{\partial \alpha} = 0, \text{ i.e., } \sum_{i=1}^{n} (\alpha + \beta\, x_i - y_i) = 0 \qquad (2.50)$$

and,
$$\frac{\partial S}{\partial \beta} = 0, \text{ i.e., } \sum_{i=1}^{n} x_i (\alpha + \beta\, x_i - y_i) = 0 \qquad (2.51)$$

These conditions give,

$$n\alpha + S_1\beta - S_{01} = 0$$
$$S_1\alpha + S_2\beta - S_{11} = 0$$

where,
$$S_1 = \sum_{i=1}^{n} x_i, \quad S_{01} = \sum_{i=1}^{n} y_i, \quad S_2 = \sum_{i=1}^{n} x_i^{\,2}, \quad S_{11} = \sum_{i=1}^{n} x_i y_i$$

Solving,

$$\frac{\alpha}{-S_1 S_{11} + S_1 S_2} = \frac{\beta}{nS_{11} - S_1 S_{01}} = \frac{1}{nS_2 - S_1^2}. \quad \text{Also} \quad \alpha = \frac{S_{01}}{n} - \beta \frac{S_1}{n}.$$

Algorithm: Fitting a straight line $y = a + bx$.

Step 1: Read n [n being the number of data points]

Step 2: Initialize : sum $x = 0$, sum $x^2 = 0$, sum $y = 0$, sum $xy = 0$

Step 3: For $j = 1$ to n compute

Begin

Read data x_j, y_j

Compute sum x = sum $x + x_j$

Compute sum $x^2 + x_j \times x_j$

Compute sum y = sum $y + y_i \times y_j$

Compute sum xy = sum $xy + x_j \times y_j$

End

Step 4: Compute $b = (n \times \text{sum } xy - \text{sum } x \times \text{sum } y)/ (n \times \text{sum } x^2 - (\text{sum } x)^2)$

Step 5: Compute x bar = sum x / n

Step 6: Compute y bar = sum y / n

Step 8: Compute $a = y$ bar $- b \times x$ bar

Step 9: Write a, b

Step 10: For $j = 1$ to n

Begin

Compute y estimate $= a + b \times x$

write x_j, y_j, y estimate

End

Step 11: Stop

Curve Fitting by a Quadratic (A Parabola): Let $g(x) = a + bx + cx^2$, be the approximating quadratic to fit a set of data (x_i, y_i), $i = 1, 2, ..., n$. Here the parameters are to be determined by the method of least squares, i.e., by minimizing the sum of the squares of the deviations given by,

$$S = \sum_{i=1}^{n} (a + bx_i + cx_i^2 - y_i)^2 \quad (2.52)$$

Thus the normal equations, $\frac{\partial S}{\partial a} = 0, \quad \frac{\partial S}{\partial b} = 0, \quad \frac{\partial S}{\partial c} = 0$, are as follows: (2.53)

$$\sum_{i=1}^{n} (a + bx_i + cx_i^2 - y_i) = 0$$

$$\sum_{i=1}^{n} x_i (a + bx_i + cx_i^2 - y_i) = 0$$

$$\sum_{i=1}^{n} x_i^2 (a + bx_i + cx_i^2 - y_i) = 0. \quad (2.54)$$

These equations can be rewritten as,

$$na + s_1 b + s_2 c - s_{01} = 0$$

$$s_1 a + s_2 b + s_3 c - s_{11} = 0$$

$$s_2 a + s_3 b + s_4 c - s_{21} = 0 \quad (2.55)$$

where, $s_1 = \sum_{i=1}^{n} x_i, \quad s_2 = \sum_{i=1}^{n} x_i^2, \quad s_3 = \sum_{i=1}^{n} x_i^3, \quad s_4 = \sum_{i=1}^{n} x_i^4$

$$s_{01} = \sum_{i=1}^{n} y_i, \quad s_{11} = \sum_{i=1}^{n} x_i \quad y_i, \quad s_{21} = \sum_{i=1}^{n} x_i^2 y_i \quad (2.56)$$

It is clear that the normal equations form a system of linear equations in the unknown parameters a, b, c. The computation of the coefficients of the normal equations can be made in a tabular form for desk computations as shown below.

	x	x_i	y_i	x_i^2	x_i^3	x_i^4	$x_i y_i$	$x_i^2 y_i$
	1	x_1	y_1	x_1^2	x_1^3	x_1^4	$x_1 y_1$	$x_1^2 y_1$
	2	x_2	y_2	x_2^2	x_2^3	x_2^4	$x_2 y_2$	$x_2^2 y_2$
	...	...	...	...	...	...	...	...
	n	x_n	y_n	x_n^2	x_n^3	x_n^4	$x_n y_n$	$x_n^2 y_n$
Sum		s_1	s_{01}	s_2	s_3	s_4	s_{11}	s_{21}

The system of linear equations can be solved by Gaussian elimination method. It may be noted that number of normal equations is equal to the number of unknown parameters.

NOTES

Example 2.28: Find the straight line fitting the following data:

x_i	4	6	8	10	12
y_1	13.72	12.90	12.01	11.14	10.31

Solution: Let $y = a + bx$, be the straight line which fits the data. We have the normal equations $\frac{\partial S}{\partial a} = 0, \frac{\partial S}{\partial b} = 0$ for determining a and b, where

$$S = \sum_{i=1}^{5} (y_i - a - bx_i)^2 .$$

Thus, $$\sum_{i=1}^{5} y_i - na - b\sum_{i=1}^{5} x_i = 0$$

and, $$\sum_{i=1}^{5} x_i y_i - a\sum_{I=1}^{5} x_i - b\sum_{i=1}^{5} x_i^{\ 2} = 0$$

The coefficients are computed in the table below.

	x_i	y_i	x_i^2	$x_i y_i$
	4	13.72	16	54.88
	6	12.90	36	77.40
	8	12.01	64	96.08
	10	11.14	100	111.40
	12	10.31	144	123.72
Sum	40	60.08	360	463.48

Thus the normal equations are,

$$5a + 40b - 60.08 = 0$$
$$40a + 360b - 463.48 = 0$$

Solving these two equations we obtain,

$$a = 15.448,\ b = 0.429$$

Thus, $y = g(x) = 15.448 - 0.429\,x$, is the straight line fitting the data.

Example 2.29: Use the method of least square approximation to fit a straight line to the following observed data:

x_i	60	61	62	63	64
y_i	40	40	48	52	55

Solution: Let the straight line fitting the data be $y = a + bx$. The data values being large, we can use a change in variable by substituting $u = x - 62$ and $v = y - 48$.

Let $v = A + B\,u$, be a straight line fitting the transformed data, where the normal equations for A and B are,

$$\sum_{i=1}^{5} v_i = 5\,A + B\sum_{i=1}^{5} u_i$$
$$\sum_{i=1}^{5} u_i v_i = A\sum_{i=1}^{5} u_i + B\sum_{i=1}^{5} u_i^2$$

The computation of the various sums are given in the table below,

	x_i	y_i	u_i	v_i	$u_i v_i$	u_i^2
	60	40	–2	–8	16	4
	61	42	–1	–6	6	1
	62	48	0	0	0	0
	63	52	1	4	4	1
	64	55	2	7	14	4
Sum			0	–3	40	10

Thus the normal equations are,

$$-3 = 5A \quad \text{and} \quad 40 = 10B$$

$$\therefore A = -\frac{3}{5} \quad \text{and} \quad B = 4$$

This gives the line, $v = -3/5 + 4u$

or, $20u - 5v - 3 = 0.$

Transforming we get the line,

$$20\,(x - 62) - 5\,(y - 48) - 3 = 0$$

or, $$20\,x - 5y - 1003 = 0$$

Curve Fitting with an Exponential Curve: We consider a two parameter exponential curve as,

$$y = ae^{-bx} \tag{2.57}$$

For determining the parameters, we can apply the principle of least squares by first using a transformation,

$z = \log y$, so that Equation (2.57) is rewritten as, (2.58)

$$z = \log a - bx \tag{2.59}$$

Thus, we have to fit a linear curve of the form $z = \alpha + \beta \quad x$, with $z-x$ variables and then get the parameters a and b as,

$$a = e^{\alpha}, \quad b = -\beta \tag{2.60}$$

Thus proceeding as in linear curve fitting,

$$\beta = \frac{n\sum_{i=1}^{n} x_i \log y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} \log y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \tag{2.61}$$

and, $\alpha = \bar{z} - p\bar{x}$, where $\bar{x} = \left(\sum_{i=1}^{n} x_i\right)\Big/ n$, $\bar{z} = \left(\sum_{i=1}^{n} \log y_i\right)\Big/ n$ (2.62)

After computing α and β, we can determine a and b given by Equation (2.59). Finally, the exponential curve fitting data set is given by Equation (2.57).

Algorithm: To fit a straight line for a given set of data points by least square error method.

Step 1: Read the number of data points, i.e., n

Step 2: Read values of data-points, i.e., Read (x_i, y_i) for $i = 1, 2, ..., n$

Step 3: Initialize the sums to be computed for the normal equations,

i.e., $sx = 0$, $sx^2 = 0$, $sy = 0$, $syx = 0$

Step 4: Compute the sums, i.e., For $i = 1$ to n do

Begin

$$sx = sx + x_i$$
$$sx^2 = sx^2 + x_i^2$$
$$sy = sy + y_i$$
$$syx = syx + x_i y_i$$

End

Step 5: Solve the normal equations, i.e., solve for a and b of the line $y = a + bx$

$$\text{Compute} \quad d = n * sx^2 - sx * sx$$
$$b = (n * sxy - sy * sx) / d$$
$$x\text{bar} = sx / n$$
$$y\text{bar} = sy / n$$
$$a = y\text{bar} - b * x\text{bar}$$

Step 6: Print values of a and b

Step 7: Print a table of values of x_i, y_i, $y_{pi} = a + bx_i$ for $i = 1, 2, ..., n$

Step 8: Stop

Algorithm: To fit a parabola $y = a + bx + cx^2$, for a given set of data points by least square error method.

Step 1: Read n, the number of data points

Step 2: Read (x_i, y_i) for $i = 1, 2, ..., n,$ the values of data points

Step 3: Initialize the sum to be computed for the normal equations,

i.e., $sx = 0$, $sx^2 = 0$, $sx^3 = 0$, $sx^4 = 0$, $sy = 0$, $sxy = 0$.

Step 4: Compute the sums, i.e., For $i = 1$ to n do

Begin

$$sx = sx + x_i$$
$$x^2 = x_i * x_i$$
$$sx^2 = sx^2 + x^2$$
$$sx^3 = sx^3 + x_i * x^2$$
$$sx^4 = sx^4 + x^2 * x^2$$
$$sy = sy + y_i$$
$$sxy = sxy + x_i * y_i$$
$$sx^2 y = sx^2 y + x^2 * y_i$$

NOTES

End

Step 5: Form the coefficients $\{a_{ij}\}$ matrix of the normal equations, i.e.,

$$a_{11} = n, \quad a_{21} = sx, \quad a_{31} = sx^2$$
$$a_{12} = sx, \quad a_{22} = sx^2, \quad a_{32} = sx^3$$
$$a_{13} = sx^2, \quad a_{23} = sx^3, \quad a_{33} = sx^4$$

Step 6: Form the constant vector of the normal equations.

$b_1 = sy, b_2 = sxy, b_3 = sx^2y$

Step 7: Solve the normal equation by Gauss-Jordan method

$$a_{12} = a_{12} / a_{11}, a_{13} = a_{13} / a_{11}, b_1 = b_1 / a_{11}$$
$$a_{22} = a_{22} - a_{21}a_{12}, a_{23} = a_{23} - a_{21}a_{13}$$
$$b_2 = b_2 - b_1a_{21}$$

$$a_{32} = a_{32} - a_{31}a_{12}$$
$$a_{33} = a_{33} - a_{31}a_{13}$$
$$b_3 = b_3 - b_1a_{31}$$
$$a_{23} = a_{23} / a_{22}$$
$$b_2 = b_2 / a_{22}$$
$$a_{33} = a_{33} - a_{23}\,a_{32}$$
$$b_3 = b_3 - a_{32}\,b_2$$
$$c = b_3 / a_{33}$$
$$b = b_2 - c\,a_{23}$$
$$a = b_1 - b\,a_{12} - c\,a_{13}$$

Step 8: Print values of a, b, c (the coefficients of the parabola)

Step 9: Print the table of values of $x_k, \;\; y_k$ and y_{pk} where $y_{pk} = a + bx_k + cx^2k$,

i.e., print $x_k, \;\; y_k, \;\; y_{pk}$ for $k = 1, \;\; 2, ..., \;\; n.$

Step 10: Stop.

2.4 TRIGONOMETRIC FUNCTIONS

Let a revolving line OP start from OX in the anticlockwise direction and trace out an angle XOP. From P draw $PM \perp OX$ (Produce OX, if necessary). Let $\angle XOP = \theta$.

Then (1) $\dfrac{MP}{OP}$ is called sine of angle θ and is written as sin θ.

(2) $\dfrac{OM}{OP}$ is called *cosine* of angle θ and is written as cos θ.

(3) $\dfrac{MP}{OM}$ is called *tangent* of angle θ and is written as tan θ.

NOTES

(4) $\frac{OM}{MP}$ is called *cotangent* of angle θ and is written as cot θ.

(5) $\frac{OP}{OM}$ is called *secant* of angle θ and is written as sec θ.

(6) $\frac{OP}{MP}$ is called *cosecant* of angle θ and is written as cosec θ.

These ratios are called *Trigonometrical Ratios* of the angle θ.

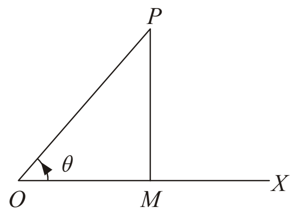


Fig. 2.2

Remarks:

1. It follows from the definition that

$$\sec\theta = \frac{1}{\cos\theta}, \quad \operatorname{cosec}\theta = \frac{1}{\sin\theta}, \quad \cot\theta = \frac{1}{\tan\theta},$$

$$\tan\theta = \frac{\sin\theta}{\cos\theta}, \quad \cot\theta = \frac{\cos\theta}{\sin\theta}.$$

2. Trigonometrical ratios are same for the same angle. For, let P' be any point on the revolving line OP. Draw $P'M' \perp OX$. Then triangles OPM and $OP'M'$ are similar, so $\frac{MP}{OP} = \frac{M'P'}{OP'}$ *i.e.*, each of these ratios is sin θ.

   Therefore, whatever be the triangle of reference (*i.e.*, ΔOPM or $\Delta OP'M'$) might be, we find that sin θ remains the same for a particular angle θ.

   It can be similarly shown that no trigonometrical ratio depends on the size of triangle of reference.

3. $(\sin\theta)^n$ is written as $\sin^n\theta$, where n is positive. Similar notation holds good for other trigonometrical ratios.

4. $\sin^{-1}\theta$ denotes that angle whose sine is θ. Note that $\sin^{-1}\theta$ does not stand for $\frac{1}{\sin\theta}$. Similar notation holds good for other trigonometrical ratios.

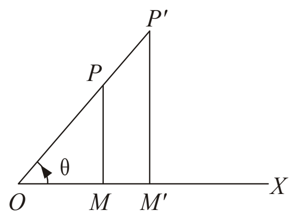


Fig. 2.3

NOTES

For Any Angle θ

1. $\sin^2\theta + \cos^2\theta = 1$
2. $\sec^2\theta = 1 + \tan^2\theta$
3. $\text{cosec}^2\theta = 1 + \cot^2\theta$

Proof. Let the revolving line OP start from OX and trace out an angle θ in the anti-clockwise direction. From P draw $PM \perp OX$. (Produce OX, if necessary.) (Refer Figure 2.2).

Then $\angle XOP = \theta$.

(1) $\sin\theta = \dfrac{MP}{OP}$, $\cos\theta = \dfrac{OM}{OP}$

Then $\sin^2\theta + \cos^2\theta = \dfrac{(MP)^2 + (OM)^2}{(OP)^2} = \dfrac{(OP)^2}{(OP)^2} = 1.$

(2) $\sec\theta = \dfrac{OP}{OM}$, $\tan\theta = \dfrac{MP}{OM}$

Then $1 + \tan^2\theta = 1 + \dfrac{(MP)^2}{(OM)^2} = \dfrac{(OM)^2 + (MP)^2}{(OM)^2}$

$$= \frac{(OP)^2}{(OM)^2} = \left(\frac{OP}{OM}\right)^2 = (\sec\theta)^2 = \sec^2\theta$$

(3) $\cot\theta = \dfrac{OM}{MP}$, $\text{cosec}\,\theta = \dfrac{OP}{MP}$.

Then $1 + \cot^2\theta = 1 + \left(\dfrac{OM}{MP}\right)^2 = \dfrac{(MP)^2 + (OM)^2}{(MP)^2}$

$$= \frac{(OP)^2}{(MP)^2} = \left(\frac{OP}{MP}\right)^2 = (\text{cosec}\,\theta)^2 = \text{cosec}^2\theta.$$

Signs of Trigonometrical Ratios

Consider four lines OX, OX', OY, OY' at right angles to each other. Let a revolving line OP start from OX in the anticlockwise direction. From P draw $PM \perp OX$ or OX'. We have the following convention of signs regarding the sides of ΔOPM.

1. OM is positive, if it is along OX.
2. OM is negative, if it is along OX'.
3. MP is negative, if it is along OY'.
4. MP is positive, if it is along OY.
5. OP is regarded always positive.

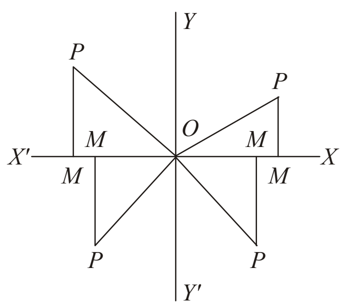


Fig. 2.4

NOTES

First Quadrant: If the revolving line *OP* is in the first quadrant, then all the sides of the triangle *OPM* are positive. Therefore, all the trigonometrical ratios are positive in the first quadrant.

Second Quadrant: If the revolving line *OP* is in the second quadrant, then *OM* is negative and the other two sides of ΔOPM are positive. Therefore, ratios involving *OM* will be negative. So, cosine, secant, tangent, cotangent of an angle in the second quadrant are negative while sine and cosecant of anlge in the second quadrant are positive.

Third Quadrant: If the revolving line is in the third quadrant, then sides *OM* and *MP* both are negative. Since *OP* is always positive, therefore, ratios involving each one of *OM* and *MP* alone will be negative. So, sine, cosine, cosecant and secant of an angle in the third quadrant are negative. Since tangent or cotangent of any angle involve both *OM* and *MP*, therefore, these will be positive. So, tangent and cotangent of an angle in the third quadrant are positive.

Fourth Quadrant: If the revolving line *OP* is in the fourth quadrant, then *MP* is negative and the other two sides of ΔOPM are positive. Therefore, ratios involving *MP* will be negative and others positive. So, sine, cosecant, tangent and cotangent of an angle in the fourth quadrant are negative while cosine and secant of an angle in the fourth quadrant are positive.

Limits to the Value of Trigonometrical Ratios

We know that $\sin^2\theta + \cos^2\theta = 1$ for any angle θ. $\sin^2\theta$ and $\cos^2\theta$ being perfect squares, will be positive. Again neither of them can be greater than 1 because then the other will have to be negative.

Thus $\sin^2\theta \le 1$, $\cos^2\theta \le 1$.

$\Rightarrow$ $\sin\theta$ and $\cos\theta$ cannot be numerically greater than 1.

Similarly, $\operatorname{cosec}\theta = \dfrac{1}{\sin\theta}$ and $\sec\theta = \dfrac{1}{\cos\theta}$ cannot be numerically less than 1.

There is no restriction on $\tan\theta$ and $\cot\theta$. They can have any value.

Example 2.30: Prove that $\sin^6\theta + \cos^6\theta = 1 - 3\sin^2\theta\cos^2\theta$.

Solution: Here LHS

$$\begin{aligned}
&= \sin^6\theta + \cos^6\theta \\
&= (\sin^2\theta)^3 + (\cos^2\theta)^3 \\
&= (\sin^2\theta + \cos^2\theta)(\sin^4\theta - \sin^2\theta\cos^2\theta + \cos^4\theta) \\
&= 1\,.\,(\sin^4\theta - \sin^2\theta\cos^2\theta + \cos^4\theta) \\
&= [(\sin^2\theta + \cos^2\theta)^2 - 3\sin^2\theta\cos^2\theta] \\
&= 1 - 3\sin^2\theta\cos^2\theta = \text{RHS}.
\end{aligned}$$

Example 2.31: Prove that $\sqrt{\dfrac{1+\cos\theta}{1-\cos\theta}} = \operatorname{cosec}\theta + \cot\theta$. Provided $\cos\theta \ne 1$.

Solution.

$$\text{LHS} = \sqrt{\frac{1+\cos\theta}{1-\cos\theta}}$$

$$= \sqrt{\frac{(1+\cos\theta)(1+\cos\theta)}{(1-\cos\theta)(1+\cos\theta)}} = -\frac{1+\cos\theta}{\sqrt{1-\cos^2\theta}}$$

$$= \frac{1+\cos\theta}{\sin\theta} = \frac{1}{\sin\theta} + \frac{\cos\theta}{\sin\theta} = \text{cosec}\,\theta + \cot\theta.$$

NOTES

Example 2.32: Prove that $(1 + \cot\theta - \text{cosec}\,\theta)(1 + \tan\theta + \sec\theta) = 2$.

Solution: LHS $= (1 + \cot\theta - \text{cosec}\,\theta)(1 + \tan\theta + \sec\theta)$

$$= \left(1 + \frac{\cos\theta}{\sin\theta} - \frac{1}{\sin\theta}\right)\left(1 + \frac{\sin\theta}{\cos\theta} + \frac{1}{\cos\theta}\right)$$

$$\text{LHS} = \frac{(\sin\theta + \cos\theta - 1)(\cos\theta + \sin\theta + 1)}{\sin\theta\cos\theta}$$

$$= \frac{(\sin\theta + \cos\theta)^2 - 1}{\sin\theta\cos\theta}$$

$$= \frac{\sin^2\theta + \cos^2\theta + 2\sin\theta\cos\theta - 1}{\sin\theta\cos\theta}$$

$$= \frac{1 + 2\sin\theta\cos\theta - 1}{\sin\theta\cos\theta} = \frac{2\sin\theta\cos\theta}{\sin\theta\cos\theta} = 2 = \text{RHS}.$$

Example 2.33: Prove that $\frac{\tan\theta}{1-\cot\theta} + \frac{\cot\theta}{1-\tan\theta} = 1 + \text{cosec}\,\theta\sec\theta$, if $\cot\theta \neq 1$, 0 and $\tan\theta \neq 1, 0$.

Solution: $\text{LHS} = \frac{\tan\theta}{1-\cot\theta} + \frac{\cot\theta}{1-\tan\theta}$

$$= \frac{\tan\theta}{1-\frac{1}{\tan\theta}} + \frac{\frac{1}{\tan\theta}}{1-\tan\theta}$$

$$= \frac{\tan^2\theta}{\tan\theta - 1} + \frac{1}{\tan\theta(1-\tan\theta)}$$

$$= \frac{\tan^2\theta}{\tan\theta - 1} - \frac{1}{\tan\theta(\tan\theta - 1)}$$

$$= \frac{\tan^3\theta - 1}{\tan\theta(\tan\theta - 1)}$$

$$= \frac{(\tan\theta - 1)(\tan^2\theta + \tan\theta + 1)}{\tan\theta(\tan\theta - 1)}$$

$$= \frac{\tan^2\theta + \tan\theta + 1}{\tan\theta} \quad \text{since } \tan\theta \neq 1$$

$$= \frac{\sec^2\theta + \tan\theta}{\tan\theta}$$

$$= \frac{\sec^2\theta}{\tan\theta} + 1 = \sec\theta\,\text{cosec}\,\theta + 1 = \text{RHS}.$$

Example 2.34: Which of the six trigonometrical ratios are positive for (i) 960º (ii) – 560º?

Solution: (*i*) 960º = 720º + 240º.

Therefore, the revolving line starting from *OX* will make two complete revolutions in the anticlockwise direction and further trace out an angle of 240º in the same direction. Thus, it will be in the third quadrant. So, the tangent and cotangent are positive and rest of trigonometrical ratios will be negative.

(*ii*) – 560º = – 360º – 200º.

NOTES

Therefore, the revolving line after making one complete revolution in the clockwise direciton, will trace out further an angle of 200º in the same direction. Thus, it will be in the second quadrant. So, only sine and cosecant are positive.

Example 2.35: In what quadrants can θ lie if $sec\ \theta = \frac{-7}{6}$?

Solution: As sec θ is negative in second and third quadrants, θ can lie in second or third quadrant only.

Example 2.36: If $\sin\theta = \frac{-12}{13}$, determine other trigonometrical ratios of θ.

Solution. $\cos^2\theta = 1 - \sin^2\theta$

$$= 1 - \frac{144}{169} = \frac{169-144}{169} = \frac{25}{169}.$$

$$\Rightarrow \quad \cos\theta = \pm\frac{5}{13}.$$

So $\tan\theta = \frac{\sin\theta}{\cos\theta} = \pm\frac{12}{5}$

$$\text{cosec}\ \theta = \frac{-13}{12},\ \sec\theta = \pm\frac{13}{5},\ \cot\theta = \pm\frac{5}{12}.$$

Example 2.37: Express all the trigonometrical ratios of θ in terms of the sin θ.

Solution: Let $\sin\theta = k$.

Then, $\cos^2\theta = 1 - \sin^2\theta = 1 - k^2 \quad \Rightarrow \quad \cos\theta = \pm\sqrt{1-k^2}$

$$\tan\theta = \frac{\sin\theta}{\cos\theta} = \pm\frac{k}{\sqrt{1-k^2}}$$

$$\cot\theta = \frac{\cos\theta}{\sin\theta} = \pm\frac{\sqrt{1-k^2}}{k}$$

$$\sec\theta = \frac{1}{\cos\theta} = \pm\frac{1}{\sqrt{1-k^2}}$$

$$\text{cosec}\ \theta = \frac{1}{\sin\theta} = \frac{1}{k}.$$

Example 2.38: Prove that $\sin\theta = a + \frac{1}{a}$ is impossible, if a is real.

Solution: $\sin\theta = a + \frac{1}{a} \quad \Rightarrow \quad \sin\theta = \frac{a^2+1}{a}$

$$\Rightarrow \quad a^2 - a\sin\theta + 1 = 0$$

$$\Rightarrow \quad a = \frac{\sin\theta \pm \sqrt{\sin^2\theta - 4}}{2}$$

For a to be real, the expression under the radical sign, must be positive or zero.

i.e., $\sin^2\theta - 4 \geq 0$

or $\sin^2\theta \geq 4 \Rightarrow \sin\theta$ is numerically greater than or equal to 2 which is impossible.

Thus, if a is real, $\sin\theta = a + \frac{1}{a}$ is impossible.

NOTES

Example 2.39: Determine the quadrant in which θ must lie if cot θ is positive and cosec θ is negative.

Solution: cot θ is positive ⇒ θ lies in first or third quadrant.

cosec θ is negative ⇒ θ lies in third or fourth quadrant.

In order that cot θ is positive and cosec θ is negatie, we see that θ must lie in third quadrant.

Example 2.40: Prove that

$$\frac{1}{cosec\,\theta + cot\,\theta} - \frac{1}{sin\,\theta} = \frac{1}{sin\,\theta} - \frac{1}{cosec\,\theta - cot\,\theta}$$

Solution:

$$\text{LHS} = \frac{1}{\operatorname{cosec}\theta + \cot\theta} - \frac{1}{\sin\theta}$$

$$= \frac{\sin\theta}{1+\cos\theta} - \frac{1}{\sin\theta}$$

$$= \frac{\sin^2\theta - (1+\cos\theta)}{(1+\cos\theta)\sin\theta}$$

$$= \frac{-(1-\sin^2\theta) - \cos\theta}{(1+\cos\theta)\sin\theta}$$

$$= \frac{-\cos^2\theta - \cos\theta}{(1+\cos\theta)\sin\theta}$$

$$= \frac{-\cos\theta\,(1+\cos\theta)}{(1+\cos\theta)\sin\theta} = -\cot\theta$$

$$\text{RHS} = \frac{1}{\sin\theta} - \frac{1}{\operatorname{cosec}\theta - \cot\theta}$$

$$= \frac{1}{\sin\theta} - \frac{\sin\theta}{1-\cos\theta}$$

$$= \frac{1-\cos\theta-\sin^2\theta}{\sin\theta(1-\cos\theta)}$$

$$= \frac{\cos^2\theta - \cos\theta}{\sin\theta\,(1-\cos\theta)}$$

$$= \frac{-\cos\theta\,(1-\cos\theta)}{\sin\theta\,(1-\cos\theta)} = -\cot\theta$$

Therefore, LHS = RHS.

Example 2.41: Prove that,

$\sin\theta(1+\tan\theta) + \cos\theta\,(1+\cot\theta) = \sec\theta + \operatorname{cosec}\theta$.

NOTES

Solution: LHS = sin θ (1 + tan θ) + cos θ (1 + cot θ)

$$= \sin\theta\left(1+\frac{\sin\theta}{\cos\theta}\right)+\cos\theta\left(1+\frac{\cos\theta}{\sin\theta}\right)$$

$$= \sin\theta+\frac{\sin^2\theta}{\cos\theta}+\cos\theta+\frac{\cos^2\theta}{\sin\theta}$$

$$= \frac{\sin^2\theta\cos\theta+\sin^3\theta+\cos^2\theta\sin\theta+\cos^3\theta}{\sin\theta\cos\theta}$$

$$= \frac{\sin^2\theta(\sin\theta+\cos\theta)+\cos^2\theta(\sin\theta+\cos\theta)}{\sin\theta\cos\theta}$$

$$= \frac{(\sin^2\theta+\cos^2\theta)(\sin\theta+\cos\theta)}{\sin\theta\cos\theta}$$

$$= \frac{\sin\theta+\cos\theta}{\sin\theta\cos\theta}$$

$$= \frac{1}{\cos\theta}+\frac{1}{\sin\theta} = \sec\theta+\text{cosec}\,\theta = \text{RHS}.$$

Example 2.42: If $\tan\theta = \frac{4}{3}$, find the value of

$$\frac{2\sin\theta+3\cos\theta}{4\cos\theta+3\sin\theta}.$$

Solution: $\frac{2\sin\theta+3\cos\theta}{4\cos\theta+3\sin\theta} = \frac{2\tan\theta+3}{4+3\tan\theta} = \frac{\frac{8}{5}+3}{4+\frac{12}{5}} = \frac{23}{32}.$

Example 2.43: State giving the reason whether the following equation is possible.

$$2\sin^2\theta - 3\cos\theta - 6 = 0$$

Solution: $2\sin^2\theta - 3\cos\theta - 6 = 0$

$\Rightarrow\quad 2(1-\cos^2\theta) - 3\cos\theta - 6 = 0$

$\Rightarrow\quad -2\cos^2\theta - 3\cos\theta - 4 = 0$

$\Rightarrow\quad 2\cos^2\theta + 3\cos\theta + 4 = 0$

$\Rightarrow\quad \cos\theta = \frac{-3\pm\sqrt{9-32}}{4} = \frac{-3\pm\sqrt{-23}}{4}$

$\Rightarrow$ cos θ is imaginary, which is not true.

Example 2.44: Prove that

$$\frac{1-\sin\theta}{1+\sec\theta}-\frac{1+\sin\theta}{1-\sec\theta} = 2\cos\theta\,(\cot\theta+\text{cosec}^2\,\theta)$$

Solution: $\text{LHS} = \frac{(1-\sin\theta)\cos\theta}{1+\cos\theta}-\frac{(1+\sin\theta)\cos\theta}{\cos\theta-1}$

$$= \cos\theta\left[\frac{(1-\sin\theta)}{(1+\cos\theta)}+\frac{(1+\sin\theta)}{(1-\cos\theta)}\right]$$

$$= \cos\theta\left[\frac{(1-\sin\theta)(1-\cos\theta) + (1+\sin\theta)(1+\cos\theta)}{1-\cos^2\theta}\right]$$

NOTES

$$= \cos\theta\left[\frac{1-\sin\theta-\cos\theta+\sin\theta\cos\theta + 1+\sin\theta+\cos\theta+\sin\theta\cos\theta}{\sin^2\theta}\right]$$

$$= \cos\theta\left[\frac{2+2\sin\theta\cos\theta}{\sin^2\theta}\right]$$

$$= 2\cos\theta\,[\text{cosec}^2\theta + \cos\theta] = \text{RHS}.$$

Example 2.45: If $\tan x = \dfrac{\sin\theta - \cos\theta}{\sin\theta + \cos\theta}$ where θ and x are both positive and acute angles, prove that

$$\sin x = \frac{1}{\sqrt{2}}(\sin\theta - \cos\theta)$$

Solution: $1 + \tan^2 x = 1 + \dfrac{\sin^2\theta + \cos^2\theta - 2\sin\theta\cos\theta}{\sin^2\theta + \cos^2\theta + 2\sin\theta\cos\theta}$

$$= 1 + \frac{(1-2\sin\theta\cos\theta)}{(1+2\sin\theta\cos\theta)} = \frac{2}{1+2\sin\theta\cos\theta}$$

Therefore, $\sec^2 x = \dfrac{2}{1+2\sin\theta\cos\theta}$

$$\Rightarrow \quad \cos^2 x = \frac{1+2\sin\theta\cos\theta}{2}$$

$$\Rightarrow \quad 1-\cos^2 x = \frac{2-(1+2\sin\theta\cos\theta)}{2}$$

$$= \frac{1-2\sin\theta\cos\theta}{2} = \frac{(\sin\theta-\cos\theta)^2}{2}$$

$$\Rightarrow \quad \sin^2 x = \frac{(\sin\theta-\cos\theta)^2}{2}$$

$$\Rightarrow \quad \sin x = \pm\frac{(\sin\theta-\cos\theta)}{\sqrt{2}}.$$

Since θ is acute and $\tan x \geq 0$, $\sin\theta \geq \cos\theta$

$\Rightarrow \sin\theta - \cos\theta \geq 0$

Also x is acute $\Rightarrow \sin x \geq 0$

$$\Rightarrow \quad \sin x = +\frac{(\sin\theta-\cos\theta)}{\sqrt{2}}.$$

Example 2.46: Exhibit $(\sin\theta-3)(\sin\theta-1)(\sin\theta+1)(\sin\theta+3)+16$ as a perfect square and examine if there is any suitable value of θ for which the above expression can vanish.

Solution: Now $(\sin\theta-3)(\sin\theta-1)(\sin\theta+1)(\sin\theta+3)+16$

$$= (\sin^2\theta-1)(\sin^2\theta-9)+16$$
$$= \sin^4\theta-10\sin^2\theta+25$$
$$= (\sin^2\theta-5)^2.$$

NOTES

This is 0 only when $\sin^2\theta-5=0$ *i.e.*, only when $\sin^2\theta=5$ which is not possible as the maximum value of $\sin^2\theta$ is 1.

Thus, there is no value of θ for which the given expression can vanish.

Example 2.47: Find the value in terms of p and q of

$$\frac{p\cos\theta+q\sin\theta}{p\cos\theta-q\sin\theta} \text{ where } \cot\theta=\frac{p}{q}.$$

Here use of any figure not allowed.

Solution: $$\frac{p\cos\theta+q\sin\theta}{p\cos\theta-q\sin\theta} = \frac{p\dfrac{\cos\theta}{\sin\theta}+q}{p\dfrac{\cos\theta}{\sin\theta}-q}$$

$$= \frac{p\cot\theta+q}{p\cot\theta-q} = \frac{\dfrac{p^2}{q}+q}{\dfrac{p^2}{q}-q} = \frac{p^2+q^2}{p^2-q^2}.$$

Example 2.48: Show that

$$\frac{\tan\theta}{\sec\theta-1}+\frac{\tan\theta}{\sec\theta+1}=2\operatorname{cosec}\theta.$$

Solution: $$\text{LHS} = \frac{\tan\theta}{\sec\theta-1}+\frac{\tan\theta}{\sec\theta+1}$$

$$= \tan\theta\left[\frac{1}{\sec\theta-1}+\frac{1}{\sec\theta+1}\right]$$

$$= \tan\theta\left[\frac{2\sec\theta}{\sec^2\theta-1}\right]$$

$$= \tan\theta\left[\frac{2\sec\theta}{\tan^2\theta}\right]$$

$$= \frac{2\sec\theta}{\tan\theta} = \frac{2}{\sin\theta} = 2\operatorname{cosec}\theta = \text{RHS}.$$

2.5 REGRESSION

The term 'Regression' was first used in 1877 by Sir Francis Galton who made a study that showed that the height of children born to tall parents will tend to move back or 'regress' toward the mean height of the population. He designated the word regression as the name of the process of predicting one variable from another variable. He coined the term multiple regression to describe the process by which several variables are used to predict another. Thus, when there is a well-established relationship between variables, it is possible to make use of this relationship in making estimates and to forecast the value of one variable (the unknown or the dependent variable) on the basis of the other variable/s (the known or the independent variable/s). A banker, for example, could predict deposits on the basis of per capita income in the trading area of bank. A marketing manager, may plan his advertising expenditures on the basis of the expected effect on total sales revenue of a change in the level of advertising expenditure. Similarly, a hospital superintendent could project his need for beds on the basis of total population. Such predictions may be made by using regression analysis. An investigator may employ regression analysis to test his theory having the cause and effect relationship. All these explain that regression analysis is an extremely useful tool especially in problems of business and industry involving predictions.

Assumptions in regression analysis

While making use of the regression techniques for making predictions, the following are always assumed:

(i) There is an actual relationship between the dependent and independent variables.

(ii) The values of the dependent variable are random but the values of the independent variable are fixed quantities without error and are chosen by the experimentor.

(iii) There is a clear indication of direction of the relationship. This means that dependent variable is a function of independent variable. (For example, when we say that advertising has an effect on sales, then we are saying that sales has an effect on advertising).

(iv) The conditions (that existed when the relationship between the dependent and independent variable was estimated by the regression) are the same when the regression model is being used. In other words, it simply means that the relationship has not changed since the regression equation was computed.

(v) The analysis is to be used to predict values within the range (and not for values outside the range) for which it is valid.

2.5.1 Linear Regression

In case of simple linear regression analysis, a single variable is used to predict another variable on the assumption of linear relationship (i.e., relationship of the type defined by $Y = a + bX$) between the given variables. The variable to be predicted is called the dependent variable and the variable on which the prediction is based is called the independent variable.

Simple linear regression model[3] (or the Regression Line) is stated as,

$$Y_i = a + bX_i + e_i$$

Where, Y_i = The dependent variable

X_i = The independent variable

e_i = Unpredictable random element (usually called residual or error term)

NOTES

(i) *a* represents the *Y*-intercept, i.e., the intercept specifies the value of the dependent variable when the independent variable has a value of zero. (However, this term has practical meaning only if a zero value for the independent variable is possible).

(ii) *b* is a constant, indicating the slope of the regression line. Slope of the line indicates the amount of change in the value of the dependent variable for a unit change in the independent variable.

If the two constants (viz., *a* and *b*) are known, the accuracy of our prediction of *Y* (denoted by $\hat{Y}$ and read as *Y*-hat) depends on the magnitude of the values of e_i. If in the model, all the e_i tend to have very large values then the estimates will not be very good, but if these values are relatively small, then the predicted values ($\hat{Y}$) will tend to be close to the true values (Y_i).

Estimating the intercept and slope of the regression model (or estimating the regression equation)

The two constants or the parameters viz., '*a*' and '*b*' in the regression model for the entire population or universe are generally unknown and as such are estimated from sample information. The following are the two methods used for estimation:

(i) Scatter diagram method

(ii) Least squares method

1. Scatter diagram method

This method makes use of the Scatter diagram also known as Dot diagram. *Scatter diagram* is a diagram representing two series with the known variable, 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 on a graph paper (Refer Figure 2.5) to get the following information illustrated in Table 2.1:

Table 2.1 *Table Derived from Scatter Diagram*

Income *X* *(Hundreds of Rupees)*	*Consumption Expenditure* *Y* *(Hundreds of Rupees)*
41	44
65	60
50	39
57	51
96	80
94	68
110	84
30	34
79	55
65	48

NOTES

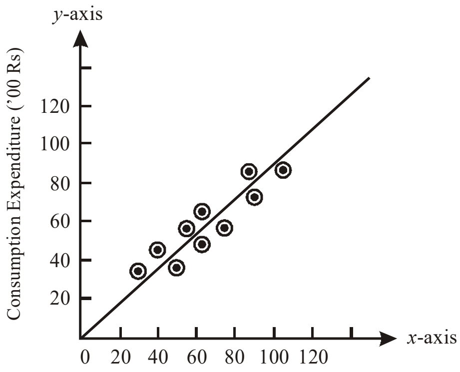


Fig. 2.5 *Scatter Diagram*

The scatter diagram by itself is not sufficient for predicting values of the dependent variable. Some formal expression of the relationship between the two variables is necessary for predictive purposes. For the purpose, one may simply take a ruler and draw a straight line through the points in the scatter diagram and this way can determine the intercept and the slope of the said line and then the line can be defined as $\hat{Y} = a + bX_i$, with the help of which we can predict *Y* for a given value of *X*. However, there are shortcomings in this approach. For example, if five different persons draw such a straight line in the same scatter diagram, it is possible that there may be five different estimates of *a* and *b,* especially when the dots are more dispersed in the diagram. Hence, the estimates cannot be worked out only through this approach. A more systematic and statistical method is required to estimate the constants of the predictive equation. The least squares method is used to draw the best fit line.

2. Least square method

The 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 way that the sum of the squares of the vertical deviations of these points from the line will be a minimum.

The meaning of the least squares criterion can be easily understood through as shown in Figure 2.6, where the earlier as shown in Figure 2.5 in scatter diagram has been reproduced along with a line which represents the least squares line to fit the data.

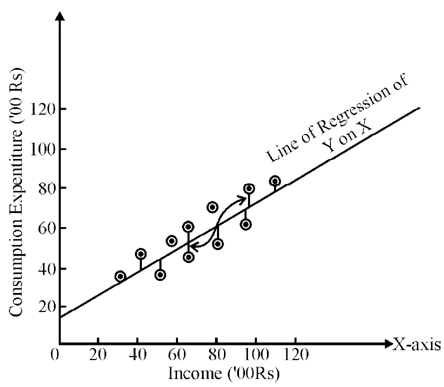


Fig. 2.6 *Scatter Diagram, Regression Line and Short Vertical Lines Representing 'e'*

NOTES

As shown in Figure 2.6, 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 will be denoted by the symbol '*e*'. The value of '*e*' varies from one point to another. In some cases it is positive, while in others it is negative. If the line drawn happens to be the least squares line, then the values of $\sum e_i$ 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 $\hat{Y}$ as $(Y-\hat{Y})$ or e_i, it is logical that we want the $\Sigma(Y-\hat{Y})$ or $\sum_{i=1}^{n} e_i$, to be as small as possible. However, mere examining $\Sigma(Y-\hat{Y})$ or $\sum_{i=1}^{n} e_i$, is inappropriate, since any e_i can be positive or negative. Large positive values and large negative values could cancel one another. However, 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. 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 when squared becomes positive). The choice of minimizing the squared sum of errors rather than the sum of the absolute values implies that there are 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*' also known as the intercept and the slope of the line. This is done with the help of the following two normal equations:

$$\Sigma Y = na + b\Sigma X$$

$$\Sigma XY = a\Sigma X + b\Sigma X^2$$

In these two equations, '*a*' and '*b*' are unknowns and all other values viz., $\sum X$, $\sum Y$, $\sum X^2$, $\sum XY$, are the sum of the products and cross products to be calculated from the sample data, and '*n*' means the number of observations in the sample.

Example 2.49 explains the Least squares method.

Example 2.49: Fit a regression line $\hat{Y} = a + bX_i$ by the method of Least squares to the following sample information.

Observations	1	2	3	4	5	6	7	8	9	10
Income (*X*) (00 ₹)	41	65	50	57	96	94	110	30	79	65
Consumption Expenditure (*Y*) (00 ₹)	44	60	39	51	80	68	84	34	55	48

Solution:

We are to fit a regression line $\hat{Y} = a + bX_i$ to the given data by the method of Least squares. Accordingly, we work out the 'a' and 'b' values with the help of the normal equations as stated above and also for the purpose, work out $\sum X$, $\sum Y$, $\sum XY$, $\sum X^2$ values from the given sample information table on summations for regression equation.

Summations for Regression Equation

Observations	*Income X (00 ₹)*	*Consumption Expenditure Y (00 ₹)*	*XY*	X^2	Y^2
1	41	44	1804	1681	1936
2	65	60	3900	4225	3600
3	50	39	1950	2500	1521
4	57	51	2907	3249	2601
5	96	80	7680	9216	6400
6	94	68	6392	8836	4624
7	110	84	9240	12100	7056
8	30	34	1020	900	1156
9	79	55	4345	6241	3025
10	65	48	3120	4225	2304
$n = 10$	$\sum X = 687$	$\sum Y = 563$	$\sum XY = 42358$	$\sum X^2 = 53173$	$\sum Y^2 = 34223$

Putting the values in the required normal equations we have,

$$563 = 10a + 687b$$

$$42358 = 687a + 53173b$$

Solving these two equations for a and b we obtain,

$$a = 14.000 \quad \text{and} \quad b = 0.616$$

Hence, the equation for the required regression line is,

$$\hat{Y} = a + bX_i$$

or,

$$\hat{Y} = 14.000 + 0.616X_i$$

This equation is known as the regression equation of Y on X from which Y values can be estimated for given values of X variable.

Checking the accuracy of equation

After finding the regression line, one can check its accuracy also. The method to be used for the purpose follows from the mathematical property of a line fitted by the method of least squares, viz., the individual positive and negative errors must sum to zero. In other words, using the estimating equation one must find out whether the term $\sum\left(Y - \hat{Y}\right)$ is zero and if this is so, then one can reasonably be sure that he has not committed any mistake in determining the estimating equation.

NOTES

The problem of prediction

When we talk about prediction or estimation, we usually imply that if the relationship $Y_i = a + bX_i + e_i$ exists, then the regression equation, $\hat{Y} = a + bX_i$ provides a base for making estimates of the value of Y which will be associated with particular values of X. In Example 2.49, we worked out the regression equation for the income and consumption data as,

$$\hat{Y} = 14.000 + 0.616X_i$$

On the basis of this equation, we can make a point estimate of Y for any given value of X. Suppose we wish to estimate the consumption expenditure of individuals with income of ₹ 10,000. We substitute $X = 100$ for the same in our equation and get an estimate of consumption expenditure as,

$$\hat{Y} = 14.000 + 0.616(100) = 75.60$$

Thus, the regression relationship indicates that individuals with ₹ 10,000 of income may be expected to spend approximately ₹ 7,560 on consumption. However, this is only an expected or an estimated value and it is possible that actual consumption expenditure of same individual with that income may deviate from this amount and if so, then our estimate will be an error, the likelihood of which will be high if the estimate is applied to any one individual. The interval estimate method is considered better and it states an interval in which the expected consumption expenditure may fall. Remember that the wider the interval, the greater the level of confidence we can have, but the width of the interval (or what is technically known as the precision of the estimate) is associated with a specified level of confidence and is dependent on the variability (consumption expenditure in our case) found in the sample. This variability is measured by the standard deviation of the error term, 'e', and is popularly known as the standard error of the estimate.

Standard error of the estimate

Standard error of estimate is a measure developed by statisticians for measuring the reliability of the estimating equation. Like the standard deviation, the Standard Error (S.E.) of $\hat{Y}$ measures the variability or scatter of the observed values of Y around the regression line. Standard Error of Estimate (S.E. of $\hat{Y}$) is worked out as,

$$\text{S.E. of } \hat{Y} \text{ (or } S_e) = \sqrt{\frac{\sum (Y - \hat{Y})^2}{n-2}} = \sqrt{\frac{\sum e^2}{n-2}}$$

where, S.E. of $\hat{Y}$ (or S_e) = Standard error of the estimate

Y = Observed value of Y

$\hat{Y}$ = Estimated value of Y

e = The error term = $(Y - \hat{Y})$

n = Number of observations in the sample

Note: In the above formula, $n - 2$ is used instead of n because of the fact that two degrees of freedom are lost in basing the estimate on the variability of the sample observations about the line with two constants viz., 'a' and 'b' whose position is determined by those same sample observations.

The square of the S_e, also known as the variance of the error term, is the basic measure of reliability. The larger the variance, the more significant are the magnitudes of the e's and the less reliable is the regression analysis in predicting the data.

Interpreting the standard error of estimate and finding the confidence limits for the estimate in large and small samples

The larger the S.E. of estimate (SE_e), the greater happens to be the dispersion, or scattering, of given observations around the regression line. However, if the S.E. of estimate happens to be zero, then the estimating equation is a 'Perfect' estimator (i.e., cent per cent correct estimator) of the dependent variable.

(i) In case of large samples, i.e., where $n > 30$ in a sample, it is assumed that the observed points are normally distributed around the regression line and we may find that,

- 68 per cent of all points lie within $\hat{Y} \pm 1$ SE_e limits.
- 95.5 per cent of all points lie within $\hat{Y} \pm 2$ SE_e limits.
- 99.7 per cent of all points lie within $\hat{Y} \pm 3$ SE_e limits.

This can be stated as,

a. The observed values of Y are normally distributed around each estimated value of $\hat{Y}$ and;

b. The variance of the distributions around each possible value of $\hat{Y}$ is the same.

(ii) In case of small samples, i.e., where $n \leq 30$ in a sample the 't' distribution is used for finding the two limits more appropriately.

This is done as follows:

$$\text{Upper limit} = \hat{Y} + \text{`}t\text{'}\,(SE_e)$$

$$\text{Lower limit} = \hat{Y} - \text{`}t\text{'}\,(SE_e)$$

Where, $\hat{Y}$ = The estimated value of Y for a given value of X.

SE_e = The standard error of estimate.

't' = Table value of 't' for given degrees of freedom for a specified confidence level.

Some other details concerning simple regression

Sometimes the estimating equation of Y also known as the regression equation of Y on X, is written as,

$$\left(\hat{Y} - \overline{Y}\right) = r\frac{\sigma_Y}{\sigma_X}\left(X_i - \overline{X}\right)$$

NOTES

or, $\hat{Y} = r\frac{\sigma_Y}{\sigma_X}\left(X_i - \overline{X}\right) + \overline{Y}$

Where, r = Coefficient of simple correlation between X and Y

σ_Y = Standard deviation of Y

σ_X = Standard deviation of X

$\overline{X}$ = Mean of X

$\overline{Y}$ = Mean of Y

$\hat{Y}$ = Value of Y to be estimated

X_i = Any given value of X for which Y is to be estimated

This is based on the formula we have used, i.e., $\hat{Y} = a + bX_i$. The coefficient of X_i is defined as,

$$\text{Coefficient of } X_i = b = r\frac{\sigma_Y}{\sigma_X}$$

(Also known as regression coefficient of Y on X or slope of the regression line of Y on X) or b_{YX}.

$$= \frac{\sum XY - n\overline{X}\,\overline{Y} \times \sqrt{\sum Y^2 - n\overline{Y}^2}}{\sqrt{\sum Y^2 - n\overline{Y^2}}\sqrt{\sum X^2 - n\overline{X}^2}\sqrt{\sum X^2 - n\overline{X}^2}}$$

$$= \frac{\sum XY - n\overline{X}\,\overline{Y}}{\sum X^2 - n\overline{X}^2}$$

and $$a = -r\frac{\sigma_Y}{\sigma_X}\overline{X} + \overline{Y}$$

$$= \overline{Y} - b\overline{X} \qquad \left(\text{since } b = r\frac{\sigma_Y}{\sigma_X}\right)$$

Similarly, the estimating equation of X, also known as the regression equation of X on Y, can be stated as,

$$\left(\hat{X} - \overline{X}\right) = r\frac{\sigma_X}{\sigma_Y}\left(Y - \overline{Y}\right)$$

or $$\hat{X} = r\frac{\sigma_X}{\sigma_Y}\left(Y - \overline{Y}\right) + \overline{X}$$

and the

Regression coefficient of X on Y (or b_{XY}) $= r\frac{\sigma_X}{\sigma_Y} = \frac{\sum XY - n\overline{X}\,\overline{Y}}{\sum Y^2 - n\overline{Y}^2}$

If we are given the two regression equations as stated above, along with the values of 'a' and 'b' constants to solve the same for finding the value of X and Y, then the values of X and Y so obtained, are the mean values of X (i.e., $\overline{X}$) and the mean value of Y (i.e., $\overline{Y}$).

NOTES

If we are given the two regression coefficients (viz., b_{XY} and b_{YX}), then we can work out the value of coefficient of correlation by just taking the square root of the product of the regression coefficients as shown,

$$r = \sqrt{b_{YX}.b_{XY}}$$

$$= \sqrt{r\frac{\sigma_Y}{\sigma_X}.r\frac{\sigma_X}{\sigma_Y}}$$

$$= \sqrt{r.r} \quad = r$$

The (±) sign of r will be determined on the basis of the sign of the given regression coefficients. If regression coefficients have minus sign then r will be taken with minus (–) sign and if regression coefficients have plus sign then r will be taken with plus (+) sign, (Remember that both regression coefficients will necessarily have the same sign, whether it is minus or plus, for their sign is governed by the sign of coefficient of correlation.) To understand it better, Refer Examples 2.50 and 2.51.

Example 2.50: Given is the following information:

	$\overline{X}$	$\overline{Y}$
Mean	39.5	47.5
Standard Deviation	10.8	17.8

Simple correlation coefficient between X and Y is = + 0.42.

Find the estimating equation of Y and X.

Solution:

Estimating equation of Y can be worked out as,

$$\because \quad \left(\hat{Y} - \overline{Y}\right) = r\frac{\sigma_Y}{\sigma_X}\left(X_i - \overline{X}\right)$$

or
$$\hat{Y} = r\frac{\sigma_Y}{\sigma_X}\left(X_i - \overline{X}\right) + \overline{Y}$$

$$= 0.42\frac{17.8}{10.8}\left(X_i - 39.5\right) + 47.5$$

$$= 0.69X_i - 27.25 + 47.5$$

$$= 0.69X_i + 20.25$$

Similarly, the estimating equation of X can be worked out as

$$\because \quad \left(\hat{X} - \overline{X}\right) = r\frac{\sigma_X}{\sigma_Y}\left(Y_i - \overline{Y}\right)$$

or
$$\hat{X} = r\frac{\sigma_X}{\sigma_Y}\left(Y_i - \overline{Y}\right) + \overline{X}$$

or
$$= 0.42\frac{10.8}{17.8}\left(Y_i - 47.5\right) + 39.5$$

$$= 0.26Y_i - 12.35 + 39.5$$

$$= 0.26Y_i + 27.15$$

NOTES

Example 2.51: The following is the given data:

Variance of $X = 9$

Regression equations:

$$4X - 5Y + 33 = 0$$
$$20X - 9Y - 107 = 0$$

Find: (i) Mean values of X and Y

(ii) Coefficient of Correlation between X and Y

(iii) Standard deviation of Y

Solution:

(i) For finding the mean values of X and Y, we solve the two given regression equations for the values of X and Y as follows:

$$4X - 5Y + 33 = 0 \quad ...(1)$$
$$20X - 9Y - 107 = 0 \quad ...(2)$$

If we multiply Equation (1) by 5, we have the following equations:

$$20X - 25Y = -165 \quad ...(3)$$
$$20X - 9Y = 107 \quad ...(2)$$
$$- \quad + \qquad -$$
$$-16Y = -272$$

Subtracting Equations (2) from (3)

or $Y = 17$

Putting this value of Y in Equation (1) we have,

$$4X = -33 + 5(17)$$

or $X = \dfrac{-33+85}{4} = \dfrac{52}{4} = 13$

Hence, $\bar{X} = 13$ and $\bar{Y} = 17$

(ii) For finding the coefficient of correlation, first of all we presume one of the two given regression equations as the estimating equation of X. Let equation $4X - 5Y + 33 = 0$ be the estimating equation of X, then we have,

$$\hat{X} = \frac{5Y_i}{4} - \frac{33}{4}$$

and

From this we can write $b_{XY} = \dfrac{5}{4}$.

The other given equation is then taken as the estimating equation of Y and can be written as,

$$\hat{Y} = \frac{20X_i}{9} - \frac{107}{9}$$

and from this we can write $b_{YX} = \dfrac{20}{9}$.

If the above equations are correct then r must be equal to,

$$r = \sqrt{5/4 \times 20/9} = \sqrt{25/9} = 5/3 = 1.6$$

which is an impossible equation, since r can in no case be greater than 1. Hence, we change our supposition about the estimating equations and by reversing it, we re-write the estimating equations as,

$$\hat{X} = \frac{9Y_i}{20} + \frac{107}{20}$$

and $$\hat{Y} = \frac{4X_i}{5} + \frac{33}{5}$$

Hence, $$r = \sqrt{9/20 \times 4/5}$$

$$= \sqrt{9/25}$$

$$= 3/5$$

$$= 0.6$$

Since, regression coefficients have plus signs, we take $r = + 0.6$

(iii) Standard deviation of Y can be calculated,

$\because$ Variance of $X = 9$ $\therefore$ Standard deviation of $X = 3$

$$\because \quad b_{YX} = r\frac{\sigma_Y}{\sigma_X} = \frac{4}{5} = 0.6\frac{\sigma_Y}{3} = 0.2\sigma_Y$$

Hence, $\sigma_Y = 4$

Alternatively, we can work it out as,

$$\because \quad b_{XY} = r\frac{\sigma_X}{\sigma_Y} = \frac{9}{20} = 0.6\frac{\sigma_Y}{3} = \frac{1.8}{\sigma_Y}$$

Hence, $\sigma_Y = 4$

2.5.2 Polynomial Regression

In statistics, polynomial regression is a form of regression analysis in which the relationship between the independent variable x and the dependent variable y is modelled as an nth degree polynomial in x. Polynomial regression fits a nonlinear relationship between the value of x and the corresponding conditional mean of y, denoted $E(y|x)$. Although polynomial regression fits a nonlinear model to the data, as a statistical estimation problem it is linear, in the sense that the regression function $E(y|x)$ is linear in the unknown parameters that are estimated from the data. For this reason, polynomial regression is considered to be a special case of multiple linear regression.

The explanatory (independent) variables resulting from the polynomial expansion of the 'Baseline' variables are known as higher-degree terms. Such variables are also used in classification settings.

Ancient Times of Polynomial regression

Polynomial regression models are usually fit using the method of least squares. The least-squares method minimizes the variance of the unbiased estimators of the coefficients, under the conditions of the Gauss–Markov theorem. The least-squares method was published in 1805 by Legendre and in 1809 by Gauss. The first design of an experiment for polynomial regression appeared in an 1815 paper of Gergonne. In the twentieth century, polynomial regression played an important role in the development of regression analysis, with a greater emphasis on issues of design and inference. More recently, the use of polynomial models has been complemented by other methods, with non-polynomial models having advantages for some classes of problems.

Definition and Example of Polynomial Regression

The goal of regression analysis is to model the expected value of a dependent variable *y* in terms of the value of an independent variable (or vector of independent variables) *x*. In simple linear regression, the model

$$y = \beta_0 + \beta_1 x + \varepsilon,$$

is used, where ε is an unobserved random error with mean zero conditioned on a scalar variable *x*. In this model, for each unit increase in the value of *x*, the conditional expectation of *y* increases by β_1 units.

In many settings, such a linear relationship may not hold. For example, if we are modeling the yield of a chemical synthesis in terms of the temperature at which the synthesis takes place, we may find that the yield improves by increasing amounts for each unit increase in temperature. In this case, we might propose a quadratic model of the form

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon.$$

In this model, when the temperature is increased from *x* to *x* + 1 units, the expected yield changes by

$$\beta_1 + \beta_2(2x + 1).$$

(This can be seen by replacing x in this equation with *x*+1 and subtracting the equation in *x* from the equation in *x*+1.) For infinitesimal changes in *x*, the effect on *y* is given by the total derivative with respect to *x*: $\beta_1 + 2\beta_2 x.$

The fact that the change in yield depends on x is what makes the relationship between x and y nonlinear even though the model is linear in the parameters to be estimated.

In general, we can model the expected value of y as an nth degree polynomial, yielding the general polynomial regression model

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \cdots + \beta_n x^n + \varepsilon.$$

Conveniently, these models are all linear from the point of view of estimation, since the regression function is linear in terms of the unknown parameters β_0, β_1, ... therefore, for least squares analysis, the computational and inferential problems of polynomial regression can be completely addressed using the techniques of multiple regression. This is done by treating x, x^2, ... as being distinct independent variables in a multiple regression model.

Matrix form and Calculation of Estimates

The polynomial regression model

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \cdots + \beta_m x_i^m + \varepsilon_i \ (i = 1, 2, \ldots, n)$$

can be expressed in matrix form in terms of a design matrix **X** is a Vandermonde matrix, the invertibility condition is guaranteed to hold if all the x_i values are distinct. This is the unique least-squares solution.

Explanation of Polynomial Regression

Although polynomial regression is technically a special case of multiple linear regression, the interpretation of a fitted polynomial regression model requires a somewhat different perspective. It is often difficult to interpret the individual coefficients in a polynomial regression fit, since the underlying monomials can be highly correlated. For example, x and x^2 have correlation around 0.97 when x is uniformly distributed on the interval (0, 1). Although the correlation can be reduced by using orthogonal polynomials, it is generally more informative to consider the fitted regression function as a whole. Point-wise or simultaneous confidence bands can then be used to provide a sense of the uncertainty in the estimate of the regression function.

Alternative Approaches of Polynomial Regression

Polynomial regression is one example of regression analysis using basis functions to model a functional relationship between two quantities. More specifically, it replaces $x \in \mathbb{R}^{d_x}$ in linear regression with polynomial basis $\varphi(x) \in \mathbb{R}^{d_\varphi}$, e.g. $[1, x] \xrightarrow{\varphi} [1, x, x^2, \ldots, x^d]$. A drawback of polynomial bases is that the basis functions are 'Non-Local', meaning that the fitted value of y at a given value $x = x_0$ depends strongly on data values with x far from x_0. In modern statistics, polynomial basis-functions are used along with new basis functions, such as splines, radial basis functions, and wavelets. These families of basis functions offer a more parsimonious fit for many types of data.

The goal of polynomial regression is to model a non-linear relationship between the independent and dependent variables (technically, between the independent variable and the conditional mean of the dependent variable). This is similar to the goal of nonparametric regression, which aims to capture non-linear regression relationships. Therefore, non-parametric regression approaches such as smoothing can be useful alternatives to polynomial regression. Some of these methods make use of a localized form of classical polynomial regression. An advantage of traditional polynomial regression is that the inferential framework of multiple regression can be used (this also holds when using other families of basis functions such as splines). A final alternative is to use kernelized models such as support vector regression with a polynomial kernel. If residuals have unequal variance, a weighted least squares estimator may be used to account for that.

2.5.3 Fitting Exponential

In this section, we consider the problem of approximating an unknown function whose values, at a set of points, are generally known only empirically and are, thus subject to inherent errors, which may sometimes be appreciably larger in many engineering and scientific problems. In these cases, it is required to derive a functional relationship using certain experimentally observed data. Here the observed data may have inherent or round-off errors, which are serious, making polynomial interpolation for approximating the function inappropriate. In polynomial interpolation the truncation error in the approximation is considered to be important. But when the data contains round-off errors or inherent errors, interpolation is not appropriate.

The subject of this section is curve fitting by least square approximation. Here we consider a technique by which noisy function values are used to generate a smooth approximation. This smooth approximation can then be used to approximate the derivative more accurately than with exact polynomial interpolation.

There are situations where interpolation for approximating function may not be efficacious procedure. Errors will arise when the function values $f(x_i)$, $i = 1, 2, \ldots, n$ are observed data and not exact. In this case, if we use the polynomial interpolation, then it would reproduce all the errors of observation. In such situations one may take a large number of observed data, so that statistical laws in effect cancel the errors introduced by inaccuracies in the measuring equipment. The approximating function is then derived, such that the sum of the squared deviation between the observed values and the estimated values are made as small as possible.

Mathematically, the problem of curve fitting or function approximation may be stated as follows:

To find a functional relationship $y = g(x)$, that relates the set of observed data values $P_i(x_i, y_i)$, $i = 1, 2, \ldots, n$ as closely as possible, so that the graph of $y = g(x)$ goes near the data points P_i's though not necessarily through all of them.

The first task in curve fitting is to select a proper form of an approximating function $g(x)$, containing some parameters, which are then determined by minimizing the total squared deviation.

For example, $g(x)$ may be a polynomial of some degree or an exponential or logarithmic function. Thus $g(x)$ may be any of the following:

(*i*) $g(x) = \alpha + \beta x$ (*ii*) $g(x) = \alpha + \beta x + \gamma x^2$

(*iii*) $g(x) = \alpha e^{\beta x}$ (*iv*) $g(x) = \alpha e^{-\beta x}$

(*v*) $g(x) = \alpha \log(\beta x)$

Here α, β, γ are parameters which are to be evaluated so that the curve $y = g(x)$, fits the data well. A measure of how well the curve fits is called the goodness of fit.

In the case of least square fit, the parameters are evaluated by solving a system of normal equations, derived from the conditions to be satisfied so that the sum of the squared deviations of the estimated values from the observed values, is minimum.

Method of Least Squares

Let $(x_1, f_1), (x_2, f_2), \ldots, (x_n, f_n)$ be a set of observed values and $g(x)$ be the approximating function. We form the sums of the squares of the deviations of the observed values f_i from the estimated values $g(x_i)$, i.e., $$S = \sum_{i=1}^{n} \{f_i - g(x_i)\}^2 \quad (2.63)$$

The function $g(x)$ may have some parameters, α, β, γ. In order to determine these parameters we have to form the necessary conditions for S to be minimum, which are

$$\frac{\partial S}{\partial \alpha} = 0, \ \frac{\partial S}{\partial \beta} = 0, \ \frac{\partial S}{\partial \gamma} = 0 \quad (2.64)$$

These equations are called normal equations, solving which we get the parameters for the best approximate function $g(x)$.

Curve Fitting by a Straight Line: Let $g(x) = \alpha + \beta x$, be the straight line which fits a set of observed data points (x_i, y_i), $i = 1, 2, ..., n$.

Let S be the sum of the squares of the deviations $g(x_i) - y_i$, $i = 1, 2, ..., n$; given by,

$$S = \sum_{i=1}^{n} (\alpha + \beta x_i - y_i)^2 \quad (2.65)$$

We now employ the method of least squares to determine α and β, so that S will be minimum. The normal equations are,

$$\frac{\partial S}{\partial \alpha} = 0, \text{ i.e., } \sum_{i=1}^{n} (\alpha + \beta \ x_i - y_i) = 0 \quad (2.66)$$

and, $$\frac{\partial S}{\partial \beta} = 0, \text{ i.e., } \sum_{i=1}^{n} x_i(\alpha + \beta \ x_i - y_i) = 0 \quad (2.67)$$

These conditions give,

$$n\alpha + S_1\beta - S_{01} = 0$$
$$S_1\alpha + S_2\beta - S_{11} = 0$$

where, $S_1 = \sum_{i=1}^{n} x_i, \quad S_{01} = \sum_{i=1}^{n} y_i, \quad S_2 = \sum_{i=1}^{n} x_i^2, \quad S_{11} = \sum_{i=1}^{n} x_i y_i$

Solving,

$$\frac{\alpha}{-S_1S_{11} + S_1S_2} = \frac{\beta}{nS_{11} - S_1S_{01}} = \frac{1}{nS_2 - S_1^2}. \quad \text{Also} \quad \alpha = \frac{S_{01}}{n} - \beta\frac{S_1}{n}.$$

Algorithm. Fitting a straight line $y = a + bx$.

Step 1. Read n [n being the number of data points]

Step 2. Initialize : sum $x = 0$, sum $x^2 = 0$, sum $y = 0$, sum $xy = 0$

Step 3. For $j = 1$ to n compute

Begin

Read data x_j, y_j

Compute sum x = sum $x + x_j$

Compute sum $x^2 + x_j \times x_j$

Compute sum y = sum $y + y_i \times y_j$

Compute sum xy = sum $xy + x_j \times y_j$

End

Step 4. Compute $b = (n \times \text{sum } xy - \text{sum } x \times \text{sum } y)/ (n \times \text{sum } x^2 - (\text{sum } x)^2)$

Step 5. Compute x bar = sum x / n

Step 6. Compute y bar = sum y / n

Step 8. Compute a = y bar – $b \times x$ bar

Step 9. Write a, b

Step 10. For $j = 1$ to n

Begin

Compute y estimate = $a + b \times x$

write x_j, y_j, y estimate

End

Step 11. Stop

Curve Fitting by a Quadratic (A Parabola): Let $g(x) = a + bx + cx^2$, be the approximating quadratic to fit a set of data (x_i, y_i), $i = 1, 2, ..., n$. Here the parameters are to be determined by the method of least squares, i.e., by minimizing the sum of the squares of the deviations given by,

$$S = \sum_{i=1}^{n} (a + bx_i + cx_i^2 - y_i)^2 \tag{2.68}$$

Thus the normal equations, $\frac{\partial S}{\partial a} = 0, \frac{\partial S}{\partial b} = 0, \frac{\partial S}{\partial c} = 0$, are as follows:

(2.69)

$$\sum_{i=1}^{n} (a + bx_i + cx_i^2 - y_i) = 0$$

$$\sum_{i=1}^{n} x_i (a + bx_i + cx_i^2 - y_i) = 0$$

$$\sum_{i=1}^{n} x_i^2 (a + bx_i + cx_i^2 - y_i) = 0. \tag{2.70}$$

These equations can be rewritten as,

$$na + s_1 b + s_2 c - s_{01} = 0$$
$$s_1 a + s_2 b + s_3 c - s_{11} = 0$$
$$s_2 a + s_3 b + s_4 c - s_{21} = 0 \tag{2.71}$$

Where, $s_1 = \sum_{i=1}^{n} x_i, \quad s_2 = \sum_{i=1}^{n} x_i^2, \quad s_3 = \sum_{i=1}^{n} x_i^3, \quad s_4 = \sum_{i=1}^{n} x_i^4$

$$s_{01} = \sum_{i=1}^{n} y_i, \quad s_{11} = \sum_{i=1}^{n} x_i y_i, \quad s_{21} = \sum_{i=1}^{n} x_i^2 y_i \tag{2.72}$$

NOTES

It is clear that the normal equations form a system of linear equations in the unknown parameters a, b, c. The computation of the coefficients of the normal equations can be made in a tabular form for desk computations as shown below,

	x	x_i	y_i	x_i^2	x_i^3	x_i^4	$x_i y_i$	$x_i^2 y_i$
	1	x_1	y_1	x_1^2	x_1^3	x_1^4	$x_1 y_1$	$x_1^2 y_1$
	2	x_2	y_2	x_2^2	x_2^3	x_2^4	$x_2 y_2$	$x_2^2 y_2$
	...	...	...	...	...	...	...	...
	n	x_n	y_n	x_n^2	x_n^3	x_n^4	$x_n y_n$	$x_n^2 y_n$
sum		s_1	s_{01}	s_2	s_3	s_4	s_{11}	s_{21}

The system of linear equations can be solved by Gaussian elimination method. It may be noted that number of normal equations is equal to the number of unknown parameters.

Example 2.52. Find the straight line fitting the following data:

x_i	4	6	8	10	12
y_1	13.72	12.90	12.01	11.14	10.31

Solution: Let $y = a + bx$, be the straight line which fits the data. We have the normal equations $\frac{\partial S}{\partial a} = 0$, $\frac{\partial S}{\partial b} = 0$ for determining a and b, where

$$S = \sum_{i=1}^{5} (y_i - a - bx_i)^2$$

Thus, $$\sum_{i=1}^{5} y_i - na - b\sum_{i=1}^{5} x_i = 0$$

and $$\sum_{i=1}^{5} x_i y_i - a\sum_{I=1}^{5} x_i - b\sum_{i=1}^{5} x_i^2 = 0$$

The coefficients are computed in the table below,

	x_i	y_i	x_i^2	$x_i y_i$
	4	13.72	16	54.88
	6	12.90	36	77.40
	8	12.01	64	96.08
	10	11.14	100	111.40
	12	10.31	144	123.72
Sum	40	60.08	360	463.48

Thus the normal equations are,

$$5a + 40b - 60.08 = 0$$
$$40a + 360b - 463.48 = 0$$

Solving these two equations we obtain,

$$a = 15.448,\ b = 0.429$$

Thus $y = g(x) = 15.448 - 0.429\,x$, is the straight line fitting the data.

Example 2.53. Use the method of least square approximation to fit a straight line to the following observed data.

NOTES

x_i	60	61	62	63	64
y_i	40	40	48	52	55

Solution: Let the straight line fitting the data be $y = a + bx$. The data values being large, we can use a change in variable by substituting $u = x - 62$, and $v = y - 48$.

Let $v = A + B\,u$, be a straight line fitting the transformed data, where the normal equations for A and B are,

$$\sum_{i=1}^{5} v_i = 5A + B\sum_{i=1}^{5} u_i$$

$$\sum_{i=1}^{5} u_i v_i = A\sum_{i=1}^{5} u_i + B\sum_{i=1}^{5} u_i^2$$

The computation of the various sums are given in the table below,

	x_i	y_i	u_i	v_i	$u_i v_i$	u_i^2
	60	40	–2	–8	16	4
	61	42	–1	–6	6	1
	62	48	0	0	0	0
	63	52	1	4	4	1
	64	55	2	7	14	4
Sum			0	–3	40	10

Thus the normal equation are,

$$-3 = 5A \quad \text{and} \quad 40 = 10B$$

$$\therefore\ A = -\frac{3}{5}, \quad \text{and} \quad B = 4$$

This gives the line, $v = -3/5 + 4u$

or, $20u - 5v - 3 = 0.$

Transforming we get the line,

$$20\,(x - 62) - 5\,(y - 48) - 3 = 0$$

or, $$20\,x - 5y - 1003 = 0$$

Curve Fitting with an Exponential Curve: We consider a two parameter exponential curve as

$$y = ae^{-bx} \tag{2.73}$$

For determining the parameters, we can apply the principle of least square by first using a transformation,

$z = \log y$, so that Equation (2.73) is rewritten as, (2.74)

i.e., $z = \log a - bx$ (2.75)

NOTES

Thus we have to fit a linear curve of the form $z = \alpha + \beta x$, with $z-x$ variables and then get the parameters a and b as,

$$a = e^{\alpha}, \; b = -\beta \tag{2.76}$$

Thus proceeding as in linear curve fitting,

$$\beta = \frac{n\sum_{i=1}^{n} x_i \log y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} \log y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \tag{6.77}$$

and, $\alpha = \bar{z} - p\bar{x}$, where $\bar{x} = \left(\sum_{i=1}^{n} x_i\right)\Big/ n, \; \bar{z} = \left(\sum_{i=1}^{n} \log \;\; y_i\right)\Big/ n$ (6.78)

After computing α and β, we can determine a and b given by equation (2.75). Finally, the exponential curve fitting the data set is given by equation (2.73).

Algorithm. To fit a straight line for a given set of data points by least square error method.

Step 1. Read the number of data points, i.e., n

Step 2. Read values of data-points, i.e., Read (x_i, y_i) for $i = 1, 2, ..., n$

Step 3 Initialize the sums to be computed for the normal equations,

i.e., $sx = 0$, $sx^2 = 0$, $sy = 0$, $syx = 0$

Step 4. Compute the sums, i.e., For $i = 1$ to n do

Begin

$$sx = sx + x_i$$
$$sx^2 = sx^2 + x_i^2$$
$$sy = sy + y_i$$
$$syx = syx + x_i y_i$$

End

Step 5. Solve the normal equations, i.e., solve for a and b of the line $y = a + bx$

Compute $d = n * sx^2 - sx * sx$

$$b = (n * sxy - sy * sx) / d$$
$$x\text{bar} = sx / n$$
$$y\text{bar} = sy / n$$
$$a = y\text{bar} - b * x\,\text{bar}$$

Step 6. Print values of a and b

Step 7. Print a table of values of $x_i, \; y_i, \; y_{pi} = a + bx_i$ for $i = 1, 2, ..., n$

Step 8. Stop

Algorithm. To fit a parabola $y = a + bx + cx^2$, for a given set of data points by least square error method.

NOTES

Step 1. Read n, the number of data points

Step 2. Read (x_i, y_i) for $i = 1, 2, ..., n$; the values of data points

Step 3. Initialize the sum to be computed for the normal equations, i.e., $sx = 0$, $sx^2 = 0$, $sx^3 = 0$, $sx^4 = 0$, $sy = 0$, $sxy = 0$.

Step 4. Compute the sums, i.e., For $i = 1$ to n do

Begin

$$sx = sx + x_i$$
$$x^2 = x_i * x_i$$
$$sx^2 = sx^2 + x^2$$
$$sx^3 = sx^3 + x_i * x^2$$
$$sx^4 = sx^4 + x^2 * x^2$$
$$sy = sy + y_i$$
$$sxy = sxy + x_i * y_i$$
$$sx^2y = sx^2y + x^2 * y_i$$

End

Step 5. Form the coefficients $\{a_{ij}\}$ matrix of the normal equations, i.e.,

$$a_{11} = n, \quad a_{21} = sx, \quad a_{31} = sx^2$$
$$a_{12} = sx, \quad a_{22} = sx^2, \quad a_{32} = sx^3$$
$$a_{13} = sx^2, \quad a_{23} = sx^3, \quad a_{33} = sx^4$$

Step 6. Form the constant vector of the normal equations.

$b_1 = sy$, $b_2 = sxy$, $b_3 = sx^2y$

Step 7. Solve the normal equation by Gauss-Jordan method

$$a_{12} = a_{12} / a_{11},\ a_{13} = a_{13} / a_{11},\ b_1 = b_1 / a_{11}$$
$$a_{22} = a_{22} - a_{21}a_{12},\ a_{23} = a_{23} - a_{21}a_{13}$$
$$b_2 = b_2 - b_1a_{21}$$

$$a_{32} = a_{32} - a_{31}a_{12}$$
$$a_{33} = a_{33} - a_{31}a_{13}$$
$$b_3 = b_3 - b_1a_{31}$$
$$a_{23} = a_{23} / a_{22}$$
$$b_2 = b_2 / a_{22}$$
$$a_{33} = a_{33} - a_{23}\,a_{32}$$
$$b_3 = b_3 - a_{32}\,b_2$$
$$c = b_3 / a_{33}$$
$$b = b_2 - c\,a_{23}$$
$$a = b_1 - b\,a_{12} - c\,a_{13}$$

Step 8. Print values of a, b, c (the coefficients of the parabola)

Step 9. Print the table of values of x_k, y_k and y_{pk} where $y_{pk} = a + bx_k + cx^2k$,

i.e., print x_k, y_k, y_{pk} for $k = 1, 2, ..., n.$

Step 10. Stop.

NOTES

Check Your Progress

9. When does an error arise in function interpolation?
10. How is approximating function found in the method of least squares?
11. Define the term first quadrant.
12. List the two methods used for estimation.
13. What are the reasons behind squaring each terms in the least square method?

2.6 ANSWERS TO 'CHECK YOUR PROGRESS'

1. In this method, we successively generate interpolating polynomials of any degree by iteratively using linear interpolating functions.
2. It can be stated explicitly as 'given a set of $(n + 1)$ values $y_0, y_1, y_2, ..., y_n$ for $x = x_0, x_1, x_2, ..., x_n$ respectively. The problem of interpolation is to compute the value of the function $y = f(x)$ for some non-tabular value of x.'
3. Lagrange's interpolation is useful for unequally spaced tabulated values.
4. For interpolation of an unknown function when the tabular values of the argument x are equally spaced, we have two important interpolation formulae, viz.,
   (a) Newton's forward difference interpolation formula
   (b) Newton's backward difference interpolation formula
5. The shift operator is denoted by E and is defined by $E f(x) = f(x + h)$.
6. The Newton's forward difference interpolation formula is a polynomial of degree less than or equal to n.
7. The interpolating polynomials are usually used for finding values of the tabulated function $y = f(x)$ for a value of x within the table. But they can also be used in some cases for finding values of $f(x)$ for values of x near to the end points x_0 or x_n outside the interval $[x_0, x_n]$. This process of finding values of $f(x)$ at points beyond the interval is termed as extrapolation.
8. The problem of inverse interpolation in a table of values of $y = f(x)$ is to find the value of x for a given y.
9. There are situations where interpolation for approximating function may not be efficacious procedure. Errors will arise when the function values $f(x_i)$, $i = 1, 2, \ldots, n$ are observed data and not exact.
10. Let $(x_1, f_1), (x_2, f_2), ..., (x_n, f_n)$ be a set of observed values and $g(x)$ be the approximating function. We form the sums of the squares of the deviations of the observed values f_i from the estimated values $g(x_i)$,

    i.e., $S = \sum_{i=1}^{n} \{f_i - g(x_i)\}^2$
11. If the revolving line OP is in the first quadrant, then all the sides of the triangle OPM are positive. Therefore, all the trigonometrical ratios are positive in the first quadrant.

NOTES

12. The following are the two methods used for estimation:
    (i) Scatter diagram method
    (ii) Least squares method
13. 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 when squared becomes positive).

2.7 SUMMARY

- The problem of interpolation is very fundamental problem in numerical analysis.
- In numerical analysis, interpolation means computing the value of a function $f(x)$ in between values of x in a table of values.
- Lagrange's interpolation is useful for unequally spaced tabulated values.
- For interpolation of an unknown function when the tabular values of the argument x are equally spaced, we have two important interpolation formulae, viz., Newton's forward difference interpolation formula and Newton's backward difference interpolation formula.
- The forward difference operator is defined by, $\Delta f(x) = f(x+h) - f(x)$.
- The backward difference operator is defined by, $\Delta f(x) = f(x+h) - f(x)$.
- We define different types of finite differences such as forward differences, backward differences and central differences, and express them in terms of operators.
- The shift operator is denoted by E and is defined by $Ef(x) = f(x+h)$.
- The first order difference of a polynomial of degree n is a polynomial of degree n–1. For polynomial of degree n, all other differences having order higher than n are zero.
- Newton's forward difference interpolation formula is generally used for interpolating near the beginning of the table while Newton's backward difference formula is used for interpolating at a point near the end of a table.
- In iterative linear interpolation, we successively generate interpolating polynomials, of any degree, by iteratively using linear interpolating functions.
- The process of finding values of a function at points beyond the interval is termed as extrapolation.
- Horner's method of synthetic substitution is used for evaluating the values of a polynomial and its derivatives for a given x.
- Descarte's rule is used to determine the number of negative roots by finding the number of changes of signs in $p_n(-x)$.
- By using the method of least squares, noisy function values are used to generate a smooth approximation. This smooth approximation can then be used to approximate the derivative more accurately than with exact polynomial interpolation.

- The term 'Regression' was first used in 1877 by Sir Francis Galton who made a study that showed that the height of children born to tall parents will tend to move back or 'Regress' toward the mean height of the population.

2.8 KEY TERMS

- **Interpolation:** Interpolation means computing the value of a function $f(x)$ in between values of x in a table of values.
- **Extrapolation:** The process of finding values of a function at points beyond the interval is termed as extrapolation.
- **Newton-Raphson method:** Newton-Raphson method is a widely used numerical method for finding a root of an equation $f(x) = 0$, to the desired accuracy.
- **First quadrant:** If the revolving line *OP* is in the first quadrant, then all the sides of the triangle *OPM* are positive. Therefore, all the trigonometrical ratios are positive in the first quadrant.
- **Scatter diagram:** A diagram representing two series with the known variable, 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 on a graph paper.

2.9 SELF-ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. What is the significance of polynomial interpolation?
2. Define the symbolic operators E and D.
3. What is the degree of the first order forward difference of a polynomial of degree n?
4. What is the degree of the nth order forward difference of a polynomial of degree n?
5. Write newton's forward and backward difference formulae.
6. State an application of iterative linear interpolation.
7. What is the advantage of extrapolation?
8. State Lagrange's formula for inverse interpolation.
9. How many roots are there in a polynomial equation of degree n?
10. How many positive real roots are there in a polynomial equation?
11. Define the term first quadrant.
12. List the basic precautions and limitations of regression and correlation analyses.
13. Differentiate between scatter diagram and least square method.

NOTES

Long-Answer Questions

1. Use Lagrange's interpolation formula to find the polynomials of least degree which attain the following tabular values:

(a)

x	–2	1	2
y	25	–8	–15

(b)

x	0	1	2	5
y	2	3	12	147

(c)

x	1	2	3	4
y	–1	–1	1	5

2. Form the finite difference table for the given tabular values and find the values of:

(a) $\Delta f(2)$

(b) $\Delta f^2(1)$

(c) $\Delta f^3(0)$

(d) $\Delta f^4(1)$

(e) $f(5)$

(f) $f(3)$

x	0	1	2	3	4	5	6
$f(x)$	3	4	13	36	79	148	249

3. How are the forward and backward differences in a table related? Prove the following:

(a) $\Delta y_i = \nabla y_{i+1}$

(b) $\Delta^2 y_i = \nabla^2 y_{i+2}$

(c) $\Delta^n y_i = \nabla^n y_{i+n}$

4. Describe Newton's forward and backward difference formulae using illustrations.
5. Explain iterative linear interpolation with the help of examples.
6. Illustrate inverse interpolation procedure.
7. Use the method of least squares to fit a straight line for the following data points:

x	–1	0	1	2	3	4	5	6
y	10	9	7	5	4	3	0	–1

8. Discuss about the trigonometric function with the help of giving examples.
9. What is regression analysis? What are the assumptions in it?
10. Explain scatter diagram and the least square method in detail. Also, mention how scatter diagram helps in studying correlation between two variables.

2.10 FURTHER READING

NOTES

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

Chandan, J.S., Jagjit Singh and K.K. Khanna. 1995. *Business Statistics*. New Delhi: Vikas Publishing House.

Elhance, D.N. 2006. *Fundamental of Statistics*. Allahabad: Kitab Mahal.

Freud, J.E., and F.J. William. 1997. *Elementary Business Statistics – The Modern Approach*. New Jersey: Prentice-Hall International.

Goon, A.M., M.K. Gupta, and B. Das Gupta. 1983. *Fundamentals of Statistics*. Vols. I & II, Kolkata: The World Press Pvt. Ltd.

Gupta, S.C. 2008. *Fundamentals of Business Statistics*. Mumbai: Himalaya Publishing House.

Kothari, C.R. 1984. *Quantitative Techniques*. New Delhi: Vikas Publishing House.

Levin, Richard. I., and David. S. Rubin. 1997. *Statistics for Management.* New Jersey: Prentice-Hall International.

Meyer, Paul L. 1970. *Introductory Probability and Statistical Applications*. Massachusetts: Addison-Wesley.

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

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

Anderson, David R., Dennis J. Sweeney and Thomas A. Williams. *Essentials of Statistics for Business and Economics*. Mumbai: Thomson Learning, 2007.

S.P. Gupta. 2021. *Statistical Methods*. Delhi: Sultan Chand and Sons.

UNIT 3 NUMERICAL DIFFERENTIATION AND INTEGRATION

NOTES

Structure

3.0 INTRODUCTION

In numerical analysis, numerical differentiation is the process of finding the numerical value of a derivative of a given function at a given point. It is the process of computing the derivatives of a function $f(x)$ when the function is not explicitly known, but the values of the function are known only at a given set of arguments $x = x_0, x_1, x_2,..., x_n$. For finding the derivatives, a suitable interpolating polynomial is used and then its derivatives are used as the formulae for the derivatives of the function. Thus, for computing the derivatives at a point near the beginning of an equally spaced table, Newton's forward difference interpolation formula is used, whereas Newton's backward difference interpolation formula is used for computing the derivatives at a point near the end of the table.

Numerical integration constitutes a broad family of algorithms for calculating the numerical value of a definite integral. The numerical computation of an integral is sometimes called quadrature. The most straightforward numerical integration technique uses the Newton-Cotes formulas, which approximate a function tabulated at a sequence of regularly spaced intervals by various degree polynomials. If the functions are known analytically instead of being tabulated at equally spaced intervals, the best numerical method of integration is called Gaussian quadrature.

The basic problem considered by numerical integration is to compute an approximate solution to a definite integral $\int_a^b f(x)\,dx$. If $f(x)$ is a smooth well performed function integrated over a small number of dimensions and the limits of integration are bounded then there are many methods of approximating the integral with arbitrary precision. Numerical integration methods can generally be described as combining evaluations of the integrand to get an approximation to the integral. The integrand is evaluated at a finite set of points called integration points and a weighted sum of these values is used to approximate the integral. The integration points and weights depend on the specific method used and the accuracy required from the approximation. Modern numerical integrations methods based on information theory have been developed to simulate information systems such as computer controlled systems, communication systems, and control systems.

An ordinary differential equation is a relation that contains functions of only one independent variable and one or more of their derivatives with respect to that variable. Ordinary differential equations are distinguished from partial differential equations, which involve partial derivatives of functions of several variables. Ordinary differential equations arise in many different contexts including geometry, mechanics, astronomy and population modelling. The Picard—Lindelöf theorem, Picard's existence theorem or Cauchy–Lipschitz theorem is an important theorem on existence and uniqueness of solutions to first-order equations with given initial conditions. The Picard method is a way of approximating solutions of ordinary differential equations. Originally, it was a way of proving the existence of solutions. It is only by advanced symbolic computing that it has become a practical way of approximating solutions. Euler's method is a first-order numerical procedure for solving ordinary differential equations with a given initial value. It is the most basic kind of explicit method for numerical integration of ordinary differential equations and is the simplest kind of Runge-Kutta method.

In this unit, you will learn about the numerical differentiation formulae, Simpson's rule, errors in integration formulae, Gaussian quadrature formulae, solving numerical differential equation, Euler's method, Taylor series method, Runge-Kutta method and higher order differential equation.

3.1 OBJECTIVES

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

- Describe numerical differentiation
- Differentiate using Newton's forward difference interpolation formula
- Differentiate using Newton's backward difference interpolation formula
- Describe numerical integration
- Identify the numerical methods for evaluating a definite integral
- Know Newton-Cotes general quadrature
- Understand Simpson's one-third and three-eighth rule
- Explain interval halving technique

NOTES

- Numerically evaluate double integrals solution of non-linear equation
- Define Picard's method of successive approximation
- Describe Euler's method and Taylor series method
- Explain Runge-Kutta and multistep methods
- Understand predictor-corrector methods
- Find numerical solution of boundary value problems
- Define higher order differential equations

3.2 NUMERICAL DIFFERENTIATION FORMULAE

Numerical differentiation is the process of computing the derivatives of a function $f(x)$ when the function is not explicitly known, but the values of the function are known only at a given set of arguments $x = x_0, x_1, x_2, ..., x_n$. For finding the derivatives, we use a suitable interpolating polynomial and then its derivatives are used as the formulae for the derivatives of the function. Thus, for computing the derivatives at a point near the beginning of an equally spaced table, Newton's forward difference interpolation formula is used, whereas Newton's backward difference interpolation formula is used for computing the derivatives at a point near the end of the table. Again, for computing the derivatives at a point near the middle of the table, the derivatives of the central difference interpolation formula is used. If, however, the arguments of the table are unequally spaced, the derivatives of the Lagrange's interpolating polynomial are used for computing the derivatives of the function.

3.2.1 Differentiation Using Newton's Forward Difference Interpolation Formula

Let the values of an unknown function $y = f(x)$ be known for a set of equally spaced values $x_0, x_1, \ldots, x_n$ of x, where $x_r = x_0 + r_h$. Newton's forward difference interpolation formula is,

$$\varphi(u) = y_0 + u \ \Delta \ y_0 + \frac{u(u-1)}{2\ !}\Delta^2 y_0 + \frac{u(u-1)(u-2)}{3\ !}\Delta^3 y_0 + ... + \frac{u(u-1)(u-2)...(u-n+1)}{n\ !}\Delta^n y_0$$

where $u = \dfrac{x - x_0}{h}$

The derivative $\dfrac{dy}{dx}$ can be evaluated as,

$$\frac{dy}{dx} = \frac{d}{dx}\{\varphi(u)\} = \frac{d\varphi}{du} \cdot \frac{du}{dx} = \frac{1}{h}\frac{d\varphi}{du}$$

Thus, $$y'(x) \simeq \frac{1}{h}\left[\Delta y_0 + \frac{2u-1}{2}\Delta^2 y_0 + \frac{3u^2-6u+2}{6}\Delta^3 y_0 + \frac{2u^3-9u^2+11u-3}{12}\Delta^4 y_0 + ...\right] \quad (3.1)$$

Similarly, $y''(x) \simeq \dfrac{1}{h^2}\varphi''(u)$

Or, $$y''(x) = \frac{1}{h^2}\left[\Delta^2 y_0 + (u-1)\Delta^3 y_0 + \frac{6u^2 - 18u + 11}{12}\Delta^4 y_0 + ...\right] \qquad (3.2)$$

NOTES

For a value of x near the beginning of a table, $u = (x - x_o)/h$ is computed first and then Equations (3.1) and (3.2) can be used to compute $f'(x)$ and $f''(x)$. At the tabulated point x_0, the value of u is zero and the formulae for the derivatives are given by,

$$y'(x_0) = \frac{1}{h}\left[\Delta y_0 - \frac{1}{2}\Delta^2 y_0 + \frac{1}{3}\Delta^3 y_0 - \frac{1}{4}\Delta^4 y_0 + \frac{1}{5}\Delta^5 y_0 - ...\right] \qquad (3.3)$$

$$y''(x_0) = \frac{1}{h^2}\left[\Delta^2 y_0 - \Delta^3 y_0 + \frac{11}{12}\Delta^4 y_0 - \frac{5}{6}\Delta^5 y_0 + ...\right] \qquad (3.4)$$

3.2.2 Differentiation Using Newton's Backward Difference Interpolation Formula

For an equally spaced table of a function, Newton's backward difference interpolation formula is,

$$\varphi(v) = y_n + v\ \nabla y_n + \frac{v(v+1)}{2\ !}\nabla^2 y_n + \frac{v(v+1)(v+2)}{3\ !}\nabla^3 y_n + \frac{v(v+1)(v+2)(v+3)}{4\ !}\nabla^4 y_n + ...$$
$$+ \frac{v(v+1)...(v+n-1)}{n\ !}\nabla^n y_n$$

where $v = \dfrac{x - x_n}{h}$

The derivatives $\dfrac{dy}{dx}$ and $\dfrac{d^2 y}{dx^2}$, obtained by differentiating the above formula are given by,

$$\frac{dy}{dx} = \frac{1}{h}\left[\nabla y_n + \frac{2v+1}{2}\nabla^2 y_n + \frac{3v^2 + 6v + 2}{6}\nabla^3 y_n + \frac{2v^3 + 9v^2 + 11v + 3}{12}\nabla^4 y_n + ...\right] \qquad (3.5)$$

$$\frac{d^2 y}{dx^2} = \frac{1}{h^2}\left[\nabla^2 y_n + (v+1)\nabla^3 y_n + \frac{6v^2 + 18v + 11}{12}\nabla^4 y_n + ...\right] \qquad (3.6)$$

For a given x near the end of the table, the values of $\dfrac{dy}{dx}$ and $\dfrac{d^2 y}{dx^2}$ are computed by first computing $v = (x - x_n)/h$ and using the above formulae. At the tabulated point x_n, the derivatives are given by,

$$y'(x_n) = \frac{1}{h}\left[\nabla y_n + \frac{1}{2}\nabla^2 y_n + \frac{1}{3}\nabla^3 y_n + \frac{1}{4}\nabla^4 y_n + ...\right] \qquad (3.7)$$

$$y''(x_n) = \frac{1}{h^2}\left[\nabla^2 y_n + \nabla^3 y_n + \frac{11}{12}\nabla^4 y_n + \frac{5}{6}\nabla^5 y_n + ...\right] \qquad (3.8)$$

Example 3.1: Compute the values of $f'(2.1)$, $f''(2.1)$, $f'(2.0)$ and $f''(2.0)$ when $f(x)$ is not known explicitly, but the following table of values is given:

x	$f(x)$
2.0	0.69315
2.2	0.78846
2.4	0.87547

Solution: Since the points are equally spaced, we form the finite difference table.

x	$f(x)$	$\Delta f(x)$	$\Delta^2 f(x)$
2.0	0.69315		
		9531	
2.2	0.78846		-83
		8701	
2.4	0.87547		

For computing the derivatives at $x = 2.1$, we have

$$f'(x) \simeq \frac{1}{h}[\Delta f_0 + \frac{2u-1}{2}\Delta^2 f_0] \quad \text{and} \quad f''(x) \simeq \frac{1}{h^2}\Delta^2 f_0$$

$$u = \frac{x - x_0}{h} = \frac{2.1 - 2.0}{0.2} = 0.5$$

$$\therefore \qquad f'(2.1) = \frac{1}{0.2}\left[0.09531 + \frac{2 \times 0.5 - 1}{2}\Delta^2 f_0\right] = 0.4765$$

$$f''(2.1) = \frac{1}{(0.2)^2} \times (-0.00083) = -0.21$$

The value of $f'(2.0)$ is given by,

$$f'(2.0) = \frac{1}{0.2}\left[\Delta f_0 - \frac{1}{2}\Delta^2 f_0\right]$$

$$= \frac{1}{0.2}\left[0.09531 + \frac{1}{2} \times 0.00083\right]$$

$$= \frac{0.09572}{0.2} = 0.4786$$

$$f''(2.0) = \frac{1}{(0.2)^2} \times (-0.0083)$$

$$= -0.21$$

Example 3.2: For the function $f(x)$ whose values are given in the table below compute values of $f'(1),\ f''(1),\ f'(5.0),\ f''(5.0)$.

x	1	2	3	4	5	6
$f(x)$	7.4036	7.7815	8.1291	8.4510	8.7506	9.0309

Solution: Since $f(x)$ is known at equally spaced points, we form the finite difference table to be used in the differentiation formulae based on Newton's interpolating polynomial.

NOTES

x	$f(x)$	$\Delta f(x)$	$\Delta^2 f(x)$	$\Delta^3 f(x)$	$\Delta^4 f(x)$	$\Delta^5 f(x)$
1	7.4036					
		0.3779				
2	7.7815		−303			
		0.3476		46		
3	8.1291		−257		−12	
		0.3219		34		8
4	8.4510		−223		−4	
		0.2996		30		
5	8.7506		−193			
		0.2803				
6	9.0309					

To calculate $f'(1)$ and $f''(1)$, we use the derivative formulae based on Newton's forward difference interpolation at the tabulated point given by,

$$f'(x_0) = \frac{1}{h}\left[\Delta f_0 - \frac{1}{2}\Delta^2 f_0 + \frac{1}{3}\Delta^3 f_0 - \frac{1}{4}\Delta^4 f_0 + \frac{1}{5}\Delta^5 f_0\right]$$

$$f''(x_0) = \frac{1}{h^2}\left[\Delta^2 f_0 - \Delta^3 f_0 + \frac{11}{12}\Delta^4 f_0 - \frac{5}{6}\Delta^5 f_0\right]$$

$$\therefore f'(1) = \frac{1}{1}\left[0.3779 - \frac{1}{2}\times(-0.0303) + \frac{1}{3}\times 0.0046 - \frac{1}{4}\times(-0.0012) + \frac{1}{5}\times 0.0008\right]$$

$$= 0.39507$$

$$f''(1) = \left[0.0303 - 0.0046 + \frac{11}{12}\times(-0.0012) - \frac{5}{6}\times 0.0008\right]$$

$$= -0.0367$$

Similarly, for evaluating $f'(5.0)$ and $f''(5.0)$, we use the following formulae

$$f'(x_n) = \frac{1}{h}\left[\nabla f_n + \frac{1}{2}\nabla^2 f_n + \frac{1}{3}\nabla^3 f_n + \frac{1}{4}\nabla^4 f_n + \frac{1}{5}\nabla^5 f_n\right]$$

$$f''(x_n) = \frac{1}{h^2}\left[\nabla^2 f_n + \nabla^3 f_n + \frac{11}{12}\nabla^4 f_n + \frac{5}{6}\nabla^5 f_n\right]$$

$$f'(5) = \left[0.2996 + \frac{1}{2}(-0.0223) + \frac{1}{3}\times 0.0034 + \frac{1}{4}(-0.0012)\right]$$

$$= 0.2893$$

$$f''(5) = [-0.0223 + 0.0034 + \frac{11}{12}\times 0.0012]$$

$$= -0.0178$$

Example 3.3: Compute the values of $y'(0)$, $y''(0.0)$, $y'(0.02)$ and $y''(0.02)$ for the function $y = f(x)$ given by the following tabular values:

x	0.0	0.05	0.10	0.15	0.20	0.25
y	0.00000	0.10017	0.20134	0.30452	0.41075	0.52110

Solution: Since the values of x for which the derivatives are to be computed lie near the beginning of the equally spaced table, we use the differentiation formulae based on Newton's forward difference interpolation formula. We first form the finite difference table.

NOTES

x	y	Δy	$\Delta^2 y$	$\Delta^3 y$	$\Delta^4 y$
0.0	0.00000				
		0.10017			
0.05	0.10017		100		
		0.10117		101	
0.10	0.20134		201		3
		0.10318		104	
0.15	0.30452		305		3
		0.10623		107	
0.20	0.41075		412		
		0.11035			
0.25	0.52110				

For evaluating $y'(0,0)$, we use the formula

$$y'(x_0) = \frac{1}{h}\left[\Delta y_0 - \frac{1}{2}\Delta^2 y_0 + \frac{1}{3}\Delta^3 y_0 - \frac{1}{4}\Delta^4 y_0\right]$$

$$\therefore \quad y'(0.0) = \frac{1}{0.05}\left(0.10017 - \frac{1}{2}\times 0.00100 + \frac{1}{3}\times 0.00101 - \frac{1}{4}\times 0.00003\right)$$

$$= 2.00000$$

For evaluating $y''(0,0)$, we use the formula

$$y''(x_0) = \frac{1}{h^2}\left(\Delta^2 y_0 - \Delta^3 y_0 + \frac{11}{12}\Delta^4 y_0\right)$$

$$= \frac{1}{(0.05)^2}\left(0.00100 - 0.00101 + \frac{11}{12}\times 0.00003\right)$$

$$= 0.007$$

For evaluating $y'(0.02)$ and $y''(0.02)$, we use the following formulae, with

$$u = \frac{0.02 - 0.00}{0.05} = 0.4$$

$$y'(0.02) = \frac{1}{h}\left[\Delta y_0 + \frac{2u-1}{2}\Delta^2 y_0 + \frac{3u^2 - 6u + 2}{6}\Delta^3 y_0 + \frac{2u^3 - 9u^2 + 11u - 3}{12}\Delta^4 y_0\right]$$

$$y''(0.02) = \frac{1}{h^2}\left[\Delta^2 y_0 + \frac{6(u-1)}{6}\Delta^3 y_0 + \frac{6u^2 - 18u + 11}{12}\times \Delta^4 y_0\right]$$

$$\therefore \quad y'(0.02) = \frac{1}{0.05}\left[0.10017 + \frac{2\times 0.4 - 1}{2}\times 0.00100 + \frac{3\times (0.4)^2 - 6\times 0.4 + 2}{6}\times 0.00101\right.$$

$$\left. + \frac{2\times 0.4^3 - 9\times 0.4^2 + 11\times 0.4 - 3}{12}\times 0.00003\right]$$

$$= 4.00028$$

$$y''(0.02) = \frac{1}{(0.05)^2}\left[0.00100 - 0.00101\times(-0.6) + \frac{6\times 0.16 - 18\times 0.4 + 11}{12}\times 0.00003\right]$$

$$= 0.800$$

Example 3.4: Compute $f'(6.0)$ and $f''(6.3)$ by numerical differentiation formulae for the function *f*(*x*) given in the following table.

x	6.0	6.1	6.2	6.3	6.4
$f(x)$	-0.1750	-0.1998	-0.2223	-0.2422	-0.2596

NOTES

Solution: We first form the finite difference table,

x	$f(x)$	$\Delta f(x)$	$\Delta^2 f(x)$	$\Delta^3 f(x)$
6.0	– 0.1750			
		– 248		
6.1	– 0.1998		23	
		– 225		3
6.2	– 0.2223		26	
		–199		–1
6.3	– 0.2422		25	
		–174		
6.4	– 0.2596			

For evaluating $f'(6.0)$, we use the formula derived by differentiating Newton's forward difference interpolation formula.

$$f'(x_0) = \frac{1}{h}\left[\Delta f_0 - \frac{1}{2}\Delta^2 f_0 + \frac{1}{3}\Delta^3 f_0\right]$$

$$\therefore \quad f'(6.0) = \frac{1}{0.1}\left[-0.0248 - \frac{1}{2}\times 0.0023 + \frac{1}{3}\times 0.0003\right]$$

$$= 10[-0.0248 - 0.00115 + 0.0001]$$

$$= -0.2585$$

For evaluating $f''(6.3)$, we use the formula obtained by differentiating Newton's backward difference interpolation formula. It is given by,

$$f''(x_n) = \frac{1}{h^2}[\nabla^2 f_n + \nabla^3 f_n]$$

$$\therefore \quad f''(6.3) = \frac{1}{(0.1)^2}[0.0026 + 0.0003] = 0.29$$

Example 3.5: Compute the values of $y'(1.00)$ and $y''(1.00)$ using suitable numerical differentiation formulae on the following table of values of x and y:

x	1.00	1.05	1.10	1.15	1.20
y	1.0000	1.02470	1.04881	1.07238	1.09544

Solution: For computing the derivatives, we use the formulae derived on differentiating Newton's forward difference interpolation formula, given by

$$f'(x_0) = \frac{1}{h}\left[\Delta y_0 - \frac{1}{2}\Delta^2 y_0 + \frac{1}{3}\Delta^3 y_0 - \frac{1}{4}\Delta^4 y_0 + \ldots\right]$$

$$f''(x_0) = \frac{1}{h^2}\left[\Delta^2 y_0 - \Delta^3 y_0 + \frac{11}{12}\Delta^4 y_0 + \ldots\right]$$

Now, we form the finite difference table.

NOTES

x	y	Δy	$\Delta^2 y$	$\Delta^3 y$	$\Delta^4 y$
1.00	1.00000				
		2470			
1.05	1.02470		−59		
		2411		5	
1.10	1.04881		−54		−2
		2357		3	
1.15	1.07238		−51		
		2306			
1.20	1.09544				

Thus with $x_0 = 1.00$, we have

$$y'(1.00) = \frac{1}{0.05}\left(0.02470 + \frac{1}{2}\times 0.00059 + \frac{1}{3}\times 0.00005 + \frac{1}{4}\times 0.00002\right)$$

$$= 0.502$$

$$y''(1.00) = \frac{1}{(0.05)^2}\left(-0.00059 - 0.00005 - \frac{11}{12}\times 0.00002\right)$$

$$= -0.26$$

Example 3.6: Using the following table of values, find a polynomial representation of $f'(x)$ and then compute $f'(0.5)$.

x	0	1	2	3
$f(x)$	1	3	15	40

Solution: Since the values of x are equally spaced we use Newton's forward difference interpolating polynomial for finding $f'(x)$ and $f'(0.5)$. We first form the finite difference table as given below:

x	$f(x)$	$\Delta f(x)$	$\Delta^2 f(x)$	$\Delta^3 f(x)$
0	1			
		2		
1	3		10	
		12		3
2	15		13	
		25		
3	40			

Taking $x_0 = 0$, we have $u = \frac{x - x_0}{h} = x$. Thus the Newton's forward difference interpolation gives,

$$f = f_0 + u\Delta f_0 + \frac{u(u-1)}{2!}\Delta^2 f_0 + \frac{u(u-1)(u-2)}{3!}\Delta^3 f_0$$

i.e., $$f(x) \approx 1 + 2x + \frac{x(x-1)}{2}\times 10 + \frac{x(x-1)\ (x-2)}{6}\times 3$$

or, $$f(x) = 1 + 3x - \frac{13}{2}x^2 + \frac{1}{2}x^3$$

$\therefore$ $$f'(x) = 3 - 13x + \frac{3}{2}x^2$$

and, $$f'(0.5) = 3 - 13\times 0.5 + \frac{3}{2}\times(0.5)^2 = -3.12$$

Example 3.7: The population of a city is given in the following table. Find the rate of growth in population in the year 2001 and in 1995.

Year x	1961	1971	1981	1991	2001
Population y	40.62	60.80	79.95	103.56	132.65

Solution: Since the rate of growth of the population is $\frac{dy}{dx}$, we have to compute $\frac{dy}{dx}$ at $x = 2001$ and at $x = 1995$. For this we consider the formula for the derivative on approximating y by the Newton's backward difference interpolation given by,

$$\frac{dy}{dx} = \frac{1}{h}\left[\nabla y_n + \frac{2u+1}{2}\nabla^2 y_n + \frac{3u^2+6u+2}{6}\nabla^3 y_n + \frac{2u^3+9u^2+11u+3}{12}\nabla^4 y_n + ...\right]$$

Where $u = \frac{x - x_n}{h}$

For this we construct the finite difference table as given below:

x	y	Δy	$\Delta^2 y$	$\Delta^3 y$	$\Delta^4 y$
1961	40.62				
		20.18			
1971	60.80		−1.03		
		19.15		5.49	
1981	79.95		4.46		−4.47
		23.61		1.02	
1991	103.56		5.48		
		29.09			
2001	132.65				

For $x = 2001$, $u = \frac{x - x_n}{h} = 0$

$$\therefore \quad \left(\frac{dy}{dx}\right)_{2001} = \frac{1}{10}\left[29.09 + \frac{1}{2}\times 5.48 + \frac{1}{3}\times 1.02 + \frac{1}{4}\times(-4.47)\right]$$

$$= 3.105$$

For $x = 1995$, $u = \frac{1995-1991}{10} = 0.4$

$$\left(\frac{dy}{dx}\right)_{1995} = \frac{1}{10}\left[23.61 + \frac{1.8}{2}\times 4.46 + \frac{3\times 0.16 + 6\times 0.4 + 2}{6}\times 5.49\right]$$

$$= 3.21$$

3.3 NUMERICAL INTEGRATION FORMULE

The evaluation of a definite integral cannot be carried out when the integrand $f(x)$ is not integrable, as well as when the function is not explicitly known but only the function values are known at a finite number of values of x. However, the value of

NOTES

the integral can be determined numerically by applying numerical methods. There are two types of numerical methods for evaluating a definite integral based on the following formula.

$$\int_a^b f(x)\, dx \tag{3.9}$$

They are termed as Newton-Cotes quadrature and Gaussian quadrature. We first confine our attention to Newton-Cotes quadrature which is based on integrating polynomial interpolation formulae. This quadrature requires a table of values of the integrand at equally spaced values of the independent variable x.

3.3.1 Newton-Cotes General Quadrature

We start with Newton's forward difference interpolation formula which uses a table of values of $f(x)$ at equally spaced points in the interval $[a, b]$. Let the interval $[a, b]$ be divided into n equal sub-intervals such that,

$$a = x_0,\ x_i = x_o + ih, \text{ for } i = 1, 2, ..., n-1, \quad x_n = b \tag{3.10}$$

so that, $nh = b–a$

Newton's forward difference interpolation formula is,

$$\varphi(s) = f_0 + s\ \Delta f_0 + \frac{s(s-1)}{2!}\Delta^2 f_0 + \frac{s(s-1)(s-2)}{3!}\Delta^3 f_0 + ... + \frac{s(s-1)(s-2)...(s-n+1)}{n!}\Delta^n f_0 \tag{3.11}$$

where, $s = \dfrac{x - x_0}{h}$

Replacing $f(x)$ by $\phi(s)$ in Equation (3.9), we get

$$\int_{x_0}^{x_n} f(x)\ dx = h\int_0^n \left[f_0 + s\Delta f_0 + \frac{s(s-1)}{2!}\Delta^2 f_0 + ... \right] ds$$

since when $x = x_0$, $s = 0$ and $x = x_n$, $s = n$ and $dx = h\ du$.

Performing the integration on the RHS we have,

$$\int_{x_0}^{x_n} f(x)dx = h\left[nf_0 + \frac{n^2}{2}\Delta f_0 + \frac{1}{2}\left(\frac{n^3}{3} - \frac{n^2}{2}\right)\Delta^2 f_0 + \frac{1}{6}\left(\frac{n^4}{4} - 3\frac{n^3}{3} - 2\frac{n^2}{2}\right)\Delta^3 f_0 \right.$$
$$\left. + \frac{1}{24}\left(\frac{n^5}{5} - \frac{3n^4}{2} + \frac{11n^3}{3} - 3n^2\right)\Delta^4 f_0 + ... \right] \tag{3.12}$$

We can derive different integration formulae by taking particular values of $n = 1, 2, 3, ...$ Again, on replacing the differences, the Newton-Cotes formula can be expressed in terms of the function values at $x_0, x_1,..., x_n$, as

$$\int_{x_0}^{x_n} f(x)dx = h\sum_{k=0}^{n} c_k f(x_k) \tag{3.13}$$

The error in the Newton-Cotes formula is given by,

$$E^n = \frac{h^{n+2}}{(n+1)!} f^{n+1}\ (\xi)\cdot\int_0^n s(s-1)\cdots(s-n)ds \tag{3.14}$$

NOTES

Trapezoidal Formula of Numerical Integration

Taking $n = 1$ in Equation (3.12), we get the trapezoidal formula given by,

$$\int_{x_0}^{x_1} f(x)dx = h\left[f_0 + \frac{1}{2}\Delta f_0\right]$$

since all other differences of higher order are absent.

Replacing Δf_0 by $f_1 - f_0$, we have

$$\int_{x_0}^{x_1} f(x)dx = \frac{h}{2}[f_0 + f_1] \qquad (3.15)$$

This is termed as trapezoidal formula of numerical integration.

This formula can be *geometrically interpreted* as the definite integral of the function $f(x)$ between the limits x_0 to x_1, as is approximated by the area of the trapezoidal region bounded by the chord joining the points (x_0, f_0) and (x_1, f_1), the x-axis and the ordinates at $x = x_0$ and at $x = x_1$. This is represented by the shaded area as shown in the Figure 3.1.

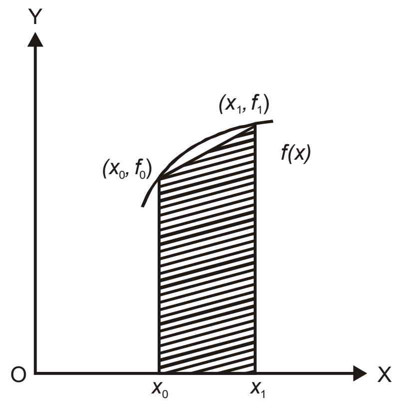


Fig. 3.1 *Trapezoidal Region*

Thus, the area under the curve $y = f(x)$ is replaced by the area under the chord joining the points.

The error in the trapezoidal formula is given by,

$$E_T = \frac{h^3}{2} f''(\xi) \times \int_0^1 s(s-1)ds = -\frac{h^3}{12} f''(\xi), \quad \text{where} \quad x_0 < \xi < x_1 \qquad (3.16)$$

Trapezoidal Rule

For evaluating the integral $\int_{x_0}^{x_n} f(x)dx$, we have to sum the integrals for each of the sub-intervals $(x_0, x_1), (x_1, x_2),..., (x_{n-1}, x_n)$. Thus,

$$\int_{x_0}^{x_n} f(x)dx = \frac{h}{2}[(f_0 + f_1) + (f_1 + f_2) + ... + (f_{n-1} + f_n)]$$

or $$\int_{x_0}^{x_n} f(x)dx = \frac{h}{2}[f_0 + 2(f_1 + f_2 + ... + f_{n-1}) + f_n] \qquad (3.17)$$

This is known as trapezoidal rule of numerical integration.

The error in the trapezoidal rule is,

$$E_T^n = \int_{x_0}^{x_n} f(x)dx - \frac{h}{2}[f_0 + 2(f_1 + f_2 + ... + f_{n-1}) + f_n]$$

$$= \frac{-h^3}{12}[f''(\xi_1) + f''(\xi_2) + ... + f''(\xi_n)]$$

where $x_0 < \xi_1 < x_1,\quad x_1 < \xi_2 < x_2, ...,\quad x_{n-1} < \xi_n < x_n$

Thus, we can write

$$E_T^n = -\frac{h^3}{12}[nf''(\xi)],\quad f''(\xi) \text{ being the mean of } f''(\xi_1), f''(\xi_2), ..., f''(\xi_n)$$

$$= -nh\frac{h^2}{12}\quad f''(\xi)$$

where $E_T^n = -\frac{h^2}{12}\quad (b-a)\quad f''(\xi),\quad \text{since}\quad nh = b - a$

or, $x_0 < \xi < x_n$

Algorithm: Evaluation of $\int_a^b f(x)dx$ by trapezoidal rule.

Step 1: Define function $f(x)$

Step 2: Initialize a, b, n

Step 3: Compute $h = (b-a)/n$

Step 4: Set $x = a$, $S = 0$

Step 5: Compute $x = x + h$

Step 6: Compute $S = S + f(x)$

Step 7: Check if $x < b$, then go to Step 4 else go to the next step

Step 8: Compute $I = h\,(S + (f(a) + f(b))/2)$

Step 9: Output I, n

Simpson's One-Third Formula

Taking $n = 2$ in the Newton-Cotes formula in Equation (3.12), we get Simpson's one-third formula of numerical integration given by,

$$\int_{x_0}^{x_2} f(x)dx = h\left[2f_0 + \frac{2^2}{2}\Delta f_0 + \frac{1}{12}(2\times 2^3 - 3\times 2^2)\Delta^2 f_0\right]$$

$$= h\left[2f_0 + 2\quad (f_1 - f_0) + \frac{1}{3}\quad (f_2 - 2f_1 + f_0)\right] \qquad (3.18)$$

$$\therefore \int_{x_0}^{x_2} f(x)dx = \frac{h}{3}[f_0 + 4f_1 + f_2]$$

This is known as Simpson's one-third formula of numerical integration.

The error in Simpson's one-third formula is defined as,

$$E_S = \int_{x_0}^{x_2} f(x)\ dx - \frac{h}{3}(f_0 + 4f_1 + f_2)$$

NOTES

Assuming $F'(x) = f(x)$, we obtain:

$$E_S = F(x_2) - F(x_0) - \frac{h}{3}(f_0 + 4f_1 + f_2)$$

Expanding $F(x_2) = F(x_0+2h), f_1 = f(x_0+h)$ and $f_2 = f(x_0+2h)$ in powers of h, we have:

$$E_S = 2hF'(x_0) + \frac{(2h)^2}{2!}F''(x) \quad (x_0) + \frac{(2h)^3}{3!}F'''(x_0) + ...$$

$$-\frac{h}{3}\left[f_0 + 4\left(f_0 + hf_0' + \frac{h^2}{2!}f''(0) + ...\right) + f_0 + 2hf_0' + \frac{(2h)^2}{2!}f''(0) + ...\right]$$

$$= 2hf_0 + 2h^2 f_0' + \frac{4}{3}h^3 f''(0) + \frac{2}{3}h^4 f'''(0) + \frac{4}{15}h^5 y_0^{iv} \quad (\xi)$$

$$-\frac{h}{3} \quad [6f_0 + 6hf_0' + 4h^2 f''(0) + 2h^3 f'''(0)...]$$

$$E_S = -\frac{h^5}{90}f^{iv}(\xi), \text{ on simplification, where } x_0 < \xi < x_2 \qquad (3.19)$$

Geometrical interpretation of Simpson's one-third formula is that the integral represented by the area under the curve is approximated by the area under the parabola through the points (x_0,f_0), (x_1,f_1) and (x_2,f_2) shown in Figure 3.2.

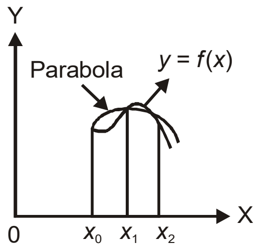


Fig. 3.2 *Simpson's One-Third Integration*

3.3.1 Simpson's One-Third Rule

On dividing the interval $[a, b]$ into $2m$ sub-intervals by points $x_0 = a, x_1 = a + h$, $x_2 = a + 2h$, ..., $x_{2m} = a+2mh$, where $b = x_{2m}$ and $h = (b–a)/(2m)$, and using Simpson's one-third formula in each pair of consecutive sub-intervals, we have

$$\int_a^b f(x)dx = \int_{x_0}^{x_2} f(x)dx + \int_{x_2}^{x_4} f(x)dx + ... + \int_{x_{2m-2}}^{x_{2m}} f(x)dx$$

$$= \frac{h}{3}\left[(f_0 + 4f_1 + f_2) + (f_2 + 4f_3 + f_4) + (f_4 + 4f_5 + f_6) + ... + (f_{2m-2} + 4f_{2m-1} + f_{2m})\right]$$

$$\int_a^b f(x)dx = \frac{h}{3}\left[f_0 + 4(f_1 + f_3 + f_5 + ... + f_{2m-1}) + 2(f_2 + f_4 + f_6 + ... + f_{2m-2}) + f_{2m}\right]$$

This is known as Simpson's one-third rule of numerical integration.

The error in this formula is given by the sum of the errors in each pair of intervals as,

$$E_S^{2m} = -\frac{h^5}{90}[f^{iv}(\xi_1) + f^{iv}(\xi_2) + \quad ... + f^{iv}(\xi_m)]$$

Which can be rewritten as,

$$E_S^{2m} = -\frac{h^5}{90}m \quad f^{iv}(\xi), \; f^{iv}(\xi) \text{ being the mean of } f^{iv}(\xi_1), \; f^{iv}(\xi_2), ..., f^{iv}(\xi_m)$$

Since $2mh = b - a$, we have

$$E_S^{2m} = -\frac{h^4}{180}(b-a)\; f^{iv}(\xi), \text{ where } a < \xi < b. \qquad (3.20)$$

Algorithm: Evaluation of $\int_a^b f(x)dx$ by Simpson's one-third rule.

Step 1: Define $f(x)$

Step 2: Input a, b, n (even)

Step 3: Compute $h = (b–a)/n$

Step 4: Compute $S_1 = f(a) + f(b)$

Step 5: Set $S_2 = 0$, $x = a$

Step 6: Compute $x = x + 2h$

Step 7: Compute $S_2 = S_2 + f(x)$

Step 8: Check If $x < b$ then go to Step 5 else go to next step

Step 9: Compute $x = a + h$

Step 10: Compute $S_4 = S_4 + f(x)$

Step 11: Compute $x = x + 2h$

Step 12: Check If $x > b$ go to next Step else go to Step 9

Step 13: Compute $I = (S_1 + 4S_4 + 2S_2)h/3$

Step 14: Write I, n

Simpson's Three-Eighth Formula

Taking $n = 3$, Newton-Cotes formula can be written as,

$$\int_{x_0}^{x_3} f(x)dx = h\int_0^3 (f_0 + u\,\Delta f_0 + \frac{u(u-1)}{2!}\Delta^2 f_0 + \frac{u(u-1)(u-2)}{3!}\Delta^3 f_0)du$$

$$= h\left[uf_0 + \frac{u^2}{2}\Delta f_0 + \frac{1}{2}\left(\frac{u^3}{3} - \frac{u^2}{2}\right)\Delta^2 f_0 + \frac{1}{6}\left(\frac{u^4}{4} - u^3 + u^2\right)\Delta^3 f_0\right]_0^3$$

$$= h\left[3y_0 + \frac{9}{2}\Delta y_0 + \frac{9}{4}\Delta^2 y_0 + \frac{3}{8}\Delta^3 y_0\right]$$

$$= h\left[3y_0 + \frac{9}{2}(y_1 - y_0) + \frac{9}{4}(y_2 - 2y_1 + y_0) + \frac{3}{8}(y_3 - 3y_2 + 3y_1 - y_0)\right]$$

$$\int_{x_0}^{x_3} f(x)\; dx = \frac{3h}{8}(y_0 + 3y_1 + y_3)$$

(3.21)

The truncation error in this formula is $-\frac{3h^5}{80} f^{iv}(\xi), x_0 < \xi < x_3$.

NOTES

This formula is known as Simpson's three-eighth formula of numerical integration.

As in the case of Simpson's one-third rule, we can write Simpson's three-eighth rule of numerical integration as,

$$\int_a^b f(x)\,dx = \frac{3h}{8}[y_0 + 3y_1 + 3y_2 + 2y_3 + 3y_4 + 3y_5 + 2y_6 + ... + 2y_{3m-3} + 3y_{3m-2} + 3y_{3m-1} + y_{3m}]$$

(3.22)

where $h = (b–a)/(3m)$; for $m = 1, 2,...$

i.e., the interval $(b–a)$ is divided into $3m$ number of sub-intervals.

The rule in Equation (3.22) can be rewritten as,

$$\int_a^b f(x)\,dx = \frac{3h}{8}[y_0 + y_{3m} + 3(y_1 + y_2 + y_4 + y_5 + ... + y_{3m-2} + y_{3m-1}) + 2(y_3 + y_6 + ... + y_{3m-3})]$$

(3.23)

The truncation error in Simpson's three-eighth rule is

$$\frac{-3h^4}{240}(b-a) f^{iv}(\xi), \quad x_0 < \xi < x_g m$$

3.3.2 Weddle's Formula

In Newton-Cotes formula with $n = 6$ some minor modifications give the Weddle's formula. Newton-Cotes formula with $n = 6$, gives

$$\int_{x_0}^{x_6} ydx = h\left[6y_0 + 18\Delta y_0 + 27\Delta^2 y_0 + 24y\Delta^3 y_0 + \frac{123}{10}\Delta^5 y_0 + \frac{41}{140}\Delta^6 y_0\right]$$

This formula takes a very simple form if the last term $\frac{41}{140}\Delta^6 y_0$ is replaced by $\frac{42}{140}\Delta^6 y_0 = \frac{3}{10}\Delta^6 y_0$. Then the error in the formula will have an additional term $\frac{1}{140}\Delta^6 y_0$. The above formula then becomes,

$$\int_{x_0}^{x_6} ydx = h\left[6y_0 + 18\Delta y_0 + 27\Delta^2 y_0 + 24\Delta^3 y_0 + \frac{123}{10}\Delta^5 y_0 + \frac{3}{10}\Delta^6 y_0\right]$$

$$\therefore \quad \int_{x_0}^{x_6} ydx = \frac{3h}{10}[y_0 + 5y_1 + y_2 + 6y_3 + y_4 + 5y_5 + y_6]$$

(3.24)

On replacing the differences in terms of y_i's, this formula is known as Weddle's formula.

The error Weddle's formula is $-\frac{1}{140}h^7 \cdot y^{(vi)}(\xi)$ (3.25)

Weddle's rule is a composite Weddle's formula, when the number of sub-intervals is a multiple of 6. One can use a Weddle's rule of numerical integration by sub-dividing the interval $(b-a)$ into $6m$ number of sub-intervals, m being a positive integer. The Weddle's rule is,

$$\int_a^b f(x)dx = \frac{3h}{10}[y_0+5y_1+y_2+6y_3+y_4+5y_5+2y_6+5y_7+y_8+6y_9+y_{10}+5y_{11}+...$$

$$+2y_{6m-6}+5y_{6m-5}+y_{6m-4}+6y_{6m-3}+y_{6m-2}+5y_{6m-1}+y_{6m}] \quad (3.26)$$

where, $b–a = 6mh$

i.e., $$\int_a^b f(x)\,dx = \frac{3h}{10}[y_0+y_{6m}+5(y_1+y_5+y_7+y_{11}+...+y_{6m-5}+y_{6m-1})+y_2+y_4+y_8+y_{10}+...$$

$$+y_{6m-4}+y_{6m-2}+6(y_3+y_9+...+y_{6m-3})+2(y_6+y_{12}+...+y_{-6})]$$

The error in Weddle's rule is given by $-\frac{1}{840}h^6(b-a)y^{(vi)}(\xi)$

(3.27)

Example 3.8: Compute the approximate value of $\int_0^2 x^4 dx$ by taking four sub-intervals and compare it with the exact value.

Solution: For four sub-intervals of [0, 2], we have $h = \frac{2}{4} = \frac{1}{2} = 0.6$. We tabulate $f(x) = x^4$.

x	0	0.5	1.0	1.5	2.0
$f(x)$	0	0.0625	1.0	5.062	16.0

By trapezoidal rule, we get

$$\int_0^2 x^4 dx \approx \frac{0.5}{2}[0+2\times(0.0625+1.0+5.062)+16.0]$$

$$\approx \frac{1}{4}[12.2690+16.0] = \frac{28.2690}{4} = 7.0672$$

By Simpson's one-third rule, we get

$$\int_0^2 x^4 dx = \frac{0\cdot 5}{3}[0+4\times(0.0625+5.062)+2\times 1.0+16.0]$$

$$= \frac{1}{6}[4\times 5.135+18.0] = \frac{38.5380}{6} = 6.4230$$

Exact value $= \frac{2^5}{5} = \frac{32}{5} = 6.4$

Error in the result by trapezoidal rule = $6.4 - 7.0672 = -0.6672$

Error in the result by Simpson's one third rule = $6.4 - 6.4230 = -0.0230$

NOTES

Example 3.9: Evaluate the following integral:

$\int_0^1 (4x - 3x^2)dx$ by taking $n = 10$ and using the following rules:

(*i*) Trapezoidal rule and (*ii*) Simpson's one-third rule. (iii) Also compare them with the exact value and find the error in each case.

Solution: We tabulate $f(x) = 4x–3x^2$, for $x = 0, 0.1, 0.2, ..., 1.0$.

x	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
$f(x)$	0.0	0.37	0.68	0.93	1.12	1.25	1.32	1.33	1.28	1.17	1.0

(*i*) Using trapezoidal rule, we have

$$\int_0^1 (4x - 3x^2)\,dx = \frac{0.1}{2}\left[0 + 2(0.37 + 0.68 + 0.93 + 1.12 + 1.25 + 1.32 + 1.33 + 1.28 + 1.17) + 1.0\right]$$

$$= \frac{0.1}{2} \times (18.90 + 1.0) = 0.995$$

(*ii*) Using Simpson's one-third rule, we have

$$\int_0^1 (4x - 3x^2)\; dx = \frac{0.1}{3}[0 + 4(0.37 + 0.93 + 1.25 + 1.33 + 1.17) + 2(0.68 + 1.12 + 1.32 + 1.28) + 1.0]$$

$$= \frac{0.1}{3}[4 \times 5.05 + 2 \times 4.40 + 1.0]$$

$$= \frac{0.1}{3} \times [30.0] = 1.00$$

(*iii*) Exact value = 1.0

Error in the result by trapezoidal rule is 0.005 and there is no error in the result by Simpson's one-third rule.

Example 3.10: Evaluate $\int_0^1 e^{-x^2}\,dx,$ using (*i*) Simpson's one-third rule with 10 sub-intervals and (*ii*) Trapezoidal rule.

Solution: (i) We tabulate values of e^{-x^2} for the 11 points $x = 0, 0.1, 0.2, 0.3, ...,$ 1.0 as given below.

x	e^{-x^2}		
0.0	1.00000		
0.1		0.990050	
0.2			0.960789
0.3		0.913931	
0.4			0.852144
0.5		0.778801	
0.6			0.697676
0.7		0.612626	
0.8			0.527292
0.9		0.444854	
1.0	0.367879		
	1.367879	3.740262	3.037901

Hence, by Simpson's one-third rule we have,

$$\int_0^1 e^{-x^2}\ dx = \frac{h}{3}\ \ [f_0 + f_{10} + 4\ \ (f_1 + f_3 + f_5 + f_7 + f_9) + 2\ \ (f_2 + f_4 + f_6 + f_8)]$$

$$= \frac{0.1}{3}\ \ [1.367879 + 4 \times 3.740262 + 2 \times 3.037901]$$

$$= \frac{0.1}{3}\ \ [1.367879 + 14.961048 + 6.075802]$$

$$= \frac{2.2404729}{3} = 0.7468243 \approx 0.746824$$

(ii) Using trapezoidal rule, we get

$$\int_0^1 e^{-x^2}\ \ dx = \frac{h}{2}[f_0 + f_{10} + 2\ \ (f_1 + f_2 + ... + f_9)]$$

$$= \frac{0.1}{2}\ \ [1.367879 + 6.778163]$$

$$= 0.4073021$$

Example 3.11: Compute the integral $I = \int_0^4 (x^3 - 2x^2 + 1)dx,$ using Simpson's one-third rule taking $h = 1$ and show that the computed value agrees with the exact value. Give reasons for this.

Solution: The values of $f(x) = x^3 – 2x^2 + 1$ are tabulated for $x = 0, 1, 2, 3, 4$ as

x	0	1	2	3	4
$f(x)$	1	0	1	10	33

The value of the integral by Simpson's one-third rule is,

$$I = \frac{1}{3}\ \ [1 + 4 \times 0 + 2 \times 1 + 4 \times 10 + 33] = 25\frac{1}{3}$$

The exact value $= \dfrac{4^4}{4} - 2 \times \dfrac{4^3}{3} + 1 \times 4 = 25\dfrac{1}{3}$

Thus, the computed value by Simpson's one-third rule is equal to the exact value. This is because the error in Simpson's one-third rule contains the fourth order derivative and so this rule gives the exact result when the integrand is a polynomial of degree less than or equal to three.

Example 3.12: Compute $\int_{0.1}^{0.5} e^x dx$ by (*i*) Trapezoidal rule and (*ii*) Simpson's one-third rule and compare the results with the exact value, by taking $h = 0.1$.

Solution: We tabulate the values of $f(x) = e^x$ for $x = 0.1$ to 0.5 with spacing $h = 0.1$.

x	0.1	0.2	0.3	0.4	0.5
$f(x)=e^x$	1.1052	1.2214	1.3498	1.4918	1.6847

NOTES

(i) The value of the integral by trapezoidal rule is,

$$I_T = \frac{0.1}{2}[1.1052+2(1.2214+1.3498+1.4918)+1.6487]$$
$$= \frac{0.1}{2}[2.7539+2\times 4.0630] = 0.5439$$

(ii) The value computed by Simpson's one-third rule is,

$$I_S = \frac{0.1}{3}[1.1052+4(1.2214+1.4918)+2\times 1.3498+1.6487]$$
$$= \frac{0.1}{3}[2.7539+4\times 2.7132+2.6996] = \frac{0.1}{3}[16.3063] = 0.5435$$

Exact value = $e^{0.5} - e^{0.1} = 1.6487 - 1.1052 = 0.5435$

The trapezoidal rule gives the value of the intergal with an error – 0.0004 but Simpson's one-third rule gives the exact value.

Example 3.13: Compute $\int_0^1 \frac{dx}{1+x}$ using (*i*) Trapezoidal rule (*ii*) Simpson's one-third rule taking 10 sub-intervals. Hence, (iii) Find $\log_{e^2}$ and compare it with the exact value up to six decimal places.

Solution: We tabulate the values of *f*(*x*) = $\frac{1}{1+x}$ for *x* = 0, 0.1, 0.2,..., 1.0 as given below:

x	y	$f(x)=\frac{1}{1+x}$		
0.0	y_0	1.000000		
0.1	y_1		0.9090909	
0.2	y_2			0.8333333
0.3	y_3		0.1692307	
0.4	y_4			0.7142857
0.5	y_5		0.6666667	
0.6	y_6			0.6250000
0.7	y_7		0.5882352	
0.8	y_8			0.5555556
0.9	y_9		0.5263157	
1.0	y_{10}	0.500000		
		1.500000	3.4595391	2.7281746

(*i*) Using trapezoidal rule, we have

$$\int_0^1 \frac{dx}{1+x} = \frac{h}{2}[f_0+f_{10}+2(f_1+f_2+f_3+f_4+...+f_0)]$$
$$= \frac{0.1}{2}[1.500000+2\times(3.4595391+2.7281745)]$$
$$= \frac{0.1}{2}[1.500000+12.3754272] = 0.6437714.$$

NOTES

(*ii*) Using Simpson's one-third rule, we get

$$\int_0^1 \frac{dx}{1+x} = \frac{h}{3}[f_0 + f_{10} + 4(f_1 + f_3 + ... + f_9) + 2(f_2 + f_4 + ... + f_8)]$$

$$= \frac{0.1}{3}[1.500000 + 4 \times 3.4595391 + 2 \times 2.7281745]$$

$$= \frac{0.1}{3}[1.5 + 13.838156 + 5.456349] = \frac{0.1}{3} \times 20.794505 = 0.6931501$$

(*iii*) Exact value:

$$\int_0^1 \frac{dx}{1+x} = \log_{e^2} = \frac{0.1}{3} \quad [1.500000 + 4 \times 3.4595391 + 2 \times 2.7281745]$$

$$= 0.6931472$$

The trapezoidal rule gives the value of the integral having an error 0.693147 – 0.6437714 = 0.0493758, while the error in the value by Simpson's one-third rule is – 0. 000029.

Example 3.14: Compute $\int_0^{\frac{\pi}{2}} \sqrt{\cos\theta}\, d\theta,$ by (*i*) Simpson's rule and (*ii*) Weddle's formula taking six sub-intervals.

Solution: Sub-division of $[0, \frac{\pi}{2}]$ into six sub-intervals will have $h = \frac{\pi}{2}. \ \frac{1}{6} = 15^\circ = 0 \cdot 26179.$ For applying the integration rules we tabulate $\sqrt{\cos\theta}$.

θ	0°	15°	30°	45°	60°	75°	90°
$\sqrt{\cos\theta}$	1	0.98281	0.93061	0.84089	0.70711	0.50874	0

(*i*) The value of the integral by Simpson's one-third rule is given by,

$$I_S = \frac{0.26179}{3}[1 + 4 \times (0.98281 + 0.84089 + 0.50874) + 2 \times (0.093061 + 0.070711) + 0)]$$

$$= \frac{0.26179}{3}[1 + 4 \times 2.33244 + 2 \times 1.63772]$$

$$= \frac{0.26179}{3} \times 13.6052 = 1.18723$$

(*ii*) The value of the integral by Weddle's formula is,

$$I_W = \frac{3}{10} \times 0.26179 \quad [1.05 + 7.45775 + 5.04534 + 0.93061 + 0.070711]$$

$$= 3 \times 0.026179 \quad [14.554411] = 1.143059 \approx 1.14306$$

Example 3.15: Evaluate the integral $\int_0^{\frac{\pi}{2}} \sqrt{1 - 0.162 \sin^2 \varphi}\, d\varphi$ by Weddle's formula.

Solution: On dividing the interval into six sub-intervals, the length of each sub-interval will be $h = \frac{1}{6} \cdot \frac{\pi}{2} = 0 \cdot 26179 = 15^\circ.$ For computing the integral by Weddle's formula, we tabulate $f(\varphi) = \sqrt{1 - 0 \cdot 162 \sin^2 \varphi}.$

φ	0°	15°	30°	45°	60°	75°	90°
$f\ (\varphi)$	1.0	0.99455	0.97954	0.95864	0.93728	0.92133	0.91542

The value of the integral by Weddle's formula is given by,

NOTES

$$I_W = \frac{3 \times 0.26179}{10} \quad [1.0 + 5 \quad (0.99455 + 0.92133) + 0.97954 + 6 \times 0.95864 + 0.93728 + 0.91542]$$
$$= 0.078537 \times 19.16348 = 1.50504$$

3.3.3 Errors in Itegration Formulae

For evaluating a definite integral correct to a desired accuracy, one has to make a suitable choice of the value of h, the length of sub-interval to be used in the formula. There are two ways of determining h, by considering the truncation error in the formula to be used for numerical integration or by successive evaluation of the integral by the technique of interval halving and comparing the results.

Truncation Error Estimation Method

In the truncation error estimation method, the value of h to be used is determined by considering the truncation error in the formula for numerical integration. Let E be the error tolerance for the integral to be evaluated. Then h is chosen by using the condition,

$$|R| < \varepsilon/2$$

As an illustration, consider the evaluation of $\int_1^2 \frac{dx}{x}$ using Simpson's one-third rule accurate up to the third decimal place. We may take $\varepsilon = 10^{-3}$.

If we wish to use Simpson's one-third rule, then the truncation error is R,

$$R = \frac{h^4}{180}(2-1) f^{iv}(\xi); \quad 1 < \xi < 2$$

Then h is determined by satisfying the condition,

$$\frac{h^4}{180} \mid f^{iv}(\xi) \mid < 0 \cdot 5 \times 10^{-3}$$

For the given problem, $f\ (x) = \frac{1}{x}$, thus $f^{iv}(x) = \frac{2 \times 3 \times 4}{x^5}$. Hence, $\max_{[1,2]} \left| f^{iv}\ (x) \right| = 24$

Thus, $h^4 \times \frac{1 \times 24}{180} < 0 \cdot 5 \times 10^{-3}$ or $h < 0.102$

But h has to be so chosen such so that the interval [1, 2] is divided into an even number of sub-intervals. Hence we may take $h = 0.1 < 0.102$, for which n = 10, i.e., there will be 10 sub-intervals.

The value of the integral is,

$$\int_1^2 \frac{dx}{x} = \frac{0.1}{3} \left[1.0 + 4 \left(\frac{1}{1.1} + \frac{1}{1.3} + \frac{1}{1.5} + \frac{1}{1.7} + \frac{1}{1.9} \right) + 2 \left(\frac{1}{1.2} + \frac{1}{1.4} + \frac{1}{1.6} + \frac{1}{1.8} \right) + \frac{1}{2} \right]$$
$$= \frac{0.1}{3} \quad [1.5 + 4 \times 3.4595 + 2 \times 2.7282]$$
$$= \frac{0.1}{3} \times 2.0749 = 0.06931 \text{ which agrees with the exact value of } \log_{e^2}.$$

Interval Halving Technique

When the estimation of the truncation error is cumbersome, the method of interval halving is used to compute an integral to the desired accuracy.

In the interval halving technique, an integral is first computed for some moderate value of h. Then, it is evaluated again for spacing $\frac{h}{2}$, i.e., with double the number of subdivisions. This requires the evaluation of the integrand at the new points of subdivision only and the previous function values with spacing h are also used. Now the difference between the integral I_h and $I_{\frac{h}{2}}$ is used to check the accuracy of the computed integral. If $|I_h - I_{h/2}| \le \varepsilon,$ where ε is the permissible error, then $I_{h/2}$ is to be taken as the computed value of the integral to the desired accuracy. If the above accuracy is not achieved, i.e., $|I_h - I_{h/2}| > \varepsilon,$ then the computation of the integral is made again with spacing $\frac{h}{4}$ and the accuracy condition is tested again. The equation of $I_{h/4}$ will require the evaluation of the integrand at the new points of sub-division only.

Notes:

1. The initial choice of h is sometimes taken as $\sqrt[m]{\varepsilon}$ where $m = 2$ for trapezoidal rule and $m = 4$ for Simpson's one-third rule.
2. The method of interval halving is widely used for computer evaluation since it enables a general choice of h together with a check on the computations.
3. The truncation error R can be estimated by using Runge's principle given by,

$R \approx \frac{1}{3}|I_h - I_{h/2}|$ for trapezoidal rule and $R \approx \frac{1}{15}|I_h - I_{h/2}|$ for Simpson's one-third rule.

Algorithm: Evaluation of an integral by Simpson's one-third rule with interval halving.

Step 1: Set/initialize a, b, ε

[a, b are limits of integration, ε is error tolerance]

Step 2: Set $h = \frac{b-a}{2}$

Step 3: Compute $S_1 = f(a) + f(b)$

Step 4: Compute $S_4 = f(a + h)$

Step 5: Set $S_2 = 0$, $I_1 = 0$

Step 6: Compute $I_2 = \frac{(S_1 + 4S_4 + S_2) \times h}{3}$

Step 7: If $(I_2 - I_1) < \varepsilon$, go to Step 17 else go to the next step

NOTES

Step 8: Set $h = \frac{h}{2},\ I_1 = I_2$

Step 9: Compute $S_2 = S_2 + S_4$

Step 10: Set $S_4 = 0$

Step 11: Set $x = a + h$

Step 12: Compute $S_4 = S_4 + f(x)$

Step 13: Set $x = x + h$

Step 14: If $x < b$, go to Step 12 else go to the next step

Step 15: Compute $I_2 = \frac{(S_1 + 2S_2 + 4S_4) \times h}{3}$

Step 16: Go to step 7

Step 17: Write I_2, h, ε

Step 18: End

Algorithm: Evaluation of an integral by trapezoidal rule with interval halving.

Step 1: Initialize/set a, b, ε [a, b are limits of integration, ε is error tolerance]

Step 2: Set $h = b – a$

Step 3: Compute $S_1 = \frac{f(a) + f(b)}{2}$

Step 4: Compute $I_1 = S_1 \times h$

Step 5: Compute $x = a + \frac{h}{2}$

Step 6: Compute $I_2 = (S_1 + f(x)) \times h$

Step 7: If $|I_2 - I_1| < \varepsilon$, go to Step 13 else go to the next step

Step 8: Set $h = \frac{h}{2}$

Step 9: Set $x = a + h$

Step 10: Set $I_2 = I_2 + h \times f(x)$

Step 11: If $x < b$, go to Step 9 else go to next step

Step 12: Go to Step 7

Step 13: Write $I_2,\ h,\ \varepsilon$

Step 14: End

Numerical Evaluation of Double Integrals

We consider the evaluation of a double integral,

$$I = \iint_R f(x, y)\,dx\,dy \qquad (3.28)$$

where R is the rectangular region $a \le x \le b, c \le y \le d$. The double integral can be transformed into a repeated integral in the following form,

$$\int_a^b dx \left[\int_c^d f(x, y) dy \right] \tag{3.29}$$

Writing $F(x) = \int_c^d f(x, y) dy$ considered as a function of x, we have (3.30)

$$I = \int_a^b F(x) dx \tag{3.31}$$

Now for numerical integration, we can divide the interval $[a, b]$ into n sub-intervals with spacing h and then use a suitable rule of numerical integration.

Trapezoidal Rule for Double Integral

By trapezoidal rule, we can write the integral Equation (3.31) as,

$$\int_a^b F(x)\ dx = \frac{h}{2}[F_0 + F_n + 2(F_1 + F_2 + F_3 + ... + F_{n-1})] \tag{3.32}$$

where $x_0 = a, x_n = b,\ h = \frac{b-a}{n}$ and

$$F_i = F(x_i) = \int_0^1 f(x_i, y)\, dy,\ x_i = a + ih \tag{3.33}$$

for $i = 0, 1, 2, ..., n$.

Each F_i can be evaluated by trapezoidal rule. For this, the interval $[c, d]$ may be divided into m sub-intervals each of length $k = \frac{c-d}{m}$. Thus we can write,

$$F_i = \frac{k}{2}[f(x_i, y_0) + f(x_i, y_m) + 2\{f(x_i, y_1) + f(x_i, y_2 + ... + f(x_i, y_{m-1})\}] \tag{3.34}$$

$y_0 = c, y_m = d, y_i = c + ik;\ i = 0, 1, ..., m.$

This Equation (3.34) can be written in a compact form,

$$F_i = \frac{k}{2}[f_{i0} + f_{im} + 2(f_{i1} + f_{i2} + ... + f_{im-1})]. \tag{3.35}$$

The relation Equations (3.32) and (3.35) together form the trapezoidal rule for evaluation of double integrals.

Simpson's One-Third Rule for Double Integrals

For the evaluation of double integrals we can write Simpson's $\frac{1}{3}$ rule. Thus we have,

$$I = \int_a^b F(x)\, dx = \frac{h}{3}[F_0 + F_n + 2(F_2 + F_4 + ... + F_{n-2}) + 4(F_1 + F_3 + ... + F_{n-1})] \tag{3.36}$$

where $h = \frac{b-a}{n}$, n is even and

$$F_i = F(x_i) = \int_c^d f(x_i, y)dy, \; x_i = a + ih, \text{ for } i = 0, 1, 2, ..., n \quad (3.37)$$

NOTES

and, $x_0 = a$ and $x_n = b$

For evaluating I, we have to evaluate each of the $(n + 1)$ integrals given in Equation (3.37). For evaluation of F_i, we can use Simpson's one-third rule by dividing $[c, d]$ into m sub-intervals. F_i can be written as,

$$F_i = \frac{k}{3}[f(x_i, y_0) + f(x_i, y_m) + 2f(x_i, y_2) + f(x_i, y_4) + ... + f(x_i, y_{m-2}) + 4\{f(x_i, y_1) + f(x_i, y_3) + ... + f(x_i, y_{m-1})\}] \quad ...(3.38)$$

Equation (3.38) can be written in a compact notation as,

$$F_i = \frac{k}{3}[f_{i0} + f_{im} + 2(f_{i2} + f_{i4} + ... + f_{in-2}) + 4(f_{i1}, f_{i3} + ... + f_{im-1})]$$

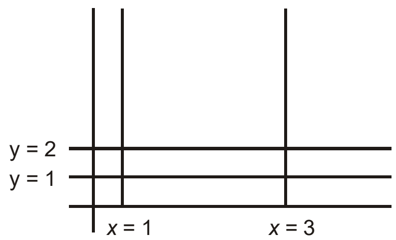
where $f_{ij} = f(x_i, y_j), j = 0, 1, 2, ..., m$.

Example 3.16: Evaluate the following double integral $\iint_R (x^2 + y^2)dx\,dy$ where R is the rectangular region $1 \le x \le 3, 1 \le y \le 2$, by Simpson's one-third rule taking $h = k = 0.5$.

Solution: We write the integral in the form of a repeated integral,

$$I = \int_1^3 dx \left[\int_1^2 (x^2 + y^2)dy\right]$$

Taking $n = 4$ sub-intervals along x, so that $h = \frac{2}{4} = 0.5$



$$\therefore I = \int_1^3 F(x)dx = \frac{0.5}{3}[F_0 + F_4 + 2F_2 + 4(F_1 + F_3)]$$

where $F(x) = \int_1^2 (x^2 + y^2)dy$

$$\therefore \quad F_i = F(x_i) = \int_1^2 (x_i^2 + y^2)dy; \; x_i = 1 + 0.5i, \text{ where } i = 0, 1, 2, 3, 4.$$

For evaluating F_i's, we take $k = \frac{1}{2} = 0.5$ and get,

$$F_0 = \int_1^2 (1 + y^2)\,dy = \frac{0.5}{3}[1 + 1^2 + 4\{1 + (1.5)^2\} + 1 + 2^2] = \frac{0.5}{3} \times 20$$

$$F_1 = \int_1^2 (1.5^2 + y^2)\,dy = \frac{0.5}{3}[(1.5)^2 + 1^2 + 4\{1.5)^2 + (1.5)^2\} + (1.5)^2 + 2^2] = \frac{0.5}{3} \times 27.50$$

NOTES

$$F_2 = \int_1^2 (2^2 + y^2)\ dy = \frac{0.5}{3}\ [2^2 + 1^2 + 4\ \ (2^2 + 1.5)^2\} + 2^2 + 2^2] = \frac{0.5}{3} \times 38$$

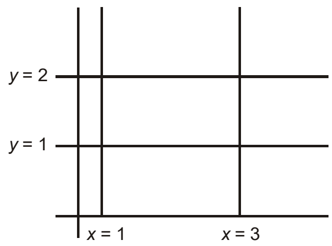
$$F_3 = \int_1^2 ((2.5)^2 + y^2) dy = \frac{0.5}{3}\ [(2.5)^2 + 1^2 + 4\ \ \{(2.5)^2 + (1.5)^2\} + (2.5)^2 + 2^2] = \frac{0.5}{3} \times 51.50$$

$$F_4 = \int_1^2 (3^2 + y^2) dy = \frac{0.5}{3}\ [3^2 + 1^2 + 4\{3^2 + (1.5)^2\} + 3^2 + 2^2] = \frac{0.5}{3} \times 68$$

$$\therefore I = \frac{0.25}{9}[20 + 68 + 2 \times 38 + 4\ (27.50 + 51.50)]$$

$$= \frac{0.25}{9} \times 480 = 13.333$$

Example 3.17: Compute $\iint_R (x^2 + y^2) dx\, dy$ by trapezoidal rule with $h = 0.5$.



Solution: $I_T = \int_1^3 F(x) dx = \frac{0.5}{2}\ [F_0 + F_4 + 2\ (F_1 + F_2 + F_3)]$

where $F_i = F(x_i) = \int_1^2 (x_i^2 + y^2) dy,\ x_i = 1 + 0.5i,\ i = 0, 1, 2, 3, 4.$

Thus, $F_0 = \int_1^2 (1 + y^2) dy = \frac{0.5}{2}\ [1^2 + 1^2 + 2\{1^2 + (1.5)^2\} + 1^2 + 2^2]$

$$= \frac{0.5}{2} \times 13.50 = 3.375$$

$$F_1 = \int_1^2 [(1.5)^2 + y^2] dy = \frac{0.5}{2}\ [(1.5)^2 + 1^2 + 2\ \ \{(1.5)^2 + (1.5)^2\} + (1.5)^2 + 2^2]$$

$$= \frac{0.5}{2} \times 18.50 = 4.625$$

$$F_2 = \int_1^2 [2^2 + y^2] dy = \frac{0.5}{2}\ [2^2 + 1^2 + 2\{2^2 + (1.5)^2\} + 2^2 + 2^2]$$

$$= \frac{0.5}{2} \times 25.50 = 6.375$$

$$F_3 = \int_1^2 [(2.5)^2 + y^2] dy = \frac{0.5}{2}\ [(2.5)^2 + 1^2 + 2\{(2.5)^2 + (1.5)^2\} + (2.5)^2 + 2^2]$$

$$= \frac{0.5}{2} \times 34.50 = 8.625$$

$$F_4 = \int_1^2 [3^2 + y^2] dy = \frac{0.5}{2}\ [3^2 + 1^2 + 2\{3^2 + (1.5)^2\} + 3^2 + 2^2]$$

$$= \frac{0.5}{2} \times 45.50 = 11.375$$

$$\therefore\ I_T = \frac{0.5}{2} \times [3.375 + 11.375 + 2(4.625 + 6.375 + 8.625)]$$

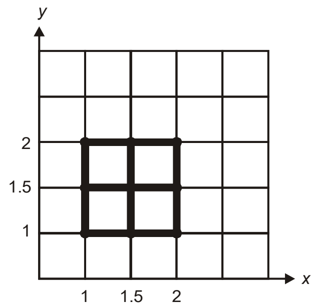
$$= \frac{1}{4}[14.750 + 2 \times 19.625]$$

$$= \frac{1}{4}\left[14.750 + 39.250\right] = \frac{1}{4} \times 54 = 13.5$$

NOTES

Example 3.18: Evaluate the following double integral using trapezoidal rule with length of sub-intervals $h = k = 0.5, \int_1^2 \int_1^2 \frac{dx\ dy}{x+y}$.

Solution: Let $f(x, y) = \frac{1}{x+y}$



By trapezoidal rule with $h = 0.5$, the integral

$I = \int_1^2 \int_1^2 dx\ \ dy\ \ f(x, y)$ is computed as,

$$I = \frac{0.5 \times 0.5}{4}\ \ [f(1,\ 1) + f(2,\ 1) + f(1,\ 2) + f(2,\ 2) + 2\{f(1.5,\ 1) + f(1,\ 1.5) + f(2,\ 1.5) + f(1.5,\ 2)\} + 4f\ \ (1.5,\ 1.5)]$$

$$= \frac{1}{16}\left[\frac{1}{2} + \frac{1}{3} + \frac{1}{3} + \frac{1}{4} + 2\ \left(\frac{2}{5} + \frac{2}{5} + \frac{2}{7} + \frac{2}{7}\right) + 4 \times \frac{1}{3}\right]$$

$$= \frac{1}{16}\left[0.666667 + 0.75 + 2 \times \frac{4 \times 12}{35} + \frac{4}{3}\right]$$

$$= \frac{1}{16}[5.492857]$$

$$= 0.343304.$$

Example 3.19: Evaluate $\int_1^2 \int_1^2 \frac{dxdy}{x+y}$ by Simpson's one-third rule. Take sub-intervals of length $h = k = 0.5$.

Solution: The value of the integral $I = \int_1^2 \int_1^2 f(x, y) dx\, dy$ by Simpson's one-third rule with $h = k = 0.5$ is,

$$I = \frac{0.5 \times 0.5}{3 \times 3}\ \ [f(1,\ 1) + f(2,\ 1) + f(1,\ 2) + f(2,\ 2) + 4\{f(1,\ 1.5) + f(1.5,\ 1) + f(2,\ 1.5) + f(1.5,\ 2)\} + 16f(1.5,\ 1.5)]$$

$$= \frac{1}{36}\ \left[\frac{1}{2} + \frac{1}{3} + \frac{1}{3} + \frac{1}{4} + 4\ \left(\frac{2}{5} + \frac{2}{5} + \frac{2}{7} + \frac{2}{7}\right) + 16 \times \frac{1}{3}\right]$$

$$= \frac{1}{36}\left[0.666667 + 0.75 + 4 \times \frac{4 \times 12}{35} + \frac{16}{3}\right]$$

$$= \frac{1}{36}\ \ [12.235714] = 0.339880$$

3.3.4 Gaussian Quadrature

NOTES

We have seen that Newton-Cotes formula of numerical integration is of the form,

$$\int_a^b f(x)dx \approx \sum_{i=0}^{n} c_i f(x_i) \tag{3.39}$$

where $x_i = a+ih$, $i = 0, 1, 2, ..., n$; $h = \dfrac{b-a}{n}$

This formula uses function values at equally spaced points and gives the exact result for $f(x)$ being a polynomial of degree less than or equal to n. Gaussian quadrature formula is similar to Equation (3.39) given by,

$$\int_{-1}^{1} F(u)du \approx \sum_{i=1}^{n} w_i F(u_i) \tag{3.40}$$

where w_i's and u_i's called weights and abscissae, respectively are derived such that above Equation (3.40) gives the exact result for $F(u)$ being a polynomial of degree less than or equal to $2n–1$.

In Newton-Cotes Equation (3.39), the coefficients c_i and the abscissae x_i are rational numbers but the weights w_i and the abscissae u_i are usually irrational numbers. Even though Gaussian quadrature formula gives the integration of $F(u)$ between the limits –1 to +1, we can use it to find the integral of $f(x)$ from a to b by a simple transformation given by,

$$x = \frac{b-a}{2}u + \frac{a+b}{2} \tag{3.41}$$

Evidently, then limits for u become –1 to 1 corresponding to $x = a$ to b and writing,

$$f(x) = f\left[\frac{b-a}{2}u + \frac{a+b}{2}\right] = F(u)$$

we have, $\displaystyle\int_a^b f(x)dx = \frac{b-a}{2}\int_{-1}^{1} F(u)du$

(3.42)

It can be shown that the u_i are the zeros of the Legendre polynomial $P_n(u)$ of degree n. These roots are real but irrational and the weights are also irrational.

Given below is a simple formulation of the relevant equations to determine u_i and w_i. Let $F(u)$ be a polynomial of the form,

$$F(u) = \sum_{k=0}^{2n-1} a_k u^k \tag{3.43}$$

Then, we can write

$$\int_{-1}^{1} F(u)du = \int_{-1}^{1}\left[\sum_{k=0}^{2n-1} a_k u^k\right]du \tag{3.44}$$

or, $$\int_{-1}^{1} F(u)du = 2a_0 + \frac{2}{3}a_2 + \frac{2}{5}a_4 + ... + \frac{2}{2n-2}a_{2n-2}$$

NOTES

(3.45)

Equation (3.40) gives,

$$\int_{-1}^{1} F(u)du = \sum_{i=1}^{n} w_i \left[\sum_{k=0}^{2n-1} a_k u_i^k \right]$$
$$= \sum_{i=1}^{n} w_i \left(a_0 + a_1 u_i + a_2 u_i^2 + ... + a_{2n-1} u_i^{2n-1} \right)$$

(3.46)

The Equations (3.45) and (3.46) are assumed to be identical for all polynomials of degree less than or equal to $2n$–1 and hence equating the coefficients of a_k on either side we obtain the following $2n$ equations for the $2n$ unknowns w_1, w_2,...,w_n and u_1, u_2,...,u_n.

$$\sum_{i=1}^{n} w_i = 2, \sum_{i=1}^{n} w_i u_i = 0, \sum_{i=1}^{n} w_i u_i^2 = \frac{2}{3}, ... \sum_{i=1}^{n} w_i u_i^{2n-1} = 0$$

(3.47)

The solution of Equation (3.47) is quite complicated. However, use of Legendre polynominals makes the labour unnecessary. It can be shown that the abscissae u_i are the zeros of the Legendre polynomial $P_n(x)$ of degree n. The weights w_i can then be easily determined by solving the first n Equations of Equations (3.47). As an illustration, we take $n = 2$. The four equations for u_1, u_2, w_1 and w_2 are,

$$w_1 + w_2 = 2$$
$$w_1 u_1 + w_2 u_2 = 0$$
$$w_1 u_1^2 + w_2 u_2^2 = \frac{2}{3}$$
$$w_1 u_1^3 + w_2 u_2^3 = 0$$

Eliminating w_1, w_2, we get

$$\frac{w_1}{w_2} = -\frac{u_2}{u_1} = -\frac{u_2^3}{u_1^3}$$

or, $u_1^3 u_2 - u_1 u_2^3 = 0$ or $u_1 u_2 (u_1^2 - u_2^2) = 0$

Since, $u_1 \neq u_2 \neq 0$, we have $u_1 = -u_2$.

Also, $w_1 = w_2 = 1$. The third equation gives, $2u_1^2 = \frac{2}{3} \Rightarrow u_1 = \frac{1}{\sqrt{3}},\ u_2 = -\frac{1}{\sqrt{3}}$

Hence, two point Gauss-Legendre quadrature formula is,

$$\int_{-1}^{1} F(u)du = F\left(\frac{1}{\sqrt{3}}\right) + F\left(-\frac{1}{\sqrt{3}}\right)$$

The Table 3.1 gives the abscissae and weights of the Gauss-Legendre quadraturen for values of n from 2 to 6.

Table 3.1 *Values of Weights and Abscissae for Gauss-Legendre Quadrature*

n	Weights	Abscissae
2	1.0	± 0.57735027
3	0.88888889	0.0
	0.55555556	± 0.77459667
4	0.65214515	± 0.33998104
	0.34785485	± 0.86113631
5	0.56888889	0.0
	0.47862867	± 0.53846931
	0.23692689	± 0.90617985
6	0.46791393	± 0.23861919
	0.36076157	± 0.66120939
	0.17132449	± 0.93246951

It is seen that the abscissae are symmetrical with respect to the origin and the weights are equal for equidistant points.

Example 3.20: Compute $\int_0^2 (1+x)dx,$ by Gauss two point quadrature formula.

Solution: Substituting $x = u + 1$, the given integral $\int_0^2 (1+x)dx$ reduces to $I = \int_{-1}^{1} (u+2)du$. Using a two point Gauss quadrature formula, we have $I = (0.57735027+2) + (-0.57735027+2) = 4.0$.

As expected, the result is equal to the exact value of the integral.

Example 3.21: Show that Gauss two-point quadrature formula for evaluating $\int_a^b f(x)dx$ can be written in the composite form as $\int_a^b f(x)dx = h\sum_{i=0}^{N}[f(r_i)+f(s_i)]$

where $r_i = x_i + hp$, $s_i = x_i + (1-p)h$, $p = \frac{1}{6}(3-\sqrt{3}).$

Solution: We subdivide the interval $[a, b]$ into N sub-intervals, each of length h, given by $h = \frac{b-a}{N}$.

Consider the integral I_i over the interval (x_i, x_{i+1}), i.e., $I_{x_i} = \int_{x_i}^{x_{i+1}} f(x)dx$.

We transform the integral I_i by putting $x = \frac{h}{2}u + \left(x_i + \frac{h}{2}\right)$, so that $x = x_i$ gives $u = -1$ and $x = x_{i+1}$ gives $u = 1$. Thus, $I_i = \frac{h}{2}\int_{-1}^{1} f\left(\frac{h}{2}u + x_i + \frac{h}{2}\right)du$.

NOTES

The Gauss two point quadrature gives,

$$I_i = \frac{h}{2}\left[f\left(\frac{h}{2}\cdot\frac{1}{\sqrt{3}}+x_i+\frac{h}{2}\right)+f\left(-\frac{h}{2\sqrt{3}}+x_i+\frac{h}{2}\right)\right]$$
$$= \frac{h}{2}[f(r_i)+f(s_i)]$$

where $r_i = x_i + ph$, $s_i = x_i + (1-p)h$, $p = \frac{1}{6}(3-\sqrt{3})$

Hence, $\int_a^b f(x)dx = \sum_{i=0}^{N-1} I_i = \frac{h}{2}\sum_{i=0}^{N-1}[f(r_i)+f(s_i)]$

Note: Instead of considering Gauss integration formula for more and more number of points for better accuracy, one can use a two point composite formula for larger number of sub-intervals.

Example 3.22: Evaluate the following integral by Gauss three point quadrature formula:

$$I = \int_0^1 \frac{dx}{1+x}$$

Solution: We first transform the interval [0, 1] to the interval (–1, 1) by substituting

$t = 2x - 1$, so that $\int_0^1 \frac{dx}{1+x} = \int_{-1}^1 \frac{dt}{t+3}$.

Now by Gauss three point quadrature we have,

$$I = \frac{1}{9}\ [8F(0) + 5F\ (3+0.77459667) + 5F\ (3.77459667)] \text{ with } F(t) = \frac{1}{t+3}$$
$$\therefore\ \ I = 0.693122$$

The exact value of $\int_0^1 \frac{dx}{1+x} = \ln 2 = 0.693147$

Error = 0.000025

Romberg's Procedure

This procedure is used to find a better estimate of an integral using the evaluation of the integral for two values of the width of the sub-intervals.

Let I_1 and I_2 be the values of an integral $I = \int_a^b f(x)\ dx$, with two different number of sub-intervals of width h_1 and h_2 respectively using the trapezoidal rule. Let E_1 and E_2 be the corresponding truncation errors. Since the errors in trapezoidal rule is of order of h_2, we can write,

$I = I_1 + Kh_1^2$ and $I = I_2 + Kh_2^2$, where K is approximately same.

NOTES

$$\therefore \qquad I_1 + Kh_1^2 = I_2 + Kh_2^2$$

$$\therefore \qquad K \approx \frac{I_1 - I_2}{h_2^2 - h_1^2}$$

Thus, $I \approx I_1 + \dfrac{I_1 - I_2}{h_2^2 - h_1^2}.h_1^2 = \dfrac{I_1 h_2^2 - I_2 h_1^2}{h_2^2 - h_1^2}$

In Romberg procedure, we take $h_2 = \dfrac{h_1}{2}$ and we then have,

$$I = \frac{I_1\left(\frac{h_1}{2}\right)^2 - I_2 h_1^2}{\left(\frac{h_1}{2}\right)^2 - h^2} = \frac{4I_2 - I_1}{3}$$

or,

$$I = I_2 + \left(\frac{I_2 - I_1}{3}\right)$$

This is known as Romberg's formula for trapezoidal integration.

The use of Romberg procedure gives a better estimate of the integral without any more function evaluation. Further, the evaluation of I_2 with $h/2$ uses the function values required in evaluation of I_1.

Example 3.23: Evaluate $I = \int_0^1 \dfrac{dx}{1+x^2}$ by trapezoidal rule with $h_1 = 0.5$ and $h_2 = 0.25$ and then use Romberg procedure for a better estimate of *I*. Compare the result with exact value.

Solution: We tabulate the value of *x* and $y = \dfrac{1}{1+x^2}$ with $h = 0.25$.

x	0	0.25	0.5	0.75	1.0
y	1	0.9412	0.80	0.64	0.5

Thus using trapezoidal rule, with $h_1 = 0.5$, we have

$$I_1 = \frac{0.5}{3} \times (1 + 0.5 + 2 \times 0.8) = 0.516$$

Similarly, with $h_2 = 0.25$,

$$I_2 = \frac{0.25}{3}[1 + 0.5 + 2 \quad (0.8 + 0.9412 + 0.64)]$$
$$= 0.5218$$

The evaluation of I_2 uses the function values for evaluation of I_1.

By Romberg formula,

NOTES

$$I \approx I_2 + \frac{1}{3}\ (I_2 - I_1)$$
$$= 0.5218 + (0.5218 - 0.516) \times \frac{1}{3}$$
$$= 0.5218 + 0.0019$$
$$= 0.5237$$

The exact integral $= \tan^{-1} x\Big|_0^1 = \frac{\pi}{4} = 0.5237.$

Thus we can take the result correct to four places of decimals.

Example 3.24: Evaluate $I = \int_1^2 \frac{dx}{x}$ by trapezoidal rule with two and four sub-intervals and then use Romberg procedure to get a better estimate of *I*.

Solution: We form a table of value of $y = \frac{1}{x}$ with spacing $h = \frac{1}{4} = 0.25.$

x	1	1.25	1.5	1.75	2.0
y	1	0.8	0.6667	0.5714	0.5

$$I_1 = \frac{0.5}{2}[1 + 0.5 + 2 \times 0.6667] = 0.7084$$
$$I_2 = \frac{0.25}{2}\ [1 + 0.5 + 2\ \ (0.8 + 0.6667 + 0.5714)] = 0.6970$$

By Romberg proecedure,

$$I = I_2 + \frac{I_2 - I_1}{3} \approx 0.6970 + \frac{1}{3}(-0.0114)$$
$$= 0.6970 - 0.0038 = 0.6932$$

Example 3.25: Compute the value of $\int_0^1 \frac{dx}{1+x}$, (*i*) By Gauss two point and (*ii*) By Gauss three point formulas.

Solution: We first transform the integral by substituting $x = \frac{b-a}{2}t + \frac{1}{2}(b+a)$

$$\int_0^1 \frac{dx}{1+x} = \frac{1}{2}\int_{-1}^1 \frac{1}{1+\frac{1}{2}+\frac{1}{2}t} = \frac{1}{2}\int_{-1}^1 \frac{2}{3+t}dt$$

(*i*) By Gauss two point quadrature $\int_{-1}^1 F(t)\,dt = F\left(\frac{1}{\sqrt{3}}\right) + F\left(-\frac{1}{\sqrt{3}}\right)$ we get,

$$\int_{-1}^1 \frac{1}{3+t}\,dt = \left(\frac{1}{3+\frac{1}{\sqrt{3}}} + \frac{1}{3-\frac{1}{\sqrt{3}}}\right) = 0.6923$$

(*ii*) By Gauss three point quadrature,

$$\int_{-1}^{1} \frac{dt}{3+t} \quad dt = \left[\frac{1}{3} \times 0.888888 + \frac{0.55555556}{3+0.77459667}\right]$$

$$= 0.443478$$

Example 3.26: Compute $\int_1^2 e^x dx$ by Gauss three point quadrature.

Solution: We first transform the integral by substituting $x = \frac{6-a}{2}t + \frac{1}{2}(b+a) = \frac{1}{2}t + \frac{3}{2}$

$$\therefore \int_1^2 e^x dx = \frac{1}{2}\int_{-1}^{1} e^{\frac{t}{2}+\frac{3}{2}} \, dt = \frac{1}{2}e^{\frac{3}{2}}\int_{-1}^{1} e^{\frac{t}{2}} \, dt$$

$$= \frac{1}{2}e^{\frac{3}{2}}\left[0.88888889 \times e^0 + 0.55555556 \times \left\{e^{\frac{1}{2}} \times 0.77459667 + e^{\frac{1}{2}} \times 0.77459667\right\}\right]$$

$$= 4.67077$$

Check Your Progress

1. Define the process of numerical differentiation.
2. Write Newton's forward difference interpolation formula.
3. Write Newton's backward difference interpolation formula.
4. How will you evaluate a definite integral?
5. Write the trapezoidal formula for numerical integration.
6. What is Simpson's one-third formula of numerical integration?
7. Define Simpson's three-eighth rule of numerical integration.
8. State Weddle's rule.
9. Why is Romberg's procedure used?

3.4 SOLVING NUMERICAL

Even though there are many methods to find an analytical solution of ordinary differential equations, for many differential equations solutions in closed form cannot be obtained. There are many methods available for finding a numerical solution for differential equations. We consider the solution of an initial value problem associated with a first order differential equation given by,

$$\frac{dy}{dx} = f(x, y) \tag{3.48}$$

With $\quad y(x_0) = y_0 \quad$ (3.49)

In general, the solution of the differential equation may not always exist. For the existence of a unique solution of the differential equation (3.48), the following conditions, known as Lipshitz conditions must be satisfied, (*i*) The function $f(x, y)$ is defined and continuous in the strip

$R: \; x_0 \le x \le b, \quad -\infty < y < \infty$

(*ii*) There exists a constant L such that for any x in (x_0, b) and any two numbers y and y_1

$$|f(x, y) - f(x, y_1)| \le L|y - y_1| \tag{3.50}$$

The numerical solution of initial value problems consists of finding the approximate numerical solution of y at successive steps $x_1, x_2,..., x_n$ of x. A number of good methods are available for computing the numerical solution of differential equations.

3.4.1 Taylor Series Method

Consider the solution of the first order differential equation,

$$\frac{dy}{dx} = f(x, y) \text{ with } y(x_0) = y_0 \tag{3.51}$$

where $f(x, y)$ is sufficiently differentiable with respect to x and y. The solution $y(x)$ of the problem can be expanded about the point x_0 by a Taylor series in the form,

$$y(x_0 + h) = y(x_0) + hy'(x_0) + \frac{h^2}{2!}y''(x_0) + ... + \frac{y^k(x_0)}{k!}h^k + \frac{h^{k+1}}{(k+1)!}(\xi)$$

The derivatives in the above expansion can be determined as follows,

$$y'(x_0) = f(x_0, y_0)$$

$$y''(x_0) = f_x(x_0, y_0) + f_y(x_0, y_0)\,y'(x_0)$$

$$y'''(x_0) = f_{xx}(x_0, y_0) + 2f_{xy}(x_0, y_0)\,y'(x_0) + f_{yy}(x_0, y_0)\{y'(x_0)\}^2 + f_y(x, y)y''(x_0)$$

where a suffix x or y denotes partial derivative with respect to x or y.

Thus the value of $y_1 = y(x_0+h)$, can be computed by taking the Taylor series expansion shown above. Usually, because of difficulties in obtaining higher order derivatives, commonly a fourth order method is used. The solution at $x_2 = x_1+h$, can be found by evaluating the derivatives at (x_1, y_1) and using the expansion; otherwise, writing $x_2 = x_0+2h$, we can use the same expansion. This process can be continued for determining y_{n+1} with known values x_n, y_n.

Note: If we take $k = 1$, we get the Euler's method, $y_1 = y_0+h\,f(x_0, y_0)$.

Thus, Euler's method is a particular case of Taylor series method.

Example 3.27: Form the Taylor series solution of the initial value problem, $\frac{dy}{dx} = xy+1,\; y(0) = 1$ up to five terms and hence compute $y(0.1)$ and $y(0.2)$, correct to four decimal places.

Solution: We have, $y' = xy+1,\; y(0) = 1$

Differentiating successively we get,

$$y''(x) = xy' + y, \;\therefore y''(0) = 1$$

$$y'''(x) = xy'' + 2y', \;\therefore y'''(0) = 2$$

$$y^{(iv)}(x) = xy''' + 3y'', \;\therefore y^{(iv)}(0) = 3$$

$$y^{(v)}(x) = xy^{(iv)} + 3y''', \;\therefore y^{(v)}(0) = 6$$

NOTES

Hence, the Taylor series solution $y(x)$ is given by,

$$y(x) \approx y(0) + xy'(0) + \frac{x^2}{2}y''(0) + \frac{x^3}{3!}y'''(0) + \frac{x^4}{4!}y^{(iv)}(0) + \frac{x^5}{5!}y^{(v)}(0)$$

$$\approx 1 + x + \frac{x^2}{2} + \frac{x^3}{6}\times 2 + \frac{x^4}{24}\times 3 + \frac{x^5}{120}\times 6 = 1 + x + \frac{x^2}{2} + \frac{x^3}{3} + \frac{x^4}{8} + \frac{x^5}{20}$$

$$\therefore y(0.1) \approx 1 + 0.1 + \frac{0.01}{2} + \frac{0.001}{3} + \frac{0.0001}{8} + \frac{0.00001}{20} = 1.1053$$

Similarly, $y\ (0.2) \approx 1 + 0.2 + \frac{0.04}{2} + \frac{0.008}{3} + \frac{0.0016}{8} + \frac{0.00032}{20} = 1.04274$

Example 3.28: Find first two non-vanishing terms in the Taylor series solution of the initial value problem $y' = x^2 + y^2,\ y(0) = 0.$ Hence, compute y (0.1), y (0.2), y (0.3) and comment on the accuracy of the solution.

Solution: We have, $y' = x^2 + y^2,\quad y\ (0) = 0$

Differentiating successively we have,

$$\begin{aligned}
y'' &= 2x + 2yy', & \therefore y''(0) &= 0\\
y''' &= 2 + 2[yy'' + (y')^2, & y'''(0) &= 2\\
y^{(iv)} &= 2(yy''' + 3y'y''), & \therefore y^{(iv)}(0) &= 0\\
y^{(v)} &= 2[yy^{(iv)} + 4y'y''' + 3(y'')^2], & \therefore y^{(v)} &= 0\\
y^{(vi)} &= 2[yy^{(v)} + 5y'y^{(iv)} + 10y''y'''], & \therefore y^{(vi)} &= 0\\
y^{(vii)} &= 2[yy^{(vi)} + 6y'y^{(v)} + 15y''y^{(iv)} + 10(y''')^2] & \therefore y^{(vi)}(0) &= 80
\end{aligned}$$

The Taylor series up to two terms is $y(x) = \frac{x^3}{6}\times 2 + \frac{x^7}{7}\times\frac{80}{7!} = \frac{1}{3}x^3 + \frac{x^7}{63}$

Example 3.29: Given $x\,y' = x - y^2$, $y(2) = 1$, evaluate $y(2.1)$, $y(2.2)$ and $y(2.3)$ correct to four decimal places using Taylor series method.

Solution: Given $y' = x - y^2$, i.e., $y' = 1 - y^2/x$ and $y = 1$ for $x = 2.$ To compute $y(2.1)$ by Taylor series method, we first find the derivatives of y at $x = 2$.

$$y' = 1 - y^2/x \qquad \therefore\ y'(2) = 1 - \frac{1}{2} = 0.5$$

$$xy'' + y' = 1 - 2yy'$$

$$2y''(2) + \frac{1}{2} = 1 - 2.\frac{1}{2} \qquad \therefore\ y''(2) = \frac{1}{4} - \frac{2}{2}\times\frac{1}{2} = -0.25$$

$$xy''' + 2y'' = -2y'^2 - 2yy''$$

$$\therefore \qquad 2y'''(2) + 2\left(-\frac{1}{4}\right) = -2\left(\frac{1}{2}\right)^2 - 2\left(-\frac{1}{4}\right)$$

Or, $$2y'''(2) = \frac{1}{2} \qquad \therefore\ y'''(2) = \frac{1}{4} = 0.25$$

$$xy^{(iv)} + 3y''' = -4y'y'' - 2y'y'' - 2yy'''$$

NOTES

$$2y''''(2) + 3 \times \frac{1}{4} = 6 \times \frac{1}{2} \times \left(\frac{1}{4}\right) - 2 \times \frac{1}{4}$$

$$y''''(2) = \left(\frac{3}{4} - \frac{3}{4} - \frac{1}{2}\right)\frac{1}{2} = -0.25$$

$$y(2.1) = y \quad (2) + 0.1 \quad y'(2) + \frac{(0.1)^2}{2} y''(2) + \frac{(0.1)^3}{3!} y'''(2) + \frac{(0.1)^4}{4!} y''''(2)$$

$$= 1 + 0.1 \times 0.5 + \frac{0.01}{2} \times (-0.25) + \frac{0.001}{6} \times 0.25 + \frac{0.0001}{24} \times (-0.25)$$

$$= 1 + 0.05 - 0.00125 + 0.00004 - 0.000001$$

$$= 1.0488$$

$$y(2.2) = 1 + 0.2 \times 0.5 + \frac{0.04}{2} \times (-0.25) + \frac{0.008}{6} \times 0.25 + \frac{0.0016}{24} \times (-0.5)$$

$$= 1 + 0.1 - 0.005 - 0.00032 - 0.00003$$

$$= 1.0954$$

$$y(2.3) = 1 + 0.3 \times 0.5 + \frac{0.09}{2} \quad (-0.25) + \frac{0.009}{2} \times 0.25 + \frac{0.0081}{24} \times (0.5)$$

$$= 1 + 0.15 - 0.01125 + 0.001125 + 0.000168$$

$$= 1.005043$$

Picard's Method of Successive Approximations

Consider the solution of the initial value problem,

$$\frac{dy}{dx} = f(x, y) \text{ with } y(x_0) = y_0$$

Taking $y = y(x)$ as a function of x, we can integrate the differential equation with respect to x from $x = x_0$ to x, in the form

$$y = y_0 + \int_{x_0}^{x} f(x, y(x))\, dx \tag{3.52}$$

The integral contains the unknown function $y(x)$ and it is not possible to integrate it directly. In Picard's method, the first approximate solution $y^{(1)}(x)$ is obtained by replacing $y(x)$ by y_0.

Thus,
$$y^{(1)}(x) = y_0 + \int_{x_0}^{x} f(x, y_0)dx \tag{3.53}$$

The second approximate solution is derived on replacing y by $y^{(1)}(x)$. Thus,

$$y^{(2)}(x) = y_0 + \int_{x_0}^{x} f(x, y^{(1)}(x))\, dx \tag{3.54}$$

The process can be continued, so that we have the general approximate solution given by,

$$y^{(n)}(x) = y_0 + \int_{x_0}^{x} f(x, y^{(n-1)}(x))dx, \text{ for } n = 2, 3... \quad (3.55)$$

This iteration formula is known as Picard's iteration for finding solution of a first order differential equation, when an initial condition is given. The iterations are continued until two successive approximate solutions $y^{(k)}$ and $y^{(k+1)}$ give approximately the same result for the desired values of x up to a desired accuracy.

Note: Due to practical difficulties in evaluating the necessary integration, this method cannot be always used. However, if $f(x, y)$ is a polynomial in x and y, the successive approximate solutions will be obtained as a power series of x.

Example 3.30: Find four successive approximate solutions for the following initial value problem: $y' = x + y$, with $y(0) = 1$, by Picard's method. Hence compute $y(0.1)$ and $y(0.2)$ correct to five significant digits.

Solution: We have, $y' = x + y$, with $y(0) = 1$.

The first approximation by Picard's method is,

$$y^{(1)}(x) = y(0) + \int_0^x [x + y(0)]\, dx$$

$$\therefore \quad y^{(1)}(x) = 1 + \int_0^x (x+1) \quad dx = 1 + x + \frac{x^2}{2}$$

The second approximation is,

$$y^{(2)}(x) = 1 + \int_0^x (x + 1 + x + \frac{x^2}{2})dx = 1 + x + x^2 + \frac{x^3}{6}$$

Similarly, the third approximation is,

$$y^{(3)}(x) = 1 + \int_0^x (1 + 2x + x^2 + \frac{x^3}{6})dx$$

$$\therefore \quad y^{(3)}(x) = 1 + x + x^2 + \frac{x^3}{3} + \frac{x^4}{24}$$

The fourth approximation is,

$$y^{(4)}x = 1 + \int_0^x (1 + 2x + x^2 + \frac{x^3}{3} + \frac{x^4}{24})dx$$

$$\therefore \quad y^{(4)}(x) = 1 + x + x^2 + \frac{x^3}{3} + \frac{x^4}{12} + \frac{x^5}{120}$$

It is clear that successive approximations are easily determined as power series of x having one degree more than the previous one. The value of $y(0.1)$ is given by,

$y\ (0.1) = 1 + 0.1 + (0.1)^2 + \frac{(0.1)^3}{3} + \frac{(0.1)^4}{4} + ... \approx 1.1103,$ correct to five significant digits.

Similarly, $y(0.2) = 1 + 0.2 + (0.2)^2 + \frac{(0.2)^3}{3} + \frac{(0.2)^4}{4} + \frac{(0.2)^5}{120} \approx 1.2431.$

NOTES

Example 3.31: Find the successive approximate solution of the initial value problem, $y' = xy + 1$, with $y(0) = 1$, by Picard's method.

Solution: The first approximate solution is given by,

$$y^{(1)}(x) = 1 + \int_0^x (x+1)dx = 1 + x + \frac{x^2}{2}$$

The second and third approximate solutions are,

$$y^{(2)}(x) = 1 + \int_0^x [x(1 + x + \frac{x^2}{2}) + 1]dx = 1 + x + \frac{x^2}{2} + \frac{x^3}{3} + \frac{x^4}{8}$$

$$y^{(3)}(x) = 1 + \int_0^x [x(1 + x + \frac{x^2}{2} + \frac{x^3}{3} + \frac{x^4}{4}) + 1]dx = 1 + x + \frac{x^2}{2} + \frac{x^3}{3} + \frac{x^4}{8} + \frac{x^5}{15} + \frac{x^6}{48}$$

Example 3.32: Compute $y(0.25)$ and $y(0.5)$ correct to three decimal places by solving the following initial value problem by Picard's method:

$$\frac{dy}{dx} = \frac{x^2}{1+y^2}, y(0) = 0$$

Solution: We have $\frac{dy}{dx} = \frac{x^2}{1+y^2}, y(0) = 0$

By Picard's method, the first approximation is,

$$y^{(1)}(x) = 0 + \int_0^x \frac{x^2}{1+0} dx = \frac{x^3}{3}$$

The second approximate solution is,

$$y^{(2)}(x) = \int_0^x \frac{x^2}{1 + [y^{(1)}(x)]^2} dx$$

$$= \int_0^x \frac{x^2}{1 + \frac{x^6}{9}} dx = \tan^{-1} \frac{x^3}{3}$$

For $x = 0.25, \quad y^{(1)} \quad (0.25) = \frac{(0.25)^2}{3} = 0.0052$

$$y^{(2)} \quad (0.25) = \tan^{-1} \quad \frac{(0.25)^2}{3} \approx 0.0052$$

$\therefore y \quad (0.25) = 0.005$, Correct to three decimal place.

Again, for $x = 0.5$, $\quad y^{(1)}(0.5) = \frac{(0.5)^2}{3} = 0.083333$

$$y^{(2)}(0.5) = \tan^{-1} \frac{(0.5)^3}{3} = 0.0416$$

Thus, correct to three decimal places, $y(0.5) = 0.042$.

NOTES

Note: For this problem we observe that, the integral for getting the third and higher approximate solution is either difficult or impossible to evalute, since

$$y^{(3)}(x) = \int_0^x \frac{x^2}{1+\left(\tan^{-1}\frac{x^3}{3}\right)^2} \text{ is not integrable.}$$

Example 3.33: Use Picard's method to find two successive approximate solutions of the initial value problem,

$$\frac{dy}{dx} = \frac{y-x}{y+x},\ y(0) = 1$$

Solution: The first approximate solution by Picard's method is given by,

$$y^{(1)}(x) = y_0 + \int_0^x f(x, y_0)dx$$

$$\therefore \quad y^{(1)}(x) = 1 + \int_0^x \frac{1-x}{1+x}dx = 1 + \int_0^x \frac{2-(1+x)}{1+x}dx$$

$$\therefore \quad y^{(1)}(x) = 1 + 2\log_e |1+x| - x$$

The second approximate solution is given by,

$$y^{(2)}(x) = y_0 + \int_0^x f(x, y^{(1)}(x))dx$$

$$= 1 + \int_0^x \frac{x - 2x + 2\log_e |1+x|}{1+2\log_e |1+x|}dx = 1 + x - 2\int_0^x \frac{x}{1+2\log_e |1+x|}dx$$

We observe that, it is not possible to obtain the integral for getting $y^{(2)}(x)$. Thus Picard's method is not applicable to get successive approximate solutions.

3.4.2 Euler's Method

This is a crude but simple method of solving a first order initial value problem:

$$\frac{dy}{dx} = f(x, y),\ y(x_0) = y_0$$

This is derived by integrating $f(x_0, y_0)$ instead of $f(x, y)$ for a small interval,

$$\therefore \quad \int_{x_0}^{x_0+h} dy = \int_{x_0}^{x_0+h} f(x_0, y_0)dx$$

$$\therefore \quad y(x_0 + h) - y(x_0) = hf(x_0, y_0)$$

Writing $y_1 = y(x_0 + h)$, we have

$$y_1 = y_0 + h f(x_0, y_0) \qquad (3.56)$$

Similarly, we can write

$$y_2 = y(x_1 + h) = y_1 + h f(x_1, y_1) \qquad (3.57)$$

where $x_1 = x_0 + h$.

NOTES

Proceeding successively, we can get the solution at any $x_n = x_0 + nh$, as

$$y_n = y_{n-1} + h\, f(x_{n-1}, y_{n-1}) \tag{3.58}$$

This method, known as Euler's method, can be geometrically interpreted, as shown in Figure 3.3.

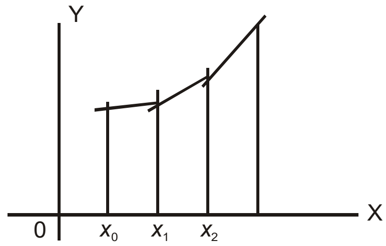


Fig. 3.3 *Euler's Method*

For small step size h, the solution curve $y = y(x)$, is approximated by the tangential line.

The local error at any x_k, i.e., the truncation error of the Euler's method is given by,

$$e_k = y(x_{k+1}) - y_{k+1}$$

where y_{k+1} is the solution by Euler's method.

$$\therefore \quad e_k = y(x_k + h) - \{y_k + hf(x_k, y_k)\}$$

$$= y_k + hy'(x_k) + \frac{h^2}{2} y''(x_k + \theta h) - y_k - hy'(x_k), \quad 0 < \theta < 1$$

$$\therefore \quad e_k = \frac{h^2}{2} y''(x_k + \theta h), \quad 0 < \theta < 1$$

Note: The Euler's method finds a sequence of values $\{y_k\}$ of y for the sequence of values $\{x_k\}$ of x, step by step. But to get the solution up to a desired accuracy, we have to take the step size h to be very small. Again, the method should not be used for a larger range of x about x_0, since the propagated error grows as integration proceeds.

Example 3.34: Solve the following differential equation by Euler's method for $x = 0.1, 0.2, 0.3$; taking $h = 0.1$; $\frac{dy}{dx} = x^2 - y,\ y(0) = 1.$ Compare the results with exact solution.

Solution: Given $\frac{dy}{dx} = x^2 - y$, with $y(0) = 1$.

In Euler's method one computes in successive steps, values of $y_1, y_2, y_3, \ldots$ at $x_1 = x_0 + h$, $x_2 = x_0 + 2h$, $x_3 = x_0 + 3h$, using the formula,

$$y_{n+1} = y_n + hf(x_n,\ \ y_n), \text{ for } n = 0,\ 1,\ 2, \ldots$$

$$\therefore \quad y_{n+1} = y_n + h\ \ (x^2_n - y_n)$$

With $h = 0.1$ and starting with $x_0 = 0$, $y_0 = 1$, we present the successive computations in the table given below.

NOTES

n	x_n	y_n	$f(x_n, y_n) = x^2_n - y_n$	$y_{n+1} = y_n + hf(x_n, y_n)$
0	0.0	1.000	−1.000	0.9000
1	0.1	0.900	−0.8900	0.8110
2	0.2	0.8110	−0.7710	0.7339
3	0.3	0.7339	−0.6439	0.6695

The analytical solution of the differential equation written as $\frac{dy}{dx} + y = x^2$, is

$$ye^x = \int x^2 e^x dx + c$$

or, $$ye^x = x^2 e^x - 2xe^x + 2e^x + c.$$

Since, $$y = 1 \text{ for } x = 0, \therefore c = -1.$$

$\therefore$ $$y = x^2 - 2x + 2 - e^{-x}.$$

The following table compares the exact solution with the approximate solution by Euler's method.

n	x_n	Approximate Solution	Exact Solution	% Error
1	0.1	0.9000	0.9052	0.57
2	0.2	0.8110	0.8213	1.25
3	0.3	0.7339	0.7492	2.04

Example 3.35: Compute the solution of the following initial value problem by Euler's method, for $x = 0.1$ correct to four decimal places, taking $h = 0.02$, $\frac{dy}{dx} = \frac{y-x}{y+x}, y(0) = 1$.

Solution: Euler's method for solving an initial value problem,

$$\frac{dy}{dx} = f(x, y),\ y(x_0) = y_0, \text{ is } y_{n+1} = y_n + h\, f(x_n, y_n), \text{ for } n = 0, 1, 2, ...$$

Taking $h = 0.02$, we have $x_1 = 0.02$, $x_2 = 0.04$, $x_3 = 0.06$, $x_4 = 0.08$, $x_5 = 0.1$.

Using Euler's method, we have, since $y(0) = 1$

$$y(0.02) = y_1 = y_0 + h\, f(x_0, y_0) = 1 + 0.02 \times \frac{1-0}{1+0} = 1.0200$$

$$y(0.04) = y_2 = y_1 + h\, f(x_1, y_1) = 1.0200 + 0.02 \times \frac{1.0200 - 0.02}{1.0200 + 0.02} = 1.0392$$

$$y(0.06) = y_3 = y_2 + h\, f(x_2, y_2) = 1.0392 + 0.02 \times \frac{1.0392 - 0.04}{1.0392 + 0.04} = 1.0577$$

$$y(0.08) = y_4 = y_3 + h\, f(x_3, y_3) = 1.0577 + 0.02 \times \frac{1.0577 - 0.06}{1.0577 + 0.06} = 1.0756$$

$$y(0.1) = y_5 = y_4 + h\, f(x_4, y_4) = 1.0756 + 0.02 \times \frac{1.0756 - 0.08}{1.0756 + 0.08} = 1.0928$$

Hence, $y(0.1) = 1.0928$.

NOTES

Modified Euler's method

In order to get somewhat moderate accuracy, Euler's method is modified by computing the derivative $y' = f(x, y)$, at a point x_n as the mean of $f(x_n, y_n)$ and $f(x_{n+1}, y^{(0)}{}_{n+1})$, where,

$$y^{(0)}{}_{n+1} = y_n + h \quad f(x_n, \ y_n)$$

$$y^{(1)}{}_{n+1} = y_n + \frac{h}{2}[f(x_n, y_n) + f(x_{n+1}, \ y^{(0)}{}_{n+1})] \tag{3.59}$$

This modified method is known as Euler-Cauchy method. The local truncation error of the modified Euler's method is of the order $O(h^3)$.

Note: Modified Euler's method can be used to compute the solution up to a desired accuracy by applying it in an iterative scheme as stated below.

$$\text{Compute } y^{(k)}{}_{n+1} = y_n + h\, f(x_n, \ y_n)$$

$$\text{Compute } \ y_{n+1}^{(k+1)} = y_n + \frac{h}{2}\left[f(x_n, y_n) + f(x_{n+1}, y_{n+1}^{(k)}\right], \text{ for } k = 0, 1, 2, ... \tag{3.60}$$

The iterations are continued until two successive approximations $y_{n+1}^{(k)}$ and $y_{n+1}^{(k+1)}$ coincide to the desired accuarcy. As a rule, the iterations converge rapidly for a sufficiently small h. If, however, after three or four iteration the iterations still do not give the necessary accuracy in the solution, the spacing h is decreased and iterations are performed again.

Example 3.36: Use modified Euler's method to compute y (0.02) for the initial value problem, $\frac{dy}{dx} = x^2 + y,$ with $y(0) = 1$, taking $h = 0.01$. Compare the result with the exact solution.

Solution: Modified Euler's method consists of obtaining the solution at successive points, $x_1 = x_0 + h, x_2 = x_0 + 2h,..., x_n = x_0 + nh$, by the two stage computations given by,

$$y_{n+1}^{(0)} = y_n + hf(x_n, y_n)$$

$$y_{n+1}^{(1)} = y_n + \frac{h}{2}\left[f(x_n, y_n) + f(x_{n+1}, y_{n+1}^{(0)})\right]$$

For the given problem, $f(x, y) = x^2 + y$ and $h = 0.01$

$$y_1^{(0)} = y_0 + h[x_0^2 + y_0] = 1 + 0.01 \times 1 = 1.01$$

$$y_1^{(1)} = 1 + \frac{0.01}{2}[1.0 + 1.01 + (0.01)^2] = 1.01005$$

i.e., $\quad y_1 = y \ (0.01) = 1.01005$

$$\begin{aligned} \text{Next, } y_2^{(0)} &= y_1 + h[x_1^2 + y_1] \\ &= 1.01005 + 0.01[(0.1)^2 + 1.01005] \\ &= 1.01005 + 0.010102 = 1.02015 \end{aligned}$$

NOTES

$$y_2^{(1)} = 1.01005 + \frac{0.01}{2}[(0.01)^2 + 1.01005 + (0.01)^2 + 1.02015]$$

$$= 1.01005 + \frac{0.01}{2} \times \ (2.02140)$$

$$= 1.01005 + 0.10107$$

$$= 1.11112$$

$$\therefore y_2 = y \ \ (0.02) = 1.11112$$

Euler's Method for a Pair of Differential Equations

Consider an initial value problem associated with a pair of first order differential equation given by,

$$\frac{dy}{dx} = f(x, y, z), \frac{dz}{dx} = g(x, y, z) \qquad (3.61)$$

$$\text{with } y(x_0) = y_0, z(x_0) = z_0 \qquad (3.62)$$

Euler's method can be extended to compute approximate values y_i and z_i of $y(x_i)$ and $z(x_i)$ respectively given by,

$$y_{i+1} = y_i + h f(x_i, y_i, z_i)$$

$$z_{i+1} = z_i + h g(x_i, y_i, z_i) \qquad (3.63)$$

starting with $i = 0$ and continuing step by step for $i = 1, 2, 3,...$ Evidently, we can also extend Euler's method for an initial value problem associated with a second order differential equation by rewriting it as a pair of first order equations.

Consider the initial value problem,

$$\frac{d^2y}{dx^2} = g\left(x, y, \frac{dy}{dx}\right), \text{ with } y(x_0) = y_0,\ y'(x_0) = y'_0$$

We write $\frac{dy}{dx} = z,$ so that $\frac{dz}{dx} = g(x, y, z)$ with $y(x_0) = y_0$ and $z(x_0) = y'_0$.

Example 3.37: Compute $y(1.1)$ and $y(1.2)$ by solving the initial value problem,

$$y'' + \frac{y'}{x} + y = 0, \text{ with } y(1) = 0.77,\ y'(1) = -0.44$$

Solution: We can rewrite the problem as $y' = z,\ z' = -\frac{z}{x} - y;$ with $y,(1) = 0.77$ and $z(1.1) = -0.44$.

Taking $h = 0.1$, we use Euler's method for the problem in the form,

$$y_{i+1} = y_i + hz_i$$

$$z_{i+1} = z_i + h\left[-\frac{z_1}{x_i} - y_i\right], i = 0, 1, 2,...$$

Thus, $y_1 = y(1.1)$ and $z_1 = z(1.1)$ are given by,

$$y_1 = y_0 + hz_0 = 0.77 + 0.1 \times (-0.44) = 0.726$$

$$z_1 = z_0 + h\left[-\frac{z_0}{x_0} - y_0\right] = -0.44 + 0.1 \times (0.44 - 0.77)$$

$$= -0.44 - 0.33 = -0.473$$

Similarly, $y_2 = y(1.2) = y_1 + hz_1 = 0.726 - 0.1\ (-0.473) =$ 0.679

$$z_2 = z(1.2) = z_1 + h \left[-\frac{z_1}{x_1} - y_1 \right]$$

$$= -0.473 + 0.1 \times \left(\frac{0.473}{1.1} - 0.726 \right)$$

$$= -0.473 + 0.1 \times -0.296 = -0.503$$

Thus, $y\ (1.1) = 0.726$ and $y\ (12) = 0.679$.

Example 3.38: Using Euler's method, compute $y\,(0.1)$ and $y\,(0.2)$ for the initial value problem,

$$y'' + y = 0,\ y(0) = 0,\ y'(0) = 1$$

Solution: We rewrite the initial value problem as $y' = z,\ z' = -y$, with $y\,(0) = 0$, $z\,(0) = 1$.

Taking $h = 0.1$, we have by Euler's method,

$$\begin{aligned}
y_1 &= y(0.1) = y_0 + hz_0 && = 0 + 0.1 \times 1 && = 0.1 \\
z_1 &= z(0.1) = z_0 + h(-y_0) && = 1 + 0.1 \times 0 && = 1.0 \\
y_2 &= y(0.2) = y_1 + hz_1 && = 0.1 + 0.1 \times 1.0 && = 0.2 \\
z_2 &= z(0.2) = z_1 - hy_1 && = 1.0 - 0.1 \times 0.1 && = 0.99
\end{aligned}$$

Example 3.39: For the initial value problem $y'' + xy' + y = 0$, $y(0) = 0$, $y'\,(0) = 1$. Compute the values of y for 0.05, 0.10, 0.15 and 0.20, having accuracy not exceeding 0.5×10^{-4}.

Solution: We form Taylor series expansion using $y\,(0)$, $y'\,(0) = 1$ and from the differential equation,

$$y'' + xy' + y = 0, \text{ we get } y''(0) = 0$$

$$y'''(x) = -xy'' - 2y' \qquad \therefore y'''(0) = -2$$

$$y^{(iv)}(x) = -xy''' - 3y'' \qquad \therefore y^{iv}(0) = 0$$

$$y^{v}(x) = -xy^{(iv)} - 4y''' \quad \therefore y^{v}(0) = 8$$

And in general, $y^{(2n)}(0) = 0$, $y^{(2n+1)}(0) = -2ny^{(2n-1)}(0) = (-1)^n 2^n.2!$

Thus, $y(x) = x - \dfrac{x^3}{3} + \dfrac{x^5}{15} - ... + (-1)^n \dfrac{2^n n! x 2^{n+1}}{(2n+1)!} + ...$

This is an alternating series whose terms decrease. Using this, we form the solution for y up to 0.2 as given below:

x	0	0.05	0.10	0.15	0.20
$y(x)$	0	0.0500	0.0997	0.1489	0.1973

3.4.3 Runge-Kutta Methods

Runge-Kutta method can be of different orders. They are very useful when the method of Taylor series is not easy to apply because of the complexity of finding higher order derivatives. Runge-Kutta methods attempt to get better accuracy and at the same time obviate the need for computing higher order derivatives. These methods, however, require the evaluation of the first order derivatives at several off-step points.

Here we consider the derivation of Runge-Kutta method of order 2.

The solution of the $(n + 1)$th step is assumed in the form,

$$y_{n+1} = y_n + ak_1 + bk_2 \tag{3.64}$$

where $k_1 = h\,f(x_n, y_n)$ and

$$k_2 = h\,f(x_n + \alpha h, y_n + \beta k_1), \text{ for } n = 0, 1, 2,... \tag{3.65}$$

The unknown parameters a, b, α, and β are determined by expanding in Taylor series and forming equations by equating coefficients of like powers of h. We have,

$$y_{n+1} = y(x_n + h) = y_n + h\,y'(x_n) + \frac{h^2}{2}y''(x_n) + \frac{h^3}{6}y'''(x_n) + 0(h^4)$$

$$= y_n + h\,f(x_n, y_n) + \frac{h^2}{2}[f_x + ff_y]_n + \frac{h^3}{6}[f_{xx} + 2ff_{yy} + f_{yy}f^2 + f_x f_y + f_y^2 f]_n + 0(h^4)$$

(3.66)

The subscript n indicates that the functions within brackets are to be evaluated at (x_n, y_n).

Again, expanding k_2 by Taylor series with two variables, we have

$$k_2 = h[f_n + ah \quad (f_x)_n + \beta k_1 (f_y)_n + \frac{\alpha^2\beta^2}{2} \quad (f_{xx})_n + \alpha\beta hk_1 \quad (f_{xy})_n + \frac{\beta^2 k_1^2}{2} \quad (f_{yy})_n + 0(h^3)]$$

(3.66) Thus on substituting the expansion of k_2, we get from Equation (3.66)

$$y_{n+1} = y_n + (a+b)h \quad f_n + bh^2 \quad (\alpha f_x + \beta ff_y)_n + bh^3 \quad \left(\frac{\alpha^2}{2}f_{xx} + \alpha\beta ff_{xx} + \frac{\beta^2}{2}f^2 f_{yy}\right) + 0(h^4)$$

On comparing with the expansion of y_{n+1} and equating coefficients of h and h^2 we get the relations,

$$a + b = 1, \quad b\alpha = b\beta = \frac{1}{2}$$

There are three equations for the determination of four unknown parameters. Thus, there are many solutions. However, usually a symmetric solution is taken by setting $a = b = \frac{1}{2}$, then $\alpha = \beta = 1$

Thus we can write a Runge-Kutta method of order 2 in the form,

$$y_{n+1} = y_n + \frac{h}{2} \quad [f(x_n, \quad y_n) + f(x_n + h, \quad y_n + h \quad f(x_n, \quad y_n))], \text{ for } n = 0, \ 1, \ 2,...$$

(3.67)

Proceeding as in second order method, Runge-Kutta method of order 4 can be formulated. Omitting the derivation, we give below the commonly used Runge-Kutta method of order 4.

NOTES

$$y_{n+1} = y_n + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4) + 0(h^5)$$

$$k_1 = h\ f(x_n, y_n)$$

$$k_2 = h f\left(x_n + \frac{h}{2},\ y_n + \frac{k_1}{2}\right)$$

$$k_3 = h f\left(x_n + \frac{h}{2},\ y_n + \frac{k_2}{2}\right)$$

$$k_4 = h f(x_n + h,\ y_n + k_3) \quad (3.68)$$

Runge-Kutta method of order 4 requires the evaluation of the first order derivative $f(x, y)$, at four points. The method is self-starting. The error estimate with this method can be roughly given by,

$$|y(x_n) - y_n| \approx \frac{y_n^* - y_n}{15} \quad (3.69)$$

where y_n^* and y_n are the approximate values computed with $\frac{h}{2}$ and h respectively as step size and $y(x_n)$ is the exact solution.

Note: In particular, for the special form of differential equation $y' = F(x)$, a function of x alone, the Runge-Kutta method reduces to the Simpson's one-third formula of numerical integration from x_n to x_{n+1}. Then,

$$y_{n+1} = y_n + \int_{x_n}^{x_{n+1}} F(x)dx$$

or,
$$y_{n+1} = y_n + \frac{h}{6}[F(x_n) + 4F(x_n + \frac{h}{2}) + F(x_n + h)]$$

Runge-Kutta methods are widely used particularly for finding starting values at steps $x_1, x_2, x_3, ...$, since it does not require evaluation of higher order derivatives. It is also easy to implement the method in a computer program.

Example 3.40: Compute values of $y(0.1)$ and $y(0.2)$ by 4th order Runge-Kutta method, correct to five significant figures for the initial value problem,

$$\frac{dy}{dx} = x + y,\ y(0) = 1$$

Solution: *We have* $\frac{dy}{dx} = x + y,\ y(0) = 1$

$$\therefore f(x,\ y) = x + y, \quad h = 0.1, \quad x_0 = 0, \quad y_0 = 1$$

By Runge-Kutta method,

$$y(0.1) = y(0) + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4)$$

where,
$$k_1 = h\ f(x_0,\ y_0) = 0.1 \times (0 + 1) = 0.1$$

$$k_2 = h\ f\left(x_0 + \frac{h}{2},\ y_0 + \frac{k_2}{2}\right) = 0.1 \times (0.05 + 1.05) = 0.11$$

$$k_3 = h\ f\left(x_0 + \frac{h}{2},\ y_0 + \frac{k_2}{2}\right) = 0.1 \times (0.05 + 1.055) = 0.1105$$

$$k_4 = h\ f(x_0 + h,\ y_0 + k_3) = 0.1 \times (0.1 + 1.1105) = 0.12105$$

$$\therefore\ y(0.1) = 1 + \frac{1}{6}\ [0.1 + 2 \times (0.11 + 0.1105 + 0.12105] = 1.130516$$

Thus, $x_1 = 0.1, \quad y_1 = 1.130516$

where,
$$y\,(0.2) = y(0.1) + \frac{1}{6}\;\;(k_1 + 2k_2 + 2k_3 + k_4)$$
$$k_1 = h\; f\,(x_1, \;\; y_1) = 0.1 \times (0.1 + 1.11034) = 0.121034$$
$$k_2 = h\; f\left(x_1 + \frac{h}{2}, y_1 + \frac{k_1}{2}\right) = 0.1\,(0.15 + 1.17086) = 0.132086$$
$$k_3 = h\; f\left(x_1 + \frac{h}{2}, y_1 + \frac{k_2}{2}\right) = 0.1\,(0.15 + 1.17638) = 0.132638$$
$$k_4 = h\; f\,(x_1 + h, y_1 + k_3) = 0.1 \;\;\; (0.2 + 1.24298) = 0.144298$$
$$y_2 = y\,(0.2) = 1.11034 + \frac{1}{6}\;\;[0.121034 + 2\,(0.132086 + 0.132638) + 0.144298] = 1.2428$$

Example 3.41: Use Runge-Kutta method of order 4 to evaluate y (1.1) and y (1.2), by taking step length $h = 0.1$ for the initial value problem,

$$\frac{dy}{dx} = x^2 + y^2, \; y\,(1) = 0$$

Solution: For the initial value problem,

$\dfrac{dy}{dx} = f\,(x, \;\; y), \; y\;(x_0) = y_0$, the Runge-Kutta method of order 4 is given as,

$$y_{n+1} = y_n + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4)$$

where

$$k_1 = h\; f\,(x_n, \;\; y_n)$$
$$k_2 = h\; f\left(x_n + \frac{h}{2}, \quad y_n + \frac{k_1}{2}\right)$$
$$k_3 = h\; f\left(x_n + \frac{h}{2}, \quad y_n + \frac{k^2}{2}\right)$$
$$k_4 = h\; f\,(x_n + h, y_n + k_3), \text{ for } n = 0, 1, 2, ...$$

For the given problem, $f\,(x, y) = x^2 + y^2$, $x_0 = 1$, $y_0 = 0$, $h = 0.1$.

Thus,

$$k_1 = h\; f\,(x_0, y_0) = 0.1 \times (1^2 + 0^2) = 0.1$$
$$k_2 = h\; f\left(x_0 + \frac{h}{2}, y_0 + \frac{k_1}{2}\right) = 0.1 \times [(1.05)^2 + (0.5)^2] = 0.13525$$
$$k_3 = h\; f\left(x_0 + \frac{h}{2}, y_0 + \frac{k_2}{2}\right) = 0.1 \times [(1.05)^2 + (0.05525)^2] = 0.13555$$

$$k_4 = h\; f\,(x_0 + h, y_0 + k_3) = 0.1 \times [(1.1)^2 + (0.13555)^2] = 0.12283$$
$$\therefore y_1 = y_0 + \frac{1}{6}\;\;(k_1 + 2k_2 + 2k_3 + k_4)$$
$$= \frac{1}{6}(0.1 + 0.2705 + 0.2711 + 0.12283) = \frac{1}{6} \times 0.76443$$
$$= 0.127405$$

NOTES

For y (1.2):

$$k_1 = 0.1[(1.1)^2 + (0.11072)^2] = 0.12226$$
$$k_2 = 0.1[(1.15)^2 + (0.17183)^2] = 0.135203$$
$$k_3 = 0.1[(1.15)^2 + (0.17832)^2] = 0.135430$$
$$k_4 = 0.1[(1.2)^2 + (0.24615)^2] = 0.150059.$$

$$\therefore \; y_2 = y\,(1.2) = 0.11072 + \frac{1}{6}(0.12226 + 0.270406 + 0.270860 + 0.150069)$$
$$= 0.24631$$

Algorithm: Solution of first order differential equation by Runge-Kutta method of order 2: $y' = f(x)$ with $y\,(x_0) = y_0$.

Step 1: Define $f(x, y)$

Step 2: Read x_0, y_0, h, x_f [h is step size, x_f is final x]

Step 3: Repeat Steps 4 to 11 until $x_1 > x_f$

Step 4: Compute $k_1 = f(x_0, y_0)$

Step 5: Compute $y_1 = y_0 + hk_1$

Step 6: Compute $x_1 = x_0 + h$

Step 7: Compute $k_2 = f(x_1, y_1)$

Step 8: Compute $y_1 = y_0 + h \times (k_1 + k_2)/2$

Step 9: Write x_1, y_1

Step 10: Set $x_0 = x_1$

Step 11: Set $y_0 = y_1$

Step 12: Stop

Algorithm: Solution of $y_1 = f(x, \; y), \; y(x_0) = y_0$ by Runge-Kutta method of order 4.

Step 1: Define $f(x, y)$

Step 2: Read x_0, y_0, h, x_f

Step 3: Repeat Step 4 to Step 16 until $x_1 > x_f$

Step 4: Compute $k_1 = h\,f(x_0, y_0)$

Step 5: Compute $x = x_0 + \frac{h}{2}$

Step 6: Compute $y = y_0 + \frac{k_1}{2}$

Step 7: Compute $k_2 = h\,f(x, y)$

Step 8: Compute $y = y_0 + \frac{k_2}{2}$

Step 9: Compute $k_3 = h\,f(x, y)$

Step 10: Compute $x_1 = x_0 + h$

Step 11: Compute $y = y_0 + k_3$

Step 12: Compute $k_4 = h\,f(x_1, y)$

Step 13: Compute $y_1 = y_0 + (k_1 + 2\,(k_2 + k_3) + k_4)/6$

Step 14: Write x_1, y_1

Step 15: Set $x_0 = x_1$

Step 16: Set $y_0 = y_1$

Step 17: Stop

Runge-Kutta Method for a Pair of Equations

Consider an initial value problem associated with a system of two first order ordinary differential equations in the form,

$$\frac{dy}{dx} = f(x, y, z), \frac{dz}{dx} = g(x, y, z)$$

with $y(x_0) = y_0$ and $z(x_0) = z_0$

The Runge-Kutta method of order 4 can be easily extended in the following form,

$$\begin{aligned} y_{i+1} &= y_i + \frac{1}{6}\;(k_1 + 2k_2 + 2k_3 + k_4) \\ z_{i+1} &= z_i + \frac{1}{6}\;(l_1 + 2l_2 + 2l_3 + l_4) \text{ for } i = 0, 1, 2, ... \end{aligned} \tag{3.70}$$

where

$$\begin{aligned} k_1 &= hf(x_i, y_i, z_i), & l_1 &= hg(x_i, y_i, z_i) \\ k_2 &= hf\left(x_i + \frac{h}{2}, y_i + \frac{k_1}{2}, z_i + \frac{l_1}{2}\right), & l_2 &= hg\left(x_i + \frac{h}{2}, y_i + \frac{k_1}{2}, z_1 + \frac{l_1}{2}\right) \\ k_3 &= hf\left(x_i + \frac{h}{2}, y_i + \frac{k_2}{2}, z_i + \frac{l_2}{2}\right), & l_3 &= hg\left(x_i + \frac{h}{2}, y_i + \frac{k_2}{2}, z_i + \frac{l_2}{2}\right) \\ k_4 &= hf(x_i + h, y_1 + k_3, z_i + l_3), & l_4 &= hf(x_i + h, y_i + k_3, z_i + l_3) \end{aligned}$$

$$y_i = y(x_i),\; z_i = z(x_i),\; i = 0, 1, 2, ...$$

The solutions for $y(x)$ and $z(x)$ are determined at successive step points $x_1 = x_0 + h$, $x_2 = x_1 + h = x_0 + 2h, ..., x_N = x_0 + Nh$.

Runge-Kutta Method for a Second Order Differential Equation

Consider the initial value problem associated with a second order differential equation,

$$\frac{d^2y}{dx^2} = g(x, y, y')$$

with $y\,(x_0) = y_0$ and $y'\,(x_0) = \alpha_0$

On substituting $z = y'$, the above problem is reduced to the problem,

$$\frac{dy}{dx} = z, \frac{dz}{dx} = g(x, y, z)$$

with $y\,(x_0) = y_0$ and $z\,(x_0) = y'\,(x_0) = \alpha_0$

NOTES

which is an initial value problem associated with a system of two first order differential equations. Thus we can write the Runge-Kutta method for a second order differential equation as,

$$y_{i+1} = y_i + \frac{1}{6} \quad (k_1 + 2k_2 + 2k_3 + k_4),$$
$$z_{i+1} = y'_{i+1} = z_i + \frac{1}{6} \quad (l_1 + 2l_2 + 2l_3 + l_4) \text{ for } i = 0, 1, 2,... \tag{3.71}$$

where $k_1 = h(z_i), \qquad l_1 = hg(x_i,\ \ y_i,\ \ z_i)$

$$k_2 = h\left(z_i + \frac{l_1}{2}\right), \quad l_2 = hg\left(x_i + \frac{h}{2},\ \ y_i + \frac{k_1}{2},\ \ z_i + \frac{l_1}{2}\right)$$
$$k_3 = h\left(z_i + \frac{l_2}{2}\right), \quad l_3 = hg\left(x_i + \frac{h}{2},\ \ y_i + \frac{k_2}{2},\ \ z_i + \frac{l_2}{2}\right)$$
$$k_4 = h(z_i + l_3), \quad l_4 = hg(x_i + h,\ \ y_i + k_3,\ \ z_i + l_3)$$

Multistep Methods

We have seen that for finding the solution at each step, the Taylor series method and Runge-Kutta methods requires evaluation of several derivatives. We shall now develop the multistep method which require only one derivative evaluation per step; but unlike the self starting Taylor series or Runge-Kutta methods, the multistep methods make use of the solution at more than one previous step points.

Let the values of y and y^1 already have been evaluated by self-starting methods at a number of equally spaced points $x_0, x_1,..., x_n$. We now integrate the differential equation,

$$\frac{dy}{dx} = f(x,\ \ y), \text{ from } x_n \text{ to } x_{n+1}$$

i.e.,
$$\int_{x_n}^{x_{n+1}} dy = \int_{x_n}^{x_{n+1}} f(x,\, y)dx$$

$\therefore$
$$y_{n+1} = y_n + \int_{x_n}^{x_{n+1}} f(x,\, y(x))\, dx$$

To evaluate the integral on the right hand side, we consider $f(x, y)$ as a function of x and replace it by an interpolating polynomial, i.e., a Newton's backward difference interpolation using the $(m + 1)$ points $x_n, x_{n+1}, x_{n-2},..., x_{n-m}$,

$$p_m(x) = \sum_{k=0}^{m} (-1)^k \binom{-s}{k} \Delta^k f_{n-k}, \text{ where } s = \frac{x - x_n}{h}$$
$$\binom{-s}{k} = -s(-s-1)(-s-2)...(-s-k+1) \cdot \frac{1}{k!}$$

Substituting $p_m(x)$ in place of $f(x, y)$, we obtain

NOTES

$$y_{n+1} = y_n + h\int_0^1 \sum_{k=0}^{m} (-1)^k \binom{-s}{k} \Delta^k f_{n-k}\, ds$$

$$= y_n + h\ [\gamma f_n + \gamma_1 \Delta f_{n-1} + \gamma_2 \Delta^2 f_{n-2} + ... + \gamma_m \Delta f_{n-m}]$$

where $\gamma_k = (-1)^k \int_0^1 \binom{-s}{k} ds$

The coefficients γ_k can be easily computed to give,

$$\gamma_0 = 1,\ \gamma_1 = \frac{1}{2},\ \gamma_2 = \frac{5}{12},\ \gamma_3 = \frac{3}{8},\ \gamma_4 = \frac{251}{720}, \text{etc.}$$

Taking $m = 3$, the above formula gives,

$$y_{n+1} = y_n + h\left[f_n + \frac{1}{2}\Delta f_{n-1} + \frac{5}{12}\Delta^2 f_{n-2} + \frac{3}{8}\Delta^3 f_{n-3}\right]$$

Substituting the expression of the differences in terms of function values given by,

$$\Delta f_{n-1} = f_n - f_{n-1},\ \Delta^2 f_{n-2} = f_n - 2f_{n-1} + f_{n-2}$$
$$\Delta^3 f_{n-3} = f_n - 3f_{n-1} + 3f_{n-2} - f_{n-3}$$

We get on arranging,

$$y_{n+1} = y_n + \frac{h}{24}[55f_n - 59f_{n-1} + 37f_{n-2} - 9f_{n-3}] \quad (3.72)$$

This is known as *Adams-Bashforth formula of order* 4. The local error of this formula is,

$$E = h^5 f^{iv}(\xi)\int_0^1 \binom{s+3}{4} ds \quad (3.73)$$

By using mean value theorem of integral calculus,

$$E = h^5 f^{iv}(\eta)\int_0^1 \binom{s+3}{4} ds$$

or,

$$E = h^5 f^{iv}(\eta).\frac{251}{720} \quad (3.74)$$

The fourth order Adams-Bashforth formula requires four starting values, i.e., the derivaties, f_3, f_2, f_1 and f_0. This is a multistep method.

Predictor-Correction Methods

These methods use a pair of multistep numerical integration. The first is the Predictor formula, which is an open-type explicit formula derived by using, in the integral, an interpolation formula which interpolates at the points $x_n, x_{n-1},..., x_{n-m}$. The second is the Corrector formula which is obtained by using interpolation formula that interpolates at the points $x_{n+1}, x_n, ..., x_{n-p}$ in the integral.

NOTES

Euler's Predictor-Corrector Formula

The simplest formula of the type is a pair of formula given by,

$$y_{n+1}^{(p)} = y_n + h\,f(x_n,\, y_n) \qquad (3.75)$$

$$y_{n+1}^{(c)} = y_n + \frac{h}{2}[f(x_n,\, y_n) + f(x_{n+1},\, y_{n+1}^{(p)})] \qquad (3.76)$$

In order to determine the solution of the problem upto a desired accuracy, the corrector formula can be employed in an iterative manner as shown below:

Step 1: Compute $y_{n+1}^{(0)}$, using Equation (3.75)

i.e., $y_{n+1}^{(0)} = y_n + h\,f(x_n, y_n)$

Step 2: Compute $y_{n+1}^{(k)}$ using Equation (3.76)

i.e., $y_{n+1}^{(k)} = y_n + \dfrac{h}{2}\;[f(x_n,\; y_n) + f(x_{n+1},\; y_{n+1}^{(k-1)})]$, for $K = 1,\; 2,\; 3,\ldots,$

The computation is continued till the condition given below is satisfied,

$$\left|\frac{y_{n+1}^{(k)} - y_{n+1}^{(k-1)}}{y_{n+1}^{(k)}}\right| < \varepsilon \qquad (3.77)$$

where ε is the prescribed accuracy.

It may be noted that the accuracy achieved will depend on step size *h* and on the local error. The local error in the predictor and corrector formula are,

$$\frac{h^2}{2}y''(\eta_1) \quad \text{and} \quad -\frac{h^3}{12}y'''(\eta_2), \text{ respectively.}$$

Milne's Predictor-Corrector Formula

A commonly used Predictor-Corrector system is the fourth order *Milne's Predictor-Corrector* formula. It uses the following as Predictor and Corrector.

$$y_{n+1}^{(p)} = y_{n-3} + \frac{4h}{3}(2f_n - f_{n-1} + 2f_{n-2}^{*})$$

$$y_{n+1}^{(c)} = y_{n-1} + \frac{h}{3}[f_{n-1} + 4f_n + f_{n+1}(x_{n+1}, y_{n+1}^{(p)})] \qquad (3.78)$$

The local errors in these formulae are respectively,

$$\frac{14}{45}h^5 y^{(v)}(\xi_1) \text{ and } -\frac{1}{90}h^5 y^{(v)}(\xi_2) \qquad (3.79)$$

Example 3.42: Compute the Taylor series solution of the problem $\frac{dy}{dx} = xy + 1,\; y(0) = 1,$ up to x^5 terms and hence compute values of $y(0.1)$, $y(0.2)$ and $y(0.3)$. Use Milne's Predictor-Corrector method to compute $y(0.4)$ and $y(0.5)$.

Solution: We have $y' = xy + 1,$ with $y(0) = 1$, $\therefore\, y'(0) = 1$

Differentiating successively, we get

$$y''(x) = xy' + y \qquad \therefore\, y''(0) = 1$$
$$y'''(x) = xy'' + y' \qquad \therefore\, y'''(0) = 2$$
$$y^{(iv)}(x) = xy''' + 3y'' \qquad \therefore\, y^{(iv)}(0) = 3$$
$$y^{(v)}(x) = xy^{(iv)} + 4y''' \qquad \therefore\, y^{(iv)}(0) = 8$$

NOTES

Thus the Taylor series solution is given by,

$$y(x) = y(0) + xy'(0) + \frac{x^2}{2}y''(0) + \frac{x^3}{3!}y'''(0) + \frac{x^4}{4!}y^{iv}(0) + \frac{x^5}{5!}y^{(v)}(0)$$

$$= 1 + x + \frac{x^2}{2} + \frac{x^3}{3!} \times 2 + \frac{x^4}{4!} \times 3 + \frac{x^5}{5!} \times 8$$

$$\therefore y(x) = 1 + x + \frac{x^2}{2} + \frac{x^3}{3} + \frac{x^4}{8} + \frac{x^5}{15}$$

$$\therefore \quad y(0.1) = 1 + 0.1 + \frac{0.01}{2} + \frac{0.001}{3} + \frac{0.0001}{8} + \frac{0.00001}{15}$$

$$= 1.1053$$

$$y(0.2) = 1 + 0.2 + \frac{0.04}{2} + \frac{0.008}{3} + \frac{0.0016}{8} + \frac{0.00032}{15}$$

$$= 1.22288$$

$$y(0.3) = 1 + 0.3 + \frac{0.09}{2} + \frac{0.027}{3} + \frac{0.0081}{8} + \frac{0.00243}{15}$$

$$= 1.35526$$

For application of Milne's Predictor-Corrector method, we compute $y'(0.1)$, $y'(0.2)$ and $y'(0.3)$.

$$y'(0.1) = 0.1 \times 1.1053 + 1 = 1.11053$$
$$y'(0.2) = 0.2 \times 1.22288 + 1 = 1.244576$$
$$y'(0.3) = 0.3 \times 1.35526 + 1 = 1.40658$$

The Predictor formula gives, $y_4 = y(0.4) = y_0 + \frac{4h}{3}(2y_1' - y_2' + 2y_3')$.

$$\therefore y_4^{(0)} = 1 + \frac{4 \times 0.1}{3}(2 \times 1.11053 - 1.24458 + 2 \times 1.40658)$$

$$= 1.50528 \qquad \therefore y_4' = 1 + 0.4 \times 1.50528 = 1.602112$$

The Corrector formula gives, $y_4^{(1)} = y_2 + \frac{h}{3}(y_2' + 4y_3' + y_4')$.

$$y(0.4) = 1.22288 + \frac{0.1}{3}(1.24458 + 4 \times 1.40658 + 1.60211)$$

$$= 1.22288 + 0.28243$$

$$= 1.50531$$

3.4.4 Higher Order Differential Equations

We consider the solution of ordinary differential equation of order 2 or more, when value of the dependent variable is given at more than one point, usually at the two ends of an interval in which the solution is required. For example, the simplest boundary value problem associated with a second order differential equation is,

$$y'' + p(x)y' + q(x)y = r(x) \qquad (3.80)$$

with boundary conditions, $y(a) = A, y(b) = B$. (3.81)

The following two methods reduce the boundary value problem into initial value problems which are then solved by any of the methods for solving such problems.

NOTES

Reduction to a Pair of Initial Value Problem

This method is applicable to linear differential equations only. In this method, the solution is assumed to be a linear combination of two solutions in the form,

$$y(x) = u(x) + \lambda \upsilon(x) \quad (3.82)$$

where λ is a suitable constant determined by using the boundary condition and $u(x)$ and $\upsilon(x)$ are the solutions of the following two initial value problems:

(*i*) $u'' + p(x)u' + q(x)u = r(x)$

$$u(a) = A,\ u'(a) = \alpha_1, \text{(say)}. \quad (3.83)$$

(*ii*) $Y'' + P(X)\,Y' + Q(X)\upsilon = R(X)$

$$\upsilon(a) = 0 \text{ and } \upsilon'(a) = \alpha_2, \text{(say)} \quad (3.84)$$

where α_1 and α_2 are arbitrarily assumed constants. After solving the two initial value problems, the constant λ is determined by satisfying the boundary condition at $x = b$. Thus,

$$B = u(b) + \lambda \upsilon(b)$$

or, $$\lambda = \frac{B - \upsilon(b)}{\upsilon(b)}, \text{ provided } \upsilon(b) \neq 0 \quad (3.85)$$

Evidently, $y(a) = A$, is already satisfied.

If $\upsilon(b) = 0$, then we solve the initial value problem for υ again by choosing $\upsilon'(a) = \alpha_3$, for some other value for which $\upsilon(b)$ will be non-zero.

Another method which is commonly used for solving boundary problems is the finite difference method discussed below.

Finite Difference Method

In this method of solving boundary value problem, the derivatives appearing in the differential equation and boundary conditions, if necessary, are replaced by appropriate difference gradients.

Consider the differential equation, $y'' + p(x)\,y' + q(x)y = r(x)$ (3.86)

with the boundary conditions, $y(a) = \alpha$ and $y(b) = \beta$

(3.87)

The interval $[a, b]$ is divided into N equal parts each of width h, so that $h = (b-a)/N$, and the end points are $x_0 = a$ and $x_n = b$. The interior mesh points x_i at which solution values $y(x_i)$ are to be determined are,

$$x_n = x_0 + nh,\ n = 1, 2, ..., N-1 \quad (3.88)$$

The values of y at the mesh points is denoted by y_n given by,

$$y_n = y(x_0 + nh),\ n = 0, 1, 2, ..., N \quad (3.89)$$

The following central difference approximations are usually used in finite difference method of solving boundary value problem,

$$y'(x_n) \approx \frac{y_{n+1} - y_{n-1}}{2h} \quad (3.90)$$

$$y''(x_n) \approx \frac{y_{n+1} - 2y_n + y_{n-1}}{h^2}$$

(3.91)

NOTES

Substituting these in the differential equation, we have

$2(y_{n+1}-2y_n+ y_{n-1}) + p_n\, h(y_{n+1}- y_{n-1}) + 2h^2 g_n y_n = 2r_n h^2,$

where $p_n = p(x_n)$, $q_n = q(x_n)$, $r_n = r(x_n)$ (3.92)

Rewriting the equation by regrouping we get,

$$(2-hp_n)y_{n-1}+(-4+2h^2q_n)y_n+(2+h^2q_n)y_{n+1} = 2r_n h^2$$

(3.93)

This equation is to be considered at each of the interior points, i.e., it is true for $n = 1, 2, ..., N-1$.

The boundary conditions of the problem are given by,

$$y_0 = \alpha, \quad y_n = \beta \tag{3.94}$$

Introducing these conditions in the relevant equations and arranging them, we have the following system of linear equations in $(N-1)$ unknowns $y_1, y_2, ..., y_{n-1}$.

$$\begin{aligned}
&(-4+2h^2q_1)y_1 + (2+hp_1)y_2 = 2r_1h^2 - (2-hp_1)\alpha \\
&(2-hp_2)y_1 + (-4+2h^2q_2)y_2 + (2+hp_2)y_3 = 2r_2h^2 \\
&(2-hp_3)y_2 + (-4+2h^2q_3)y_3 + (2+hp_3)y_4 = 2r_3h^2 \\
&\ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \\
&(2-hp_{N-2}) + (-4+2h^2q_{N-2})y_{N-2} \quad (2+hp_{N-2})y_{N-1} = 2r_{N-2}h^2 \\
&(2-hp_{N-1}) \quad y_{N-2} + (-4+2h^2q_{N-1})y_{N-1} = 2r_{N-1}h^2 - (2-hp_{N-1})\beta
\end{aligned} \tag{3.95}$$

The above system of $N-1$ equations can be expressed in matrix notation in the form

$$Ay = b$$

(3.96)

where the coefficient matrix A is a tridiagonal one, of the form

$$A = \begin{bmatrix} B_1 & C_1 & 0 & 0\ldots & 0 & 0 & 0 \\ A_2 & B_2 & C_2 & 0\ldots & 0 & 0 & 0 \\ 0 & A_3 & B_3 & C_3\ldots & 0 & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & 0\ldots & A_{N-2} & B_{N-2} & C_{N-2} \\ 0 & 0 & 0 & 0\ldots & 0 & A_{N-1} & B_{N-1} \end{bmatrix} \tag{3.97}$$

Where,

$$\begin{aligned}
B_i &= -4+2h^2q_i, \quad i = 1, 2, \ldots, N-1 \\
C_i &= 2+hp_i, \qquad i = 1, 2, \ldots, N-2 \\
A_i &= 2-hp_i, \qquad i = 2, 3, \ldots, N-1
\end{aligned}$$

(3.98)

NOTES

The vector b has components,

$$b_1 = 2\gamma_1 h^2 - (2 - hp_1)\alpha$$
$$b_i = 2\gamma_i h^2, \text{for } i = 2, 3, ..., N-2$$
$$b_{N-1} = 2\gamma_{N-1} + h^2 - (2 - hlp_{N-1})\beta \quad (3.99)$$

The system of linear equations can be directly solved using suitable methods.

Example 3.43: Compute values of $y(1.1)$ and $y(1.2)$ on solving the following initial value problem, using Runge-Kutta method of order 4:

$$y'' + \frac{y'}{x} + y = 0, \text{ with } y(1) = 0.77,\ y'(1) = -0.44$$

Solution: We first rewrite the initial value problem in the form of pair of first order equations.

$$y' = z,\ z' = \frac{-z}{x} - y$$

with $y(1) = 0.77$ and $z(1) = -0.44$.

We now employ Runge-Kutta method of order 4 with $h = 0.1$,

$$y(1.1) = y(1) + \frac{1}{6}\ (k_1 + 2k_2 + 2k_3 + k_4)$$
$$y'(1.1) = z(1.1) = 1 + \frac{1}{6}(l_1 + 2l_2 + 2l_3 + l_4)$$
$$k_1 = -0.44 \times 0.1 = -0.044$$
$$l_1 = 0.1 \times \left(\frac{0.44}{1} - 0.77\right) = -0.033$$
$$k_2 = 0.1 \times \left(-0.44 - \frac{0.033}{2}\right) = 0.04565$$
$$l_2 = 0.1 \times \left(\frac{0.4565}{1.05} - 0.748\right) = -0.031323809$$

$$k_3 = 0.1 \times \left(-0.44 + \frac{-0.03123809}{2}\right) = -0.0455661904$$
$$l_3 = 0.1 \times \left[\frac{0.0455661904}{1.05} - 0.747175\right] = -0.031321128$$
$$k_4 = 0.1 \times (-0.47132112) = -0.047132112$$
$$l_4 = 0.1 \times \left(\frac{0.047132112}{1.1} - 0.72443381\right) = -0.068158643$$
$$\therefore\ y(1.1) = 0.77 + \frac{1}{6}[-0.044 + 2 \times (-0.045661904) - 0.029596005] = 0.727328602$$
$$y'(1.1) = -0.44 + \frac{1}{6}\ [-0.033 + 2(-0.031323809) + 2(-0.031321128) - 0.029596005]$$
$$= -0.44 + \frac{1}{6}[-0.33 - 0.062647618 - 0.062642256 - 0.029596005]$$
$$= -0.526322021$$

NOTES

Example 3.44: Compute the solution of the following initial value problem for $x = 0.2$, using Taylor series solution method of order 4: n.l.

$$\frac{d^2y}{dx^2} = y + x\frac{dy}{dx}, \quad y(0) = 1, \quad y'(0) = 0$$

Solution: Given $y'' = y + xy'$, we put $z = y'$, so that

$z' = y + xz,\ y' = z$ and $y(0) = 1,\ z(0) = 0.$

We solve for y and z by Taylor series method of order 4. For this we first compute $y''(0),\ y'''(0),\ y^{iv}(0),...$

We have,

$$y''(0) = y(0) + 0 \times y'(0) = 1, \quad z'(0) = 1$$
$$y'''(0) = z''(0) = y'(0) + z(0) + 0.z'(0) = 0$$
$$y^{iv}(0) = z'''(0) = y''(0) + 2z'(0) + 0.z''(0) = 3$$
$$z^{iv}(0) = 4z''(0) + 0.z'''(0) = 0$$

By Taylor series of order 4, we have

$$y(0 + x) = y(0) + xy'(0) + \frac{x^2}{2!}y''(0) + \frac{x^3}{3!}y'''(0) + \frac{x^4}{4!}y^{iv}(0)$$

or,
$$y(x) = 1 + \frac{x^2}{2!} + \frac{x^4}{4!} \times 3$$

$$\therefore y(0.2) = 1 + \frac{(0.2)^2}{2!} + \frac{(0.2)^4}{8} = 1.0202$$

Similarly, $y'(0.2) = z(0.2) = 0.2 + \dfrac{(0.2)^3}{4!} \times 3 = 0.204$

Example 3.45: Compute the solution of the following initial value problem for $x = 0.2$ by fourth order Runge -Kutta method: n.l. $\dfrac{d^2y}{dx^2} = xy,\ y(0) = 1,\ y'(0) = 1$

Solution: Given $y'' = xy$, we put $y' = z$ and the simultaneous first order problem,

$y' = z = f(x, y, z)$, say $z' = xy = g(x, y, z)$, say with $y(0) = 1$ and $z(0) = 1$

We use Runge-Kutta 4th order formulae, with $h = 0.2$, to compute $y(0.2)$ and $y'(0.2)$, given below.

$$k_1 = h\, f(x_0,\ y_0,\ z_0) = 0.2 \times 1 = 0.2$$
$$l_1 = h\, g(x_0,\ y_0,\ z_0) = 0.2 \times 0 = 0$$
$$k_2 = h\, f\left(x_0 + \frac{h}{2},\ y_0 + \frac{k_1}{2},\ z_0 + \frac{l_1}{2}\right) = 0.2 \times (1 + 0) = 0.2$$
$$l_2 = h\, g\left(x_0 + \frac{h}{2},\ y_0 + \frac{k_1}{2},\ z_0 + \frac{l_1}{2}\right) = 0.2 \times \frac{0.2}{2}\left(1 + \frac{0.2}{2}\right) = 0.022$$
$$k_3 = h\, f\left(x_0 + \frac{h}{2},\ y_0 + \frac{k_2}{2}, z_0 + \frac{l_2}{2}\right) = 0.2 \times 1.011 = 0.2022$$
$$l_3 = h\, g\left(x_0 + \frac{h}{2},\ y_0 + \frac{k_2}{2},\ z_0 + \frac{l_2}{2}\right) = 0.2 \times 0.1 \times 1.1 = 0.022$$
$$k_4 = h\, f(x_0 + h,\ y_0 + k_3,\ z_0 + l_3) = 0.2 \times 1.022 = 0.2044$$
$$l_4 = h\ \ g(x_0 + h,\ y_0 + k_3,\ z_0 + l_3) = 0.2 \times 0.2 \times 1.2022 = 0.048088$$
$$y(0.2) = 1 + \frac{1}{6}\ \ (0.2 + 2(0.2 + 0.2022) + 0.2044) = 1.2015$$
$$y'(0.2) = 1 + \frac{1}{6}\ \ (0 + 2\ \ (0.022 + 0.022) + 0.048088) = 1.02268$$

NOTES

Check Your Progress

10. How are Euler's method and Taylor's method related?
11. Define Picard's method of successive approximation.
12. Why should we not use Euler's method for a larger range of x?
13. When are Runge-Kutta methods applied?
14. What is a predictor formula?
15. What are local errors in Milne's predictor-corrector formulae?
16. Where can the method of reduction to a pair of initial value problem be applied?

3.5 ANSWERS TO 'CHECK YOUR PROGRESS'

1. Numerical differentiation is the process of computing the derivatives of a function $f(x)$ when the function is not explicitly known, but the values of the function are known for a given set of arguments $x = x_0, x_1, x_2, ..., x_n$. To find the derivatives, we use a suitable interpolating polynomial and then its derivatives are used as the formulae for the derivatives of the function.
2. Newton's forward difference interpolation formula is,

$$\varphi(u) = y_0 + u\ \Delta\ y_0 + \frac{u(u-1)}{2\,!}\Delta^2 y_0 + \frac{u(u-1)(u-2)}{3\,!}\Delta^3 y_0 + ... + \frac{u(u-1)(u-2)...(u-n+1)}{n\,!}\Delta^n y_0$$

where $u = \frac{x - x_0}{h}$

3. Newton's backward difference interpolation formula is,

$$\varphi(v) = y_n + v\ \nabla y_n + \frac{v(v+1)}{2\,!}\nabla^2 y_n + \frac{v(v+1)(v+2)}{3\,!}\nabla^3 y_n + \frac{v(v+1)(v+2)(v+3)}{4\,!}\nabla^4 y_n + ...$$
$$+ \frac{v(v+1)...(v+n-1)}{n\,!}\nabla^n y_n$$

Where $v = \frac{x - x_n}{h}$

4. The evaluation of a definite integral cannot be carried out when the integrand $f(x)$ is not integrable, as well as when the function is not explicitly known but only the function values are known at a finite number of values of x. There are two types of numerical methods for evaluating a definite integral based on the following formula:

$$\int_a^b f(x)\ dx$$

5. The formula is, $\int_{x_0}^{x_1} f(x)dx = \frac{h}{2}[f_0 + f_1]$.
6. The formula is, $\int_{x_0}^{x_2} f(x)dx = \frac{h}{3}[f_0 + 4f_1 + f_2]$.

7. Simpson's three-eighth rule of numerical integration is, $\int_a^b f(x)\,dx = \frac{3h}{8}$ [y_0 $+ 3y_1 + 3y_2 + 2y_3 + 3y_4 + 3y_5 + 2y_6 + \ldots + 2y_{3m-3} + 3y_{3m-2} + 3y_{3m-1} + y_{3m}$]
where $h = (b–a)/(3m)$; for $m = 1, 2, ...$

8. The Weddle's rule is, $\int_a^b f(x)dx = \frac{3h}{10}$ [$y_0 + 5y_1 + y_2 + 6y_3 + y_4 + 5y_5 + 2y_6 + 5y_7 + y_8 + 6y_9 + y_{10} + 5y_{11} + ... + 2y_{6m-6} + 5y_{6m-5} + y_{6m-4} + 6y_{6m-3} + y_{6m-2} + 5y_{6m-1} + y_{6m}$], where $b - a = 6mh$.

9. This procedure is used to find a better estimate of an integral using the evaluation of the integral for two values of the width of the sub-intervals.

10. If we take $k = 1$, we get the Euler's method, $y_1 = y_0 + h f(x_0, y_0)$.

11. In Picard's method the first approximate solution $y^{(1)}(x)$ is obtained by replacing $y(x)$ by y_0. Thus, $y^{(1)}(x) = y_0 + \int_{x_0}^{x} f(x, y_0)dx$. The second approximate solution is derived on replacing y by $y(1)$ (x). Thus,

$$y^{(2)}(x) = y_0 + \int_{x_0}^{x} f(x,\, y^{(1)}(x))\, dx.$$

This iteration formula is known as Picard's iteration for finding solution of a first order differential equation, when an initial condition is given. The iterations are continued until two successive approximate solutions y^k and y^{k+1} give approximately the same result for the desired values of x up to a desired accuracy.

12. The method should not be used for a larger range of x about x_0, since the propagated error grows as integration proceeds.

13. Runge-Kutta methods are very useful when the method of Taylor series is not easy to apply because of the complexity of finding higher order derivatives.

14. A predictor formula is an open-type explicit formula derived by using, in the integral, an interpolation formula which interpolates at the points $x_n, x_{n-1}, ..., x_{n-m}$.

15. The local errors in these formulae are $\frac{14}{45}h^5 y^{(v)}(\xi_1)$ and $-\frac{1}{90}h^5 y^{(v)}(\xi_2)$.

16. This method is applicable to linear differential equations only.

3.6 SUMMARY

- Numerical differentiation is the process of computing the derivatives of a function $f(x)$ when the function is not explicitly known, but the values of the function are known only at a given set of arguments $x = x_0, x_1, x_2, ..., x_n$.

NOTES

- For finding the derivatives, we use a suitable interpolating polynomial and then its derivatives are used as the formulae for the derivatives of the function.
- For computing the derivatives at a point near the beginning of an equally spaced table, Newton's forward difference interpolation formula is used, whereas Newton's backward difference interpolation formula is used for computing the derivatives at a point near the end of the table.
- Let the values of an unknown function $y = f(x)$ be known for a set of equally spaced values $x_0, x_1, \ldots, x_n$ of x, where $x_r = x_0 + r_h$. Newton's forward difference interpolation formula is,

$$\varphi(u) = y_0 + u \quad \Delta \quad y_0 + \frac{u(u-1)}{2\ !}\Delta^2 y_0 + \frac{u(u-1)(u-2)}{3\ !}\Delta^3 y_0 + \ldots + \frac{u(u-1)(u-2)\ldots(u-n+1)}{n\ !}\Delta^n y_0$$

where $u = \dfrac{x - x_0}{h}$.

- At the tabulated point x_0, the value of u is zero and the formulae for the derivatives are given by,

$$y'(x_0) = \frac{1}{h}\left[\Delta y_0 - \frac{1}{2}\Delta^2 y_0 + \frac{1}{3}\Delta^3 y_0 - \frac{1}{4}\Delta^4 y_0 + \frac{1}{5}\Delta^5 y_0 - \ldots\right]$$

$$y''(x_0) = \frac{1}{h^2}\left[\Delta^2 y_0 - \Delta^3 y_0 + \frac{11}{12}\Delta^4 y_0 - \frac{5}{6}\Delta^5 y_0 + \ldots\right]$$

- For a given x near the end of the table, the values of $\dfrac{dy}{dx}$ and $\dfrac{d^2 y}{dx^2}$ are computed by first computing $v = (x - x_n)/h$ and using the above formulae. At the tabulated point x_n, the derivatives are given by,

$$y'(x_n) = \frac{1}{h}\left[\nabla y_n + \frac{1}{2}\nabla^2 y_n + \frac{1}{3}\nabla^3 y_n + \frac{1}{4}\nabla^4 y_n + \ldots\right]$$

$$y''(x_n) = \frac{1}{h^2}\left[\nabla^2 y_n + \nabla^3 y_n + \frac{11}{12}\nabla^4 y_n + \frac{5}{6}\nabla^5 y_n + \ldots\right]$$

- For computing the derivatives at a point near the middle of the table, the derivatives of the central difference interpolation formula is used.
- If the arguments of the table are unequally spaced, then the derivatives of the Lagrange's interpolating polynomial are used for computing the derivatives of the function.
- Numerical differentiation is the process of computing the derivatives of a function $f(x)$ when the function is not explicitly known, but the values of the function are known only at a given set of arguments $x = x_0, x_1, x_2, \ldots, x_n$.
- For computing the derivatives at a point near the beginning of an equally spaced table, Newton's forward difference interpolation formula is used,

NOTES

whereas Newton's backward difference interpolation formula is used for computing the derivatives at a point near the end of the table.

- Numerical methods can be applied to determine the value of the integral when the integrand is not integrable as well as when the function is not explicitly known but only the function values are known.
- The two types of numerical methods for evaluating a definite integral are Newton-Cotes quadrature and Gaussian quadrature.
- Taking $n = 2$ in the Newton-Cotes formula, we get Simpson's one-third formula of numerical integration while taking $n = 3$, we get Simpson's three-eighth formula of numerical integration.
- In Newton-Cotes formula with $n = 6$ some minor modifications give the Weddle's formula.
- For evaluating a definite integral correct to a desired accuracy, one has to make a suitable choice of the value of h, the length of sub-interval to be used in the formula.
- There are two ways of determining h, by considering the truncation error in the formula to be used for numerical integration or by successive evaluation of the integral by the technique of interval halving and comparing the results.
- In the truncation error estimation method, the value of h to be used is determined by considering the truncation error in the formula for numerical integration.
- When the estimation of the truncation error is cumbersome, the method of interval halving is used to compute an integral to the desired accuracy.
- Numerical evaluation of double integrals is done by applying trapezoidal rule and Simpson's one-third rule.
- This procedure is used to find a better estimate of an integral using the evaluation of the integral for two values of the width of the sub-intervals.
- There are many methods available for finding a numerical solution for differential equations.
- Picard's iteration is a method of finding solutions of a first order differential equation when an initial condition is given.
- Euler's method is a crude but simple method for solving a first order initial value problem.
- Euler's method is a particular case of Taylor's series method.
- Runge-Kutta methods are useful when the method of Taylor series is not easy to apply because of the complexity of finding higher order derivatives.
- For finding the solution at each step, the Taylor series method and Runge-Kutta methods require evaluation of several derivatives.
- The multistep method requires only one derivative evaluation per step; but unlike the self starting Taylor series on Runge-Kutta methods, the multistep methods make use of the solution at more than one previous step points.
- These methods use a pair of multistep numerical integration. The first is the predictor formula, which is an open-type explicit formula derived by using, in the integral, an interpolation formula which interpolates at the points

NOTES

$x_n, x_{n-1}, ..., x_{n-m}$. The second is the corrector formula which is obtained by using interpolation formula that interpolates at the points $x_{n+1}, x_n, ..., x_{n-p}$ in the integral.

- The solution of ordinary differential equation of order 2 or more, when values of the dependent variable is given at more than one point, usually at the two ends of an interval in which the solution is required.
- The methods used to reduce the boundary value problem into initial value problems are reduction to a pair of initial value problem and finite difference method.

3.7 KEY TERMS

- **Numerical differentiation:** It is the process of computing the derivatives of a function $f(x)$ when the function is not explicitly known, but the values of the function are known for a given set of arguments $x = x_0, x_1, x_2, ..., x_n$.
- **Newton's forward difference interpolation formula:** The Newton's forward difference interpolation formula is used for computing the derivatives at a point near the beginning of an equally spaced table.
- **Newton's backward difference interpolation formula:** Newton's backward difference interpolation formula is used for computing the derivatives at a point near the end of the table.
- **Central difference interpolation formula:** For computing the derivatives at a point near the middle of the table, the derivatives of the central difference interpolation formula is used.
- **Newton-Cotes quadrature:** This is based on integrating polynomial interpolation formulae and requires a table of values of the integrand at equally spaced values of the independent variable x.
- **Trapezoidal formula:** The trapezoidal formula of numerical integration is defined using the definite integral of the function $f(x)$ between the limits x_0 to x_1, as it is approximated by the area of the trapezoidal region bounded by the chord joining the points (x_0, f_0) and (x_1, f_1), the x-axis and the ordinates at $x = x_0$ and at $x = x_1$.
- **Romberg's procedure:** This procedure is used to find a better estimate of an integral using the evaluation of the integral for two values of the width of the sub-intervals.
- **Weddle's rule:** It is a composite Weddle's formula and is used when the number of sub-intervals is multiple of 6.
- **Predictor formula:** It is an open-type explicit formula derived by using, in the integral, an interpolation formula which interpolates at the points x_n, $x_{n-1}, ..., x_{n-m}$.
- **Corrector formula:** It is obtained by using interpolation formula that interpolates at the points $x_{n+1}, x_n, ..., x_{n-p}$ in the integral.

3.8 SELF-ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. Define the term numerical differentiation.
2. Give the differentiation formula for Newton's forward difference interpolation.
3. How the derivative $\frac{dy}{dx}$ can be evaluated?
4. Give the formulae for the derivatives at the tabulated point $x0$ where the value of u is zero.
5. Give the differentiation formula for Newton's backward difference interpolation.
6. Give the Newton's backward difference interpolation formula for an equally spaced table of a function.
7. State Newton-Cotes formula.
8. State the trapezoidal rule.
9. What is the difference between Simpson's one-third formula and one-third rule?
10. What is the error in Weddle's rule?
11. Give the truncation error in Simpson's one-third rule.
12. Where is interval halving technique used?
13. Name the methods used for numerical evaluation of double integrals.
14. State the Gauss quadrature formula.
15. State an application of Romberg's procedure.
16. What are ordinary differential equations?
17. Name the methods for computing the numerical solution of differential equations.
18. What is the significance of Runge-Kutta methods of different orders?
19. When is multistep method used?
20. Name the predictor-corrector methods.
21. How will you find the numerical solution of boundary value problems?

Long-Answer Questions

1. Discuss numerical differentiation using Newton's forward difference interpolation formula and Newton's backward difference interpolation formula.

NOTES

2. Use the following table of values to compute $\int_0^3 f(x)\,dx$:

x	0	1	2	3
$f(x)$	1.6	3.8	8.2	15.4

3. Use suitable formulae to compute $y'(1.4)$ and $y''(1.4)$ for the function $y = f(x)$, given by the following tabular values:

x	1.4	1.8	2.2	2.6	3.0
y	0.9854	0.9738	0.8085	0.5155	0.1411

4. Compute $\frac{dy}{dx}$ and $\frac{d^2y}{dx^2}$ for $x = 1$ where the function $y = f(x)$ is given by the following table:

x	1	2	3	4	5	6
y	1	8	27	64	125	216

5. A rod is rotating in a plane about one of its ends. The following table gives the angle θ (in radians) through which the rod has turned for different values of time t seconds. Find its angular velocity $\frac{d\theta}{dt}$ and angular acceleration $\frac{d^2\theta}{dt^2}$ at $t = 1.0$.

t secs	0.0	0.2	0.4	0.6	0.8	1.0
θ radians	0.0	0.12	0.48	1.10	2.00	3.20

6. Find $\frac{dy}{dx}$ and $\frac{d^2y}{dx^2}$ at $x = 1$ and at $x = 3$ for the function $y = f(x)$, whose values in [1, 6] are given in the following table:

x	1	2	3	4	5	6
y	2.7183	3.3210	4.0552	4.9530	6.0496	7.3891

7. Find $\frac{dy}{dx}$ and $\frac{d^2y}{dx^2}$ at $x = 0.96$ and at $x = 1.04$ for the function $y = f(x)$ given in the following table:

x	0.96	0.98	1.0	1.02	1.04
y	0.7825	0.7739	0.7651	0.7563	0.7473

8. Use suitable formulae to compute $y'(1.4)$ and $y''(1.4)$ for the function $y = f(x)$, given by the following tabular values.

x	1.4	1.8	2.2	2.6	3.0
y	0.9854	0.9738	0.8085	0.5155	0.1411

9. Compute $\frac{dy}{dx}$ and $\frac{d^2y}{dx^2}$ for $x = 1$ where the function $y = f(x)$ is given by the following table:

x	1	2	3	4	5	6
y	1	8	27	64	125	216

10. Compute $\int_0^{20} f(x)\, dx$ by Simpson's one-third rule, where:

x	0	5	10	15	20
$f(x)$	1.0	1.6	3.8	8.2	15.4

11. Compute $\int_0^4 x^3\, dx$ by Simpson's one-third formula and comment on the result:

x	0	2	4
x^3	0	8	64

12. Compute $\int_0^2 x^3\, dx$ by Simpson's one-third formula and comment on the result:

13. Compute $\int_0^2 e^x\, dx$ by Simpson's one-third formula and compare with the exact value, where $e^0 = 1$, $e^1 = 2.72$, $e^2 = 7.39$.

14. Compute an approximate value of π, by integrating $\int_0^1 \frac{dx}{1+x^2}$, by Simpson's one-third formula.

15. A rod is rotating in a plane about one of its ends. The following table gives the angle θ (in radians) through which the rod has turned for different values of time t seconds. Find its angular velocity $\frac{d\theta}{dt}$ and angular acceleration $\frac{d^2\theta}{dt^2}$ at $t = 1.0$.

t secs	0.0	0.2	0.4	0.6	0.8	1.0
θ radius	0.0	0.12	0.48	1.10	2.00	3.20

16. Find $\frac{dy}{dx}$ and $\frac{d^2y}{dx^2}$ at $x = 1$ and at $x = 3$ for the function $y = f(x)$, whose values are given in the following table:

x	1	2	3	4	5	6
y	2.7183	3.3210	4.0552	4.9530	6.0496	7.3891

NOTES

17. Find $\frac{dy}{dx}$ and $\frac{d^2y}{dx^2}$ at $x = 0.96$ and at $x = 1.04$ for the function $y = f(x)$ given in the following table:

x	0.96	0.98	1.0	1.02	1.04
y	0.7825	0.7739	0.7651	0.7563	0.7473

18. Compute $\int_0^2 (x+1)dx,$ by trapezoidal rule by taking four sub-intervals and comment on the result by comparing it with the exact value.
19. Compute $\int_1^{1.4} (x^3+2)dx,$ by Simpson's one-third rule by taking four sub-intervals and find the error in the result.
20. Evaluate $\int_0^1 \cos x \, dx,$ correct to three significant figures taking five equal sub-intervals.
21. Compute the value of the integral $\int_0^1 \frac{xdx}{1+x}$ correct to three significant figures by Simpson's one-thrid rule with six sub-intervals.
22. Compute the integral $\int_0^1 \frac{dx}{1+x},$ by Simpson's one-third rule taking four sub-intervals and use it to compute the approximate value of $\prod$.
23. Compute $\int_0^4 e^x dx$ by Simpson's rule correct to four significant digits taking four sub-intervals and compare it with the exact value.
24. Compute the approximate value of $\int_1^2 \frac{dx}{x}$ by Simpson's one-third rule with four sub-intervals.
25. Evaluate $\int_0^1 \sqrt{1-x^3} \, dx$ by trapezoidal rule, taking four sub-intervals; give the result upto four decimal places.
26. Compute the following integral by Simpson's one-third rule taking $h = 0.05$ correct to five significant digits $\int_1^{1.3} \sqrt{x} \, dx.,$
27. Compute the integral $\int_0^{\frac{\pi}{2}} \sin x \, dx$ by (*a*) Trapezoidal rule and (*b*) Simpson's one-third rule taking six sub-intervals and compare the results with the exact value.
28. Evaluate the following integrals by Weddle's rule,

    (a) $\int_0^1 \frac{dx}{1+x^2},$ taking $n = 12$

    (b) $\int_0^1 \frac{x^2+1}{x^2+1} \, dx,$ taking $n = 12$

29. Compute $\int_0^1 xe^{-x}dx$ by Gauss-Legendre two point and three point formula, and compare with the exact value.
30. Evaluate the following integrals by Gauss-Legendre three point formula:
    (a) $\int_0^{\frac{\pi}{2}} \sin x\, dx$
    (b) $\int_0^1 \cos xe^{-x^2} dx$
    (c) $\int_0^1 \frac{dx}{1+x^2}$
31. Illustrate Romberg's procedure.
32. Use Picard's method to compute values of $y(0.1)$, $y(0.2)$ and $y(0.3)$ correct to four decimal places, for the problem, $y' = x + y,\ \ y(0) = 1$.
33. Compute values of y at $x = 0.02$, by Euler's method taking $h = 0.01$, given y is the solution of the following initial value problem: $\frac{dy}{dx} = x^3 + y, y(0) = 1$.
34. Evaluate $y(0.02)$ by modified Euler's method, given $y' = x^2 + y, y(0) = 1$, correct to four decimal places.
35. Given $y' = \frac{1}{x^2 + y}$, $y(4) = 4$, compute $y(4.2)$ by Taylor series method, taking $h = 0.1$.
36. Using Runge-Kutta method of order 4, compute $y(0.1)$ for each of the following problems:
    (a) $\frac{dy}{dx} = x + y,\ y(0) = 1$
    (b) $\frac{dy}{dx} = x + y^2,\ y(0) = 1$
37. Compute solution of the following initial value problem by Runge-Kutta method of order 4 taking $h = 0.2$ upto $x = 1$; $y' = x - y, y(0) = 1.5$.
38. Given $\frac{dy}{dx} = \frac{1}{2}(1+x^2)y^2$, and $y(0) = 1, y(0.1) = 1.06, y(0.2) = 1.12, y(0.3) = 1.21$. Compute $y(0.4)$ by Milne's predictor-corrector method.

3.9 FURTHER READING

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

Chandan, J.S., Jagjit Singh and K.K. Khanna. 1995. *Business Statistics*. New Delhi: Vikas Publishing House.

Elhance, D.N. 2006. *Fundamental of Statistics*. Allahabad: Kitab Mahal.

Freud, J.E., and F.J. William. 1997. *Elementary Business Statistics – The Modern Approach*. New Jersey: Prentice-Hall International.

NOTES

Goon, A.M., M.K. Gupta, and B. Das Gupta. 1983. *Fundamentals of Statistics*. Vols. I & II, Kolkata: The World Press Pvt. Ltd.

Gupta, S.C. 2008. *Fundamentals of Business Statistics*. Mumbai: Himalaya Publishing House.

Kothari, C.R. 1984. *Quantitative Techniques*. New Delhi: Vikas Publishing House.

Levin, Richard. I., and David. S. Rubin. 1997. *Statistics for Management.* New Jersey: Prentice-Hall International.

Meyer, Paul L. 1970. *Introductory Probability and Statistical Applications*. Massachusetts: Addison-Wesley.

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

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

Anderson, David R., Dennis J. Sweeney and Thomas A. Williams. *Essentials of Statistics for Business and Economics*. Mumbai: Thomson Learning, 2007.

S.P. Gupta. 2021. *Statistical Methods*. Delhi: Sultan Chand and Sons.

UNIT 4 STATISTICAL COMPUTATION AND PROBABILITY DISTRIBUTION

NOTES

Structure

4.0 INTRODUCTION

Statistics is the discipline that concerns the collection, organization, analysis, interpretation, and presentation of data. Every day we are confronted with some form of statistical information through different sources. All raw data cannot be

NOTES

termed as statistics. Similarly, single or isolated facts or figures cannot be called statistics as these cannot be compared or related to other figures within the same framework. Hence, any quantitative and numerical data can be identified as statistics when it possesses certain identifiable characteristics according to the norms of statistics.

In statistics, the term statistical computation specifies the method through which the quantitative data have a tendency to cluster approximately about some value. A measure of statistical computation is any precise method of specifying this 'Central Value'. In the simplest form, the measure of statistical computation is an average of a set of measurements, where the word average refers to as mean, median, mode or other measures of location. Typically the most commonly used measures are arithmetic mean, mode and median. The measures of dispersion, which in itself is a very important property of a distribution and needs to be measured by appropriate statistics. Hence, this unit has taken into consideration several aspects of dispersion. It describes absolute and relative measures of dispersion. It deals with range, the crudest measure of dispersion. It also explains quartile deviation, mean deviation and standard deviation. The standard deviation is the most useful measure of dispersion.

The subject of probability in itself is a cumbersome 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 have similar meanings. 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. The probability theory helps a decision-maker to analyse a situation and decide accordingly. We study why all these uncertainties require knowledge of probability so that calculated risks can be taken. 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.

The probability distribution of a discrete random variable is a list of probabilities associated with each of its possible values. It is also sometimes called the probability function or the probability mass function. The probability density function of a continuous random variable is a function which can be integrated to obtain the probability that the random variable takes a value in a given interval. The binomial distribution is used in finite sampling problems where each observation is one of two possible outcomes ('Success' or 'Failure'). The Poisson distribution is used for modelling rates of occurrence. The 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.

In this unit, you will learn about the history and meaning of statistics, scope of statistics, various measures of statistical computations, measures of dispersion, standard deviation, probability and standard probability distribution.

NOTES

4.1 OBJECTIVES

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

- Examine the functions and meaning of statistics
- Understand the various measures of statistical data
- Analysis the absolute and relative measures of distribution
- Discuss the meaning, uses and merits of range in statistical presentation
- Define standard deviation
- Understand the basic concept of probability
- Do random experiment
- Explain about the concepts of probability distribution
- Describe the Poisson distribution
- Analyse Poisson distribution as an approximation of binomial distribution
- Understand exponential distribution
- Learn about the uniform distribution (discrete random and continuous variable)

4.2 HISTORY AND MEANING OF STATISTICS

The term statistics is used to mean either statistical data or statistical method. When it is used in the sense of statistical data it refers to *quantitative* aspects of things, and is a numerical description. Thus, the distribution of family incomes is a quantitative description, as also the annual production figures of various industries. These quantities are numerical to begin with. But there are also some quantities which are not in themselves numerical, but can be made so by *counting.* The sex of a baby is not a number, but by counting the number of boys, we can associate a numerical description to the sex of all new-born babies, for example, when saying that 54 per cent of all live-born babies are boys. This information, then, comes within the realm of statistics. Likewise, the statistics of students of a college include count of, the number of students, and separate counts of numbers of various kinds, as males and females, married and unmarried, of postgraduates and undergraduates. They may also include such measurements as their heights and weights. In addition, there may also be numbers computed on the basis of these measurements or counts, e.g., the proportion of female students; their average height or average weight. An example of statistical data is given in Table 4.1.

NOTES

Table 4.1 *Statistics of Students of a College where Total Number of Students is 1,000*

Sex-wise distribution			***Class-wise distribution***		
Sex	Number	Percentage	Course	Number	Percentage
Male	900	90	Undergraduate	800	80
Female	100	10	Postgraduate	200	20
Total	1,000	100	Total	1,000	100

Distribution According to Height

Height	***Number***
From 160 cm to 170	600
From 170 cm to 180	390
From 180 cm to 190	10
Total	1,000

The other aspect of statistics is as a body of theories and techniques employed in analysing the *numerical information* and using it to make informed decisions. It is a branch of the scientific method used in dealing with those phenomena which can be described numerically, either by measurements or by counts. For example, if a preliminary test using a new vaccine shows that in a sample of 100 cases the incidence of the disease is reduced to 8 while that in the unvaccinated population it is 10, can we say if the vaccine is effective or not? Surely, we can select 100 individuals in the unvaccinated population in which the incidence is 8 or even less. Is the observed reduction due to chance causes or does it show the effect of the vaccine? The statistical method provides with theories and techniques for checking this out. In this text, we will be primarily concerned with the statistical method, its theories and its techniques.

Consider the following situations typical of what are faced by decision-makers. In each of these the method of arriving at policy decision consists of first understanding the parameters of the problem, for which statistical data is called for.

- A large technical university experiences a fall in the number of persons seeking admission. Is this fall due to factors which are peculiar to that particular university, or is it a country-wide trend? The exact course of policy action will depend upon the answer to this question.
- At what age should one retire airline pilots? Does the increase in age affect the safety record ? How is it balanced by the increasing experience, if at all?
- In certain circles it is taken as axiomatic that more the people travel in the country, the more emotional integration is achieved. Recently there has been some doubt cast on this premise. The implications of this are wide ranging, affecting the government attitude towards promotion of tourism. How does one verify the truth of the contention?

- More than 50 per cent of heart-transplant patient die within a year. Is heart-transplant surgery beneficial? Would not a heart-patient be better off without a transplant than with it? Would he live longer?
- Is advertising on television more cost effective than advertising in print? Are street-corner hoardings effective at all?

Anybody would see immediately that we need data to answer any of these questions. But mere data would not help. The data will have to be systematically collected and analysed so that our answers are not affected by other factors. For example, how does one separate the effect of age and experience when one knows that experience increases with age? Also in each of these there are a whole lot of chance factors interacting with the outcome. Therefore, we have to isolate the effect that we want to study, and for this specialized statistical methods are called for.

The statistical method, when used properly, helps in understanding phenomena using numerical evidence. As a further example, suppose we want to understand the factors that affect the yield of farms. We may note that various factors such as rainfall, soil fertility, quality of seed, soil nutrients used, method of cultivation, etc. are all more or less important. One can never for sure predict the influence of one parameter, because we cannot control all of these independently. But it is possible to design experiments and collect data so that one is able to, more or less, isolate each effect to a predetermined level of certainty. The procedures for doing so are provided by statistical method.

As a further example, let us suppose we are interested in studying the level of income of the people living in a certain village.

For this purpose the following procedure may be adopted with advantages:

(i) *Collect data:* Information should be collected regarding
   - the number of persons living in the village
   - the number of persons who are getting income
   - the daily income of each earning member

(ii) Organize the data obtained above so as to show the number of persons within different income groups, and in that way reduce bulk.

(iii) Present this information by means of diagrams or other visual aids.

(iv) Analyse the data to determine the 'Average' income of the people and the *extent* of disparities that exist.

(v) On the basis of the above it would be possible to have an understanding of the phenomenon (income of people), and one would know (a) The average income of people', and (b) The extent of disparity in the level of incomes.

(vi) All this may lead to a policy decision for improvement of the existing situation.

4.2.1 Scope of Statistics

The proper function of statistics is to enlarge our knowledge of complex phenomena, and to lend precision to our ideas that would otherwise remain vague and indeterminate. Our knowledge of such things as 'National Income,' 'Population,'

NOTES

'National Resources', etc., would not have been so definite and precise, if there were no reliable statistics pertaining to each one of these. To say that the per capita income in India is low is a vague statement. The term 'Low' may mean one thing to one individual while to another it might mean something altogether different. One may take it to be near about ₹ 100 while someone else may think it to be in the neighborhood of ₹ 5,000. But the moment we say that our per capita income is ₹ 750 we make a statement which is precise and convincing. Again a statement, viz., the per capita income in agricultural sector is lower than in the industrial sector, is vague and indefinite. But if the per capita incomes for both these sectors are ascertained; the comparison would be easier and even a layman would be able to appreciate the difference in the productivity of these two sectors. It can thus be said that 'Statistics increases the field of mental vision, as an opera glass or telescope increases the field of physical vision. Statistics is able to widen our knowledge because of the following services that it renders.

It presents facts in a definite form. It is the quality of definiteness which is responsible for the growing universal application of statistical methods. The conclusions stated numerically are definite and hence more convincing than conclusions stated qualitatively. This fact can be readily understood by a simple example. In an advertisement, statements expressed numerically have greater attraction and are more appealing than those expressed in a qualitative manner. The caption, 'We have sold *more* cars this year' is certainly less attractive than 'Record sale of 10,000 cars in 1985 as compared to 6,000 in 1984'. The latter statement emphasizes in a much better manner the growing popularity of the advertiser's cars.

Statistics simplifies unwieldy and complex mass of data and presents them in such a manner that they at once become intelligible: The complex data may be reduced to totals; averages, percentages, etc., and presented either graphically or diagrammatically. These devices help us to understand quickly the significant characteristics of the numerical data, and consequently save us from a lot of mental strain. Single figures in the form of averages and percentages can be grasped more easily than a mass of statistical data comprising thousands of facts. Similarly, diagrams and graphs, because of their greater appeal to the eye and imagination tender valuable assistance in the proper understanding of numerical data. Time and energy of business executives are thus economized, if the statistician supplies them with the results of production, sales and finances in a condensed form.

Statistics classifies numerical facts: The procedure of classification brings into relief the salient features of the variable that is under investigation. This can be clearly illustrated by an example. If we are given the marks in mathematics of each individual student of a class and if it is desired to judge the performance of the class on **the** basis of these data it will not be an easy matter. Human mind has its limitations and cannot easily grasp a multitude of figures. But if the students are classified i.*e.,* if we put into one group all those boys who get more than second division marks, in still another group those who get third division marks, and have a separate group of those who fait to get pass marks, it will be easier for us to form **a** more precise idea about the performance of the class.

Statistics furnishes a technique of comparison: The facts, having been classified, are now in a shape when they can be used for purposes of comparisons and contrasts. Certain facts, by themselves, may be meaningless unless they are capable of being compared with similar facts at other places or at other periods in time. We estimate the national income of India not essentially for the value of that fact itself, but' mainly in order that we may compare the income of today with that of the past and thus draw conclusions as to whether the standard of living' of the people is on the increase, decrease or is stationary. Statistics affords suitable technique for comparison. It is with the help of statistics that the cost accountant is able to compare the actual accomplishment (in terms of cost) with programmes laid out (in terms of standard cost). Some of the modes of comparison provided by statistics include totals, ratios, averages or measure of central tendencies, graphs and diagrams, and coefficients. Statistics thus 'Serves as a scale in which facts in various combinations are weighed and valued,

Statistics endeavours *to* interpret conditions: Like an artist statistics renders useful service in presenting an attractive picture of the phenomenon under investigation- But it frequently does far more than this by enabling the interpretation of condition, by developing possible causes for the results described. If the production manager discovers that a certain machine is turning out some articles which are not up to the standard specifications, he will be able to find statistically if this condition is due to some defect in the machine or whether such a condition is normal.

Statistical Method

Statistical approach to a problem may broadly be summarized as: (i) Collection of facts: (ii) Organization of facts; (iii) Analysis of facts; and (iv) Interpretation of facts.

A detailed discussion of the various methods of collection, presentation, analysis and interpretation of facts is given later in the unit. Here the intention is to give only a bird's eye-view of the entire statistical procedure,

(i) Collection of facts is the first step in the statistical treatment of a problem. Numerical facts are the raw materials upon which the statistician is to work and just as in a manufacturing concern the quality of a finished product depends, *inter alia,* upon the quality of the raw material, in the same manner, the validity of statistical conclusions will be governed, among other considerations, by the quality of data used. Assembling of the facts is thus a very important process and no pains should be spared to see that the data collected are accurate, reliable and thorough. One thing that should be noted here is that the work of collecting facts should be undertaken in a planned manner. Without proper planning the facts collected may not be suitable for the purpose and a lot of time and money may be wasted.

(ii) The data so collected will more often than not be a huge mass of facts running into hundreds and thousands of figures. Human mind has its limitations. No One can appreciate at a glance or even after a careful study hold in mind the information contained in a hundred or a thousand schedules. For a proper understanding of the data their irregularities must be brushed

NOTES

off and their bulk be reduced, *i.e.,* some process of condensation must take place. Condensation implies the organization, classification, tabulation and presentation of the data in a suitable form.

(iii) The process of statistical analysis is a method of abstracting significant facts from the collected mass of numerical data. This process includes such things as 'measures of central tendency'—the determination, of Mean, Median and Mode—'measures of dispersion' and the determination of trends and tendencies, etc. This is more or less a mechanical process involving the use of elementary mathematics.

(iv) The interpretation of the various statistical constants obtained through a process of statistical analysis is the final phase or the finishing process of the statistical technique. It involves those methods by which judgments are formed and inferences obtained. To make estimates of the population parameters on the basis of sample statistics in an example of the problem of interpretation. For the interpretation of results a knowledge of advanced mathematics is essential.

Characteristics of Statistical Data

Even a casual look at Table 4.2 would lead us to the conclusion that statistical data always denotes 'Figures', *i.e.,* numerical descriptions. Whereas this is true, it must be remembered that all numerical descriptions are not statistical data. In order that numerical descriptions may be called statistics they must possess the following characteristics:

(i) They must be in aggregates.

(ii) They must be affected to a marked extent by a multiplicity of causes.

(iii) They must be enumerated or estimated according to reason able standard of accuracy,

(iv) They must have been collected in a systematic manner for a predetermined purpose.

(v) They must be placed in relation to each other.

Let us explain these characteristics:

Statistics are aggregates of facts: This means that statistics are a 'number of facts.' A single fact,' even though numerically stated, cannot be called statistics. 'A single death, an accident, a sale, a shipment does not constitute statistics. Yet numbers of deaths, accidents; sales and shipments are statistics.' Observe carefully Table 4.2 containing information about the population of India. Column (a) states the population only for one year whereas Column (b) gives population figures for seven different years. The data given in Column (b) are statistics whereas the figure given in Column (a) is not so, for the simple reason that it is a single solitary figure.

Table 4.2 *Characteristics of Statistical Data*

Column (a)		Column (b)	
Year	*Population (in lakhs)*	*Year*	*Population (in lakhs)*
1951	3,569	1911	2,490
		1921	2,481
		1931	2,755
		1941	3,128*
		1951	3,569
		1961	4,390
		1971	5,470

*After deducting estimated amount of inflation of returns in West Bengal and Punjab (20 lakhs).

They must be affected to a marked extent by a multiplicity of causes: The term statistical data can be used only when we cannot predict exactly the values of the various physical quantities. This means that the numerical value of any quantity at any particular moment is the result of the action and interaction of a number of forces, differing amongst themselves and it is not possible to say as to how much of it is due *to* any one particular cause. Thus, the volume of wheat production is attributable to a number of factors, viz., rainfall, soil, fertility, quality of seed, methods of cultivation, etc. All these factors acting jointly determine the amount of the yield and it is not possible for any one to assess the individual contribution of any one of these factors.

Statistics must be enumerated or estimated according to reasonable standards of accuracy: This means that if aggregates of numerical facts are to be called' statistics' they must be reasonably accurate. This is necessary because statistical data are to serve as a basis for statistical investigations. If the basis happens to be incorrect the results are bound to be misleading. It must, however, be clearly stated that it is not 'mathematical accuracy, but only reasonable accuracy' that is necessary in statistical work. What standard of accuracy is to be regarded as reasonable will depend upon the aims and objects of inquiry. Where precision is required' accuracy is necessary; where generel impressions are sufficient, appreciable errors may be tolerated. Again, whatever standard of accuracy is once adopted, it should be uniformly maintained throughout the inquiry.

Statistics are collected in a systematic manner for a predetermined purpose: Numerical data can be called statistics only if they have been compiled in a properly planned manner and for a purpose about which the enumerator had a definite idea. So long as the compiler is not clear about the object for which facts are to be collected, he will not be able to distinguish between facts that are relevant and those that are unnecessary; and as such the data collected will, in all probability, be a heterogeneous mass of unconnected facts. Again, the procedure of data collection must be properly planned, *i.e.,* it must be decided beforehand as to what kind of information is to be collected and the method that is to be applied in obtaining it. This involves decisions on matters like 'statistical unit,' 'standard of accuracy,' 'list of questions,' etc. Facts collected in an unsystematic manner, and without a complete awareness of the object, will be confusing and cannot be made the basis of valid conclusions.

NOTES

Statistics should be placed in relation to each other. Numerical facts may be placed in relation to each other either in point of time, space or condition. The phrase 'placed in relation to each other' suggests that the facts should be comparable. Facts are comparable in point of time when we have measurements of the same object, obtained in an identical manner, for different periods. They are said to be related in point of space or condition when we have the measurements of the same phenomenon at different places or in different conditions, but at the same time. Numerical facts will be comparable, if they pertain to the same inquiry and have been compiled in a systematic manner for a predetermined purpose.

Putting all these characteristics together, Secrets has defined statistics (numerical descriptions) as: 'Aggregates of facts, affected to a marked extent by multiplicity of causes, numerically expressed, enumerated or estimated, according to reasonable standard of accuracy, collected in a systematic manner, for a predetermined purpose, and placed in relation to each other.'

Some Other Definitions of Statistics

As numerical data

Waster has defined statistics as 'Classified facts respecting the condition of the people in a state especially those facts which can be stated in numbers or in tables or in any other tabular or classified arrangement.' No doubt, this definition was correct at a time when statistics were collected only for purposes of internal administration or for knowing, for purposes of war, the wealth of the State. The scope of statistics is now considerably wider and it has almost a universal application. Obviously, therefore, the definition is inadequate.

Bowley defines statistics as 'numerical statements of facts in any department of inquiry placed in relation to each other.' This is somewhat more accurate. It means that if numerical facts do not pertain to a department of inquiry or if such facts are not related to each other they cannot be called statistics. The leads us to the conclusion that 'all statistics are numerical facts but all numerical facts are not statistics.' This definition is certainly better than the previous one But it is not comprehensive enough in as much as it does not give any importance either to the nature of facts or the standard of accuracy.

As Statistical Methods

Bowley has called it 'The science of measurement of the social organism, regarded as a whole, in all its manifestations.' This definition is too narrow as it confines the scope of statistics only to human activities. Statistics in fact has a much wider application and is not confined only to the social organism. Besides, statistics is not only the technique of measuring but also of analysing and interpreting. Again, statistics, strictly speaking, is not a science but a scientific method. It is a device of inferring knowledge and not knowledge itself.

Bowley has also called statistics 'the science of counting,' and 'the science of average.' These definitions are again incomplete in the sense that they pertain to only a limited field. True, statistical work includes counting and averaging, but it also includes many other processes of treating quantitative data. In fact, while dealing with large numbers, actual count becomes illusory and only estimates are made. Thus these definitions can also be discarded on the ground of inadequacy.

NOTES

Origin of Statistics

Statistics originated from two quite dissimilar fields, viz., games of chance and political states. These two different fields are also termed as two distinct disciplines—one primarily analytical and the other essentially descriptive. The former is associated with the concept of chance and probability and the latter is concerned with the collection of data.

The theoretical development of the subject has its origin in the mid-17 century and many mathematicians and gamblers of France, Germany and England are credited for its development. Notable amongst them are Pascal (1623–1662), who investigated the properties of the coefficients of binomial expansion and James Bernoulli (1654–1705). who wrote the first treatise on the theory of probability.

As regards the descriptive side of statistics it may be stated that statistics is as old as statecraft. Since time immemorial men must have been compiling information about wealth and manpower for purpose of peace and war. This activity considerably expanded at each upsurge of social and political development and received added impetus in periods of war.

The development of statistics can be divided into the following three stages:

The empirical stage (1600): During this, the primitive stage of the subject, numerical facts were utilized by the rulers, principally as an aid in the administration of Government. Information was gathered about the number of people and the amount of property held by them—the former serving the ruler as an index of human fighting strength and the latter as an indication of actual and potential taxes.

The comparative stage (1600–1800): During this period statisticians frequently made comparisons between nations with a view to judging their relative strength and prosperity. In some countries enquiries were instituted to judge the economic and social conditions of their people. Colbert introduced in France a 'mercantile' theory of government whose basis was essentially statistical in character. In 1719, Frederick William I began gathering information about population occupation, house-taxes, city finance, etc., which helped to study the condition of the people.

*The modern st*age (1800 up to date): During this period statistics is viewed as a way of handling numerical facts rather than a mere device of collecting numerical data. Besides, there has been a considerable extension of the field of its applicability. It has now become a useful tool and statistical methods of analysis are now being increasingly used in biology, psychology, education, economics and business.

4.3 VARIOUS MEASURES OF STATISTICAL COMPUTATIONS

For quantitative data, the mean, median, mode, percentiles, range, variance, and standard deviation are the most widely used numerical measurements. The mean, also known as the average, is calculated by summing all of a variable's data values and dividing the total by the number of data values.

NOTES

4.3.1 Average

In statistics, the term central tendency specifies the method through which the quantitative data have a tendency to cluster approximately about some value. A measure of central tendency is any precise method of specifying this 'Central Value'. In the simplest form, the measure of central tendency is an average of a set of measurements, where the word *average* refers to as mean, median, mode or other measures of location. Typically the most commonly used measures are arithmetic mean, mode and median. These values are very useful not only in presenting the overall picture of the entire data but also for the purpose of making comparisons among two or more sets of data. As an example, questions like 'How hot is the month of June in Delhi?' can be answered, generally by a single figure of the average for that month. Similarly, suppose we want to find out if boys and girls at age 10 years differ in height for the purpose of making comparisons. Then, by taking the average height of boys of that age and average height of girls of the same age, we can compare and record the differences.

While arithmetic mean is the most commonly used measure of central location, mode and median are more suitable measures under certain set of conditions and for certain types of data. However, each measure of central tendency should meet the following requisites.

1. It should be easy to calculate and understand.
2. It should be rigidly defined. It should have only one interpretation so that the personal prejudice or bias of the investigator does not affect its usefulness.
3. It should be representative of the data. If it is calculated from a sample, then the sample should be random enough to be accurately representing the population.
4. It should have sampling stability. It should not be affected by sampling fluctuations. This means that if we pick 10 different groups of college students at random and compute the average of each group, then we should expect to get approximately the same value from each of these groups.
5. It should not be affected much by extreme values. If few very small or very large items are present in the data, they will unduly influence the value of the average by shifting it to one side or other, so that the average would not be really typical of the entire series. Hence, the average chosen should be such that it is not unduly affected by such extreme values.

All these measures of central tendency are discussed in this section.

4.3.2 Mean

There are several commonly used measures, such as arithmetic mean, mode and median. These values are very useful not only in presenting the overall picture of the entire data, but also for the purpose of making comparisons among two or more sets of data.

As an example, questions like 'How hot is the month of June in Delhi?' can be answered generally by a single figure of the average for that month. Similarly, suppose we want to find out if boys and girls of age 10 years differ in height for the purpose of making comparisons. Then, by taking the average height of boys of that age and the average height of girls of the same age, we can compare and record the differences.

While arithmetic mean is the most commonly used measure of central tendency, mode and median are more suitable measures under certain set of conditions and for certain types of data. However, each measure of central tendency should meet the following requisites:

(*i*) It should be easy to calculate and understand.

(*ii*) It should be rigidly defined. It should have only one interpretation so that the personal prejudice or the bias of the investigator does not affect its usefulness.

(*iii*) It should be representative of the data. If it is calculated from a sample, the sample should be random enough to be accurately representing the population.

(*iv*) It should have a sampling stability. It should not be affected by sampling fluctuations. This means that if we pick ten different groups of college students at random and compute the average of each group, then we should expect to get approximately the same value from each of these groups.

(*v*) It should not be affected much by extreme values. If few, very small or very large items are present in the data, they will unduly influence the value of the average by shifting it to one side or other, so that the average would not be really typical of the entire series. Hence, the average chosen should be such that it is not unduly affected by such extreme values.

Arithmetic mean is also commonly known as the mean. Even though average, in general, means measure of central tendency, when we use the word average in our daily routine, we always mean the arithmetic average. The term is widely used by almost everyone in daily communication. We speak of an individual being an average student or of average intelligence. We always talk about average family size or average family income or Grade Point Average (GPA) for students, and so on.

For discussion purposes, let us assume a variable X which stands for some value such as the ages of students. Let the ages of 5 students be 19, 20, 22, 22 and 17 years. Then variable X would represent these ages as,

X: 19, 20, 22, 22, 17

Placing the Greek symbol Σ(Sigma) before X would indicate a command that all values of X are to be added together. Thus,

$\Sigma X = 19 + 20 + 22 + 22 + 17$

The mean is computed by adding all the data values and dividing it by the number of such values. The symbol used for sample average is $\overline{X}$, so that,

$$\overline{X} = \frac{19 + 20 + 22 + 22 + 17}{5}$$

In general, if there are n values in the sample, then,

$$\overline{X} = \frac{X_1 + X_2 + ... + X_n}{n}$$

In other words,

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}, \qquad i = 1,\ 2\ ...\ n$$

According to this formula, mean can be obtained by adding all values of X_i, where the value of i starts at 1 and ends at n with unit increments so that $i = 1, 2, 3, ...\ n$.

If instead of taking a sample, we take the entire population in our calculations of the mean, then the symbol for the mean of the population is μ(mu) and the size of the population is N, so that,

$$\mu = \frac{\sum_{i=1}^{N} X_i}{N}, \qquad i = 1,\ 2\ ...N$$

If we have the data in grouped discrete form with frequencies, then the sample mean is given by,

$$\overline{X} = \frac{\Sigma f(X)}{\Sigma f}$$

Here, Σf = Summation of all frequencies
= n
$\Sigma f(X)$ = Summation of each value of X multiplied by its corresponding frequency (f)

Example 4.1: Let us take the ages of 10 students as follows:

19, 20, 22, 22, 17, 22, 20, 23, 17, 18

Solution: This data can be arranged in a frequency distribution as follows:

(X)	(f)	$f(X)$
17	2	34
18	1	18
19	1	19
20	2	40
22	3	66
23	1	23
	Total = 10	200

In this case, we have $\Sigma f = 10$ and $\Sigma f(X) = 200$, so that,

$$\overline{X} = \frac{\Sigma f(X)}{\Sigma f}$$

$$= 200/10 = 20$$

Characteristics of the Mean

The arithmetic mean has some interesting properties. These are as follows:

(*i*) The sum of the deviations of individual values of X from the mean will always add up to zero. This means that if we subtract all the individual values from their mean, then some values will be negative and some will be positive, but if all these differences are added together then the sum will be zero. In other words, the positive deviations must balance the negative deviations. Or symbolically,

$$\sum_{i=1}^{n}(X_i - \overline{X}) = 0, i = 1, 2, ... n$$

(*ii*) The second important characteristic of mean is that it is very sensitive to extreme values. Since the computation of mean is based upon inclusion of all values in the data, an extreme value in the data would shift the mean towards it, thus making the mean unrepresentative of the data.

(*iii*) The third property of mean is that the sum of squares of the deviations about the mean is minimum. This means that if we take the differences between individual values and the mean and square these differences individually and then add these squared differences, then the final figure will be less than the sum of the squared deviations around any other number other than the mean. Symbolically, it means that

$$\sum_{i=1}^{n}(X_i - \overline{X})^2 = \text{Minimum}, i = 1, 2, ... n$$

(*iv*) The product of the arithmetic mean and the number of values on which the mean is based is equal to the sum of all given values. In other words, if we replace each item in series by the mean, then the sum of these substitutions will equal the sum of individual items. Thus, if we take random figures as an example like 3, 5, 7, 9, and if we substitute the mean for each item 6, 6, 6, 6 then the total is 24, both in the original series and in the substitution series.

This can be shown like,

Since, $\overline{X} = \frac{\Sigma X}{N}$

$\therefore$ $N\overline{X} = \Sigma X$

For example, if we have a series of values 3, 5, 7, 9, the mean is 6. The squared deviations will be:

X	$X - \overline{X} = X'$	X'^2
3	$3 - 6 = -3$	9
5	$5 - 6 = -1$	1
7	$7 - 6 = 1$	1
9	$9 - 6 = 3$	9
		$\Sigma X'^2 = 20$

This property provides a test to check if the computed value is the correct arithmetic mean.

NOTES

Example 4.2: The mean age of a group of 100 persons (grouped in intervals 10–, 12–,..., etc.) was found to be 32.02. Later, it was discovered that age 57 was misread as 27. Find the corrected mean.

Solution: Let the mean be denoted by X. So, putting the given values in the formula of arithmetic mean, we have,

$$32.02 = \frac{\sum X}{100}, \text{ i.e., } \sum X = 3202$$

$$\text{Correct } \sum X = 3202 - 27 + 57 = 3232$$

$$\therefore \quad \text{Correct AM} = \frac{3232}{100} = 32.32$$

Example 4.3: The mean monthly salary paid to all employees in a company is ₹ 500. The monthly salaries paid to male and female employees average ₹ 520 and ₹ 420, respectively. Determine the percentage of males and females employed by the company.

Solution: Let N_1 be the number of males and N_2 be the number of females employed by the company. Also, let x_1 and x_2 be the monthly average salaries paid to male and female employees and $\bar{x}$ be the mean monthly salary paid to all the employees.

$$\bar{x} = \frac{N_1 x_1 + N_2 x_2}{N_1 + N_2}$$

$$\text{or} \quad 500 = \frac{520 N_1 + 420 N_2}{N_1 + N_2} \quad \text{or} \quad 20N_1 = 80N_2$$

$$\text{or} \quad \frac{N_1}{N_2} = \frac{80}{20} = \frac{4}{1}$$

Hence, the males and females are in the ratio of 4 : 1 or 80 per cent are males and 20 per cent are females in those employed by the company.

Short-Cut Methods for Calculating Mean

We can simplify the calculations of mean by noticing that if we subtract a constant amount A from each item X to define a new variable $X' = X - A$, the mean $\bar{X}'$ of X' differs from $\bar{X}$ by A. This generally simplifies the calculations and we can then add back the constant A, termed as the assumed mean such as,

$$\bar{X} = A + \bar{X}' = A + \frac{\Sigma f(X')}{\Sigma f}$$

Table 4.3 illustrates the procedure of calculation by short-cut method using the data given in Example 4.1.

The choice of A is made in such a manner as to simplify calculation the most, and is generally in the region of concentration of data.

Table 4.3 *Short-Cut Method of Calculating Mean*

X	*(f)*	*Deviation from Assumed Mean (13) X′*	*f(X′)*
9	1	–4	–4
10	2	–3	–6
11	3	–2	–6
12	6	–1	–6
			–22
13	10	0	
14	11	+1	+11
15	7	+2	+14
16	3	+3	+9
17	2	+4	+8
18	1	+5	+ 5
			+47
			–22
	Σf = 46		ΣfX′ = 25

The mean,

$$\bar{X} = A + \frac{\sum f(X')}{\sum f} = 13 + \frac{25}{46} = 4.54$$

This mean is the same as calculated in Example 4.1.

In the case of grouped frequency data, the variable *X* is replaced by midvalue *m*, and in the short-cut technique, we subtract a constant value *A* from each *m*, so that the formula becomes:

$$\bar{X} = A + \frac{\sum f(m - A)}{\sum f}$$

In cases where the class intervals are equal, we may further simplify calculation by taking the factor *i* from the variable *m*–A defining,

$$X' = \frac{m - A}{i}$$

where *i* is the class width. It can be verified that when *X′* is defined, then the mean of the distribution is given by

$$\bar{X} = A + \frac{\sum f(X')}{\sum f} \times i$$

Example 4.2 will illustrate the use of short-cut method.

Example 4.4: The ages of twenty husbands and wives are given in the following table. Form frequency tables showing the relationship between the ages of husbands and wives with class intervals 20 – 24; 25 – 29; etc.

Calculate the arithmetic mean of the two groups after the classification.

S.No.	Age of Husband	Age of Wife
1	28	23
2	37	30
3	42	40
4	25	26
5	29	25
6	47	41
7	37	35
8	35	25
9	23	21
10	41	38
11	27	24
12	39	34
13	23	20
14	33	31
15	36	29
16	32	35
17	22	23
18	29	27
19	38	34
20	48	47

Solution:

Calculation of Arithmetic Mean of Husbands' Age

Class Intervals	Midvalues m	Husband Frequency (f_1)	$x_1' = \frac{m-37}{5}$	f_1x_1'
20–24	22	3	–3	–9
25–29	27	5	–2	–10
30–34	32	2	–1	–2
				– 21
35–39	37	6	0	0
40–44	42	2	1	2
45–49	47	2	2	4
				6
		$\Sigma f_1 = 20$		$\Sigma f_1 x_1' = -15$

Arithmetic mean of husbands age,

$$\bar{x} = \frac{\Sigma f_1 x_1'}{N} \times i + A = \frac{-15}{20} \times 5 + 37 = 33.25$$

Calculation of Arithmetic Mean of Wives' Age

Class Intervals	Midvalues m	Wife Frequency (f_2)	$x_2' = \frac{m-37}{5}$	f_2x_2'
20–24	22	5	–3	–15
25–29	27	5	–2	–10
30–34	32	4	–1	–4
35–39	37	3	0	0
40–44	42	2	1	2
45–49	47	1	2	2
		$\Sigma f_2 = 20$		$\Sigma f_2 x_2' = -25$

Arithmetic mean of wife's age,

$$\bar{x} = \frac{\sum f_2 x_2'}{N} \times i + A = \frac{-25}{20} \times 5 + 37 = 30.75$$

The Weighted Arithmetic Mean

In the computation of arithmetic mean we had given equal importance to each observation in the series. This equal importance may be misleading if the individual values constituting the series have different importance as in Example 4.5.

Example 4.5: The Raja Toy shop sells

Toy Cars at	₹ 3 each
Toy Locomotives at	₹ 5 each
Toy Aeroplanes at	₹ 7 each
Toy Double Decker at	₹ 9 each

What will be the average price of the toys sold, if the shop sells 4 toys, one of each kind?

Solution:

Mean Price, i.e., $\bar{x} = \frac{\sum x}{4} = ₹\,\frac{24}{4} = ₹\,6$

In this case, the importance of each observation (price quotation) is equal in as much as one toy of each variety has been sold. In the computation of the arithmetic mean, this fact has been taken care of by including 'once only' the price of each toy.

If , however the shop sells 100 toys: 50 cars, 25 locomotives, 15 aeroplanes and 10 double deckers, the importance of the four price quotations to the dealer is **not equal** as a source of earning revenue. In fact, their respective importance is equal to the number of units of each toy sold, i.e.,

The importance of Toy Car	50
The importance of Locomotive	25
The importance of Aeroplane	15
The importance of Double Decker	10

It may be noted that 50, 25, 15, 10 are the quantities of the various classes of toys sold. It is for these quantities that the term 'weights' is used in statistical language. Weight is represented by symbol 'w', and Σw represents the sum of weights.

While determining the 'Average price of toy sold', these weights are of great importance and are taken into account in the manner as shown,

$$\bar{x} = \frac{w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4}{w_1 + w_2 + w_3 + w_4} = \frac{\sum wx}{\sum w}$$

When w_1, w_2, w_3, w_4 are the respective weights of x_1, x_2, x_3, x_4 which in turn represent the price of four varieties of toys, viz., car, locomotive, aeroplane and double decker, respectively.

$$\bar{x} = \frac{(50 \times 3) + (25 \times 5) + (15 \times 7) + (10 \times 9)}{50 + 25 + 15 + 10}$$

$$= \frac{(150) + (125) + (105) + (90)}{100} = \frac{470}{100} = ₹\,4.70$$

NOTES

The following table summarizes the steps taken in the computation of the weighted arithmetic mean.

Weighted Arithmetic Mean of Toys Sold by Raja Toy Shop

Toys	*Price per Toy* ₹*x*	*Number Sold* *w*	*Price × Weight* *xw*
Car	3	50	150
Locomotive	5	25	125
Aeroplane	7	15	105
Double Decker	9	10	90
		$\Sigma w = 100$	$\Sigma xw = 470$

$$\Sigma w = 100; \quad \Sigma wx = 470$$

$$\bar{x} = \frac{\sum wx}{\sum w} = \frac{470}{100} = 4.70$$

The weighted arithmetic mean is particularly useful where we have to compute the mean of means. If we are given two arithmetic means, one for each of two different series, in respect of the same variable, and are required to find the arithmetic mean of the combined series, the weighted arithmetic mean is the only suitable method of its determination (Refer Example 4.6).

Example 4.6: The arithmetic mean of daily wages of two manufacturing concerns A Ltd. and B Ltd. is ₹ 5 and ₹ 7, respectively. Determine the average daily wages of both concerns if the number of workers employed were 2,000 and 4,000, respectively.

Solution:

(*i*) Multiply each average (viz., 5 and 7), by the number of workers in the concern it represents.

(*ii*) Add up the two products obtained in (i).

(*iii*) Divide the total obtained in (ii) by the total number of workers.

Weighted Mean of Mean Wages of A Ltd. and B Ltd.

Manufacturing Concern	*Mean Wages* *x*	*Workers Employed* *w*	*Mean Wages × Workers Employed* *wx*
A Ltd.	5	2,000	10,000
B Ltd.	7	4,000	28,000
		$\sum w = 6{,}000$	$\sum wx = 38{,}000$

$$\bar{x} = \frac{\sum wx}{\sum w} = \frac{38{,}000}{6{,}000} = ₹\,6.33$$

These examples explain that 'Arithmetic Means and Percentage' are not original data. They are derived figures and their importance is relative to the original data from which they are obtained. This relative importance must be taken into account by weighting while averaging them (means and percentage).

Advantages of Mean

(*i*) Its concept is familiar to most people and is intuitively clear.

(*ii*) Every data set has a mean, which is unique and describes the entire data to some degree. For example, when we say that the average salary of a professor is ₹ 25,000 per month, it gives us a reasonable idea about the salaries of professors.

(*iii*) It is a measure that can be easily calculated.

(*iv*) It includes all values of the data set in its calculation.

(*v*) Its value varies very little from sample to sample taken from the same population.

(*vi*) It is useful for performing statistical procedures, such as computing and comparing the means of several data sets.

Disadvantages of Mean

(*i*) It is affected by extreme values, and hence are not very reliable when the data set has extreme values especially when these extreme values are on one side of the ordered data. Thus, a mean of such data is not truly a representative of such data. For example, the average age of three persons of ages 4, 6 and 80 years gives us an average of 30.

(*ii*) It is tedious to compute for a large data set as every point in the data set is to be used in computations.

(*iii*) We are unable to compute the mean for a data set that has open-ended classes either at the high or at the low-end of the scale.

(*iv*) The mean cannot be calculated for qualitative characteristics, such as beauty or intelligence, unless these can be converted into quantitative figures such as intelligence into IQs.

4.3.3 Median

The second measure of central tendency that has a wide usage in statistical works is the median. Median is that value of a variable which divides the series in such a manner that the number of items below it is equal to the number of items above it. Half the total number of observations lie below the median, and half above it. The median is thus a positional average.

The median of ungrouped data is found easily if the items are first arranged in order of the magnitude. The median may then be located simply by counting, and its value can be obtained by reading the value of the middle observations. If we have five observations whose values are 8, 10, 1, 3 and 5, the values are first arrayed: 1, 3, 5, 8 and 10. It is now apparent that the value of the median is 5, since two observations are below that value and two observations are above it.

When there is an even number of cases, there is no actual middle item and the median is taken to be the average of the values of the items lying on either side of (N+ 1)/2, where N is the total number of items. Thus, if the values of six items of a series are 1, 2, 3, 5, 8 and 10, then the median is the value of item number (6 + 1)/2 = 3.5, which is approximated as the average of the third and the fourth items, i.e., (3+5)/2 = 4.

Thus, the steps required for obtaining median are as follows:

(*i*) Arrange the data as an array of increasing magnitude.

(*ii*) Obtain the value of the (N+ l)/2th item.

Even in the case of grouped data, the procedure for obtaining median is straightforward as long as the variable is discrete or non-continuous as is clear from Example 4.7.

Example 4.7: Obtain the median size of shoes sold from the following data:

Number of Shoes Sold by Size in One Year

Size	*Number of Pairs*	*Cumulative Total*
5	30	30
$5\frac{1}{2}$	40	70
6	50	120
$6\frac{1}{2}$	150	270
7	300	570
$7\frac{1}{2}$	600	1170
8	950	2120
$8\frac{1}{2}$	820	2940
9	750	3690
$9\frac{1}{2}$	440	4130
10	250	4380
$10\frac{1}{2}$	150	4530
11	40	4570
$11\frac{1}{2}$	39	4609
	Total	4609

Solution: Median is the value of $\frac{(N+1)}{2}$th $= \frac{4609+1}{2}$th $=$ 2305th item. Since the items are already arranged in ascending order (size-wise), the size of 2305th item is easily determined by constructing the cumulative frequency. Thus, the median size of shoes sold is 8½, the size of 2305th item.

In the case of grouped data with continuous variable, the determination of median is a bit more involved. Consider the following table where the data relating to the distribution of male workers by average monthly earnings is given. Clearly the median of 6291 is the earnings of (6291 + l)/2 = 3l46th worker arranged in ascending order of earnings.

From the cumulative frequency, it is clear that this worker has his income in the class interval 67.5 – 72.5. However, it is impossible to determine his exact income. We therefore, resort to approximation by assuming that the 795 workers of this class are distributed uniformly across the interval 67.5 – 72.5. The median worker is (3146–2713) = 433rd of these 795, and hence, the value corresponding to him can be approximated as,

$$67.5 + \frac{433}{795} \times (72.5 - 67.5) = 67.5 + 2.73 = 70.23$$

Distribution of Male Workers by Average Monthly Earnings

Group No.	Monthly Earnings (₹)	No. of Workers	Cumulative No. of Workers
1	27.5–32.5	120	120
2	32.5–37.5	152	272
3	37.5–42.5	170	442
4	42.5–47.5	214	656
5	47.5–52.5	410	1066
6	52.5–57.5	429	1495
7	57.5–62.5	568	2063
8	62.5–67.5	650	2713
9	67.5–72.5	795	3508
10	72.5–77.5	915	4423
11	77.5–82.5	745	5168
12	82.5–87.5	530	5698
13	87.5–92.5	259	5957
14	92.5–97.5	152	6109
15	97.5–102.5	107	6216
16	102.5–107.5	50	6266
17	107.5–112.5	25	6291
		Total	6291

The value of the median can thus be put in the form of the formula,

$$Me = l + \frac{\frac{N+1}{2} - C}{f} \times i$$

Where l is the lower limit of the median class, i its width, f its frequency, C the cumulative frequency upto (but not including) the median class, and N is the total number of cases.

Finding Median by Graphical Analysis

The median can quite conveniently be determined by reference to the ogive which plots the cumulative frequency against the variable. The value of the item below which half the items lie, can easily be read from the ogive as is shown in Example 4.8.

NOTES

Example 4.8: Obtain the median of data given in the following table:

Monthly Earnings	*Frequency*	*Less Than*	*More Than*
27.5	—	0	6291
32.5	120	120	6171
37.5	152	272	6019
42.5	170	442	5849
47.5	214	656	5635
52.5	410	1066	5225
57.5	429	1495	4796
62.5	568	2063	4228
67.5	650	2713	3578
72.5	795	3508	2783
77.5	915	4423	1868
82.5	745	5168	1123
87.5	530	5698	593
92.5	259	5957	334
97.5	152	6109	182
102.5	107	6216	75
107.5	50	6266	25
112.5	25	6291	0

Solution: It is clear that this is grouped data. The first class is 27.5 – 32.5, whose frequency is 120, and the last class is 107.5 – 112.5 whose frequency is 25. Figure 4.1 shows the ogive of less than cumulative frequency. The median is the value below which *N*/2 items lie, is 6291/2 = 3145.5 items lie, which is read of from Figure 4.2 as about 70. More accuracy than this is unobtainable because of the space limitation on the earning scale.

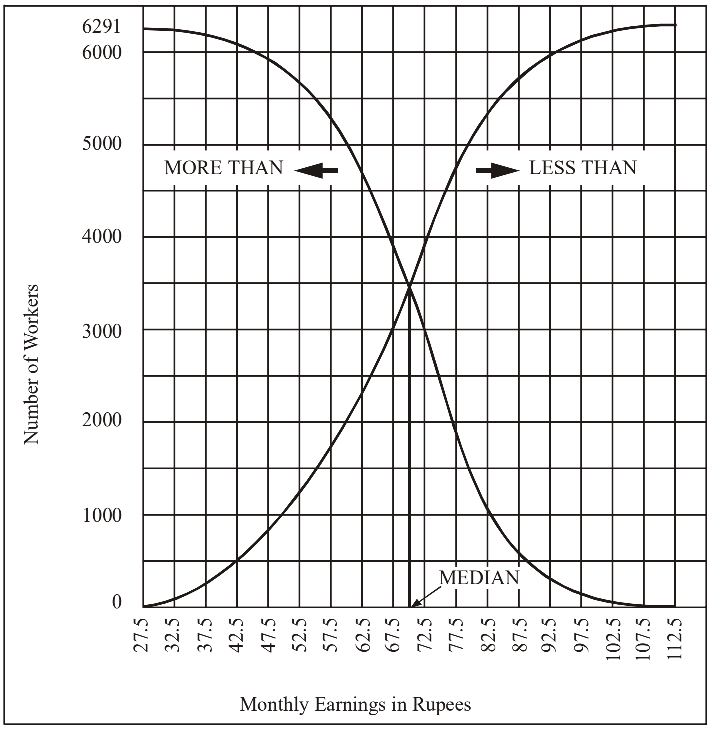


Fig. 4.1 *Median Determination by Plotting Less than and More than Cumulative Frequency*

NOTES

The median can also be determined by plotting both 'Less Than' and 'More than' cumulative frequency as shown in Figure 4.1. It should be obvious that the two curves should intersect at the median of the data.

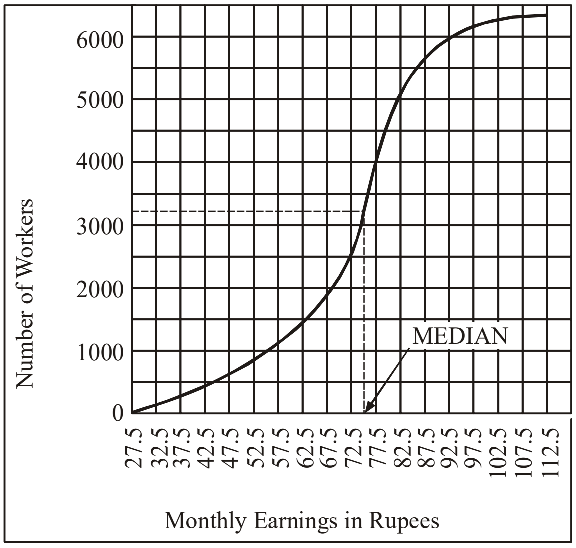


Fig. 4.2 *Median*

Advantages of Median

(*i*) Median is a positional average and hence the extreme values in the data set do not affect it as much as they do to the mean.

(*ii*) Median is easy to understand and can be calculated from any kind of data, even from grouped data with open-ended classes.

(*iii*) We can find the median even when our data set is qualitative and can be arranged in the ascending or the descending order, such as average beauty or average intelligence.

(*iv*) Similar to mean, median is also unique, meaning that, there is only one median in a given set of data.

(*v*) Median can be located visually when the data is in the form of ordered data.

(*vi*) The sum of absolute differences of all values in the data set from the median value is minimum. This means that, it is less than any other value of central tendency in the data set, which makes it more central in certain situations.

Disadvantages of Median

(*i*) The data must be arranged in order to find the median. This can be very time consuming for a large number of elements in the data set.

(*ii*) The value of the median is affected more by sampling variations. Different samples from the same population may give significantly different values of the median.

(*iii*) The calculation of median in case of grouped data is based on the assumption that the values of observations are evenly spaced over the entire class interval and this is usually not so.

(*iv*) Median is comparatively less stable than mean, particularly for small samples, due to fluctuations in sampling.

(*v*) Median is not suitable for further mathematical treatment. For example, we cannot compute the median of the combined group from the median values of different groups.

NOTES

4.2.4 Mode

Mode is that value of the variable which occurs or repeats itself the greatest number of times. The mode is the most 'Fashionable' size in the sense that it is the most common and typical, and is defined by Zizek as 'the value occurring most frequently in a series (or group of items) and around which the other items are distributed most densely'.

The mode of a distribution is the value at the point around which the items tend to be most heavily concentrated. It is the most frequent or the most common value, provided that a sufficiently large number of items are available, to give a smooth distribution. It will correspond to the value of the maximum point (ordinate), of a frequency distribution if it is an 'ideal' or smooth distribution. It may be regarded as the most typical of a series of values. The modal wage, for example, is the wage received by more individuals than any other wage. The modal 'hat' size is that, which is worn by more persons than any other single size.

It may be noted that the occurrence of one or a few extremely high or low values has no effect upon the mode. If a series of data are unclassified, not have been either arrayed or put into a frequency distribution, the mode cannot be readily located.

Taking first an extremely simple example, if seven men receive daily wages of ₹ 5, 6, 7, 7, 7, 8 and 10, it is clear that the modal wage is ₹ 7 per day. If we have a series such as 2, 3, 5, 6, 7, 10 and 11, it is apparent that there is no mode.

There are several methods of estimating the value of the mode. However, it is seldom that the different methods of ascertaining the mode give us identical results. Consequently, it becomes necessary to decide as to which method would be most suitable for the purpose in hand. In order that a choice of the method may be made, we should understand each of the methods and the differences that exist among them.

The four important methods of estimating mode of a series are: (i) Locating the most frequently repeated value in the array; (ii) Estimating the mode by interpolation; (iii) Locating the mode by graphic method; and (iv) Estimating the mode from the mean and the median. Only the last three methods are discussed in this unit.

Estimating the Mode by Interpolation: In the case of continuous frequency distributions, the problem of determining the value of the mode is not so simple as it might have appeared from the foregoing description. Having located the modal class of the data, the next problem in the case of continuous series is to interpolate the value of the mode within this 'modal' class.

The interpolation is made by the use of any one of the following formulae:

$$(i)\ Mo = l_1 + \frac{f_2}{f_0 + f_2} \times i; \quad (ii)\ Mo = l_2 - \frac{f_0}{f_0 + f_2} \times i$$

$$(iii)\ Mo = l_1 + \frac{f_1 - f_0}{(f_1 - f_0) + (f_1 - f_2)} \times i$$

Where l_1 is the lower limit of the modal class, l_2 is the upper limit of the modal class, f_0 equals the frequency of the preceding class in value, f_1 equals the frequency of the modal class in value, f_2 equals the frequency of the following class (class next to modal class) in value, and *i* equals the interval of the modal class. Example 4.9 explains the method of estimating mode.

Example 4.9: Determine the mode for the data given in the following table:

Wage Group	*Frequency (f)*
14 — 18	6
18 — 22	18
22 — 26	19
26 — 30	12
30 — 34	5
34 — 38	4
38 — 42	3
42 — 46	2
46 — 50	1
50 — 54	0
54 — 58	1

Solution:

In the given data, 22 – 26 is the modal class since it has the largest frequency. The lower limit of the modal class is 22, its upper limit is 26, its frequency is 19, the frequency of the preceding class is 18, and of the following class is 12. The class interval is 4. Using the various methods of determining mode, we have,

$$(i)\ Mo = 22 + \frac{12}{18 + 12} \times 4 \qquad (ii)\ Mo = 26 - \frac{18}{18 + 12} \times 4$$

$$= 22 + \frac{8}{5} \qquad\qquad = 26 - \frac{12}{5}$$

$$= 23.6 \qquad\qquad = 23.6$$

$$(iii)\ Mo = 22 + \frac{19 - 18}{(19 - 18) + (19 - 12)} \times 4 = 22 + \frac{4}{8} = 22.5$$

In formulae (i) and (ii), the frequency of the classes adjoining the modal class is used to pull the estimate of the mode away from the midpoint towards either the upper or lower class limit. In this particular case, the frequency of the class preceding the modal class is more than the frequency of the class following and therefore, the estimated mode is less than the midvalue of the modal class. This seems quite logical. If the frequencies are more on one side of the modal class than on the other it can be reasonably concluded that the items in the modal class are concentrated more towards the class limit of the adjoining class with the larger frequency.

NOTES

Formula (iii) is also based on a logic similar to that of formulae (i) and (ii). In this case, to interpolate the value of the mode within the modal class, the differences between the frequency of the modal class, and the respective frequencies of the classes adjoining it are used. This formula usually gives results better than the values obtained by the other and exactly equals results obtained by graphic method. The formulae (i) and (ii) give values which are different from the value obtained by formula (iii) and are more close to the central point of modal class. If the frequencies of the class adjoining the modal are equal, the mode is expected to be located at the midvalue of the modal class, but if the frequency on one of the sides is greater, the mode will be pulled away from the central point. It will be pulled more and more if the difference between the frequencies of the classes adjoining the modal class is higher and higher. In given example in this book, the frequency of the modal class is 19 and that of the preceding class is 18. So, the mode should be quite close to the lower limit of the modal class. The midpoint of the modal class is 24 and the lower limit of the modal class is 22.

Locating the Mode by the Graphic Method: The method of graphic interpolation is shown in Figure 4.3. The upper corners of the rectangle over the modal class have been joined by straight lines to those of the adjoining rectangles as shown in Figure 4.3; the right corner to the corresponding one of the adjoining rectangle on the left, etc. If a perpendicular is drawn from the point of intersection of these lines, we have a value for the mode indicated on the base line. The graphic approach is, in principle, similar to the arithmetic interpolation explained earlier.

The mode may also be determined graphically from an ogive or cumulative frequency curve. It is found by drawing a perpendicular to the base from that point on the curve where the curve is most nearly vertical, i.e., steepest (in other words, where it passes through the greatest distance vertically and smallest distance horizontally). The point where it cuts the base gives us the value of the mode. How accurately this method determines the mode is governed by: (i) The shape of the ogive, (ii) The scale on which the curve is drawn.

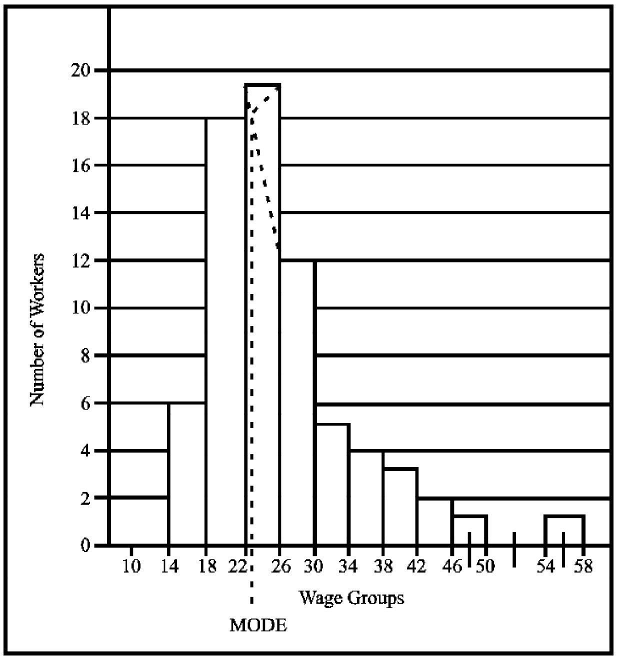


Fig. 4.3 Method of Mode Determination by Graphic Interpolation

Estimating the Mode from the Mean and the Median. There usually exists a relationship among the mean, median and mode for moderately asymmetrical distributions. If the distribution is symmetrical, the mean, median and mode will have identical values, but if the distribution is skewed (moderately) the mean, median and mode will pull apart. If the distribution tails off towards higher values, the mean and the median will be greater than the mode. If it tails off towards lower values, the mode will be greater than either of the other two measures. In either case, the median will be about one-third as far away from the mean as the mode is. This means that,

$$\begin{aligned}\text{Mode} &= \text{Mean} - 3\,(\text{Mean} - \text{Median})\\ &= 3\ \text{Median} - 2\ \text{Mean}\end{aligned}$$

Consider Example 4.10 to better understand the calculation of mode.

Example 4.10: Consider the mean to be 68.53 and the median to be 70.2 Calculate the mode using the formula discussed.

Solution:

In the case of the average monthly earnings, the mean is 68.53 and the median is 70.2. If these values are substituted in the above formula, we get,

$$\begin{aligned}\text{Mode} &= 68.5 - 3(68.5 - 70.2)\\ &= 68.5 + 5.1 = 73.6\end{aligned}$$

According to the formula used earlier,

$$\begin{aligned}\text{Mode} &= l_1 + \frac{f_2}{f_0 + f_2} \times i\\ &= 72.5 + \frac{745}{795 + 745} \times 5\\ &= 72.5 + 2.4 = 74.9\end{aligned}$$

or

$$\begin{aligned}\text{Mode} &= l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i\\ &= 72.5 + \frac{915 - 795}{2 \times 915 - 795 - 745} \times 5\\ &= 72.5 + \frac{120}{290} \times 5 = 74.57\end{aligned}$$

The difference between the two estimates is due to the fact that the assumption of relationship between the mean, median and mode may not always be true which is obviously not valid in this case.

Example 4.11: (*i*) In a moderately symmetrical distribution, the mode and mean are 32.1 and 35.4 respectively. Calculate the median.

(*ii*) If the mode and median of moderately asymmetrical series are respectively 16" and 15.7", what would be its most probable median?

(*iii*) In a moderately skewed distribution, the mean and the median are respectively 25.6 and 26.1 inches. What is the mode of the distribution?

NOTES

Solution:

(*i*) We know,

$$\text{Mean} - \text{Mode} = 3\,(\text{Mean} - \text{Median})$$

or $$3\text{ Median} = \text{Mode} + 2\text{ Mean}$$

or $$\text{Median} = \frac{32.1 + 2 \times 35.4}{3} = \frac{102.9}{3} = 34.3$$

(*ii*) $$2\text{ Mean} = 3\text{ Median} - \text{Mode}$$

or $$\text{Mean} = \frac{1}{2}(3 \times 15.7 - 16.0) = \frac{31.1}{2} = 15.55$$

(*iii*) $$\text{Mode} = 3\text{ Median} - 2\text{ Mean} = 3 \times 26.1 - 2 \times 25.6 = 78.3 - 51.2 = 27.1$$

Advantages of Mode

(*i*) Similar to median, the mode is not affected by extreme values in the data.

(*ii*) Its value can be obtained in open-ended distributions without ascertaining the class limits.

(*iii*) It can be easily used to describe qualitative phenomenon. For example, if most people prefer a certain brand of tea, then this will become the modal point.

(*iv*) Mode is easy to calculate and understand. In some cases, it can be located simply by observation or inspection.

Disadvantages of Mode

(*i*) Quite often, there is no modal value.

(*ii*) It can be bi-modal or multi-modal, or it can have all modal values making its significance more difficult to measure.

(*iii*) If there is more than one modal value, the data is difficult to interpret.

(*iv*) A mode is not suitable for algebraic manipulations.

(*v*) Since the mode is the value of maximum frequency in the data set, it cannot be rigidly defined if such frequency occurs at the beginning or at the end of the distribution.

(*vi*) It does not include all observations in the data set, and hence, less reliable in most of the situations.

4.3.5 Geometric Mean

If α, β, γ are in GP, then β is called a *geometric mean* between α and γ, written as GM.

If $a_1, a_2, ...,a_n$ are in GP, then $a_2, ..., a_{n-1}$ are called *geometric means* between a_1 and a_n.

Thus, 3, 9, 27 are three geometric means between 1 and 81.

Let $G_1, G_3, ... G_n$ be n geometric means between a and b. Thus, a, G_1, $G_2, ... G_n$, b is a GP, b being $(n + 2)$th term $= ar^{n+1}$, where r is the common ratio of GP

Thus, $$b = ar^{n+1} \Rightarrow r = \left(\frac{b}{a}\right)^{\frac{1}{n+1}}$$

So, $$G_1 = ar = a\left(\frac{b}{a}\right)^{\frac{1}{n+1}} = \left(a^n b\right)^{\frac{1}{n+1}}$$

$$G_2 = ar^2 = a\left(\frac{b}{a}\right)^{\frac{2}{n+1}} = \left(a^{n-1} b^2\right)^{\frac{1}{n+1}}$$

... .

$$G_n = ar^{n-1} = a\left(\frac{b}{a}\right)^{\frac{n-1}{n+1}} = \left(a^2 b^{n-1}\right)^{\frac{1}{n+1}}$$

Example 4.12: Find 7 GM's between 1 and 256.

Solution: Let $G_1, G_2, ... G_7$, be 7 GM's between 1 and 256

Then, 256= 9th term of GP,

= 1. r^8, where r is the common ratio of the GP

This gives that, $r^8 = 256 \Rightarrow r = 2$

Thus,
$$G_1 = ar = 1.2 = 2$$
$$G_2 = ar^2 = 1.4 = 4$$
$$G_3 = ar^3 = 1.8 = 8$$
$$G_4 = ar^4 = 1.16 = 16$$
$$G_5 = ar^5 = 1.32 = 32$$
$$G_6 = ar^6 = 1.64 = 64$$
$$G_7 = ar^7 = 1.128 = 128$$

Hence, required GM's are 2, 4, 8, 16, 32, 64, 128.

Example 4.13: Sum the series $1 + 3x + 5x^2 + 7x^3 + ...$ up to n terms, $x \neq 1$.

Solution: Note that nth term of this series $= (2n - 1)\, x^{n-1}$

Let $S_n = 1 + 3x + 5x^2 + ... + (2n - 1)\, x^{n-1}$

Then, $xS_n = x + 3x^2 + ... + (2n - 3)\, x^{n-1} + (2n - 1)\, x^n$

Subtracing, we get

$$S_n(1 - x) = 1 + 2x + 2x^2 + ... + 2x^{n-1} - (2n - 1)\, x^n$$

$$= 1 + 2x\left(\frac{1 - x^{n-1}}{1 - x}\right) - (2n - 1)\, x^n$$

$$= \frac{1 - x + 2\,x - 2x^n - (2n - 1)\, x^n (1 - x)}{1 - x}$$

$$= \frac{1 + x - 2\,x^n - (2n - 1)\, x^n + (2n - 1)\, x^{n+1}}{1 - x}$$

$$= \frac{1 + x - (2n+1)\, x^n + (2n-1)\, x^{n+1}}{1 - x}$$

NOTES

Hence, $$S = \frac{1 + x - (2n+1)\, x^n + (2n-1)\, x^{n+1}}{(1 - x)^2}$$

Example 4.14: Sum the series 5 + 55 + 555 + ... up to n terms.

Solution: Let $S_n = 5 + 55 + 555 + \ldots.$

$$S_n = 5\,(1 + 11 + 111 + \ldots.)$$

$$= \frac{5}{9}\,(9 + 99 + 999 + \ldots)$$

$$= \frac{5}{9}\,[(10 - 1) + (100 - 1) + (1000 - 1) + \ldots]$$

$$= \frac{5}{9}\,[(10 + 10^2 + 10^3 + \ldots + 10^n) - (1 + 1 + \ldots.n \text{ terms})]$$

$$= \frac{5}{9}\,[(10 + 10^2 + 10^3 + \ldots + 10^n) - n]$$

$$= \frac{5}{9}\left[\frac{10(1 - 10^n)}{1 - 10} - n\right]$$

$$= \frac{5}{9}\left[\frac{10(10^n - 1)}{9} - n\right]$$

$$= \frac{50}{81}(10^n - 1) - \frac{5n}{9}$$

Example 4.15: Three numbers are in GP. Their product is 64 and sum is $\frac{124}{5}$. Find them.

Solution: Let the numbers be $\frac{a}{r}, a, ar$

Since, $\frac{a}{r} + a + ar = \frac{124}{5}$ and $\frac{a}{r} \times a \times ar = 64$,

we have, $a^3 = 64 \Rightarrow a = 4$

This gives, $\frac{4}{r} + 4 + 4r = \frac{124}{5}$

$\Rightarrow \quad \frac{1}{r} + 1 + r = \frac{31}{5}$

$\Rightarrow \quad \frac{r^2 + 1}{r} = \frac{26}{5}$

$\Rightarrow \quad 5r^2 + 5 = 26r$

$\Rightarrow \quad 5r^2 - 26r + 5 = 0$

$\Rightarrow \quad 5r^2 - 25r - r + 5 = 0$

$\Rightarrow \quad 5r\,(r - 5) - 1\,(r - 5) = 0$

$\Rightarrow \quad (r - 5)\,(5r - 1) = 0$

$\Rightarrow \quad r = \frac{1}{5}$ or 5

In either case, numbers are $\frac{4}{5}$, 4 and 20.

NOTES

Example 4.16: Sum to n terms the series

$$0.7 + 0.77 + 0.777 + \ldots$$

Solution: Given series,

$$= 0.7 + 0.77 + 0.777 + \ldots \text{ up to } n \text{ terms}$$
$$= 7\,(0.1 + 0.11 + 0.111 + \ldots \text{ up to } n \text{ terms})$$
$$= \frac{7}{9}\,(0.9 + 0.99 + 0.999 + \ldots \text{ up to } n \text{ terms})$$
$$= \frac{7}{9}\left[\left(1-\frac{1}{10}\right)+\left(1-\frac{1}{10^2}\right)+\left(1-\frac{1}{10^3}\right)+\ldots\right]$$
$$= \frac{7}{9}\left[n-\frac{1}{10}-\frac{1}{10^2}-\ldots \text{ up to } n \text{ terms}\right]$$
$$= \frac{7}{9}\left[n-\frac{\frac{1}{10}\,(1-1/10^n)}{1-\frac{1}{10}}\right]$$
$$= \frac{7}{9}\left[n-\frac{1}{9}\left(1-\frac{1}{10^n}\right)\right]$$
$$= \frac{7}{9}\left[n-\frac{1}{9}\left(1-\frac{1}{10^n}\right)\right]$$

Example 4.17: A manufacturer reckons that the value of a machine which costs him ₹ 18750 will depreciate each year by 20%. Find the estimated value at the end of 5 years.

Solution: At the end of first year the value of machine is

$$= 18750 \times \frac{80}{100}$$
$$= \frac{4}{5}\,(18750)$$

At the end of 2nd year it is equal to $\left(\frac{4}{5}\right)^2$ (18750); proceeding in this manner, the estimated value of machine at the end of 5 years is $\left(\frac{4}{5}\right)^5$ (18750)

$$= \frac{64 \times 16}{125 \times 25} \times 18750$$
$$= \frac{1024}{125} \times 750$$
$$= 1024 \times 6$$
$$= 6144 \text{ rupees}$$

Example 4.18: Show that a given sum of money accumulated at 20 % per annum, more than doubles itself in 4 years at compound interest.

Solution: Let the given sum be a rupees. After 1 year it becomes $\frac{6a}{5}$ (it is increased by $\frac{a}{5}$).

NOTES

At the end of two years it becomes $\frac{6}{5}\left(\frac{6a}{5}\right) = \left(\frac{6}{5}\right)^2 a$

Proceeding in this manner, we get that at the end of 4th year, the amount will be $\left(\frac{6}{5}\right)^4 a = \frac{1296}{625}\ a$

Now, $\frac{1296}{625}a - 2a = \frac{46}{625}a,$ since a is a + ve quantity, so the amount after 4 years is more than double of the original amount.

4.3.6 Harmonic Mean

If a, b, c are in HP, then b is called a *Harmonic Mean* between a and c, written as HM

Let $H_1, H_2, H_3, ..., H_n$ be the required Harmonic Means. Then

$a, H_1, H_2, ..., H_n, b$ are in HP

i.e., $\frac{1}{a}, \frac{1}{H_1}, \frac{1}{H_2}, ..., \frac{1}{H_n}, \frac{1}{b}$ are in AP

Then, $\frac{1}{b} = (n+2)$th term of an AP

$$= \frac{1}{a} + (n+1)d$$

Where d is the common difference of AP.

This gives, $d = \frac{a-b}{(n+1)ab}$

Now, $$\frac{1}{H_1} = \frac{1}{a} + d = \frac{1}{a} + \frac{a-b}{(n+1)\,ab}$$

$$= \frac{nb+b+a-b}{(n+1)\,ab} = \frac{a+nb}{(n+1)\,ab}$$

So, $$\frac{1}{H_1} = \frac{a+nb}{(n+1)\,ab}$$

$$\Rightarrow \quad H_1 = \frac{(n+1)\,ab}{a+nb}$$

Again, $$\frac{1}{H_2} = \frac{1}{a} + 2d = \frac{1}{a} + \frac{2\,(a-b)}{(n+1)\,ab}$$

$$= \frac{nb+b+2a-2b}{(n+1)\,ab} = \frac{2a-b+nb}{(n+1)\,ab}$$

$$\Rightarrow \quad H_2 = \frac{(n+1)\,ab}{2a-b+nb}$$

Similarly, $$\frac{1}{H_3} = \frac{1}{a} + 3d = \frac{3a-2b+nb}{(n+1)\,ab}$$

$$\Rightarrow \quad H_3 = \frac{(n+1)\,ab}{3a-2b+nb}$$ and so on,

$$\frac{1}{H_n} = \frac{1}{a} + nd = \frac{1}{a} + \frac{n\,(a-b)}{(n+1)\,ab}$$

$$= \frac{nb+b+na-nb}{(n+1)\,ab}$$

$$= \frac{na+b}{(n+1)\,ab} \Rightarrow H_n = \frac{(n+1)\,ab}{na+b}$$

NOTES

Example 4.19: Find the 5th term of 2, $2\frac{1}{2}$, $3\frac{1}{3}$, ...

Solution: Let 5th term be x. Then, $\frac{1}{x}$ is 5th term of corresponding AP $\frac{1}{2}, \frac{2}{5}, \frac{3}{10}, \ldots\ldots$

Then, $$\frac{1}{x} = \frac{1}{2} + 4\left(\frac{2}{5} - \frac{1}{2}\right) = \frac{1}{2} + 4\left(\frac{-1}{10}\right)$$

$$\Rightarrow \quad \frac{1}{x} = \frac{1}{2} - \frac{2}{5} = \frac{1}{10} \Rightarrow x = 10$$

Example 4.20: Insert two harmonic means between $\frac{1}{2}$ and $\frac{4}{17}$.

Solution: Let H_1, H_2 be two harmonic means between $\frac{1}{2}$ and $\frac{4}{17}$

Thus, 2, $\frac{1}{H_1}, \frac{1}{H_2}, \frac{17}{4}$ are in AP Let d be their common difference

Then, $$\frac{17}{4} = 2 + 3d$$

$$\Rightarrow \quad 3d = \frac{9}{4} \quad \Rightarrow \quad d = \frac{3}{4}$$

Thus, $$\frac{1}{H_1} = 2 + \frac{3}{4} = \frac{11}{4} \quad \Rightarrow \quad H_1 = \frac{4}{11}$$

$$\frac{1}{H_2} = 2 + 2 \times \frac{3}{4} = \frac{7}{2} \quad \Rightarrow \quad H_2 = \frac{2}{7}$$

Required harmonic means are $\frac{4}{11}, \frac{2}{7}$.

4.3.7 Quartiles, Percentiles and Deciles

Some measures, other than the measures of central tendency, are often employed when summarizing or describing a set of data where it is necessary to divide the data into equal parts. These are positional measures and are called quantiles and consist of quartiles, deciles and percentiles. The quartiles divide the data into four equal parts. The deciles divide the total ordered data into ten equal parts and the percentiles divide the data into 100 equal parts. Consequently, there are three quartiles, nine deciles and 99 percentiles. The quartiles are denoted by the symbol Q, which can be fractioned as Q_1, Q_2, Q_3, ..., and so on. Here, Q_1 will be such point in the ordered data which has 25 per cent of the data below and Q_2 will represent 75 per cent of the data above it. In other words, Q_1 is the value corresponding to $\left(\frac{n+1}{4}\right)$th ordered observation. Similarly, Q_2 divides the data in the middle, and is also equal to the median and its value, Q_2 is given by:

Q_2 = The value of $2\left(\frac{n+1}{4}\right)$th ordered observation in the data.

Similarly, we can calculate the values of various deciles. For instance,

$D_1 = \left(\frac{n+1}{10}\right)$th observaton in the ordered data, and

$D_7 = 7\left(\frac{n+1}{10}\right)$th observation in the ordered data.

NOTES

Percentiles are generally used in the research area of education where people are given standard tests and it is desirable to compare the relative position of the subject's performance on the test. Percentiles are similarly calculated as,

$P_7 = 7\left(\frac{n+1}{100}\right)$th observation in the ordered data.

and,

$P_{69} = 69\left(\frac{n+1}{100}\right)$th observation in the ordered data.

Quartiles

The formula for calculating the values of quartiles for grouped data is given as,

$Q = L + (j/f)C$

Where,

Q = The quartile under consideration.

L = Lower limit of the class interval which contains the value of Q.

j = The number of units we lack from the class interval which contains the value of Q, in reaching the value of Q.

f = Frequency of the class interval containing Q.

C = Size of the class interval.

Let us assume, we took the data of the ages of 100 students and a frequency distribution for this data has been constructed as shown in Table 4.4.

Table 4.4 *The Frequency Distribution Ages of 100 Students*

Ages (*CI*)	*Mid-point* (*X*)	(*f*)	*f*(*X*)	$f(X)^2$
16 and upto 17	16.5	4	66	1089.0
17 and upto 18	17.5	14	245	4287.5
18 and upto 19	18.5	18	333	6160.5
19 and upto 20	19.5	28	546	10647.0
20 and upto 21	20.5	20	410	8405.0
21 and upto 22	21.5	12	258	5547.0
22 and upto 23	22.5	4	90	2025.0
		Total = 100	1948	38161

In our case, in order to find Q_1, where Q_1 is the cut-off point so that 25 per cent of the data is below this point and 75 per cent of the data is above, we see that the first group has 4 students and the second group has 14 students, making a total of 18 students. Since Q_1 cuts off at 25 students, it is the third class interval which contains Q_1. This means that the value of L in our formula is 18.

Since we already have 18 students in the first two groups, we need 7 more students from the third group to make it a total of 25 students, which is the value of Q_1. Hence, the value of (j) is 7. Also, since the frequency of this third class interval which contains Q_1 is 18, the value of (f) in our formula is 18. The size of the class interval C is given as 1. Substituting these values in the formula for Q, we get,

$$Q_1 = 18 + (7/18)1$$
$$= 18 + 0.38 = 18.38$$

This means that 25 per cent of the students are below 18.38 years of age and 75 per cent are above this age.

Similarly, we can calculate the value of Q_2, using the same formula. Hence,

$$Q_2 = L + (j/f)C$$
$$= 19 + (14/28)1$$
$$= 19.5$$

This also happens to be the median.

By using the same formula and the same logic we can calculate the values of all deciles as well as percentiles.

We have defined the median as the value of the item which is located at the centre of the array. We can define other measures which are located at other specified points. Thus, the Nth percentile of an array is the value of the item such that N per cent items lie below it. Clearly then, the N_{th} percentile P_n of grouped data is given by,

$$P_n = l + \frac{\frac{nN}{100} - C}{f} \times i$$

Here, l is the lower limit of the class in which $nN/100$th item lies, i its width, f its frequency, C the cumulative frequency upto (but not including) this class, and N is the total number of items.

We can similarly define the Nth decile as the value of the item below which ($nN/10$) items of the array lie. Clearly,

$$D_n = P_{10n} = l + \frac{\frac{nN}{10} - C}{f} \times i$$

where the symbols have the obvious meanings.

The other most commonly referred to measures of location are the quartiles. Thus, nth quartile is the value of the item which lies at the $n(N/5)$th item. Clearly, Q_2, the second quartile, is the median for grouped data.

$$Q_n = P_{25n} = l + \frac{\frac{nN}{4} - C}{f} \times i$$

4.3.8 Box Plot

In descriptive statistics, a box plot or boxplot, also called as box-and-whisker diagram or plot, is the most appropriate method to depict clusters of numerical data graphically by means of the following:

1. The Smallest Observation (Sample Minimum).
2. The Lower Quartile (Q_1).
3. The Median (Q_2).
4. The Upper Quartile (Q_3).
5. The Largest Observation (Sample Maximum).

NOTES

A box plot also specifies that which observations can be considered as outliers. Box plots are non-parametric and can be drawn either horizontally or vertically.

The box plots are formed as follows:

- Vertical Axis: Response Variable.
- Horizontal Axis: The Factor of Interest.

To draw a box plot, perform the following steps:

Step 1: Calculate the median and the quartiles. The lower quartile is the 25th percentile and the upper quartile is the 75th percentile.

Step 2: Plot a symbol at the median or draw a line and then draw a box between the lower and upper quartiles. This box represents the middle 50% of the data which is the body of the data.

Step 3: Draw first line from the lower quartile to the minimum point and another line from the upper quartile to the maximum point. Typically, a symbol is drawn at these minimum and maximum points, although this is optional.

Step 4: Calculate the interquartile range, i.e., the difference between the upper and lower quartile, called IQ.

Step 5: Now calculate the following points:

$$L_1 = \text{Lower Quartile} - 1.5*IQ$$

$$L_2 = \text{Lower Quartile} - 3.0*IQ$$

$$U_1 = \text{Upper Quartile} + 1.5*IQ$$

$$U_2 = \text{Upper Quartile} + 3.0*IQ$$

Step 6: The line from the lower quartile to the minimum can be drawn from the lower quartile to the smallest point that is greater than L_1. Similarly, the line from the upper quartile to the maximum can be drawn to the largest point smaller than U_1.

Step 7: Points between L_1 and L_2 or between U_1 and U_2 can be drawn as small circles. Points less than L_2 or greater than U_2 can be drawn as large circles

Thus the box plot identifies the middle 50% of the data, the median and the extreme points. A single box plot can be drawn for one set of data with no distinct groups. Alternatively, multiple box plots can be drawn together to compare multiple data sets or to compare groups in a single data set. For a single box plot, the width of the box is arbitrary. For multiple box plots, the width of the box plot can be set proportional to the number of points in the given group or sample.

Check Your Progress

1. Define statistics.
2. How does statistics classify numerical facts?
3. What is the first step in the statistical treatment of a problem?
4. What is central tendency in statistics?
5. Define the term arithmetic mean.
6. When is weighted arithmetic mean used?
7. Define the term median.
8. What is mode?
9. What are the four important methods of estimating mode of a series?
10. What are positional measures?

NOTES

4.4 MEASURES OF DISPERSION

A measure of dispersion, or simply dispersion may be defined as statistics signifying the extent of the scatteredness of items around a measure of central tendency.

A measure of dispersion may be expressed in an 'Absolute form', or in a 'Relative form'. It is said to be in an absolute form when it states the actual amount by which the value of an item on an average deviates from a measure of central tendency. Absolute measures are expressed in concrete units, i.e., units in terms of which the data have been expressed, e.g., rupees, centimetres, kilograms, etc., and are used to describe frequency distribution.

A relative measure of dispersion computed is a quotient by dividing the absolute measures by a quantity in respect to which absolute deviation has been computed. It is as such a pure number and is usually expressed in a percentage form. Relative measures are used for making comparisons between two or more distributions.

A measure of dispersion should possess all those characteristics which are considered essential for a measure of central tendency, viz.

- It should be based on all observations.
- It should be readily comprehensible.
- It should be fairly easily calculated.
- It should be affected as little as possible by fluctuations of sampling.
- It should be amenable to algebraic treatment.

The following are some common measures of dispersion:

(*i*) The range, (*ii*) the Semi-interquartile range or the quartile deviation, (*iii*) The mean deviation, and (*iv*) The standard deviation. Of these, the standard deviation is the best measure. We describe these measures in the following sections.

NOTES

4.4.1 Range

The crudest measure of dispersion is the range of the distribution. The range of any series is the difference between the highest and the lowest values in the series. If the marks received in an examination taken by 248 students are arranged in ascending order, then the range will be equal to the difference between the highest and the lowest marks.

In a frequency distribution, the range is taken to be the difference between the lower limit of the class at the lower extreme of the distribution and the upper limit of the class at the upper extreme.

Table 4.5 *Weekly Earnings of Labourers in Four Workshops of the Same Type*

Weekly earnings ₹	*No. of workers*			
	Workshop A	*Workshop B*	*Workshop C*	*Workshop D*
15–16	...	...	2	...
17–18	...	2	4	...
19–20	...	4	4	4
21–22	10	10	10	14
23–24	22	14	16	16
25–26	20	18	14	16
27–28	14	16	12	12
29–30	14	10	6	12
31–32	...	6	6	4
33–34	...	...	2	2
35–36	...	...	...	...
37–38	...	...	4	...
Total	80	80	80	80
Mean	25.5	25.5	25.5	25.5

Consider the data on weekly earning of worker on four workshops given in the Table 4.5. We note the following:

Workshop	*Range*
A	9
B	15
C	23
D	15

From these figures, it is clear that the greater the range, the greater is the variation of the values in the group.

The range is a measure of absolute dispersion and as such cannot be usefully employed for comparing the variability of two distributions expressed in different units. The amount of dispersion measured, say, in pounds, is not comparable with dispersion measured in inches. So the need of measuring relative dispersion arises.

An absolute measure can be converted into a relative measure if we divide it by some other value regarded as standard for the purpose. We may use the mean of the distribution or any other positional average as the standard.

For Table 4.5, the relative dispersion would be:

$$\text{Workshop } A = \frac{9}{25.5} \qquad \text{Workshop } C = \frac{23}{25.5}$$

$$\text{Workshop } B = \frac{15}{25.5} \qquad \text{Workshop } D = \frac{15}{25.5}$$

NOTES

An alternate method of converting an absolute variation into a relative one would be to use the total of the extremes as the standard. This will be equal to dividing the difference of the extreme items by the total of the extreme items. Thus,

$$\text{Relative Dispersion} = \frac{\text{Difference of extreme items, i.e, Range}}{\text{Sum of extreme items}}$$

The relative dispersion of the series is called the coefficient or ratio of dispersion. In our example of weekly earnings of workers considered earlier, the coefficients would be:

$$\text{Workshop } A = \frac{9}{21+30} = \frac{9}{51} \qquad \text{Workshop } B = \frac{15}{17+32} = \frac{15}{49}$$

$$\text{Workshop } C = \frac{23}{15+38} = \frac{23}{53} \qquad \text{Workshop } D = \frac{15}{19+34} = \frac{15}{53}$$

Merits and limitations of range

Merits

Of the various characteristics that a good measure of dispersion should possess, the range has only two, viz (i) It is easy to understand, and (ii) Its computation is simple.

Limitations

Besides the aforesaid two qualities, the range does not satisfy the other test of a good measure and hence it is often termed as a crude measure of dispersion.

The following are the limitations that are inherent in the range as a concept of variability:

(i) Since it is based upon two extreme cases in the entire distribution, the range may be considerably changed if either of the extreme cases happens to drop out, while the removal of any other case would not affect it at all.

(ii) It does not tell anything about the distribution of values in the series relative to a measure of central tendency.

(iii) It cannot be computed when distribution has open-end classes.

(iv) It does not take into account the entire data. These can be illustrated by the following illustration. Consider the data given in Table 4.6.

NOTES

Table 4.6 *Distribution with the Same Number of Cases, but Different Variability*

Class	*No. of students*		
	Section A	*Section B*	*Section C*
0–10	...	...	...
10–20	1	...	...
20–30	12	12	19
30–40	17	20	18
40–50	29	35	16
50–60	18	25	18
60–70	16	10	18
70–80	6	8	21
80–90	11	...	...
90–100	...	...	...
Total	110	110	110
Range	80	60	60

The table is designed to illustrate three distributions with the same number of cases but different variability. The removal of two extreme students from section *A* would make its range equal to that of *B* or *C*.

The greater range of *A* is not a description of the entire group of 110 students, but of the two most extreme students only. Further, though sections *B* and *C* have the same range, the students in section B cluster more closely around the central tendency of the group than they do in section *C*. Thus, the range fails to reveal the greater homogeneity of *B* or the greater dispersion of *C*. Due to this defect, it is seldom used as a measure of dispersion.

Specific uses of range

In spite of the numerous limitations of the range as a measure of dispersion, there are the following circumstances when it is the most appropriate one:

(i) In situations where the extremes involve some hazard for which preparation should be made, it may be more important to know the most extreme cases to be encountered than to know anything else about the distribution. For example, an explorer, would like to know the lowest and the highest temperatures on record in the region he is about to enter; or an engineer would like to know the maximum rainfall during 24 hours for the construction of a storem water drain.

(ii) In the study of prices of securities, range has a special field of activity. Thus to highlight fluctuations in the prices of shares or bullion it is a common practice to indicate the range over which the prices have moved during a certain period of time. This information, besides being of use to the operators, gives an indication of the stability of the bullion market, or that of the investment climate.

(iii) In statistical quality control the range is used as a measure of variation. We, e.g., determine the range over which variations in quality are due to random causes, which is made the basis for the fixation of control limits.

4.4.2 Quartile Deviation

Another measure of dispersion, much better than the range, is the semi-interquartile range, usually termed as 'Quartile Deviation'. As stated in the previous unit, quartiles are the points which divide the array in four equal parts. More precisely, Q_1 gives the value of the item 1/4th the way up the distribution and Q_3 the value of the item 3/4th the way up the distribution. Between Q_1 and Q_3 are included half the total number of items. The difference between Q_1 and Q_3 includes only the central items but excludes the extremes. Since under most circumstances, the central half of the series tends to be fairly typical of all the items, the interquartile range ($Q_3 - Q_1$) affords a convenient and often a good indicator of the absolute variability. The larger the interquartile range, the larger the variability.

Usually, one-half of the difference between Q_3 and Q_1 is used and to it is given the name of quartile deviation or semi-interquartile range. The interquartile range is divided by two for the reason that half of the interquartile range will, in a normal distribution, be equal to the difference between the median and any quartile. This means that 50 per cent items of a normal distribution will lie within the interval defined by the median plus and minus the semi-interquartile range.

Symbolically:

$$Q.D. = \frac{Q_3 - Q_1}{2} \qquad ...(4.1)$$

Let us find quartile deviations for the weekly earnings of labour in the four workshop whose data is given in Table 4.5. The computations are as shown in Table 4.7.

As shown in the table, Q.D. of workshop *A* is ₹ 2.12 and median value in 25.3. This means that if the distribution is symmetrical the number of workers whose wages vary between (25.3–2.1) = ₹ 23.2 and (25.3 + 2.1) = ₹ 27.4, shall be just half of the total cases. The other half of the workers will be more than ₹ 2.1 removed from the median wage. As this distribution is not symmetrical, the distance between Q_1 and the median Q_2 is not the same as between Q_3 and the median. Hence the interval defined by median plus and minus semi inter-quartile range will not be exactly the same as given by the value of the two quartiles. Under such conditions the range between ₹ 23.2 and ₹ 27.4 will not include precisely 50 per cent of the workers.

If quartile deviation is to be used for comparing the variability of any two series, it is necessary to convert the absolute measure to a coefficient of quartile deviation. To do this the absolute measure is divided by the average size of the two quartile.

Symbolically:

$$\text{Coefficient of quartile deviation} = \frac{Q_3 - Q_1}{Q_3 + Q_1} \qquad ...(4.2)$$

Applying this to our illustration of four workshops, the coefficients of Q.D. are as given below.

NOTES

Table 4.7 *Calculation of Quartile Deviation*

		Workshop A	Workshop B	Workshop C	Workshop D
Location of Q_2	$\frac{N}{2}$	$\frac{80}{2}=40$	$\frac{80}{2}=40$	$\frac{80}{2}=40$	$\frac{80}{2}=40$
	Q_2	$24.5+\frac{40-30}{22}\times 2$	$24.5+\frac{40-30}{18}\times 2$	$24.5+\frac{40-30}{16}\times 2$	$24.5+\frac{40-30}{16}\times 2$
		= 24.5 + 0.9	= 24.5 + 1.1	= 24.5 + 0.75	= 24.5 + 0.75
		= 25.4	= 25.61	= 25.25	= 25.25
Location of Q_1	$\frac{N}{4}$	$\frac{80}{4}=20$	$\frac{80}{4}=20$	$\frac{80}{4}=20$	$\frac{80}{4}=20$
	Q_1	$22.5+\frac{20-10}{22}\times 2$	$22.5+\frac{20-16}{14}\times 2$	$20.5+\frac{20-10}{10}\times 2$	$22.5+\frac{20-18}{16}\times 2$
		= 22.5 + .91	= 22.5 + .57	= 20.5 + 2	= 22.5 + .25
		= 23.41	= 23.07	= 22.5	= 22.75
Location of Q_3	$\frac{3N}{4}$	$3\times\frac{80}{4}=60$	60	60	60
	Q_3	$26.5+\frac{60-52}{14}\times 2$	$26.5+\frac{60-48}{16}\times 2$	$26.5+\frac{60-50}{12}\times 2$	
$26.5+\frac{60-50}{12}\times 2$					
		= 26.5 + 1.14	= 26.5 + 1.5	= 26.5 + 1.67	= 26.5 + 1.67
		= 27.64	= 28.0	= 28.17	= 28.17
Quartile Deviation	$\frac{Q_3-Q_1}{2}$	$\frac{27.64-23.41}{2}$	$\frac{28-23.07}{2}$	$\frac{28.17-22.5}{2}$	$\frac{28.17-22.75}{2}$
		$=\frac{4.23}{2}$ = ₹ 2.12	$=\frac{4.93}{2}$ = ₹ 2.46	$=\frac{5.67}{2}$ = ₹ 2.83	$=\frac{5.42}{2}$ = ₹. 2.71
Coefficient of quartile deviation	$\frac{Q_3-Q_1}{Q_3+Q_1}$ =	$\frac{27.64-23.41}{27.64+23.41}$	$\frac{28-23.07}{28+23.07}$	$\frac{28.17-22.5}{28.17+22.5}$	$\frac{28.17-22.75}{28.17+22.75}$
		= 0.083	= 0.097	= 0.112	= 0.106

Characteristics of quartile deviation

- The size of the quartile deviation gives an indication about the uniformity or otherwise of the size of the items of a distribution. If the quartile deviation is small it denotes large uniformity. Thus, a coefficient of quartile deviation may be used for comparing uniformity or variation in different distributions.
- Quartile deviation is not a measure of dispersion in the sense that it does not show the scatter around an average, but only a distance on scale. Consequently, quartile deviation is regarded as a measure of partition.
- It can be computed when the distribution has open-end classes.

Limitations of quartile deviation

Except for the fact that its computation is simple and it is easy to understand, a quartile deviation does not satisfy any other test of a good measure of variation.

4.4.3 Mean Deviation

A weakness of the measures of dispersion discussed earlier, based upon the range or a portion thereof, is that the precise size of most of the variants has no effect on the result. As an illustration, the quartile deviation will be the same whether the variates between Q_1 and Q_3 are concentrated just above Q_1 or they are spread uniformly from Q_1 to Q_3. This is an important defect from the viewpoint of measuring the divergence of the distribution from its typical value. The mean deviation is employed to answer the objection.

Mean deviation also called average deviation, of a frequency distribution is the mean of the absolute values of the deviation from some measure of central tendency. In other words, mean deviation is the arithmetic average of the variations (deviations) of the individual items of the series from a measure of their central tendency.

We can measure the deviations from any measure of central tendency, but the most commonly employed ones are the median and the mean. The median is preferred because it has the important property that the average deviation from it is the least.

Calculation of the mean deviation then involves the following steps:

(*a*) Calculate the median (or the mean) Me (or $\overline{X}$).

(*b*) Record the deviations $|d| = |x - Me|$ of each of the items, ignoring the sign.

(*c*) Find the average value of deviations.

$$\text{Mean Deviation} = \frac{\Sigma|d|}{N} \qquad ...(4.3)$$

Example 4.21: Calculate the mean deviation from the following data giving marks obtained by 11 students in a class test.

14, 15, 23, 20, 10, 30, 19, 18, 16, 25, 12.

Solution: $\text{Median} = \text{Size of } \frac{11+1}{2} \text{ th item}$

$= \text{size of 6th item} = 18.$

Serial No.	*Marks*	$\lvert x - \text{Median}\rvert$ $\lvert d\rvert$
1	10	8
2	12	6
3	14	4
4	15	3
5	16	2
6	18	0
7	19	1
8	20	2
9	23	5
10	25	7
11	30	12
		$\Sigma\lvert d\rvert = 50$

NOTES

Mean deviation from median $= \dfrac{\sum |d|}{N}$

$= \dfrac{50}{11} = 4.5$ marks.

For grouped data, it is easy to see that the mean deviation is given by

$$\text{Mean deviation, M.D.} = \frac{\sum f|d|}{\sum f} \quad ...(1)$$

where $|d| = |x - \text{median}|$ for grouped discrete data, and $|d| = M - \text{median}|$ for grouped continuous data with M as the mid-value of a particular group. The following examples illustrate the use of this formula.

Example 4.22: Calculate the mean deviation from the following data

Size of item	6	7	8	9	10	11	12
Frequency	3	6	9	13	8	5	4

Solution:

Size	*Frequency* f	*Cumulative frequency*	*Deviations from median (9)* $\lvert d \rvert$	$f\lvert d \rvert$
6	3	3	3	9
7	6	9	2	12
8	9	18	1	9
9	13	31	0	0
10	8	39	1	8
11	5	44	2	10
12	4	48	3	12
	48			60

Median = the size of $\dfrac{48+1}{2} = 24.5$th item which is 9.

Therefore, deviations d are calculated from 9, i.e., $|d| = |x - 9|$.

$$\text{Mean deviation} = \frac{\sum f|d|}{\sum f} = \frac{60}{48} = 1.25$$

Example 4.23: Calculate the mean deviation from the following data:

x	0–10	10–20	20–30	30–40	40–50	50–60	60–70	70–80
f	18	16	15	12	10	5	2	2

Solution:

This is a frequency distribution with continuous variable. Thus, deviations are calculated from mid-values.

x	*Mid-value*	f	*Less than c.f.*	*Deviation from median* $\|d\|$	$f\|d\|$
0–10	5	18	18	19	342
10–20	15	16	34	9	144
20–30	25	15	49	1	15
30–40	35	12	61	11	132
40–50	45	10	71	21	210
50–60	55	5	76	31	155
60–70	65	2	78	41	82
70–80	75	2	80	51	102
		80			1182

NOTES

$$\text{Median} = \text{the size of } \frac{80}{2} \text{ th item}$$

$$= 20 + \frac{6}{15} \times 10 = 24$$

and then, mean deviation $= \dfrac{\sum f|d|}{\sum f}$

$$= \frac{1182}{80} = 14.775.$$

Merits and demerits of the mean deviation

Merits

- It is easy to understand.
- As compared to standard deviation (discussed later), its computation is simple.
- As compared to standard deviation, it is less affected by extreme values.
- Since it is based on all values in the distribution, it is better than range or quartile deviation.

Demerits

- It lacks those algebraic properties which would facilitate its computation and establish its relation to other measures.
- Due to this, it is not suitable for further mathematical processing.

Coefficient of mean deviation

The coefficient or relative dispersion is found by dividing the mean deviations recorded. Thus,

$$\text{Coefficient of M.D.} = \frac{\text{Mean Deviation}}{\text{Mean}} \quad ...(4.4)$$

(when deviations were recorded from the mean)

$$= \frac{\text{M.D.}}{\text{Median}} \quad ...(4.5)$$

(when deviations were recorded from the median)

Applying the above formula to Example 4.23.

Coefficient of Mean deviation $= \frac{14.775}{24}$

$= 0.616$

NOTES

4.5 STANDARD DEVIATION

By far the most universally used and the most useful measure of dispersion is the standard deviation or root mean square deviation about the mean. We have seen that all the methods of measuring dispersion so far discussed are not universally adopted for want of adequacy and accuracy. The range is not satisfactory as its magnitude is determined by most extreme cases in the entire group. Further, the range is notable because it is dependent on the item whose size is largely matter of chance. Mean deviation method is also an unsatisfactory measure of scatter, as it ignores the algebraic signs of deviation. We desire a measure of scatter which is free from these shortcomings. To some extent standard deviation is one such measure.

The calculation of standard deviation differs in the following respects from that of mean deviation. First, in calculating standard deviation, the deviations are squared. This is done so as to get rid of negative signs without committing algebraic violence. Further, the squaring of deviations provides added weight to the extreme items, a desirable feature for certain types of series.

Secondly, the deviations are always recorded from the arithmetic mean, because although the sum of deviations is the minimum from the median, the sum of squares of deviations is minimum when deviations are measured from the arithmetic average. The deviation from $\bar{x}$ is represented by d.

Thus, standard deviation, σ (sigma) is defined as the square root of the mean of the squares of the deviations of individual items from their arithmetic mean.

$$\sigma = \sqrt{\frac{\Sigma(x-\bar{x})^2}{N}} \qquad ...(4.6)$$

For grouped data (discrete variables)

$$\sigma = \sqrt{\frac{\Sigma f(x-\bar{x})^2}{\Sigma f}} \qquad ...(4.7)$$

and, for grouped data (continuous variables)

$$\sigma = \sqrt{\frac{\Sigma f(M-\bar{x})}{\Sigma f}} \qquad ...(4.8)$$

where M is the mid-value of the group.

The use of these formulae is illustrated by the following examples.

Example 4.24: Compute the standard deviation for the following data:

11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21.

NOTES

Solution:

Here Equation (4.6) is appropriate. We first calculate the mean as $\bar{x} = \sum x/N =$ 176/11= 16, and then calculate the deviation as follows:

x	$(x - \bar{x})$	$(x - \bar{x})^2$
11	–5	25
12	–4	16
13	–3	9
14	–2	4
15	–1	1
16	0	0
17	+1	1
18	+2	4
19	+3	9
20	+4	16
21	+5	25
176		10

Thus by Equation (4.6).

$$\sigma = \sqrt{\frac{110}{11}} = \sqrt{10} = 3.16$$

Example 4.25: Find the standard deviation of the data in the following distributions:

x	12	13	14	15	16	17	18	20
f	4	11	32	21	15	8	6	4

Solution:

For this discrete variable grouped data, we use formula 8. Since for calculation of $\bar{x}$, we need $\sum fx$ and then for σ we need $\sum f(x-\bar{x})^2$, the calculations are conveniently made in the following format.

x	f	fx	$d = x - \bar{x}$	d^2	fd^2
12	4	48	–3	9	36
13	11	143	–2	4	44
14	32	448	–1	1	32
15	21	315	0	0	0
16	15	240	1	1	15
17	8	136	2	4	32
18	5	90	3	9	45
20	4	80	5	25	100
	100	1500			304

Here $\bar{x} = \sum fx / \sum f = 1500/100 = 15$

and $\sigma = \sqrt{\frac{\sum fd^2}{\sum f}}$

$$= \sqrt{\frac{304}{100}} = \sqrt{3.04} = 1.74$$

NOTES

Example 4.26: Calculate the standard deviation of the following data.

Class	1–3	3–5	5–7	7–19	9–11	11–13	13–15
frequency	1	9	25	35	17	10	3

Solution: This is an example of continuous frequency series and formula 9 seems appropriate.

Class	*Mid-point* x	*Frequency* f	fx	*Deviation of mid-point x from mean (8)*	*Squared deviation* d^2	*Squared deviation times frequency* d^2
1–3	2	1	2	–6	36	36
3–5	4	9	36	–4	16	144
5–7	6	25	150	–2	4	100
7–9	8	35	280	0	0	0
9–11	10	17	170	2	4	68
11–13	12	10	120	4	16	160
13–15	14	3	42	6	36	108
		100	800			616

First the mean is calculated as

$$\bar{x} = \sum fx/\sum x = 800/100 = 8.0$$

Then the deviations are obtained from 8.0. The standard deviation

$$\sigma = \sqrt{\frac{\sum f(M-\bar{x})^2}{\sum f}}$$

$$\sigma = \sqrt{\frac{\sum fd^2}{\sum f}} = \sqrt{\frac{616}{100}}$$

$$= 2.48$$

4.5.1 Calculation of Standard Deviation by Short-cut Method

The three examples worked out above have one common simplifying feature, namely $\bar{x}$ in each, turned out to be an integer, thus, simplifying calculations. In most cases, it is very unlikely that it will turn out to be so. In such cases, the calculation of d and d^2 becomes quite time-consuming. Short-cut methods have consequently been developed. These are on the same lines as those for calculation of mean itself.

In the short-cut method, we calculate deviations x' from an assumed mean A. Then,

for ungrouped data

$$\sigma = \sqrt{\frac{\sum x'^2}{N} - \left(\frac{\sum x'}{N}\right)^2} \qquad ...(4.9)$$

and for grouped data

$$\sigma = \sqrt{\frac{\sum fx'^2}{\sum f} - \left(\frac{fx'}{\sum f}\right)^2} \qquad ...(4.10)$$

NOTES

This formula is valid for both discrete and continuous variables. In case of continuous variables, x in the equation $x' = x - A$ stands for the mid-value of the class in question.

Note that the second term in each of the formulae is a correction term because of the difference in the values of A and $\bar{x}$. When A is taken as $\bar{x}$ itself, this correction is automatically reduced to zero.

Example 4.27: Compute the standard deviation by the short-cut method for the following data:

11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21

Solution: Let us assume that $A = 15$.

	$x' = (x - 15)$	x'^2
11	–4	16
12	–3	9
13	–2	4
14	–1	1
15	0	0
16	1	1
17	2	4
18	3	9
19	4	16
20	5	25
21	6	36
$N = 11$	$\sum x' = 11$	$\sum x'^2 = 121$

$$\sigma = \sqrt{\frac{\sum x'^2}{N} - \left(\frac{\sum x'}{N}\right)^2}$$

$$= \sqrt{\frac{121}{11} - \left(\frac{11}{11}\right)^2}$$

$$= \sqrt{11 - 1}$$

$$= \sqrt{10}$$

$$= 3.16.$$

Another method

If we assumed A as zero, then the deviation of each item from the assumed mean is the same as the value of item itself. Thus, 11 deviates from the assumed mean of zero by 11, 12 deviates by 12, and so on. As such, we work with deviations without having to compute them, and the formula takes the following shape:

NOTES

x	x^2
11	121
12	144
13	169
14	196
15	225
16	256
17	289
18	324
19	361
20	400
21	441
176	2,926

$$\sigma = \sqrt{\frac{\sum x^2}{N} - \left(\frac{\sum x}{N}\right)^2}$$

$$= \sqrt{\frac{2926}{11} - \left(\frac{176}{11}\right)^2} = \sqrt{266 - 256} = 3.16$$

Example 4.28: Calculate the standard deviation of the following data by short method.

Person	1	2	3	4	5	6	7
Monthly income (Rupees)	300	400	420	440	460	480	580

Solution: In this data, the values of the variable are very large making calculations cumbersome. It is advantageous to take a common factor out. Thus, we use $x' = \frac{x - A}{20}$. The standard deviation is calculated using x' and then the true value of σ is obtained by multiplying back by 20. The effective formula then is

$$\sigma = C \times \sqrt{\frac{\sum x'^2}{N} - \left(\frac{\sum x'}{N}\right)^2}$$

where C represents the common factor.

Using $x' = (x - 420)/20$.

x	*Deviation from Assumed mean* $x' = (x - 420)$	x'	x'^2
300	–120	–6	36
400	–20	–1	1
420	0	0	0
		–7	
440	20	1	1
460	40	2	4
480	60	3	9
580	160	8	64
		+ 14	
$N = 7$		7	115

$$\sigma = 20\times\sqrt{\frac{\sum x'^2}{N}-\left(\frac{\sum x'}{N}\right)^2}$$

$$= 20\sqrt{\frac{115}{7}-\left(\frac{7}{7}\right)^2}$$

$$= 78.56$$

Example 4.29: Calculate the standard deviation from the following data:

Size	6	9	12	15	18
Frequency	7	12	19	10	2

Solution:

x	*Frequency f*	*Deviation from assumed mean 12*	*Deviation divided by common factor 3 x′*	*x′ times frequency fx′*	*x′² times frequency fx′²*
6	7	–6	–2	–14	28
9	12	–3	–1	–12	12
12	19	0	0	0	0
15	10	3	1	10	10
18	2	6	2	4	8
	$N = 50$			$\sum fx' = -12$	$\sum fx'^2 = 58$

Since deviations have been divided by a common factor, we use

$$\sigma = C\sqrt{\frac{\sum fx'^2}{N}-\left(\frac{\sum fx'}{N}\right)^2}$$

$$= 3\sqrt{\frac{58}{50}-\left(\frac{-12}{50}\right)^2}$$

$$= 3\sqrt{1.1600-.0576} = 3\times 1.05 = 3.15.$$

Example 4.30: Obtain the mean and standard deviation of the first N natural numbers, i.e., of 1, 2, 3, ..., N – 1, N.

Solution: Let *x* denote the variable which assumes the values of the first *N* natural numbers.

Then

$$\bar{x} = \frac{\sum_1^N x}{N} = \frac{\frac{N(N+1)}{2}}{N} = \frac{N+1}{2}$$

because $\sum_1^N x = 1+2+3+...+(N-1)+N$

$$= \frac{N(N+1)}{2}$$

To calculate the standard deviation σ, we use 0 as the assumed mean A. Then

$$\sigma = \sqrt{\frac{\sum x^2}{N} - \left(\frac{\sum x}{N}\right)^2}$$

But $\sum x^2 = 1^2 + 2^2 + 3^2 + ... (N-1)^2 + N^2 = \frac{N(N+1)(2N+1)}{6}$

Therefore

$$\sigma = \sqrt{\frac{N(N+1)(2N+1)}{6N} - \frac{N^2(N+1)^2}{4N^2}}$$

$$= \sqrt{\frac{(N+1)}{2}\left[\frac{2N+1}{3} - \frac{N+1}{2}\right]} = \sqrt{\frac{(N+1)(N-1)}{12}}$$

Thus for first 11 natural numbers

$$\bar{x} = \frac{11+1}{2} = 6$$

and $\sigma = \sqrt{\frac{(11+1)(11-1)}{12}} = \sqrt{10} = 3.16$

Example 4.31:

	Mid-point x	*Frequency* f	*Deviation from class of assumed mean* x'	*Deviation time frequency* fx'	*Squared deviation times frequency* fx'^2
0–10	5	18	–2	–36	72
10–20	15	16	–1	–16	16
				–52	
20–30	25	15	0	0	0
30–40	35	12	1	12	12
40–50	45	10	2	20	40
50–60	55	5	3	15	45
60–70	65	2	4	8	32
70–80	75	1	5	5	25
				– 60	
		79		60	242
				–52	
				$\sum fx' = 8$	

Solution: Since the deviations are from assumed mean and expressed in terms of class-interval units,

$$\sigma = i \times \sqrt{\frac{\sum x'^2}{N} - \left(\frac{\sum fx'}{N}\right)^2}$$

$$= 10 \times \sqrt{\frac{242}{79} - \left(\frac{8}{79}\right)^2}$$

$$= 10 \times 1.75 = 17.5.$$

NOTES

4.5.2 Combining Standard Deviations of Two Distributions

If we were given two sets of data of N_1 and N_2 items with means $\bar{x}_1$ and $\bar{x}_2$ and standard deviations σ_1 and σ_2 respectively, we can obtain the mean and standard deviation $\bar{x}$ and σ of the combined distribution by the following formulae:

$$\bar{x} = \frac{N_1\bar{x}_1 + N_2\bar{x}_2}{N_1 + N_2} \qquad ...(4.11)$$

and $$\sigma = \sqrt{\frac{N_1\sigma_1^2 + N_2\sigma_2^2 + N_1(\bar{x} - \bar{x}_1)^2 + N_2(\bar{x} - \bar{x}_2)^2}{N_1 + N_2}} \qquad ...(4.12)$$

Example 4.32: The mean and standard deviations of two distributions of 100 and 150 items are 50, 5 and 40, 6 respectively. Find the standard deviation of all taken together.

Solution: Combined mean

$$\bar{x} = \frac{N_1\bar{x}_1 + N_2\bar{x}_2}{N_1 + N_2} = \frac{100 \times 50 + 150 \times 40}{100 + 150}$$

$$= 44$$

Combined standard deviation

$$\sigma = \sqrt{\frac{N_1\sigma_1^2 + N_2\sigma_2^2 + N_1(\bar{x} - \bar{x}_1)^2 + N_2(\bar{x} - \bar{x}_2)^2}{N_1 + N_2}}$$

$$= \sqrt{\frac{100 \times (5)^2 + 150\,(6)^2 + 100\,(44 - 50)^2 + 150\,(44 - 40)^2}{100 + 150}}$$

$$= 7.46.$$

Example 4.33: A distribution consists of three components with 200, 250, 300 items having mean 25, 10 and 15 and standard deviation 3, 4 and 5, respectively. Find the standard deviation of the combined distribution.

Solution: In the usual notations, we are given here

$N_1 = 200,\ N_2 = 250,\ N_3 = 300$

$\bar{x}_1 = 25,\ \bar{x}_2 = 10,\ \bar{x}_3 = 15$

The Equations (4.11) and (4.12) can easily be extended for combination of three series as

$$\bar{x} = \frac{N_1\bar{x}_1 + N_2\bar{x}_2 + N_3\bar{x}_3}{N_1 + N_2 + N_3}$$

$$= \frac{200 \times 25 + 250 \times 10 + 300 \times 15}{200 + 250 + 300}$$

$$= \frac{12000}{750} = 16$$

NOTES

and

$$\sigma = \sqrt{\frac{N_1\sigma_1^2 + N_2\sigma_2^2 + N_3\sigma_3^2 + N_1(\bar{x}-\bar{x}_1)^2 + N_2(\bar{x}-\bar{x}_2)^2 + N_3(\bar{x}-\bar{x}_3)^2}{N_1+N_2+N_3}}$$

$$= \sqrt{\frac{200\times 9 + 250\times 16 + 300\times 25 + 200\times 81 + 250\times 36 + 300\times 1}{200+250+300}}$$

$$= \sqrt{51.73} = 7.19.$$

4.5.3 Comparison of Various Measures of Dispersion

The range is the easiest to calculate the measure of dispersion, but since it depends on extreme values, it is extremely sensitive to the size of the sample, and to the sample variability. In fact, as the sample size increases the range increases dramatically, because the more the items one considers, the more likely it is that some item will turn up which is larger than the previous maximum or smaller than the previous minimum. So, it is, in general, impossible to interpret properly the significance of a given range unless the sample size is constant. It is for this reason that there appears to be only one valid application of the range, namely in statistical quality control where the same sample size is repeatedly used, so that comparison of ranges are not distorted by differences in sample size.

The quartile deviations and other such positional measures of dispersions are also easy to calculate but suffer from the disadvantage that they are not amenable to algebraic treatment. Similarly, the mean deviation is not suitable because we cannot obtain the mean deviation of a combined series from the deviations of component series. However, it is easy to interpret and easier to calculate than the standard deviation.

The standard deviation of a set of data, on the other hand, is one of the most important statistics describing it. It lends itself to rigorous algebraic treatment, is rigidly defined and is based on all observations. It is, therefore, quite insensitive to sample size (provided the size is 'Large Enough') and is least affected by sampling variations.

It is used extensively in testing of hypothesis about population parameters based on sampling statistics.

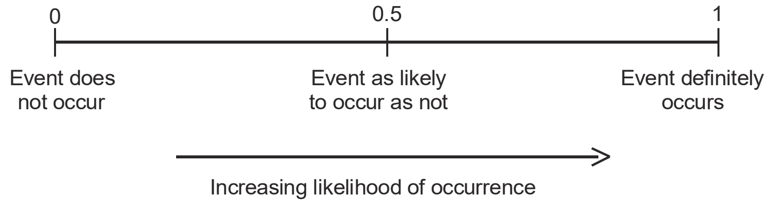
In fact, the standard deviations has such stable mathematical properties that it is used as a standard scale for measuring deviations from the mean. If we are told that the performance of an individual is 10 points better than the mean, it really does not tell us enough, for 10 points may or may not be a large enough difference to be of significance. But if we know that the s for the score is only 4 points, so that on this scale, the performance is 2.5s better than the mean, the statement becomes meaningful. This indicates an extremely good performance. This sigma scale is a very commonly used scale for measuring and specifying deviations which immediately suggest the significance of the deviation.

The only disadvantages of the standard deviation lies in the amount of work involved in its calculation, and the large weight it attaches to extreme values because of the process of squaring involved in its calculations.

NOTES

4.6 PROBABILITY

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 a likelihood where the probability of *zero* indicates that the given event cannot occur and the probability of *one* assures certainty of such an occurrence. For example, if a radio weather report indicates *a near-zero probability of rain*, it can be interpreted as no chance of rain and if a 90 per cent probability of rain is reported, then our understanding is, that the rain is most likely to occur. A 50 per cent, probability or chance of rain indicates that rain is just as likely to occur as not. This likelihood can be shown as follows:



Probability theory provides us with a mechanism for measuring and analysing uncertainties associated with future events. Probability can be subjective or objective. Subjective probability is purely individualistic so that an individual can assign a probability to the outcome of a particular event based upon whatever information regarding this event is available to him along with his personal feelings, experience, judgement and expectations. Two different individuals may assign two different probabilities for the outcome of the same event.

The objective probability of an event, on the other hand, can be defined as the *relative frequency* of its occurrence in the long run. In other words, the probability of an outcome in which we are interested, known as favourable outcome or successful outcome can be calculated as the number of favourable outcomes divided by the total number of outcomes. For example, if (s) defines the number of successful outcomes and (n) is the total number of outcomes, then the probability of a successful outcome is given as (s/n).

Experiment: An experiment is any activity that generates data. For example, tossing of a fair coin is considered as a statistical experiment. An experiment is identified by two properties.

(*i*) Each experiment has several possible outcomes and all these outcomes are known in advance.

(*ii*) None of these outcomes can be predicted with certainty.

For example, while tossing a fair coin, we are not certain whether the outcome will be a head or a tail. Some of the experiments and their possible outcomes can be given as:

Experiment	*Possible Outcomes*
1. Tossing of a fair coin	Head, tail
2. Rolling a die	1, 2, 3, 4, 5, 6
3. Selecting an item from a production lot	Good, bad
4. Introducing a new product	Success, failure

NOTES

4.6.1 Probability Distribution of a Random Variable

An experiment is said to be *random* if we cannot predict the outcome before the experiment is carried out. A random experiment is one which can be repeated, practically or theoretically, any number of times. We can toss a coin or roll a die any number of times and study the outcomes. The result of any outcome may or may not influence that of succeeding outcomes. Thus any throw of a coin or dice is independent of all earlier throws. But if a card is drawn from a deck of cards and not replaced, the experiment made for a second draw will be influenced by the result of the first.

The consideration of the random character of a phenomenon is inevitable because of the complicated characteristics of the laws of nature, ignorance about the relevant laws and the theoretical or practical difficulty allowing for the effect of a large number of factors under consideration or disturbing factors or other shortcomings of our tools of work.

A random variable is a *probability variable.* A specified degree of probability is *attached* to every value of a random variable. What is meant by this statement is that when we conduct an experiment, say by tossing a coin, we cannot definitely say what will come up. Random forces will decide the outcome, although each outcome may have an equal chance of probability of showing up. A random variable is a quantity which in different observations can assume different values.

4.6.2 Axiomatic or Modern Approach to 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

Classical Theory of Probability

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, $p(\text{club}) + p(\text{king}) - p(\text{king of clubs})$.
- 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.

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,

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

NOTES

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 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:

HTHHTTHHHHTTTTHHHHHHHHTTTTTTTTHHHHHHHHHHHHH HHHTTTTTTTTTTTTTTTT...

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.

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.

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 approximately1/6 of the time. The study of empirical probabilities is known as *statistics.*

4.6.3 Theorems on Probability

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

Events [A 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 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:

Event [A] = 50

Event [B] = 30

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,

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

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:

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

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.

4.6.4 Counting Techniques

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 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,

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

NOTES

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 4.34: 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)
Demand Probability		0.2	0.5	0.3
Number of Rooms	$R_1 = (50)$	25	35	50
	$R_2 = (100)$	–10	40	70
	$R_3 = (150)$	–30	20	100

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 4.8.

Table 4.8 *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, 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,

State 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:

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

NOTES

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

NOTES

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 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.

4.6.5 Mean and Variance of Random Variables

Mean

A discrete random variable's mean x_i represents a weighted average of the random variable's possible values. The mean of a random variable, unlike the sample mean of a series of observations, which gives each observation equal weight, weights each result p_i according to its likelihood, μ. The mean (also known as the expected value of X) is denoted by the symbol W, which is technically defined by

$$\mu_x = x_1p_1 + x_2p_2 + \cdots + x_kp_k$$
$$= \sum x_ip_i$$

The mean of a random variable is the variable's long-run average, or the expected average outcome over a large number of observations.

The mean of a continuous random variable is determined by the distribution's density curve. The mean of a symmetric density curve, such as the normal density, is in the curve's centre.

The rule of large numbers asserts that as the number of observations of a random variable grows larger, the observed random mean will always approach the distribution mean μ.

That is, as the number of observations grows larger, the mean of these data approaches the true mean of the random variable. This isn't to say that short-term averages will always reflect the mean.

The total of the means of two random variables, X and Y, is the mean:

$$\mu_{X+Y} = \mu_X + \mu_Y$$

Variance

The variance of a discrete random variable X is defined by and measures the spread, or variability, of the distribution:

$$\sigma_X^2 = \sum (x_i - \mu_x)^2 p_i$$

The square root of the variance is the standard deviation σ.

Properties of Variance

The variance is modified as follows when a random variable *X* is adjusted by multiplying by the value *b* and adding the value *a*:

$$\sigma_{a+bX}^2 = b^2\sigma^2$$

The value *a* is ignored since adding or removing a constant has no effect on the spread of the distribution. Because the variance is a sum of squared terms, any multiplier value *b* used to alter the variance must also be squared.

The variance of their total or difference is the sum of their variances for independent random variables *X* and *Y*:

$$\sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2$$
$$\sigma_{X-Y}^2 = \sigma_X^2 + \sigma_Y^2$$

Because the variation in each variable adds to the variation in each case, variances are added for both the sum and difference of two independent random variables. Variability in one variable is connected to variability in the other if the variables are not independent. As a result, applying the preceding formula to compute the variance of their total or difference may not be possible.

Assume that variable *X* represents the amount of money (in dollars) spent on lunch by a group of people, and variable *Y* represents the amount of money spent on supper by the same group of people. Because *X* and *Y* are not considered independent variables, the variance of the sum *X* + *Y* cannot be calculated as the sum of the variances.

4.7 STANDARD PROBABILITY DISTRIBUTION

Once the random variable of interest is defined and the probabilities are assigned to all its values, it is called a probability distribution. Table 4.9 showns the probability distribution for various sales levels (sales level being the random variable represented as *X*) for a new product as stated by the sales manager:

NOTES

Table 4.9 *Probability Distribution for Various Sales Levels*

Sales (in units) X_i	*Probability* $pr.\ (X_i)$
X_1 50	0.10
X_2 100	0.30
X_3 150	0.30
X_4 200	0.15
X_5 250	0.10
X_6 300	0.05
Total	1.00

Sometimes, the probability distribution may be presented in the form called a cumulative probability distribution. The probability distribution given in Table 4.9 can also be presented in the form of cumulative probability distribution as in Table 4.10.

Table 4.10 *Cumulative Probability Distribution*

Sales (in units) $(X_i) \propto$	*Probability* $pr\ (X_i)$	*Cumulative Probabilities* $pr\ (X_i \leq \propto)$
X_1 50	0.10	0.10
X_2 100	0.30	0.40
X_3 150	0.30	0.70
X_4 200	0.15	0.85
X_5 250	0.10	0.95
X_6 300	0.05	1.00

The meaning of probability distribution can be made more clear if you remember the following:

- An *observed frequency distribution* (often called simply as a frequency distribution) is a listing of the observed frequencies of all the outcomes of an experiment that actually occurred while performing the experiment.
- A *probability distribution* is a listing of the probabilities of all the possible outcomes that could result if the experiment is performed. The assignment of probabilities may be based either on theoretical considerations or it may be a subjective assessment or may be based on experience.
- A *theoretical frequency distribution* is a probability distribution that describes how outcomes are expected to vary. In other words, it enlists the expected values (i.e., observed values multiplied by corresponding probabilities) of all the outcomes.

Types of Probability Distributions

Probability distributions can be classified as either discrete or continuous. In a *discrete probability distribution*, the variable under consideration is allowed to take only a limited number of discrete values along with corresponding probabilities. The two important discrete probability distributions are: the binomial probability distribution and the Poisson probability distribution. In a *continuous probability distribution*, the variable under consideration is allowed to take on any value within a given range. Important continuous probability distributions are exponential probability distribution and normal probability distribution.

Important disrete and continuous probability distributions are discussed later in this unit.

Probability Functions

In probability distribution, it is not always necessary to calculate probabilities for each and every outcome in the sample space. There exist many mathematical formulae for many commonly encountered problems which can assign probabilities to the values of random variables. Such formulae are generally termed as probability functions. In fact, a probability function is a mathematical way of describing a given probability distribution. To select a suitable probability function that best fits in the given situation, you should work out the values of its parameters. Once you have worked out the values of the parameters, you can then assign the probabilities, if required, using the appropriate probability function to the values of random variables. Various probability functions will be explained shortly while describing the various probability distributions.

Techniques of Assigning Probabilities

You can assign probability values to the random variables. Since the assignment of probabilities is not an easy task, you should observe certain rules in this context as follows:

(*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:

(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. But at times the use of this technique gives ridiculous conclusions. For example, you have to assign probability to the event that a person of age 35 will live up to 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.

NOTES

(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.

4.7.1 Binomial Distribution

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 the two mutually exclusive outcomes (often denoted as 'Success' and 'Failure') can result on each trial.

The 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, e.g., '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 follows:

$$f(X=r) = {}^nC_r p^r q^{n-r}$$

$$r = 0, 1, 2, \ldots, n$$

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

NOTES

Parameters of Binomial Distribution

Binomial 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 the definition of the problem. The probability of the happening of exactly r events in n trials can be found out using the previously stated binomial function.

The value of p also determines the general appearance of the binomial distribution, if shown graphically. In this context, the usual generalizations are as follows:

(i) When p is small (say 0.1), the binomial distribution is skewed to the right, i.e., the graph takes the form as shown in Figure 4.4.

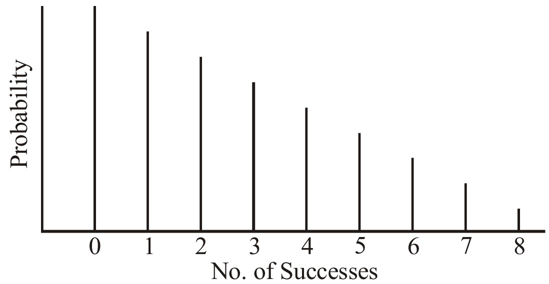


Fig. 4.4 *Bonimial Distribution Skewed to the Right*

(ii) When p is equal to 0.5, the binomial distribution is symmetrical and the graph takes the form as shown in Figure 4.5.

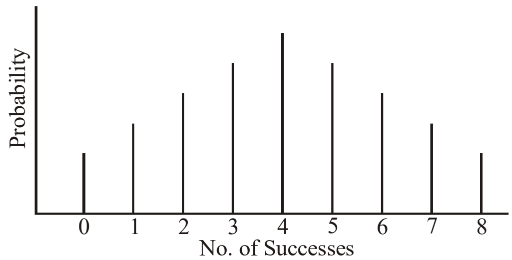


Fig. 4.5 *Symmetrical Binomial Distribution*

(iii) 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 4.6.

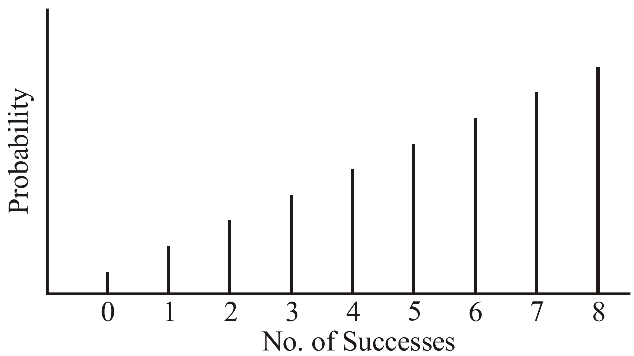


Fig. 4.6 *Bonimial Distribution Skewed to the Left*

NOTES

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 4.7.

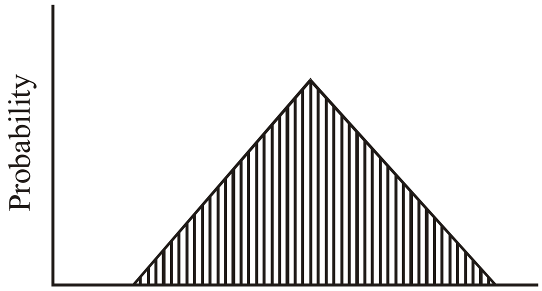


Fig. 4.7 *The Bell-Shaped Binomial Distribution*

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 np and the variance of random variable is equal to npq or $np\,(1-p)$. Accordingly, the standard deviation of binomial distribution is equal to $\sqrt{npq}$. The other important measures relating to binomial distribution are as under:

$$\text{Skewness} = \frac{1-2p}{\sqrt{npq}}$$

$$\text{Kurtosis} = 3 + \frac{1-6p+6q^2}{npq}$$

When to use binomial distribution

The use of binomial distribution is most appropriate in situations fulfilling the previously stated conditions. Two such situations, for example, can be described as follows:

(i) When you have to find the probability of 6 heads in 10 throws of a fair coin.

(ii) When you 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 4.35: 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: $\bar{X} = np$ and variance: $\sigma^2 = npq$.

Solution: Since the coin is fair and so, when thrown, it can come either with head upwards or tail upwards. Hence, p (head) $= \frac{1}{2}$ and q (no head) $= \frac{1}{2}$. The required probability function is:

$$f(X = r) = {}^nC_r\, p^r q^{n-r}$$
$$r = 0, 1, 2, ..., 10$$

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	${}^{10}C_0 p^0 q^{10}$	=	$1/1024^6$	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
			$\Sigma\bar{X}$ = $\bar{X}$ = 5	5120/1024		Variance = σ^2 = $\Sigma\ (X_i - \bar{X})^2 pr_i$ =	2560/1024 = 2.5

The mean of the binomial distribution is given by $np = 10 \times \frac{1}{2} = 5$ and the variance of this distribution is equal to $npq = 10 \times \frac{1}{2} \times \frac{1}{2} = 2.5$

These values are exactly the same as we have found them in the preceding table.

Hence, these values stand verified with the calculated values of the two measures as shown in the table.

Fitting a binomial distribution

When a binomial distribution is to be fitted to the given data, the following procedure is adopted:

(i) Determine the values of 'p' and 'q' keeping in view that $\bar{X} = np$ and $q = (1 - p)$.

(ii) Find the probabilities for all possible values of the given random variable applying the binomial probability function, namely

$$f(X_i = r) = {}^nC_r p^r q^{n-r}$$
$$r = 0, 1, 2, \ldots, n$$

(iii) Work out the expected frequencies for all values of random variable by multiplying N (the total frequency) with the corresponding probability as worked out in case (*ii*).

The expected frequencies so calculated constitute the fitted binomial distribution to the given data.

4.7.2 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. It is frequently used in the context of operations research, and, for this reason, has a great significance for management people. It plays an important role in *queuing* theory, inventory control problems and 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, especially 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 (namely np) 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 the number of occurrences of a given event during a given interval (interval may be time, distance, area, etc.). You 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:

$$f(X_i = x) = \frac{\lambda^i e^{-\lambda}}{x}$$

$$x = 0,\ 1,\ 2,\ldots$$

where λ = Average number of occurrences per specified interval. 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.

The Poisson process

The poisson distribution applies in the case of the Poisson process which has the 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 you divide the given interval into very very small intervals you will find the following:
  (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 with in a very small interval is so small that you 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

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 in cases when you do not know the value of '*n*' or when '*n*' cannot 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, if 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 you watch carefully, you 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 you use Poisson distribution. The number of deaths 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, etc. are a few other examples where Poisson probability distribution is generally used.

Example 4.36: 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 4 defects.

Solution: If the product has 2 defects per unit of product inspected. Hence, λ = 2.

Poisson probability function is as follows:

$$f\left(X_i = x\right) = -\frac{\lambda^x e^{-\lambda}}{x}$$

$$x = 0, 1, 2, \ldots$$

Using this probability function, you will find the required probabilities as follows:

$$P(\text{without any defects, i.e., } x = 0) = \frac{2^0 e^{-2}}{0}$$

$$= \frac{1.(0.13534)}{1} = 0.13534$$

$$P(\text{with 3 defects, i.e., } x = 3) = \frac{2^3 e^{-2}}{3} = \frac{2 \times 2 \times 2(0.13534)}{3 \times 2 \times 1}$$

NOTES

$$= \frac{0.54136}{3} = 0.18045$$

$$P(\text{with 4 defects, i.e., } x = 4) = \frac{2^4 e^{-2}}{4} = \frac{2 \times 2 \times 2 \times 2(0.13534)}{4 \times 3 \times 2 \times 1}$$

$$= \frac{0.27068}{3} = 0.09023$$

Fitting a Poisson distribution

When a Poisson distribution is to be fitted to the given data, then the following procedure is adopted:

(i) Determine the value of λ, the mean of the distribution.

(ii) Find the probabilities for all possible values of the given random variable using the Poisson probability function, namely

$$f(X_i = x) = \frac{\lambda^x . e^{-\lambda}}{x}$$

$$x = 0, 1, 2, \ldots$$

(iii) Work out the expected frequencies as follows:

$$np(X_i = x)$$

The result of case (iii) 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 infinity, and p is small, approaching 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 the cases where these two conditions are fulfilled, you can use mean of the binomial distribution (namely np) in place of the mean of Poisson distribution (namely λ) so that the probability function of Poisson distribution becomes as follows:

$$f(X_i = x) = \frac{(np)^x e^{-(np)}}{x}$$

You can explain Poisson distribution as an approximation of the binomial distribution with the help of following example.

Example 4.37: Given are 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.

What is the probability that exactly three 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 np in place of λ)
(since $n \geq 20$ and $p \leq 0.05$)

$$f(X_i = x) = \frac{(np)^x e^{-np}}{x}$$

Where x means number of machines becoming out of order on the same day.

$$P(X_i = 3) = \frac{(20 \times 0.02)^3 e^{-(20 \times 0.02)}}{3}$$

$$= \frac{(0.4)^3 .(0.67032)}{3 \times 2 \times 1} = \frac{(0.064)(0.67032)}{6}$$

$$= 0.00715$$

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 three 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, you can state that in the given case Poisson distribution appears to be a good approximation of binomial distribution.

Example 4.38: 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: Given:

$$n = 100 \text{ and } p = 0.1$$

$$\therefore \quad \lambda = n.p = 100 \times 0.1 = 10$$

To find the required probability, the Poisson probability function can be used as an approximation to the binomial probability function as follows:

$$f(X_i = x) = \frac{\lambda^x e^{-\lambda}}{x} = \frac{(np)^x e^{-(np)}}{x}$$

or

$$P(5)^7 = \frac{10^5 e^{-10}}{5} = \frac{(100000)(0.00005)}{5 \times 4 \times 3 \times 2 \times 1} = \frac{5.00000}{5 \times 4 \times 3 \times 2 \times 1}$$

$$= \frac{1}{24} = 0.042$$

4.7.3 Exponential Distribution

Exponential probability distribution is the probability distribution of time (say t), between events and as such it is continuous probability distribution concerned with the continuous random variable that takes on any value between zero and

NOTES

positive infinity. In the exponential distribution, you often ask the question: What is the probability that it will take x trials before the first occurrence? This distribution plays an important role in describing a large class of phenomena, particularly in the area of reliability theory and in *queuing* models.

The probability function of the exponential distribution is as follows:

$$F(x) = \mu e^{-\mu x} \; x \geq 0$$

where μ = The average length of the interval between two occurrences[10]

e = 2.7183 being the basis of natural logarithms

The only parameter of the exponential distribution is μ.

The expected value or mean of the exponential distribution is $1/\mu$ and its variance is $1/\mu$.[11]

The cumulative distribution (less than type) of the exponential is

$$F(x) = P(X \leq x) = 1 - e^{-\mu x}, \; x \geq 0$$
$$= 0, \text{ elsewhere}$$

Thus, for instance, the probability that $x \leq 2$ is $1 - e^{-2\mu}$.

Example 4.39: In an industrial complex, the average number of fatal accidents per month is one-half. The number of accidents per month is adequately described by a Poisson distribution. What is the probability that four months will pass without a fatal accident?

Solution: You have been given that the average number of fatal accidents per month is one-half and the number of accidents per month is well described by a Poisson distribution.

Hence, $\lambda = 0.5$

$\therefore$ The average length of the time interval between two accidents $= \dfrac{1}{\lambda} = \dfrac{1}{0.5} = 2$ months, assuming exponential distribution.

Now, by using the cumulative distribution of the exponential, we can find the required probability that four months will pass without a fatal accident (i.e. $x > 4$) as follows:

$$\because \quad F(x) = P(X \leq x) = 1 - e^{-\mu x}$$

$$\therefore \quad P(X > x) = e^{-\mu x}$$

$$\therefore \quad P(X > x) = e^{-2(4)} = e^{-8} = 0.00034$$

Thus, 0.00034 is the required probability that 4 months will pass without a fatal accident.

4.7.4 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, the 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, you get a curve normal probability smooth and symmetrical. Such curves are called normal probability curves (or at times known as normal curves of error) and represent the normal distributions.

The probability function in the case of normal probability distribution[13] is given as follows:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

where μ = The mean of the distribution

σ^2 = Variance of the distribution

The normal distribution is thus defined by two parameters, namely μ and σ^2. This distribution can be represented graphically (Refer Figure 4.8).

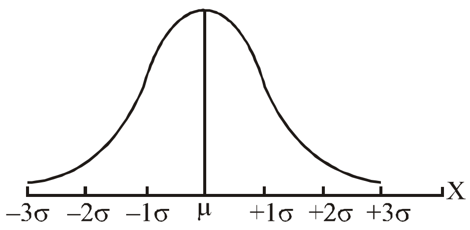


Fig. 4.8 *Curve Representing Normal Distribution*

Characteristics of Normal Distribution

The characteristics of the normal distribution or those of a normal curve are as follows:

(i) It is symmetric distribution.

(ii) 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 per cent 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.

(iii) The curve is asymptotic to the baseline which means that it continues to approach but never touches the horizontal axis.

(iv) The variance (σ^2) defines the spread of the curve.

(v) Area enclosed between mean ordinate and an ordinate at a distance of one standard deviation from the mean is always 34.134 per cent of the total

NOTES

area of the curve. It means that the area enclosed between two ordinates at one Sigma Distance (SD) from the mean on either side would always be 68.268 per cent of the total area. This is shown in Figure 4.9.

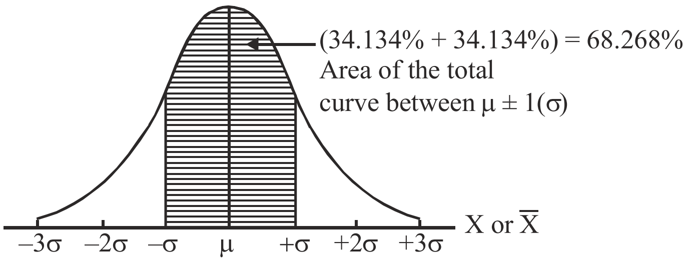


Fig. 4.9 *An Area Enclosed between Two Ordinates at One SD*

Similarly, the other area relationships are as given in Table 4.11.

Table 4.11 *Area Relationships*

Between		*Area Covered to Total Area of the Normal Curve*[15]
$\mu \pm 1$	SD	68.27%
$\mu \pm 2$	SD	95.45%
$\mu \pm 3$	SD	99.73%
$\mu \pm 1.96$	SD	95%
$\mu \pm 2.578$	SD	99%
$\mu \pm 0.6745$	SD	50%

(vi) The normal distribution has only one mode since the curve has a single peak. In other words, it is always a unimodal distribution.

(vii) The maximum ordinate divides the graph of normal curve into two equal parts.

(viii) 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 or curves

You can have several normal probability distributions but each particular normal distribution is being defined by its two parameters, namely the mean (μ) and the standard deviation (σ). There is, thus, not a single normal curve but rather a family of normal curves. Figures 4.10–4.12 exhibit some of these normal curves:

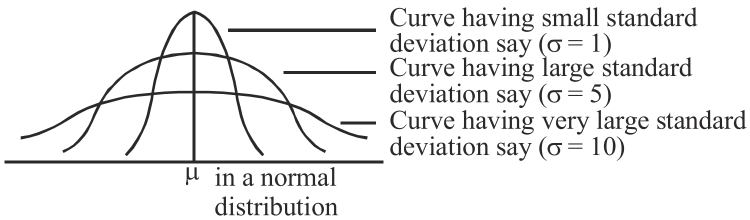


Fig. 4.10 *Normal Curves with Identical Means but Different Standard Deviations*

NOTES

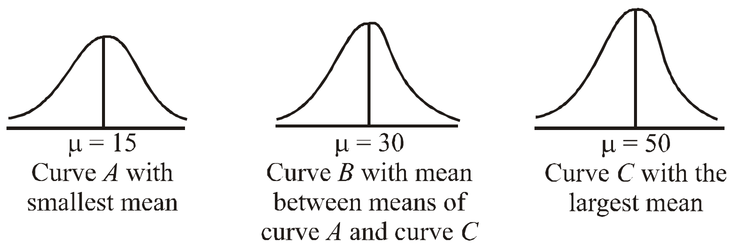


***Fig. 4.11** Normal Curves with Identical Standard Deviation but Each with Different Means*

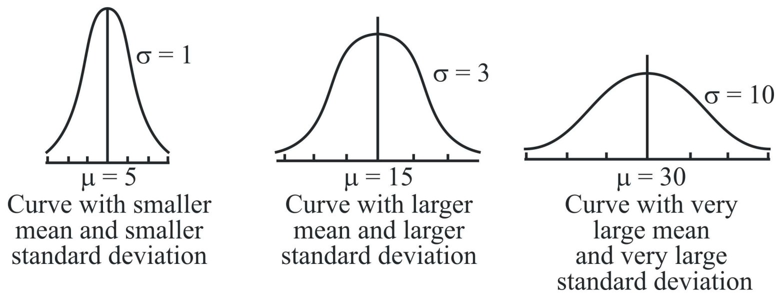


***Fig. 4.12** Normal Curves Each with Different Standard Deviations and Different Means*

How to measure the area under the normal curve?

You have learned about the 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? You can make use of the statistical tables constructed by mathematicians for the purpose. Using these tables, you 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, you 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)

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.

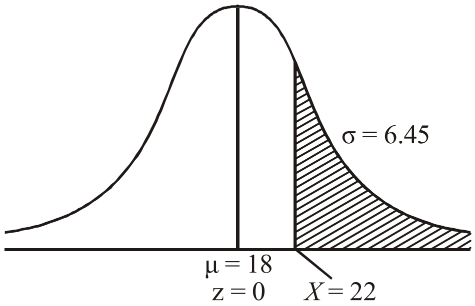
NOTES

Example 4.40: 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 questions:

(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, you are interested in the area of the portion of the normal curve as shaded and shown in the following diagram:

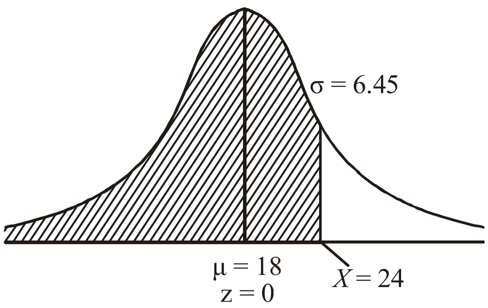


Calculate Z as under:

$$Z = \frac{X-\mu}{\sigma} = \frac{22-18}{6.45} = 0.62$$

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, you are interested in the area of the portion of the normal curve as shaded and shown in the following figure:



Calculate Z as under:

$$Z = \frac{24-18}{6.45} = 0.93$$

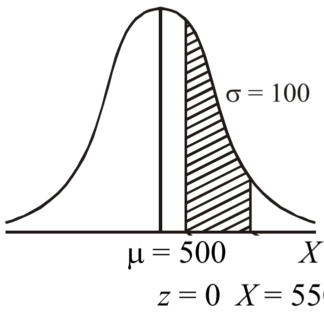
The value from the concerning table, when $Z = 0.93$, is 0.3238 which refers to the area of the curve between $\mu = 18$ and $X = 24$. The area of the entire left hand portion of the curve is 0.5 as usual.

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 4.41: Regarding a certain normal distribution concerning the income of the individuals, you 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, you are interested in the area of the portion of the normal curve as shaded and shown in the following figure:



For finding the area of the curve between $X = 550$ to 650, 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

$$Z = \frac{650 - 500}{100} = \frac{150}{100} = 1.5$$

Corresponding to which the area between $\mu = 500$ and $X = 650$ in the curve as per table is equal to 0.4332

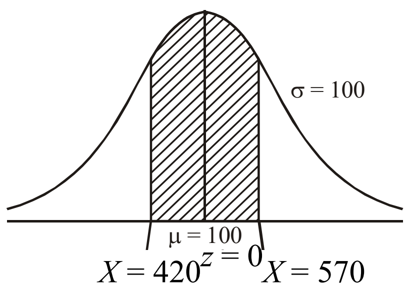
Hence, the area of the curve that lies between $X = 550$ and $X = 650$ is

$$(0.4332) - (0.1915) = 0.2417$$

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, you are interested in the area of the portion of the normal curve as shaded and shown in the following figure:

To find the area of the shaded portion we make the following calculations:



NOTES

$$Z = \frac{570 - 500}{100} = 0.70$$

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.

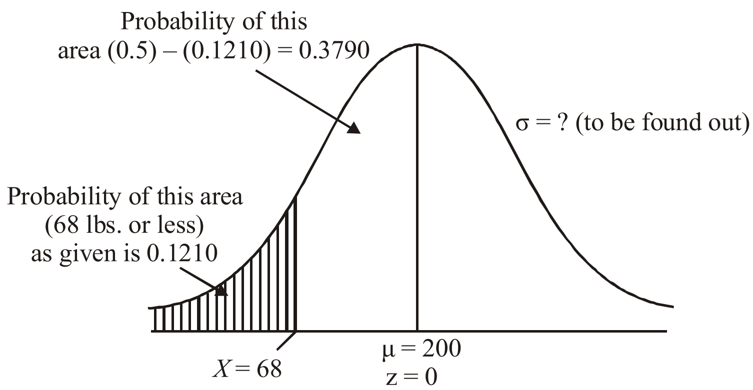
Hence, the required area in the curve between X = 420 and X = 570 is:

$$(0.2580) + (0.2881) = 0.5461$$

This is the required probability that an individual selected at random will belong to income group of Rs 420 to Rs 570.

Example 4.42: 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.

Solution: Given information can be depicted in a normal curve as shown in the following figure:



If the probability of the area falling within μ = 200 and X = 68 is 0.3790 as stated above, the corresponding value of Z as per the table 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, you can write:

$$Z = \frac{X - \mu}{\sigma}$$

or $$-1.17 = \frac{68 - 200}{\sigma}$$

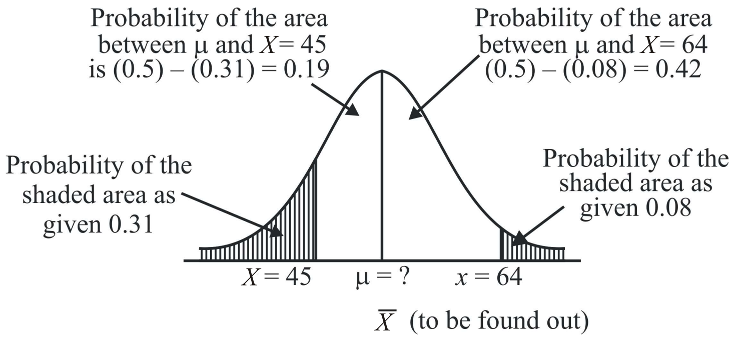
or $$-1.17\sigma = -132$$

or $$\sigma = 112.8 \text{ lbs approx.}$$

Thus, the required standard deviation is 112.8 lbs approximately.

Example 4.43: In a normal distribution, 31 per cent items are below 45 and 8 per cent are above 64. Find the $\bar{X}$ and σ of this distribution.

Solution: You can depict the given information in a normal curve as follows:



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, you are in the left portion of the curve so we can express this as under,

$$-0.50 = \frac{45 - \mu}{\sigma} \quad (1)$$

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, you are in the right portion of the curve, so you can express this as under:

$$1.41 = \frac{64 - \mu}{\sigma} \quad (2)$$

If you solve Equations (1) and (2) above to obtain the value of μ or $\bar{X}$, you have:

$$-0.5\,\sigma = 45 - \mu \quad (3)$$

$$1.41\,\sigma = 64 - \mu \quad (4)$$

By subtracting the Equation (4) from (3), you have:

$$-1.91\,\sigma = -19$$

$$\therefore \quad \sigma = 10$$

Putting σ = 10 in Equation (3), you have:

$$-5 = 45 - \mu$$

$$\therefore \quad \mu = 50$$

Hence, $\bar{X}$ (or μ)=50 and σ =10 for the concerning normal distribution.

4.7.5 Uniform Distribution (Discrete Random and Continous Variable)

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.

NOTES

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.

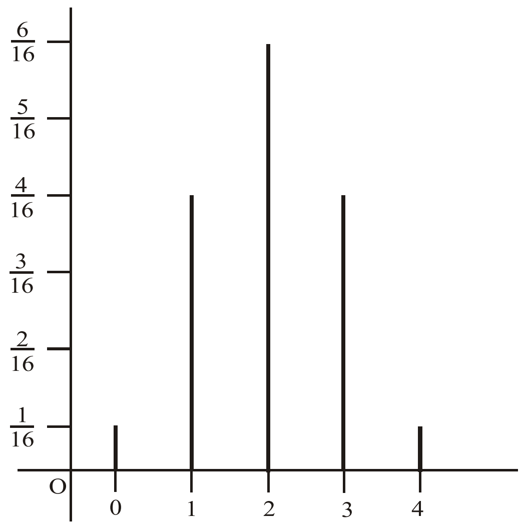
$$p(X = 0) = \left(\frac{1}{2}\right)^4 = \frac{1}{16}$$

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

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

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

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



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

This is a discrete probability distribution (Refer Example 4.44).

Example 4.44: 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	$2a$

NOTES

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

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

Discrete Distributions

There are several discrete distributions. Some other discrete distributions are described as follows:

(i) 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 values it is a case of a discrete variable where,

NOTES

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, then

(i) The probability of the success of the last trial is p.

(ii) 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 (i) and (ii) happening together is,

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

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

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

This can be summed up as,

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 be the Poisson distribution.

(ii) 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, then

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

We get the geometric distribution,

$$p(x) = {}^{x}C_0 p^1 q^x = pq^x$$

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

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

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

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

Refer Example 4.45 to understand it better.

Example 4.45: 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,

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

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

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

Since $p = 1 - q$.

(iii) 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 discussed case, is called the hypergeometic 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.

It can be shown that,

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

NOTES

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

Example 4.46: 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 of no prize $p(0) = \dfrac{{}^{3}C_0\,{}^{17}C_5}{{}^{20}C_5}$

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

Example 4.47: Examine the nature of the distibution if 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,

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

(iv) 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, $\displaystyle\sum_{i=1}^{k} n_i = n$

Characteristic Features of the Binomial Distribution

The following are the characteristics of binomial distribution:

(i) It is a discrete distribution.

(ii) It gives the probability of x successes and $n-x$ failures in a specific order.

(iii) The experiment consists of n repeated trials.

(iv) Each trial results in a success or a failure.

(v) The probability of success remains constant from trial to trial.

(vi) The trials are independent.

(vii) The success probability p of any outcome remains constant over time. This condition is usually not fully satisfied in situations involving management and economics, e.g., 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 mind set as used for the first informant.

(viii) The binomial distribution depends on two parameters, n and p. Each set of different values of n, p has a different binomial distribution.

(ix) 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.

(x) The binomial distribution can be approximated by the normal. As n becomes large and p is close to 0.5, the approximation becomes better.

Through the following examples you can understand multinomial better.

Example 4.48: 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 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 4.49: 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, $$\Sigma p(x) = 1, 0 + a + \frac{a}{2} + \frac{a}{2} + \frac{a}{4} + \frac{a}{4} = 1$$

NOTES

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

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

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

Example 4.50: 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 4.51: 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 4.52: 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	$p(1) = e^{-m} \dfrac{m}{1!} = p(0).m$

NOTES

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 4.53: 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 $\mu'_3 = m^3 + 3m^2 + m$

(v) 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 $p(z_1 \le X \le z_2)$.

This probability corresponds to the shaded area under the curve in Figure 4.13.

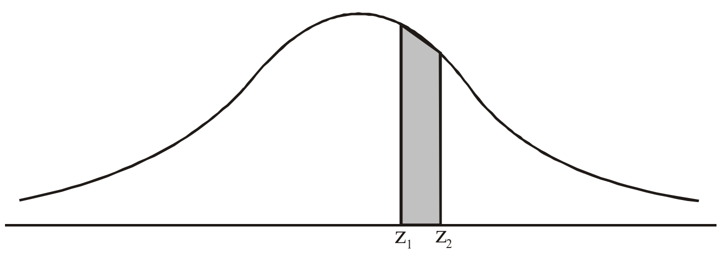


Fig. 4.13 Continuous Probability Distribution

NOTES

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

11. What is absolute measure of dispersion?
12. What is relative measure of dispersion?
13. Define range.
14. Calculate standard deviation for the series 1, 2, 3, 5, 7.
15. Define the terms 'Simple Probability' and 'Joint Probability'.
16. What are the types of probability distributions?
17. Write the probability function of binomial distribution.
18. What are the different parameters of binomial distribution?
19. Under what circumstances would you use binomial distribution?
20. What is Poisson distribution?
21. What is the use of exponential distribution?
22. Define a normal distribution.
23. When is a distribution said to be symmetrical?

4.8 ANSWERS TO 'CHECK YOUR PROGRESS'

1. The term statistics is used to mean either statistical data or statistical method. When it is used in the sense of statistical data it refers to quantitative aspects of things, and is a numerical description.
2. The procedure of classification brings into relief the salient features of the variable that is under investigation. This can be clearly illustrated by an example. If we are given the marks in mathematics of each individual student of a class and if it is desired to judge the performance of the class on the basis of these data it will not be an easy matter. Human mind has its limitations and cannot easily grasp a multitude of figures. But if the students are classified

i.e., if we put into one group all those boys who get more than second division marks, in still another group those who get third division marks, and have a separate group of those who fait to get pass marks, it will be easier for us to form a more precise idea about the performance of the class.

3. Collection of facts is the first step in the statistical treatment of a problem. Numerical facts are the raw materials upon which the statistician is to work and just as in a manufacturing concern the quality of a finished product depends, inter alia, upon the quality of the raw material, in the same manner, the validity of statistical conclusions will be governed, among other considerations, by the quality of data used.
4. In statistics, the term central tendency specifies the method through which the quantitative data have a tendency to cluster approximately about some value. A measure of central tendency is any precise method of specifying this 'central value'.
5. Arithmetic mean is also commonly known as the mean. Even though average, in general, means measure of central tendency, when we use the word average in our daily routine, we always mean the arithmetic average. The term is widely used by almost everyone in daily communication.
6. The weighted arithmetic mean is particularly useful where we have to compute the mean of means. If we are given two arithmetic means, one for each of two different series, in respect of the same variable, and are required to find the arithmetic mean of the combined series, the weighted arithmetic mean is the only suitable method of its determination.
7. Median is that value of a variable which divides the series in such a manner that the number of items below it is equal to the number of items above it. Half the total number of observations lie below the median, and half above it. The median is thus a positional average.
8. Mode is that value of the variable which occurs or repeats itself the greatest number of times. The mode is the most 'Fashionable' size in the sense that it is the most common and typical, and is defined by Zizek as 'The value occurring most frequently in a series (or group of items) and around which the other items are distributed most densely'.
9. The four important methods of estimating mode of a series are: (i) Locating the most frequently repeated value in the array; (ii) Estimating the mode by interpolation; (iii) Locating the mode by graphic method; and (iv) Estimating the mode from the mean and the median.
10. Some measures, other than the measures of central tendency, are often employed when summarizing or describing a set of data where it is necessary to divide the data into equal parts. These are positional measures and are called quantiles and consist of quartiles, deciles and percentiles. The quartiles divide the data into four equal parts. The deciles divide the total ordered data into ten equal parts and the percentiles divide the data into 100 equal parts. Consequently, there are three quartiles, nine deciles and 99 percentiles.

NOTES

11. Absolute measure of dispersion states the actual amount by which an item on an average deviates from a measure of central tendency.
12. Relative measure of dispersion is a quotient computed by dividing the absolute measures by a quantity in respect to which absolute deviation has been computed.
13. The range of a set of numbers is the difference between the maximum and minimum values. It indicates the limits within which the values fall.
14. 2.15
15. Simple Probability: 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.

    Joint Probability: 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.
16. There are two types of probability distributions, discrete and continuous probability distributions. In discrete probability distribution, the variable under consideration is allowed to take only a limited number of discrete values along with corresponding probabilities. On the other hand, in a continuous probability distribution, the variable under consideration is allowed to take on any value within a given range.
17. The probability function of binomial distribution is written as follows:

    $$f(X = r) = {}^nC_r p^r q^{n-r}$$

    $$r = 0, 1, 2, \ldots, n$$

    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
18. 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.
19. The use of binomial distribution is needed 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 per cent defective items on average, will be defective.
20. 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.

21. Exponential distribution is used for describing a large class of phenomena, particularly in the area of reliability theory and in queuing models.
22. Normal distribution is the most important and frequently used continuous probability distribution among all the probability distributions. This is so because this distribution fits well 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.
23. If $p = 0.5$, the distribution is symmetrical.

4.9 SUMMARY

- Statistics influence the operations of business and management in many dimensions.
- Statistical applications include the area of production, marketing, promotion of product, financing, distribution, accounting, marketing research, manpower planning, forecasting, research and development, and so on.
- In statistics, the term central tendency specifies the method through which the quantitative data have a tendency to cluster approximately about some value.
- A measure of central tendency is any precise method of specifying this 'central value'. In the simplest form, the measure of central tendency is an average of a set of measurements, where the word average refers to as mean, median, mode or other measures of location. Typically the most commonly used measures are arithmetic mean, mode and median.
- While arithmetic mean is the most commonly used measure of central location, mode and median are more suitable measures under certain set of conditions and for certain types of data.
- There are several commonly used measures, such as arithmetic mean, mode and median. These values are very useful not only in presenting the overall picture of the entire data, but also for the purpose of making comparisons among two or more sets of data.
- Arithmetic mean is also commonly known as the mean. Even though average, in general, means measure of central tendency, when we use the word average in our daily routine, we always mean the arithmetic average. The term is widely used by almost everyone in daily communication.
- The weighted arithmetic mean is particularly useful where we have to compute the mean of means. If we are given two arithmetic means, one for each of two different series, in respect of the same variable, and are required to find the arithmetic mean of the combined series, the weighted arithmetic mean is the only suitable method of its determination.

NOTES

- Median is that value of a variable which divides the series in such a manner that the number of items below it is equal to the number of items above it. Half the total number of observations lie below the median, and half above it. The median is thus a positional average.
- The median of ungrouped data is found easily if the items are first arranged in order of the magnitude. The median may then be located simply by counting, and its value can be obtained by reading the value of the middle observations.
- The median can quite conveniently be determined by reference to the ogive which plots the cumulative frequency against the variable. The value of the item below which half the items lie, can easily be read from the ogive.
- Median is a positional average and hence the extreme values in the data set do not affect it as much as they do to the mean.
- The mode of a distribution is the value at the point around which the items tend to be most heavily concentrated. It is the most fre-quent or the most common value, provided that a sufficiently large number of items are available, to give a smooth distribution.
- The measures of dispersion bring out this inequality. In engineering problems too the variability is an important concern.
- The amount of variability in dimensions of nominally identical components is critical in determining whether or not the components of a mass-produced item will be really interchangeable.
- 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 a 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 provides us with a mechanism for measuring and analysing uncertainties associated with future events. Probability can be subjective or objective.
- The objective probability of an event, on the other hand, can be defined as the relative frequency of its occurrence in the long run.
- Binomial distribution is probably the best known of discrete distributions. The normal distribution, or Z-distribution, is often used to approximate the binomial distribution.
- If the sample size is very large, the Poisson distribution is a philosophically more correct alternative to binomial distribution than normal distribution.
- One of the main differences between the Poisson distribution and the binomial distribution is that in using the binomial distribution all eligible phenomena are studied, whereas in the Poisson, only the cases with a particular outcome are studied.
- Exponential distribution is a very commonly used distribution in reliability engineering. The reason for its widespread use lies in its simplicity, so much that it has even been employed in cases to which it does not apply directly.

- Amongst all types of distributions, the normal probability distribution is by far the most important and frequently used distribution because it fits well in many types of problems.

NOTES

4.10 KEY TERMS

- **Statistics:** Numerical statements of facts in any department of inquiry placed in relation to each other.
- **Median:** Measure of central tendency and it appears in the centre of an ordered data.
- **Mode:** A form of average that can be defined as the most frequently occurring value in the data.
- **The weighted arithmetic mean:** The weighted arithmetic mean is particularly useful where we have to compute the mean of means.
- **Mean deviation:** The mean deviation of a series of values is the arithmetic mean of their absolute deviations.
- **Standard deviation:** The square root of the average of the squared deviations from their mean of a set of observations.
- **Range:** The difference between the maximum and minimum values of a set of number. It indicates the limits within which the values fall.
- **Classical theory of probability:** It is the theory of probability based on the number of favourable outcomes and the number of total outcomes.
- Binomial distribution: It is also called the Bernoulli process and is used to describe a discrete random variable.
- **Poisson distribution:** It is used to describe the empirical results of past experiments relating to the problem and plays an important role in queuing theory, inventory control problems and risk models.
- **Exponential distribution:** It is a continuous probability distribution and is used to describe the probability distribution of time between two events.
- **Normal distribution:** It is referred to as most important and frequently used continuous probability distribution as it fits well in many types of problems.

4.11 SELF-ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. State the significance of statistical methods.
2. How does statistics aid in interpreting conditions?
3. List the various characteristics of statistical data.
4. Write a short note on the origin of statistics.

5. Define the term arithmetic mean.
6. Differentiate between a mean and a mode.
7. Write three characteristics of mean.
8. What is the importance of arithmetic mean in statistics?
9. Define the term median with the help of an example.
10. Differentiate between geometric and harmonic mean.
11. Define the terms quartiles, percentiles and deciles.
12. Write the definition and formula of quartile deviation.
13. How will you calculate the mean deviation of a given data?
14. What is standard deviation? Why is it used in statistical evaluation of data?
15. Define the concept of probability.
16. What are the different theories of probability? Explain briefly.
17. Define probability distribution and probability functions.
18. What do you mean by the binomial distribution and its measures?
19. How can a binomial distribution be fitted to given data?
20. How will define the Poisson distribution and its important measures?
21. Poisson distribution can be an approximation of binomial distribution. Define.
22. When is the Poisson distribution used?
23. What is exponential distribution?
24. Define any six characteristics of normal distribution.
25. Write the formula for measuring the area under the curve.
26. How will you define the circumstances when the normal probability distribution can be used?
27. What is CPF?

Long-Answer Questions

1. Give a detailed description on the various functions of statistics.
2. Describe the various features of the statistical procedure.
3. According to Bowley statistics is 'The science of counting'. Do you agree? Give reasons.
4. An elevator is designed to carry a maximum load of 3200 pounds. If 18 passengers are riding in the elevator with an average weight of 154 pounds, is there any danger that the elevator might be overloaded?
5. In a car assembly plant, the cars were diagnostically checked after assembly and before shipping them to the dealers. All such cars with any defect were returned for extra work. The number of such defective cars returned in one day of a 16-days period is given as follows:

   30, 34, 10, 16, 28, 9, 22, 2, 6, 23, 25, 10, 15, 10, 8, 24

NOTES

(*i*) Find the average number of defective cars returned for extra work per day.

(*ii*) Find the median for defective cars per day.

(*iii*) Find the mode for defective cars per day.

(*iv*) Find Q_2.

(*v*) Find D_2.

(*vi*) Find P_{70}.

6. Calculate mean deviation and its coefficient about median, arithmetic mean and mode for the following figures, and show that the mean deviation about the median is least.

   (103, 50, 68, 110, 108, 105, 174, 103, 150, 200, 225, 350, 103)

7. A group has $\sigma = 10$, $N = 60$, $\sigma^2 = 4$. A subgroup of this has $\overline{X}_1 = 11$, $N_1 =$ 4 0 , $\sigma_1^2 = 2.25$. Find the mean and the standard deviation of the other subgroup.

8. The following are some of the particulars of the distribution of weights of boys and girls in a class

	Boys	*Girls*
Number	100	50
Mean weight	60 kg	45 kg
Variance	9	4

(*i*) Find the standard deviation of the combined data.

(*ii*) Which of the two distributions is more variable?

9. Find the Q.D. and coefficient of Q.D. for the following data:

Marks	*No. of Students*
35 – 40	4
40 – 45	8
45 – 50	12
50 – 55	7
55 – 60	2

10. Find the Q.D. from the mean for the series 5, 7, 10, 12, 6.

11. (*i*) Calculate the mean deviation from the mean for the following data. What light does it throw on the social conditions of the community? Data showing differences in ages of husbands and wives.

Difference in years	*Frequency*
0 – 5	499
5 – 10	705
10 – 15	507
15– 20	281
20 – 25	109
25 – 30	52
30 – 35	164

NOTES

(*ii*) The age distribution of 100 life insurance policy-holders is as follows:

Age	*No. of policy holders*
17 – 19	9
20 – 25	16
26 – 35	12
36 – 40	26
41 – 50	14
51 – 55	12
56 – 60	5
61 – 70	5

12. Calculate the mean deviation from the mean and the median and their coefficients for the following data.

Size of shoes:	3	6	11	2	4	10	5	7	8	9
No. of pairs sold:	10	15	25	6	4	3	2	8	9	4

13. Discuss briefly about the measures of dispersion with the help of giving examples and characteristics.
14. Explain briefly about the standard deviation. Give appropriate examples.
15. 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.)
16. 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
17. 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.
18. (a) Explain the meaning of the Bernoulli process pointing out its main characteristics.

    (b) Give a few examples narrating some situations wherein binomial *pr* distribution can be used.
19. 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.

20. Explain the circumstances when the following probability distributions are used:
    (a) Binomial distribution
    (b) Poisson distribution
    (c) Exponential distribution
    (d) Normal distribution
21. Certain articles have been produced of which 0.5 per cent are defective and the articles are packed in cartons each containing 130 articles. What proportion of cartons are free from defective articles? What proportion of cartons contain two or more defective articles?

    (Given $e^{-0.5}$=0.6065).
22. 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 given data and test the goodness of fit.
23. 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?
24. 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?
25. Fit a normal distribution to the following data:

Height in inches	*Frequency*
60–62	5
63–65	18
66–68	42
69–71	27
72–74	8

26. Analyse the types of discrete distributions with the help of giving examples.

4.12 FURTHER READING

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

Chandan, J.S., Jagjit Singh and K.K. Khanna. 1995. *Business Statistics*. New Delhi: Vikas Publishing House.

Elhance, D.N. 2006. *Fundamental of Statistics*. Allahabad: Kitab Mahal.

Freud, J.E., and F.J. William. 1997. *Elementary Business Statistics – The Modern Approach*. New Jersey: Prentice-Hall International.

Goon, A.M., M.K. Gupta, and B. Das Gupta. 1983. *Fundamentals of Statistics*. Vols. I & II, Kolkata: The World Press Pvt. Ltd.

Gupta, S.C. 2008. *Fundamentals of Business Statistics*. Mumbai: Himalaya Publishing House.

Kothari, C.R. 1984. *Quantitative Techniques*. New Delhi: Vikas Publishing House.

Levin, Richard. I., and David. S. Rubin. 1997. *Statistics for Management.* New Jersey: Prentice-Hall International.

Meyer, Paul L. 1970. *Introductory Probability and Statistical Applications*. Massachusetts: Addison-Wesley.

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

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

Anderson, David R., Dennis J. Sweeney and Thomas A. Williams. *Essentials of Statistics for Business and Economics*. Mumbai: Thomson Learning, 2007.

S.P. Gupta. 2021. *Statistical Methods*. Delhi: Sultan Chand and Sons.

UNIT 5 ESTIMATION AND HYPOTHESIS TESTING

Structure

5.0 INTRODUCTION

In statistics, quality assurance, and survey methodology, sampling is the selection of a subset (a statistical sample) of individuals from within a statistical population to estimate characteristics of the whole population. Statisticians attempt to collect samples that are representative of the population in question. Sampling has lower costs and faster data collection than measuring the entire population and can provide insights in cases where it is infeasible to sample an entire population. Each observation measures one or more properties (such as weight, location, colour) of independent objects or individuals. In survey sampling, weights can be applied to the data to adjust for the sample design, particularly in stratified sampling. Results from probability theory and statistical theory are employed to guide the practice. In business and medical research, sampling is widely used for gathering information about a population. Acceptance sampling is used to determine if a production lot of material meets the governing specifications.

Single or isolated facts or figures cannot be called statistics as these cannot be compared or related to other figures within the same framework. Hence, any quantitative and numerical data can be identified as statistics when it possesses certain identifiable characteristics as per the norms of statistics. The area of statistics can be split up into two identifiable sub-areas. These sub-areas constitute descriptive statistics and inferential statistics. This unit will describe some of the terms used extensively in the field of statistics for scientific measurement. Statistical investigation is a comprehensive process and requires systematic collection of data about some group of people or objects, describing and organizing the data, analysing the data with the help of various statistical methods, summarizing the analysis and using the results for making judgements, decisions and predictions.

A sampling distribution or finite-sample distribution is the probability distribution of a given random-sample-based statistic. If an arbitrarily large number of samples, each involving multiple observations (data points), were separately used in order to compute one value of a statistic (such as, for example, the sample mean or sample variance) for each sample, then the sampling distribution is the probability distribution of the values that the statistic takes on. In many contexts, only one sample is observed, but the sampling distribution can be found theoretically. Sampling distributions are important in statistics because they provide a major simplification en route to statistical inference. More specifically, they allow analytical considerations to be based on the probability distribution of a statistic, rather than on the joint probability distribution of all the individual sample values.

Estimation (or estimating) is the process of finding an estimate, or approximation, which is a value that is usable for some purpose even if input data may be incomplete, uncertain, or unstable. The value is nonetheless usable because it is derived from the best information available. Typically, estimation involves 'Using the value of a statistic derived from a sample to estimate the value of a corresponding population parameter'. The sample provides information that can be projected, through various formal or informal processes, to determine a range most likely to describe the missing information. An estimate that turns out to be incorrect will be an overestimate if the estimate exceeded the actual result, and an underestimate if the estimate fell short of the actual result.

Statistical hypothesis test is a method of statistical inference used to determine a possible conclusion from two different, and likely conflicting, hypotheses. In a statistical hypothesis test, a null hypothesis and an alternative hypothesis is proposed for the probability distribution of the data. If the sample obtained has a probability of occurrence less than the pre-specified threshold probability, the significance level, given the null hypothesis is true, the difference between the sample and the null hypothesis in deemed statistically significant. The hypothesis test may then lead to the rejection of null hypothesis and acceptance of alternate hypothesis. The process of distinguishing between the null hypothesis and the alternative hypothesis is aided by considering Type I error and Type II error which are controlled by the pre-specified significance level. Hypothesis tests based on statistical significance are another way of expressing confidence intervals (more precisely, confidence sets). In other words, every hypothesis test based on significance can be obtained via a confidence interval, and every confidence interval can be obtained via a hypothesis test based on significance.

In this unit, you will learn about the sampling theory, sampling distribution of the number of successes, the student's distribution theory of estimation, theory of estimation and hypothesis testing.

5.1 OBJECTIVES

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

- Explain about the sampling theory
- Discuss the methods of sampling
- Learn about parameter and statistics

- Explain the concept of population in statistics
- Know about estimation
- Understand hypothesis distribution and test of significance

5.2 SAMPLING THEORY

A universe is the complete group of items about which knowledge is sought. The universe may be finite or infinite. Finite universe is one which has a definite and certain number of items but when the number of items is uncertain and infinite, the universe is said to be an infinite universe. Similarly the universe may be hypothetical or existent. In the former case the universe in fact does not exist and we can only imagine the items constituting it. Tossing of a coin or throwing of a dice are examples of hypothetical universes. Existent universe is a universe of concrete objects, i.e., the universe where the items constituting it really exist. On the other hand, the term sample refers to that part of the universe which is selected for the purpose of investigation. *The theory of sampling studies the relationships that exist between the universe and the sample or samples drawn from it.*

5.2.1 Parameter and Statistic

It would be appropriate to explain the meaning of two terms viz., parameter and statistic. All the statistical measures based on all items of the universe are termed as parameters whereas statistical measures worked out on the basis of sample studies are termed as sample statistics. Thus, a sample mean or a sample standard deviation is an example of statistic whereas the universe mean or universe standard deviation is an example of a parameter.

The main problem of sampling theory is the problem of relationship between a parameter and a statistic. The theory of sampling is concerned with estimating the properties of the population from those of the sample and also with gauging the precision of the estimate. This sort of movement from particular Sample towards general Universe is what is known as statistical induction or statistical inference. In more clear terms, 'From the sample we attempt to draw inferences concerning the universe. In order to be able to follow this inductive method, we first follow a deductive argument which is that we imagine a population or universe (finite or infinite) and investigate the behaviour of the samples drawn from this universe applying the laws of probability.' The methodology dealing with all this is known as Sampling Theory.

Objects of Sampling Theory

Sampling theory is to attain one or more of the following objectives:

(a) *Statistical Estimation:* Sampling theory helps in estimating unknown population quantities or what are called parameters from a knowledge of statistical measures based on sample studies often called as 'Statistic'. In other words, to obtain the estimate of parameter from statistic is the main objective of the sampling theory. The estimate can either be a point estimate or it may be an interval estimate. *Point estimate* is a single estimate

NOTES

expressed in the form of a single figure but *interval estimate* has two limits, the upper and lower limits. Interval estimates are often used in statistical induction.

(b) *Tests of Hypotheses or Tests of Significance:* The second objective of sampling theory is to enable us to decide whether to accept or reject hypotheses or to determine whether observed samples differ significantly from expected results. The sampling theory helps in determining whether observed differences are actually due to chance or whether they are really significant. Tests of significance are important in the theory of decisions.

(c) *Statistical Inference:* Sampling theory helps in making generalization about the universe from the studies based on samples drawn from it. It also helps in determining the accuracy of such generalizations.

5.2.2 Sampling Distribution of Sample Mean

In sampling theory we are concerned with what is known as the sampling distribution. For this purpose we can take certain number of samples and for each sample we can compute various statistical measures such as mean, standard deviation etc. It is to be noted that each sample will give its own value for the statistic under consideration. All these values of the statistic together with their relative frequencies with which they occur, constitute the sampling distribution. We can have sampling distribution of means or the sampling distribution of standard deviations or the sampling distribution of any other statistical measure. The sampling distribution tends quite closer to the normal distribution if the number of samples is large. *The significance of sampling distribution follows from the fact that the mean of a sampling distribution is the same as the mean of the universe. Thus, the mean of the sampling distribution can be taken as the mean of the universe.*

The Concept of Standard Error (or S.E.)

The standard deviation of sampling distribution of a statistic is known as its standard error and is considered the key to sampling theory. The utility of the concept of standard error in statistical induction arises on account of the following reasons:

(a) The standard error helps in testing whether the difference between observed and expected frequencies could arise due to chance. The criterion usually adopted is that if a difference is upto 3 times the S.E. then the difference is supposed to exist as a matter of chance and if the difference is more than 3 times the S.E., chance fails to account for it, and we conclude the difference as significant difference. This criterion is based on the fact that at $x \pm 3$(S.E.), the normal curve covers an area of 99.73 per cent. The product of the critical value at certain level of significance and the S. E. is often described as the Sampling Error at that particular level of significance. We can test the difference at certain other levels of significance as well depending upon our requirement.

NOTES

(b) The standard error gives an idea about the reliability and precision of a sample. If the relationship between the standard deviation and the sample size is kept in view, one would find that the standard error is smaller than the standard deviation. The smaller the S.E. the greater the uniformity of the sampling distribution and hence greater is the reliability of sample. Conversely, the greater the S.E., the greater the difference between observed and expected frequencies and in such a situation the unreliability of the sample is greater. The size of S.E. depends upon the sample size; the greater the number of items included in the sample the smaller the error to be expected and vice versa.

(c) The standard error enables us to specify the limits, maximum and minimum, within which the parameters of the population are expected to lie with a specified degree of confidence. Such an interval is usually known as confidence interval. The degree of confidence with which it can be asserted that a particular value of the population lies within certain limits is known as the level of confidence.

Procedure of Significance Testing

The following sequential steps constitute, in general, the procedure of significance testing:

(a) *Statement of the Problem:* First, the problem has to be stated in clear terms. It should be quite clear as to in respect of what the statistical decision has to be taken. The problem may be, Whether the hypothesis is to be rejected or accepted? Is the difference between a parameter and a statistic significant? or the like ones.

(b) *Defining the Hypothesis:* Usually, we start with the null hypothesis according to which it is presumed that there is no difference between a parameter and a statistic. If we are take a decision whether the students have been benefited from the extra coaching and if we start with the supposition that they have not been benefited then this supposition would be termed as null hypothesis which in symbolic form is denoted by H_0. As against null hypothesis, the researcher may as well start with some alternative hypothesis, (symbolically H_1) which specifies those values that the researcher believes to hold true and then may test such hypothesis on the basis of sample data. Only one alternative hypothesis can be tested at one time against the null hypothesis.

(c) *Selecting the Level of Significance:* The hypothesis is examined on a pre-determined level of significance. Generally, either 5 per cent level or 1 per cent level of significance is adopted for the purpose. However, it can be stated here that the level of significance must be adequate keeping in view the purpose and nature of enquiry.

(d) *Computation of the Standard Error.* After determining the level of significance the standard error of the concerning statistic (mean, standard deviation or any other measure) is computed. There are different formulae for computing the standard errors of different statistic. For example, the

NOTES

Standard Error of Mean = $\dfrac{\text{Standard Deviation}}{\sqrt{n}}$, the standard error of Standard Deviation = $\dfrac{\text{Standard Deviation}}{\sqrt{2n}}$, the standard error of Karl Pearson's Coefficient of Correlation = $\dfrac{1-r^2}{\sqrt{n}}$ and so on. (A detailed description of important standard errors formulae has been given on the pages that follow).

(e) *Calculation of the Significance Ratio:* The significance ratio, symbolically described as *z, t, f* etc. depending on the test we use, is often calculated by diving the difference between a parameter and a statistic by the standard error concerned. Thus, in context of mean, of small sample when population variance is not known, in context of $t = \dfrac{|\bar{x} - \mu|}{\text{SE}_x}$ and in context of difference between two sample means $t = \dfrac{|X_1 - X_2|}{\text{SE}_{\text{diff}\, x_1 - x_2}}$. (All this has been fully explained while explaining sampling theory in respect of small samples of variables later in this unit).

(f) *Deriving the Inference:* The significance ratio is then compared with the predetermined critical value. If the ratio exceeds the critical value then the difference is taken as significant but if the ratio is less than the critical value, the difference is considered insignificant. For example, the critical value at 5 per cent level of significance is 1.96. If the computed value exceeds 1.96 then the inference would be that the difference at 5 per cent level is significant and this difference is not the result of sampling fluctuations but the difference is a real one and should be understood as such.

5.3 SAMPLING DISTRIBUTION OF THE NUMBER OF SUCCESSES

Population

A population, in statistical terms, is the totality of things under consideration. It is the collection of all values of the variable that is under study. For instance, if we are interested in knowing as to how much on an average an American bachelor spends on his clothes per year, then all American bachelors would constitute the population. Similarly, if we want to know the percentage of adult American travellers who go to Europe, then only those adult Americans who travel are considered as population.

The amount paid by parents in one year for an average Class I day-boarding public school student can be evaluated by calculating the fee structure, such as admission fee, day-boarding fees, tuition fees and annual charges. Thus, Class I day-boarding students would constitute the specific population group.

Another example we can consider about is the population of coal mines workers, who are suffering from pneumoconiosis throughout the country. To evaluate this, we collect information on all cases of pneumoconiosis in different coal mines in the country.

A summary measure that describes any given characteristic of the population is known as a parameter. For example, the measure, the average income of American professors, would be considered as a parameter since it describes the characteristic of income of the population of American professors.

Sample

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 of 10,000 for the purpose of our study. 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

A random sample 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 into 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 above example is known as sampling without replacement, as each person can only be selected once. This is because 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 a 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.

Random Number Tables

For a sample to be truly representative of the population, it must truly be random. To make the random selection easier, we can make use of tables of random numbers which are generated by computers. A perfect random number table would be one in which every digit has been entered randomly. This means that no matter where you start within the table and no matter in which direction you move, the probability of encountering any one of the ten digits (0, 1, 2,...9) would be the same. This means that the chance of any one of these digits being at any place in the table is exactly one out of ten. Similarly, if these digits are grouped in pairs (00, 01, 02,...99), then each of these pairs has the same chance of occurring at any place so that each pair would have a chance of occurring of one out of a hundred.

Table 5.1 illustrates a random number table.

Table 5.1 *A Random Number Table*

Column Number

Row Number	1	2	3	4	5
1	81625	42372	07090	23422	10742
2	20891	27833	93079	16274	92818
3	62882	48722	39630	96434	09895
4	59882	84713	82521	29026	08591
5	17932	14360	42933	89380	68191
6	67732	36772	09281	26898	30919
7	58198	87824	47958	04701	17369
8	57041	47778	02361	86939	61463
9	05264	49678	02067	58121	61822
10	84935	60407	16547	21359	58913

As an example of use of random number tables, let us assume that we have to select a random sample from a finite population. The population cannot be infinite due to the limitation as to how far the random numbers can go. Let there be 100 students in the population from which we have to draw a sample of five students. Now we assign a two-digit number to each member of the population so that each member is known as 00, 01, 02 ... 99. For selecting five students at random from this population, we go to the random number table with groups of two digits each and starting at any point and moving in any direction we pick the five groups of numbers. Suppose that the numbers picked up are 07, 22, 23, 58 and 78. Then those members of the population to whom these numbers are assigned constitute the random sample. In case we want to use a random number table in which groups of five digits are arranged, as in Table 5.1, then we can use only the first two digits or any two digits out of the five and reach the same conclusion of randomness. In Table 5.1, suppose we pick row 5 and go across and pick up the first two digits from each group of five, we get the following numbers: 17, 14, 42, 89 and 68. Thus, those five members of the population to whom these numbers are assigned constitute the random sample.

Sample Selection

Selecting an adequate sample is one of the steps in the primary data collection process. 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, 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.

There are many reasons behind sampling. 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 the sample size:

NOTES

(i) The larger the size of the population, the larger should be the sample size.

(ii) If the resources available do not put a heavy constraint on the sample size, a larger sample would be desirable.

(iii) If the samples are selected by scientific methods, a larger sample size would ensure greater degree of accuracy in conclusions.

(iv) 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 vs Sampling

Under the census or complete enumeration survey method, data is collected for all units (e.g., 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 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:

(i) 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 item of the population has an equal chance of being selected.

(ii) 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.

NOTES

Methods of Sampling

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: (i) Simple random sampling, (ii) Stratified random sampling, (iii) Systematic sampling, and (iv) Cluster sampling.

Non-probability sampling methods 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 non-random sampling process. The non-probability methods include: (i) Judgement sampling, (ii) Quota sampling, and (iii) Convenience sampling.

As shown in Figure 5.1 depicts the broad classification and sub-classification of various methods of sampling.

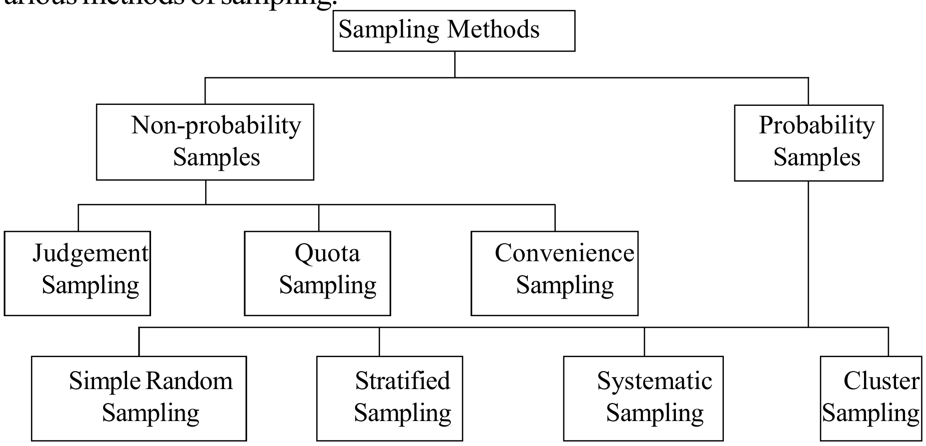


***Fig. 5.1** Methods of Sampling*

Non-Probability Sampling Methods

The following are the non-probability sampling methods:

(i) Judgement Sampling

In judgement 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.

NOTES

(ii) 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.

(iii) 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, 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 selecting 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:

(i) Simple Random Sampling

In simple random sampling each 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.

(ii) Stratified Random Sampling

In stratified random sampling, 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.

(iii) Systematic Sampling

Systematic sampling 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 interval until the sample has been formed. This method is popularly used in 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. The subsequent items are selected by taking every Kth 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.

(iv) Multistage or Cluster Sampling

In multistage or cluster sampling, the primary, intermediate and final (or the ultimate) units are randomly selected 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

NOTES

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 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 errors 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.

Parameter and Statistics

Parameter

A parameter is a numeric quantity that describes a certain population characteristic. Parameters are in general represented by Greek letters. The most common parameters are the population mean and variance, represented by the Greek letters μ and σ^2, respectively. For example, the population mean is a parameter that is often used to indicate the average value of a quantity.

Parameters are often estimated, in view of the fact that their value is generally unknown, especially when the population is large enough that it is impossible or impractical to obtain measurements for all people.

Statistics

A statistic is a quantity, calculated from a sample of data, used to estimate a parameter. For example, the average of the data in a sample is used to give information about the overall average in the population from which that sample was drawn. Statistics is usually represented by Latin letters with other symbols. The sample mean and variance, two of the most common statistics derived from samples, are denoted by the symbols $\bar{X}_n$ and σ^2, respectively.

It is possible to draw more than one sample from the same population, and each sample will have its own value for any statistic used to estimate a particular parameter. For example, the mean of the data in a sample is used to provide information about the overall mean in the population from which that sample was drawn. However, the sample means for two independent samples, drawn from the same population, will not necessarily be equal. Each sample mean is still an estimate of the underlying population mean.

Check Your Progress

1. What would you use to test the validity of hypothesis?
2. What is one very important aspect of sampling theory?
3. Define a sample.
4. What are the two types of sampling?
5. Define non-probability sampling.
6. What is probability sampling?
7. Name the probability sampling method.

NOTES

5.4 THE STUDENT'S DISTRIBUTION

One of the major objectives of statistical analysis is to know the 'True' values of different parameters of the population. Since it is not possible due to time, cost and other constraints to take the entire population for consideration, random samples are taken from the population. These samples are analysed properly and they lead to generalizations that are valid for the entire population. The process of relating the sample results to population is referred to as, 'Statistical Inference' or 'Inferential Statistics'.

In general, a single sample is taken and its mean X is considered to represent the population mean. However, in order to use the sample mean to estimate the population mean, we should examine every possible sample (and its mean, etc.) that could have occurred, because a single sample may not be representative enough. If it was possible to take all the possible samples of the same size, then the distribution of the results of these samples would be referred to as, 'Sampling Distribution'. The distribution of the means of these samples would be referred to as, 'Sampling Distribution; of the Means'.

The relationship between the sample means and the population mean can best be illustrated by Example 5.1.

Example 5.1: Suppose a babysitter has 5 children under her supervision with average age of 6 years. However, individually, the age of each child be as follows:

$$X_1 = 2$$
$$X_2 = 4$$
$$X_3 = 6$$
$$X_4 = 8$$
$$X_5 = 10$$

Now these 5 children would constitute our entire population, so that N = 5.

Solution:

The population mean $\mu = \frac{\Sigma X}{N}$

$$= \frac{2+4+6+8+10}{5} = 30/5 = 6$$

and the standard deviation is given by the formula:

NOTES

$$\sigma = \sqrt{\frac{\Sigma(\times - \mu)^2}{N}}$$

Now, let us calculate the standard deviation.

X	μ	$(X-\mu)^2$
2	6	16
4	6	4
6	6	0
8	6	4
10	6	16
		Total = $\Sigma (X-\mu)^2 = 40$

Then,

$$\sigma = \sqrt{\frac{40}{5}} = \sqrt{8} = 2.83$$

Now, let us assume the sample size, $n = 2$, and take all the possible samples of size 2, from this population. There are 10 such possible samples. These are as follows, along with their means.

$X_1,\ X_2$	(2, 4)	$\bar{X}_1 = 3$
$X_1,\ X_3$	(2, 6)	$\bar{X}_2 = 4$
$X_1,\ X_4$	(2, 8)	$\bar{X}_3 = 5$
$X_1,\ X_5$	(2,10)	$\bar{X}_4 = 6$
$X_2,\ X_3$	(4, 6)	$\bar{X}_5 = 5$
$X_2,\ X_4$	(4, 8)	$\bar{X}_6 = 6$
$X_2,\ X_5$	(4, 10)	$\bar{X}_7 = 7$
$X_3,\ X_4$	(6, 8)	$\bar{X}_8 = 7$
$X_3,\ X_5$	(6, 10)	$\bar{X}_9 = 8$
$X_4,\ X_5$	(8, 10)	$\bar{X}_{10} = 9$

Now, if only the first sample was taken, the average of the sample would be 3. Similarly, the average of the last sample would be 9. Both of these samples are totally unrepresentative of the population. However, if a grand mean X of the distribution of these sample means is taken, then,

$$\bar{\bar{X}} = \frac{\sum_{i=1}^{10} \bar{X}_i}{10}$$

$$\frac{3+4+5+6+5+6+7+7+8+9}{10} = 60/10 = 6$$

NOTES

This grand mean has the same value as the mean of the population. Let us organize this distribution of sample means into a frequency distribution and probability distribution.

Sample mean	*Freq.*	*Rel.freq.*	*Prob.*
3	1	1/10	.1
4	1	1/10	.1
5	2	2/10	.2
6	2	2/10	.2
7	2	2/10	.2
8	1	1/10	.1
9	1	1/10	.1
			1.00

This probability distribution of the sample means is referred to as 'sampling distribution of the mean.'

Sampling Distribution of the Mean

The sampling distribution of the mean can thus be defined as, 'A probability distribution of all possible sample means of a given size, selected from a population'.

Accordingly, the sampling distribution of the means of the ages of children as tabulated in Example 5.1, has 3 predictable patterns. These are as follows:

(i) The mean of the sampling distribution and the mean of the population are equal. This can be shown as follows:

Sample mean ($\bar{X}$)	*Prob.* $P(\bar{X})$
3	.1
4	.1
5	.2
6	.2
7	.2
8	.1
9	.1
	1.00

Then,

$$\mu = \Sigma \bar{X}P(\bar{X}) = (3 \times .1) + (4 \times .1) + (5\times.2) + (6 \times .2) + (7 \times .2) + (8 \times .1) + 9 \times .1) = 6$$

This value is the same as the mean of the original population.

(ii) The spread of the sample means in the distribution is smaller than in the population values. For example, the spread in the distribution of sample means above is from 3 to 9, while the spread in the population was from 2 to 10.

(iii) The shape of the sampling distribution of the means tends to be, 'Bell-shaped' and approximates the normal probability distribution, even when the population is not normally distributed. This last property leads us to the 'Central Limit Theorem'.

NOTES

Central Limit Theorem

Central Limit Theorem states that, 'Regardless of the shape of the population, the distribution of the sample means approaches the normal probability distribution as the sample size increases.'

The question now is how large should the sample size be in order for the distribution of sample means to approximate the normal distribution for any type of population. In practice, the sample sizes of 30 or larger are considered adequate for this purpose. This should be noted however, that the sampling distribution would be normally distributed, if the original population is normally distributed, no matter what the sample size.

As we can see from our sampling distribution of the means, the grand mean $\bar{\bar{X}}$ of the sample means or $\mu_{\bar{X}}$ equals μ , the population mean. However, realistically speaking, it is not possible to take all the possible samples of size n from the population. In practice only one sample is taken, but the discussion on the sampling distribution is concerned with the proximity of 'a' sample mean to the population mean.

It can be seen that the possible values of sample means tend towards the population mean, and according to Central Limit Theorem, the distribution of sample means tend to be normal for a sample size of n being larger than 30. Hence, we can draw conclusions based upon our knowledge about the characteristics of the normal distribution.

For example, in the case of sampling distribution of the means, if we know the grand mean $\mu_{\bar{X}}$ of this distribution, which is equal to p, and the standard deviation of this distribution, known as 'Standard error of free mean' and denoted by $\sigma_{\bar{X}}$, then we know from the normal distribution that there is a 68.26 per cent chance that a sample selected at random from a population, will have a mean that lies within one standard error of the mean of the population mean. Similarly, this chance increases to 95.44 per cent, that the sample mean will lie within two standard errors of the mean ($\sigma_{\bar{X}}$) of the population mean. Hence, knowing the properties of the sampling distribution tells us as to how close the sample mean will be to the true population mean.

Standard Error

Standard error of the mean ($\sigma_{\bar{X}}$)

Standard error of the mean ($\sigma_{\bar{X}}$)is a measure of dispersion of the distribution of sample means and is similar to the standard deviation in a frequency distribution and it measures the likely deviation of a sample mean from the grand mean of the sampling distribution.

If all sample means are given, then ($\sigma_{\bar{X}}$)can be calculated as follows:

$$\sigma_{\bar{x}} = \sqrt{\frac{\Sigma(\bar{x} - \mu_{\bar{x}})}{N}} \quad \text{where N = Number of sample means}$$

Thus we can calculate $\sigma_{\bar{X}}$ for Example 5.1 of the sampling distribution of the ages of 5 children as follows:

$\bar{X}$	$(\mu_{\bar{x}})$	$(\bar{X}-\mu_{\bar{x}})^2$
3	6	9
4	6	4
5	6	1
6	6	0
7	6	1
8	6	4
9	6	9
		$\Sigma\ (\bar{X}-\mu_{\bar{x}})^2 = 28$

Then,

$$\sigma_{\bar{x}} = \sqrt{\frac{\Sigma(\bar{x}-\mu_{\bar{x}})}{N}}$$

$$= \sqrt{\frac{28}{7}}$$

$$= \sqrt{4} = 2$$

However, since it is not possible to take all possible samples from the population, we must use alternate methods to compute $\sigma_{\bar{x}}$.

The standard error of the mean can be computed from the following formula, if the population is finite and we know the population mean. Hence,

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{(N-n)}{(N-1)}}$$

Where,

σ = population standard deviation
N = population size
n = sample size

This formula can be made simpler to use by the fact that we generally deal with very large populations, which can be considered infinite, so that if the population size A' is very large and sample size n is small, as for example in the case of items tested from assembly line operations, then,

$$\sqrt{\frac{(N-n)}{(N-1)}} \text{ would approach 1.}$$

Hence,

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

The factor $\sqrt{\frac{(N-n)}{(N-n)}}$ is also known as the 'finite correction factor', and should be used when the population size is finite.

As this formula suggests, $\sigma_{\bar{x}}$ decreases as the sample size (w) increases, meaning that the general dispersion among the sample means decreases, meaning further that any single sample mean will become closer to the population mean, as the value of ($\sigma_{\overline{X}}$) decreases. Additionally, since according to the property of the normal curve, there is a 68.26 per cent chance of the population mean being within one $\sigma_{\overline{X}}$ of the sample mean, a smaller value of $\sigma_{\overline{X}}$ will make this range shorter; thus making the population mean closer to the sample mean (Refer Example 5.2).

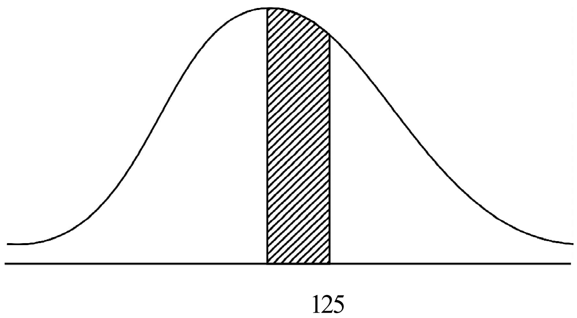
Example 5.2: The *IQ* scores of college students are normally distributed with the mean of 120 and standard deviation of 10.

(a) What is the probability that the *IQ* score of any one student chosen at random is between 120 and 125?

(b) If a random sample of 25 students is taken, what is the probability that the mean of this sample will be between 120 and 125.

Solution:

(a) Using the standardized normal distribution formula,



$\mu = 120$

$\sigma = 10$

$$Z = \frac{(X - \mu)}{\sigma}$$

$$Z = \frac{125 - 120}{10} = 5/10 = .5$$

The area for Z = .5 is 19.15.

This means that there is a 19.15 per cent chance that a student picked up at random will have an IQ score between 120 and 125.

(b) With the sample of 25 students, it is expected that the sample mean will be much closer to the population mean, hence it is highly likely that the sample mean would be between 120 and 125.

The formula to be used in the case of standardized normal distribution for sampling distribution of the means is given by,

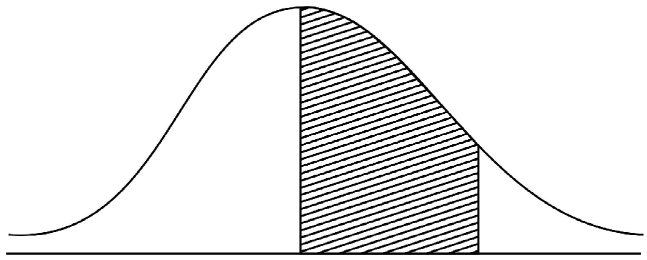
$$Z = \frac{\overline{X} - \mu}{\sigma_{\overline{X}}}$$

where,

$$\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}}$$

NOTES

Hence,



$$Z = \frac{\overline{X} - \mu}{\sigma_{\overline{x}}}$$

where, $\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}} = \frac{10}{\sqrt{25}} = \frac{10}{5} = 2$

Then,

$$Z = \frac{125 - 120}{2} = 5/2 = 2.5$$

The area for Z = 2.5 is 49.38.

This shows that there is a chance of 49.38 per cent that the sample mean will be between 120 and 125. As the sample size increases further, this chance will also increase. It can be noted that the probability of a sample mean being between 120 and 125 is much higher than the probability of an individual student having an IQ between 120 and 125.

5.5 THEORY OF ESTIMATION

The best estimator should be highly reliable and have such desirable properties as unbiasedness, consistency, efficiency and sufficiency. These criteria are described as follows:

(i) Unbiasedness: An estimator is a random variable since it is always a function of the sample values. For example, the value of the sample average would depend upon the values of the sample and may differ from sample to sample. The expected value of the sample average is considered to be an unbiased estimator if it equals the population mean, which is being estimated. This means that:

$$E(\overline{X}) = \mu$$

(Since sampling distribution is a probability distribution, we refer to the average, as expected value instead of simply the average).

(ii) Consistency: This refers to the effect of the sample size on the accuracy of the estimator. A statistic is said to be consistent estimator of the population parameter, if it approaches the parameter as the sample size increases, so that in the case of the mean:

$$\overline{X} \to \mu \text{ as } n \to N$$

NOTES

(iii) Efficiency: An estimator is considered to be efficient if its value remains stable from sample to sample. The best estimator would be the one which would have the least variance from sample to sample taken randomly from the same population. From the three point estimators of central tendency, namely the mean, the mode and the median, the mean is considered to be the least variant and hence, a better estimator.

(iv) Sufficiency: An estimator is said to be sufficient if it uses all the information about the population parameter contained in the sample. For example, the statistic mean uses all the sample values in its computation, while the mode and the median do not. Hence, the mean is a better estimator in this sense.

Some of the parameters of the population and their estimators are as follows:

$$\mu = \overline{X} = \frac{\sum X}{n}$$

$$\sigma = s = \sqrt{\frac{\sum (X - \overline{X})^2}{n-1}}$$

$$p = p_s = \left(\frac{X}{n}\right), \text{ where } p_s \text{ is the sample proportion}$$

5.5.1 Point Estimation

The theory of estimation is a very commonly used 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 the following:

- Finding linear models and generalized linear models.
- 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.

A point estimate uses a single sample value to estimate the desired population parameter. For example, a sample mean $\bar{X}$ is considered as a point estimate of the population mean μ. Similarly, a sample standard deviation s, is a point estimate of population standard deviation σ. For instance, if we want to know the Grade Point Average (GPA) of seniors majoring in Business Administration at Medgar Evers College, then we take a random sample of business major seniors and calculate the sample mean $\bar{X}$ of the sample. Then, the value of this $\bar{X}$ would be considered as a point estimate of μ which is the grade point average of the entire population of students majoring in business administration. Similarly, the sample variance s^2 is the point estimate of the population variance σ^2.

In point estimate, we seek the sample statistic, such as $\bar{X}$, computed from sample observations, which is the best estimate of the corresponding population parameter, such as μ. But, how do we know that the sample statistic that we computed from sample observations is the best estimator of the population parameter? By *best* we mean that the value of the sample statistic should be as close to the population parameter as possible. For example, if the sample mean grade point average for business students is calculated as 3.5 out of 4, then the population average grade point average should also be 3.5 or very close to, it in order for sample average to be a good estimator of population average. Since the population parameter is always inferred from sample statistic, it is necessary and important that such sample statistic should be as highly reliable as an estimator for population parameter, as possible. For example, there are three measures of central tendency, namely mean, mode and median for a sample that can be used as point estimators for the population average. It is important to know as to which one of these measures best represents the population mean. As an illustration, suppose that we want to find out the average time that a salesman of a company spends with the customer. Suppose further that we took a sample and found out that on an average, a salesman spent 60 minutes with a customer (mean). However, most salesmen spent 45 minutes (the mode) and the median was 65 minutes. The question now is to establish as to which of these measures would best describe the population parameter as to how much time on an average a salesman spends with the customer?

5.5.2 Interval Estimation

Point estimator, though simplistic in nature, has some drawbacks. First, a point estimator from the sample may not exactly locate the population parameter resulting in some margin of uncertainty. The average of a sample for example, may or may not be equal or close to the average of the population. If the sample average is different from the population average, the point estimator does not indicate the extent of the possible error, even though this error can be reduced by increasing the sample size. Second, a point estimate does not specify as to how confident we can be that the estimate is close to the parameter it is estimating.

To reasonably overcome these drawbacks, statisticians use another type of estimation known as interval estimation. In this method, we first find a point estimate. Then we use this estimate to construct an interval on both sides of the point estimate, within which we can be reasonably confident that the true parameter will lie. For example, suppose that we wanted to find out the average salary of full professors at a university who had served at least five years at that rank. Suppose further, that a random sample was taken and the average of the sample was computed to be \$55,000. It is quite possible that the actual average salary of all university professors is \$55,000. However, it is equally possible that the sample was not true representative of the population and the average of the population is quite far off the average of the sample. Accordingly, it is much more likely that the average salary of all the professors lies somewhere, let us say, between \$50,000 and \$60,000 than exactly at \$55,000. Of course, the greater the range of interval around the sample mean, the more likely it is that the population mean lies in that range. This degree of likelihood is known as the confidence level and the range around the sample mean is known as the confidence interval at a given confidence level. (It is, of course, assumed that the sample is large enough so that the Central Limit Theorem holds.)

NOTES

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:

$$Z = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}}$$

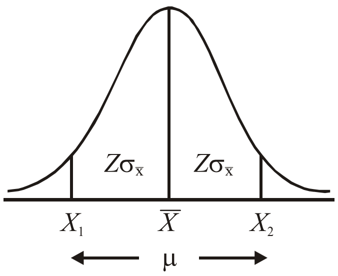
$$or\ \bar{X} - \mu = Z\sigma_{\bar{X}}$$

$$or\ \mu = \bar{X} - Z\sigma_{\bar{X}}$$

Since μ falls within a range of values equidistant from $\bar{X}$,

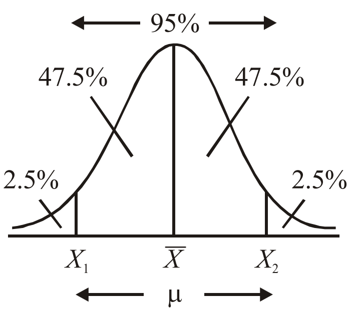
$$\mu = \bar{X} \pm Z\sigma_{\bar{X}}$$

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 in the following illustration:



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 Appendix) 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:

(i) 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.

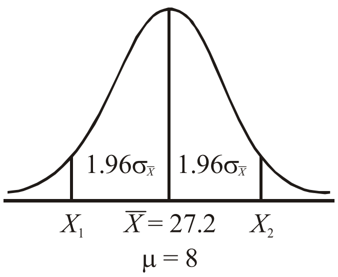
(ii) If we took a random sample of size n from a given population, the probability is 0.95 that the population mean would lie between the interval X_1 and X_2 around the sample mean, as shown.

(iii) 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 per cent, 95 per cent or 99 per cent is taken into consideration for which the Z score values are 1.68, 1.96 and 2.58, respectively.

Refer Examples 5.3 and 5.4 to understand internal estimation better.

Example 5.3: 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 (σ) 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,

$$\overline{X} \pm Z\sigma_{\overline{X}} \qquad \text{or } \overline{X} - Z\sigma_{\overline{X}} < \mu < \overline{X} + Z\sigma_{\overline{X}}$$

where $\sigma_{\overline{X}} = \dfrac{\sigma}{\sqrt{n}}$

NOTES

Accordingly, we need only four values, namely $\overline{X}$, Z, σ and n. In our case:

$$\overline{X} = 27.2$$
$$Z = 1.96$$
$$\sigma = 8$$
$$n = 100$$

Hence $\sigma_{\overline{X}} = \dfrac{\sigma}{\sqrt{n}} = \dfrac{8}{\sqrt{100}} = \dfrac{8}{10} = 0.8$

Then,

$$X_1 = \overline{X} - Z\sigma_{\overline{X}}$$
$$= 27.2 - (1.96 \times 0.8) = 27.2 - 1.568$$
$$= 25.632$$

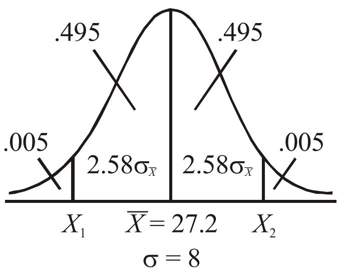
and

$$X_2 = \overline{X} + Z\sigma_{\overline{X}}$$
$$= 27.2 + (1.96 \times 0.8) = 27.2 + 1.568$$
$$= 28.768$$

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 5.4: Calculate the confidence interval in the previous problem, if we want to increase our confidence level from 95 per cent to 99 per cent. 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,

$$Z = 2.58$$
$$\sigma_{\overline{X}} = 0.8$$

Then

$$X_1 = \overline{X} - Z\sigma_{\overline{X}}$$
$$= 27.2 - (2.58 \times 0.8) = 27.2 - 2.064$$
$$= 25.136$$

and

$$X_2 = \overline{X} + Z\sigma_{\overline{X}}$$
$$= 27.2 + 2.064$$
$$= 29.264$$

(The value of Z is established from the table of Z scores against the area of 0.495 or a figure closest to it. The table shows that the area close to 0.495 is 0.4949 for which the Z score is 2.57 or 0.4951 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.)

5.6 HYPOTHESIS TESTING

A hypothesis is an approximate assumption that a researcher wants to test for its logical or empirical consequences. Hypothesis refers to a provisional idea whose merit needs evaluation, but has no specific meaning, though it is often referred as a convenient mathematical approach for simplifying cumbersome calculation. Setting up and testing hypothesis is an integral art of statistical inference. Hypotheses are often statements about population parameters like variance and expected value. During the course of hypothesis testing, some inference about population like the mean and proportion are made. Any useful hypothesis will enable predictions by reasoning including deductive reasoning. According to Karl Popper, a hypothesis must be falsifiable and that a proposition or theory cannot be called scientific if it does not admit the possibility of being shown false. Hypothesis might predict the outcome of an experiment in a lab, setting the observation of a phenomenon in nature. Thus, hypothesis is an explanation of a phenomenon proposal suggesting a possible correlation between multiple phenomena.

The characteristics of hypothesis are as follows:

- **Clear and accurate:** Hypothesis should be clear and accurate so as to draw a consistent conclusion.
- **Statement of relationship between variables:** If a hypothesis is relational, it should state the relationship between different variables.
- **Testability:** A hypothesis should be open to testing so that other deductions can be made from it and can be confirmed or disproved by observation. The researcher should do some prior study to make the hypothesis a testable one.
- **Specific with limited scope:** A hypothesis, which is specific, with limited scope, is easily testable than a hypothesis with limitless scope. Therefore, a researcher should pay more time to do research on such kind of hypothesis.
- **Simplicity:** A hypothesis should be stated in the most simple and clear terms to make it understandable.

NOTES

- **Consistency:** A hypothesis should be reliable and consistent with established and known facts.
- **Time limit:** A hypothesis should be capable of being tested within a reasonable time. In other words, it can be said that the excellence of a hypothesis is judged by the time taken to collect the data needed for the test.
- **Empirical reference:** A hypothesis should explain or support all the sufficient facts needed to understand what the problem is all about.

A hypothesis is a statement or assumption concerning a population. For the purpose of decision-making, a hypothesis has to be verified and then accepted or rejected. This is done with the help of observations. We test a sample and make a decision on the basis of the result obtained. Decision-making plays a significant role in different areas such as marketing, industry and management.

Statistical Decision-Making

Testing a statistical hypothesis on the basis of a sample enables us to decide whether the hypothesis should be accepted or rejected. The sample data enables us to accept or reject the hypothesis. Since the sample data gives incomplete information about the population, the result of the test need not be considered to be final or unchallengeable. The procedure, on the basis of which sample results, enables to decide whether a hypothesis is to be accepted or rejected. This is called Hypothesis Testing or Test of Significance.

Note 1: A test provides evidence, if any, against a hypothesis, usually called a null hypothesis. The test cannot prove the hypothesis to be correct. It can give some evidence against it.

The test of hypothesis is a procedure to decide whether to accept or reject a hypothesis.

Note 2: The acceptance of a hypotheses implies, if there is no evidence from the sample that we should believe otherwise.

The rejection of a hypothesis leads us to conclude that it is false. This way of putting the problem is convenient because of the uncertainty inherent in the problem. In view of this, we must always briefly state a hypothesis that we hope to reject.

A hypothesis stated in the hope of being rejected is called a *null hypothesis* and is denoted by H_0.

If H_0 is rejected, it may lead to the acceptance of an alternative hypothesis denoted by H_1.

For example, a new fragrance soap is introduced in the market. The null hypothesis H_0, which may be rejected, is that the new soap is not better than the existing soap.

Similarly, a dice is suspected to be rolled. Roll the dice a number of times to test.

The null hypothesis H_0: $p = 1/6$ for showing six.

The alternative hypothesis H_1: $p \neq 1/6$.

NOTES

For example, skulls found at an ancient site may all belong to race X or race Y on the basis of their diameters. We may test the hypothesis, that the mean is μ of the population from which the present skulls came. We have the hypotheses.

$$H_0 : \mu = \mu_x, H_1 : \mu = \mu_y$$

Here, we should not insist on calling either hypothesis null and the other alternative since the reverse could also be true.

Committing Errors: Type I and type II

Types of Errors

There are two types of errors in statistical hypothesis, which are as follows:

- **Type I error:** In this type of error, you may reject a null hypothesis when it is true. It means rejection of a hypothesis, which should have been accepted. It is denoted by α (alpha) and is also known alpha error.
- **Type II error:** In this type of error, you are supposed to accept a null hypothesis when it is not true. It means accepting a hypothesis, which should have been rejected. It is denoted by β (beta) and is also known as beta error.

  Type I error can be controlled by fixing it at a lower level. For example, if you fix it at 2 per cent, then the maximum probability to commit Type I error is 0.02. However, reducing Type I error has a disadvantage when the sample size is fixed, as it increases the chances of Type II error. In other words, it can be said that both types of errors cannot be reduced simultaneously. The only solution of this problem is to set an appropriate level by considering the costs and penalties attached to them or to strike a proper balance between both types of errors.

In a hypothesis test, a Type I error occurs when the null hypothesis is rejected when it is in fact true; that is, H_0 is wrongly rejected. For example, in a clinical trial of a new drug, the null hypothesis might be that the new drug is no better, on average, than the current drug; that is H_0: there is no difference between the two drugs on average. A Type I error would occur if we concluded that the two drugs produced different effects, when in fact there was no difference between them.

In a hypothesis test, a Type II error occurs when the null hypothesis H_0 is not rejected, when it is in fact false. For example, in a clinical trial of a new drug, the null hypothesis might be that the new drug is no better, on average, than the current drug; that is H_0: there is no difference between the two drugs on average. A Type II error would occur if it were concluded that the two drugs produced the same effect, that is, there is no difference between the two drugs on average, when in fact they produced different ones.

In how many ways can we commit errors?

We reject a hypothesis when it may be true. This is Type I Error.

We accept a hypothesis when it may be false. This is Type II Error.

The other true situations are desirable: We accept a hypothesis when it is true. We reject a hypothesis when it is false.

NOTES

	Accept H_0	Reject H_0
H_0 True	Accept True H_0 Desirable	Reject True H_0 Type I Error
H_1 False	Accept False H_0 Type II Error	Reject False H_0 Desirable

The level of significance implies the probability of Type I error. A 5 per cent level implies that the probability of committing a Type I error is 0.05. A 1 per cent level implies 0.01 probability of committing Type I error.

Lowering the significance level and hence the probability of Type I error is good but unfortunately, it would lead to the undesirable situation of committing Type II error.

To Sum Up:

- **Type I Error**: Rejecting H_0 when H_0 is true.
- **Type II Error**: Accepting H_0 when H_0 is false.

Note: The probability of making a Type I error is the level of significance of a statistical test. It is denoted by α.

Where, α = Prob. (Rejecting H_0 / H_0 true)

$1-\alpha$ = Prob. (Accepting H_0 / H_0 true)

The probability of making a Type II error is denoted by β.

Where, β = Prob. (Accepting H_0 / H_0 false)

1– = Prob. (Rejecting H_0 / H_0 false) = Prob. (The test correctly rejects H_0 when H_0 is false)

$1-\beta$ is called the power of the test. It depends on the level of significance α, sample size n and the parameter value.

5.6.1 Test of Hypothesis Concerning Mean and Proportion

Test of Significance

Tests for a sample mean $\overline{X}$

We have to test the null hypothesis that the population mean has a specified value μ, i.e., H_0: $\overline{X} = \mu$. For large n, if H_0 is true then,

$$z = \left| \frac{\overline{X} - \mu}{SE(\overline{X})} \right|$$

is approximately nominal. The theoretical region for z depending on the desired level of significance can be calculated.

For example, a factory produces items, each weighing 5 kg with variance 4. Can a random sample of size 900 with mean weight 4.45 kg be justified as having been taken from this factory?

$$n = 900$$
$$\bar{X} = 4.45$$
$$\mu = 5$$
$$\sigma = \sqrt{4} = 2$$
$$z = \left|\frac{\bar{X}-\mu}{SE(\bar{X})}\right| = \left|\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right| = \left|\frac{4.45-5}{2/30}\right| = 8.25$$

We have $z > 3$. The null hypothesis is rejected. The sample may not be regarded as originally from the factory at 0.27 per cent level of significance (corresponding to 99.73 per cent acceptance region).

Test for equality of two proportions

If P_1, P_2 are proportions of some characteristic of two samples of sizes n_1, n_2, drawn from populations with proportions P_1, P_2, then we have H_0: $P_1 = P_2$ vs H_1:P_1 ^{1} P_2 .

• **Case (I):** If H_0 is true, then let $P_1 = P_2 = p$

Where, p can be found from the data,

$$p = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2}$$
$$q = 1 - p$$

p is the mean of the two proportions.

$$SE(P_1 - P_2) = \sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

$$z = \frac{P_1 - P_2}{SE(P_1 - P_2)}, \ P \text{ is approximately normal } (0,1)$$

We write $z \sim N(0, 1)$

The usual rules for rejection or acceptance are applicable here.

• **Case (II):** If it is assumed that the proportion under question is not the same in the two populations from which the samples are drawn and that P_1, P_2 are the true proportions, we write,

$$SE(P_1 - P_2) = \sqrt{\left(\frac{P_1 q_1}{n_1} + \frac{P_2 q_2}{n_2}\right)}$$

We can also write the confidence interval for $P_1 - P_2$.

For two independent samples of sizes n_1, n_2 selected from two binomial populations, the 100 (1 – a) per cent confidence limits for $P_1 - P_2$ are,

$$(P_1 - P_2) \pm z_{\alpha/2}\sqrt{\left(\frac{P_1 q_1}{n_1} + \frac{P_2 q_2}{n_2}\right)}$$

The 90% confidence limits would be [with a = 0.1, 100 (1 – a) = 0.90]

$$(P_1 - P_2) \pm 1.645\sqrt{\left(\frac{P_1q_1}{n_1} + \frac{P_2q_2}{n_2}\right)}$$

NOTES

Consider Example 5.5 to further understand the test for equality.

Example 5.5: Out of 5000 interviewees, 2400 are in favour of a proposal, and out of another set of 2000 interviewees, 1200 are in favour. Is the difference significant?

Where, $P_1 = \frac{2400}{5000} = 0.48 \quad P_2 = \frac{1200}{2000} = 0.6$

Solution:

Given, $P_1 = \frac{2400}{5000} = 0.48 \quad P_2 = \frac{1200}{2000} = 0.6$

$n_1 = 5000 \qquad n_2 = 2000$

$$SE = \sqrt{\left(\frac{0.48 \times 0.52}{5000} + \frac{0.6 \times 0.4}{2000}\right)} = 0.013 \text{ (using Case (II))}$$

$$z = \left|\frac{P_1 - P_2}{SE}\right| = \frac{0.12}{0.013} = 9.2 > 3$$

The difference is highly significant at 0.27 per cent level.

5.6.2 Test of Hypothesis Conerning Standard Deviation

Large sample test for equality of two means $\bar{X}_1, \bar{X}_2$

Suppose two samples of sizes n_1 and n_2 are drawn from populations having means μ_1, μ_2 and standard deviations σ_1, σ_2.

To test the equality of means $\bar{X}_1, \bar{X}_2$ we write,

$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$

If we assume H_0 is true, then

$$z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$, approximately normally distributed with mean 0, and S.D. = 1.

We write $z \sim N(0, 1)$

As usual, if $|z| > 2$ we reject H_0 at 4.55% level of significance, and so on (Refer Example 5.6).

Example 5.6: Two groups of sizes 121 and 81 are subjected to tests. Their means are found to be 84 and 81 and standard deviations 10 and 12. Test for the significance of difference between the groups.

NOTES

Solution:

$\bar{X}_1 = 84$ $\bar{X}_2 = 81$ $n_1 = 121$ $n_2 = 81$

$\sigma_1 = 10$ $\sigma_2 = 12$

$$z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}, \quad z = \frac{84 - 81}{\sqrt{\frac{100}{121} + \frac{144}{81}}} = 1.86 < 1.96$$

The difference is not significant at the 5 per cent level of significance.

Small sample tests of significance

The sampling distribution of many statistics for large samples is approximately normal. For small samples with $n < 30$, the normal distribution, as shown in Example 5.4, can be used only if the sample is from a normal population with known σ.

If σ is not known, we can use student's t distribution instead of the normal. We then replace σ by sample standard deviation σ with some modification as shown.

Let $x_1, x_2, ..., x_n$ be a random sample of size n drawn from a normal population with mean μ and S.D. σ. Then,

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n-1}}.$$

Here, t follows the student's t distribution with $n - 1$ degrees of freedom.

Note: For small samples of $n < 30$, the term $\sqrt{n-1}$, in SE $= s/\sqrt{n-1}$, corrects the bias, resulting from the use of sample standard deviation as an estimator of σ.

Also,

$$\frac{s^2}{S^2} = \frac{n-1}{n} \text{ or } s = S\sqrt{\frac{n-1}{n}}$$

Procedure: Small samples

To test the null hypothesis $H_0 : \mu = \mu_0$, against the alternative hypothesis $H_1 : \mu \neq \mu_0$

Calculate $|t| = \frac{\bar{X} - \mu}{SE(\bar{X})}$ and compare it with the table value with $n - 1$ degrees of freedom (d.f.) at level of significance 1 per cent.

If this value > table value, reject H_0

If this value < table value, accept H_0

(Significance level idea same as for large samples)

We can also find the 95% (or any other) confidence limits for μ.

For the two-tailed test (use the same rules as for large samples; substitute t for z) the 95% confidence limits are,

$$\bar{X} \pm t_\alpha s/\sqrt{n-1} \quad \alpha = 0.025$$

NOTES

Rejection region

At a per cent level for two-tailed test if $| t | > t_{a/2}$ reject.

For one-tailed test, (right) if $t > t_a$ reject

(left) if $t > -t_a$ reject

At 5 per cent level the three cases are,

If $| t | > t_{0.025}$ reject two-tailed

If $t > t_{0.05}$ reject one-tailed right

If $t £ t_{0.05}$ reject one-tailed left

For proportions, the same procedure is to be followed.

Example 5.7: A firm produces tubes of diameter 2 cm. A sample of 10 tubes is found to have a diameter of 2.01 cm and variance 0.004. Is the difference significant? Given $t_{0.05,9} = 2.26$.

Solution:

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n-1}}$$

$$= \frac{2.01 - 2}{\sqrt{0.004/(10-1)}}$$

$$= \frac{0.01}{0.021}$$

$$= 0.48$$

Since, $|t| < 2.26$, the difference is not significant at 5 per cent level.

Check Your Progress

8. What are the objects of sample distribution theory?
9. Define the central limit theorem.
10. Give the name of some properties of estimation.
11. What is a hypothesis?
12. How will you define the characteristics of hypothesis?
13. How many types of statistical errors?

5.7 ANSWERS TO 'CHECK YOUR PROGRESS'

1. To test the validity of the hypothesis we would make use of sample observations and statistics.
2. A very important aspect of sampling theory is the study of test of significance which gives us a ground to decide the deviation between the observed sample statistic and the hypothetical parameter value or the deviation between two independent sample statistics.
3. A sample is a portion of the total population that is considered for study and analysis.

NOTES

4. The various methods of sampling can be grouped under two broad categories—probability (or random) sampling and non-probability (or non-random) sampling.
5. Non-probability sampling methods do not provide every item in the universe with a known chance of being included in the sample.
6. Probability sampling methods are those in which every item in the universe has a known chance, or probability of being chosen for the sample.
7. The following are the probability sampling method:
    - Simple random sampling
    - Stratified random sampling
    - Systematic sampling
    - Multistage or cluster sampling
8. One of the major objectives of statistical analysis is to know the 'true' values of different parameters of the population. Since it is not possible due to time, cost and other constraints to take the entire population for consideration, random samples are taken from the population.
9. Central Limit Theorem states that, 'Regardless of the shape of the population, the distribution of the sample means approaches the normal probability distribution as the sample size increases.'
10. The best estimator should be highly reliable and have such desirable properties as unbiasedness, consistency, efficiency and sufficiency.
11. A hypothesis is an approximate assumption that a researcher wants to test for its logical or empirical consequences. Hypothesis refers to a provisional idea whose merit needs evaluation, but as no specific meaning.
12. The characteristics of hypothesis are as follows:
    - Clear and accurate: Hypothesis should be clear and accurate so as to draw a consistent conclusion.
    - Statement of relationship between variables: If a hypothesis is relational, it should state the relationship between different variables.
    - Testability: A hypothesis should be open to testing so that other deductions can be made from it and can be confirmed or disproved by observation. The researcher should do some prior study to make the hypothesis a testable one.
    - Specific with limited scope: A hypothesis, which is specific, with limited scope, is easily testable than a hypothesis with limitless scope. Therefore, a researcher should pay more time to do research on such kind of hypothesis.
13. There are two types of errors in statistical hypothesis, which are as follows:
    - Type I error: In this type of error, you may reject a null hypothesis when it is true. It means rejection of a hypothesis, which should have been accepted. It is denoted by a (alpha) and is also known alpha error.

- Type II error: In this type of error, you are supposed to accept a null hypothesis when it is not true. It means accepting a hypothesis, which should have been rejected. It is denoted by b (beta) and is also known as beta error.

5.8 SUMMARY

- In statistics, population does not mean human population alone. A complete set of objects under study, living or non-living is called population or universe; for example, graduates in Pondy, Bajaj tube lights, Ceat car tyres etc.
- Each individual or object is called unit or member or element of that population. If there are a finite number of elements then it is called finite population. If the number of elements is infinite then it is called infinite population.
- A population, in statistical terms, is the totality of things under consideration. It is the collection of all values of the variable that is under study.
- A sample is a portion of the total population that is considered for study and analysis.
- A random sample 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.
- Selecting an adequate sample is one of the steps in the primary data collection process.
- The various methods of sampling can be grouped under two broad categories—probability (or random) sampling and non-probability (or non-random) sampling.
- A parameter is a numeric quantity that describes a certain population characteristic.
- The best estimator should be highly reliable and have such desirable properties as unbiasedness, consistency, efficiency and sufficiency.
- A hypothesis is an approximate assumption that a researcher wants to test for its logical or empirical consequences. Hypothesis refers to a provisional idea whose merit needs evaluation, but having no specific meaning.
- A hypothesis should be stated in the most simple and clear terms to make it understandable.
- A hypothesis should be reliable and consistent with established and known facts.
- A hypothesis should be capable of being tested within a reasonable time. In other words, it can be said that the excellence of a hypothesis is judged by the time taken to collect the data needed for the test.
- In type I error, you may reject a null hypothesis when it is true. It means rejection of a hypothesis, which should have been accepted. It is denoted by a (alpha) and is also known alpha error.

- In type II error, you are supposed to accept a null hypothesis when it is not true. It means accepting a hypothesis, which should have been rejected. It is denoted by b (beta) and is also known as beta error.

NOTES

5.9 KEY TERMS

- **Population (in statistics):** It is a complete set of objects under study, living or non-living.
- **Standard error of mean:** Measures the likely deviation of a sample mean from the grand mean of the sampling distribution.
- **Efficiency:** An estimator is considered to be efficient if its value remains stable from sample to sample. The best estimator would be the one which would have the least variance from sample to sample taken randomly from the same population. From the three point estimators of central tendency, namely the mean, the mode and the median, the mean is considered to be the least variant and hence, a better estimator.
- **Null hypothesis:** A hypothesis stated in the hope of being rejected.

5.10 SELF-ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. What is meant by statistical estimation?
2. What do you understand by computation of the standard error?
3. Describe the terms sample and sampling.
4. What are the two laws of statistics?
5. How do sampling and non-sampling errors arise?
6. What is estimation?
7. What are the characteristics of a hypothesis?

Long-Answer Questions

1. A company claims that 5% of its products are defective. In a sample of 400 items 320 are good. Test whether the claim is valid.
2. Discuss why sampling is necessary with the help of giving examples.
3. Write an explanatory note on census and sampling.
4. Explain the various non-probability sampling methods.
5. Discuss the various types of probability sampling methods.
6. Briefly explain about the estimation with the help of giving examples.
7. Explain the two types of errors in statistical hypothesis.

5.11 FURTHER READING

NOTES

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

Chandan, J.S., Jagjit Singh and K.K. Khanna. 1995. *Business Statistics*. New Delhi: Vikas Publishing House.

Elhance, D.N. 2006. *Fundamental of Statistics*. Allahabad: Kitab Mahal.

Freud, J.E., and F.J. William. 1997. *Elementary Business Statistics – The Modern Approach*. New Jersey: Prentice-Hall International.

Goon, A.M., M.K. Gupta, and B. Das Gupta. 1983. *Fundamentals of Statistics*. Vols. I & II, Kolkata: The World Press Pvt. Ltd.

Gupta, S.C. 2008. *Fundamentals of Business Statistics*. Mumbai: Himalaya Publishing House.

Kothari, C.R. 1984. *Quantitative Techniques*. New Delhi: Vikas Publishing House.

Levin, Richard. I., and David. S. Rubin. 1997. *Statistics for Management.* New Jersey: Prentice-Hall International.

Meyer, Paul L. 1970. *Introductory Probability and Statistical Applications*. Massachusetts: Addison-Wesley.

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

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

Anderson, David R., Dennis J. Sweeney and Thomas A. Williams. *Essentials of Statistics for Business and Economics*. Mumbai: Thomson Learning, 2007.

S.P. Gupta. 2021. *Statistical Methods*. Delhi: Sultan Chand and Sons.