Bachelor of Computer Application

(B.C.A.)

Computer Applications in Statistics

Semester-II

Author- Surendra B. Ghatpande

SURESH GYAN VIHAR UNIVERSITY

Centre for Distance and Online Education
Mahal, Jagatpura, Jaipur-302025

Published by:

S. B. Prakashan Pvt. Ltd.

WZ-6, Lajwanti Garden, New Delhi: 110046

Tel.: (011) 28520627 | Ph.: 9205476295

Email: info@sbprakashan.com | Web.: www.sbprakashan.com

Designed & Graphic by : S. B. Prakashan Pvt. Ltd.

Printed at :

Syllabus
Computer Application in Statistics

Learning Objectives

- The students will learn the basic components of statistics and use of computer in the same.
- Develop the base acumen for using computer in real world statistical problems. \
- The students will develop the calculation skills with a base acumen of using computer application for the same.
- Will be able to resolve the statistical problem in the field of management, IT and ITES.

Unit 1

Statistical Population, Census, Sampling, Advantages of Sampling, Classification, Objectives of Classification, Basis of Classification, Types of Classification, Spreadsheet, Preparation of Frequency Distribution, Bar Chart, Pie Charl, Frequency Curves and Ogive Curves, Methods of Counting, Counting Techniques, Fundamental Principle of Counting, Multiplication Principle, Addition Principle.

Unit 2

Factorial Notation. Permutations, Combinations, Some Standard Results, Solved Examples.

Unit 3

Elements of Probability Theory, Sample Space and Events, Discrete Sample Space, Events, Occurrence of an event, Algebra of Events, Complementary Event, Union and intersection of Two Events, Union of Three Events, intersection of Three Events, De Morgan's Laws. Classical Approach of Probability, Probability of an Event: Equally likely Outcomes, Limitations of the Classical Definition, Axiomatic Approach, Probability of an Event: Properties, Independence of Standard Discrete Distributions.

Unit 4

Concept of a Random Variable, Discrete Random Variable, Probability Mass Function (PMF), Cumulative Distribution Function (CDF), Mathematical Expectation, Properties of Expectation, Variance, Standard Probability Distributions, Discrete Uniform Distribution, Cumulative Distribution Function (CDF), Bernoulli Distribution, Binomial Distribution.

Unit 5

Simulation Techniques, Random Number Generator, Monte Carlo Simulation, Model Sampling, from Discrete Distributions, Computer Aided Simulation, Merits and Demerits of Simulation.

References

1. R. A. Becker, J. M. Chambers, and A. R. Wilks. The New S Language. Chapman & Hall, Boca Raton, FL, 1988.
2. W. John Braun and Duncan J. Murdoch. A First Course in Statistical Programming with R. Cambridge University Press, Cambridge, 2007.
3. Cynthia A. Brewer. Designing Better Maps: A Guide for GIS Users. ESRI Press, Redlands, CA, 2005.
4. John M. Chambers. Data Management in S. Technical report, AT&T Bell Laboratories Statistics Research Report No. 99, 1991.
5. John M. Chambers. Programming with Data: A Guide to the S Language. Springer, New York, 1998.

Contents

***

Chapter 1

Introduction To Statistical Functions of Excel

1. Introduction

The word 'Statistics' is well known to us. We all are aware that, it has something to do with numerical figures, tables, averages, rates and so on. The usage of the word ***Statistics*** is now common in our daily life. Many a times we express the situation with the help of numbers and take appropriate decisions on the basis of available data.

For example,

1. We express the weather conditions by reporting minimum and maximum temperature of a day.
2. We talk about performance of a student using marks secured in the examination.
3. A cricket team is selected with the help of performance of batsmen, bowlers in various domestic and international cricket tournaments.
4. A person prefers a two wheeler automobile based upon the fuel efficiency that is, based upon the mileage per litre of petrol.

5. We come across statistics related to crimes, accidents as well as about literacy, population etc. through different media, by which we judge the socio-economic status of localities.

6. The individuals interested in investing money in stocks are keen about the prices of stocks as well as about the stock exchange indices.

7. 16 cricket players have scored over 8000 runs in One-Day Internationals.

8. 5,50,000 people in India die of cancer every year.

9. Percentage of pass students of S.S.C. board, March 2007 examination in Maharashtra increased by 6.

10. Seven percent of the world's population is lefties.

The above list can be expanded limitlessly. All the above quoted examples are statistical situations. In fact, in all walks of life the use of *Statistics* is inevitable. We are showered by lot of statistical information and it is necessary to make use of this information for the benefit of all. Thus, understanding and learning Statistics and statistical techniques is the need of today. We start with the historical background of Statistics.

2. Definitions

- "Statistics are the classified facts representing the collections of the people in a State ... specially those facts which can be stated in number or in tables of numbers or in any tabular or classified arrangements". ***Webster***

- "Statistics are numerical statement of facts in any department of enquiry placed in relation to each other". ***Bowley***

- "By Statistics we mean quantitative data affected to a marked extent by multiplicity of causes". ***Yule and Kendall***

- "The science of Statistics is essentially a branch of applied mathematics and may be regarded as mathematics applied to observational data". ***Sir R. A. Fisher***

- "Statistics may be defined as the aggregate of facts affected to a marked extent by multiplicity of causes, numerically expressed, enumerated or estimated according to a reasonable standard of accuracy, collected in a systematic manner, for a predetermined purpose and placed in relation to each other". ***Prof. Horace Secrist***

- "Statistics is the science of collection, presentation, analysis and interpretation". ***Croxton & Cowden***

3. Statistical Population

In any statistical investigation, it is important to identify a group or aggregate or total mass of individual items which is to be observed. This identification of group depends on the concrete definition of the problem. Such as,

- To estimate percentage of defectives in the mass production, the total production is the group to be observed.
- To study the socio-economic status of families in city, all the families residing in the city is the required group.
- To study performance of a motorcycle with respect to fuel efficiency of a particular company, all motorcycles running on the road of the company may be considered as the required total mass.
- In biodiversity studies, the total vegetation in a particular area is the required totality.

We can list number of such examples. In all above examples, total groups or totality are called as populations. In common dialect, the word population is used in limited sense and it means that an aggregate of particular kind of animates, e.g., number of lions in India, number of persons in India, number of students enrolled for a particular course, etc. But in Statistics, the word population is used in wider sense. It means an aggregate of individuals or objects or results of an experiment etc. under consideration. Hence statistical population may consists of group of men, animals or the set of non-living objects like villages, houses, cars or any manufactured product or agriculture produce. (Refer to above mentioned examples). Sometimes the word 'UNIVERSE' is used in place of word population.

Definition

An aggregate or totality of objects or individuals under consideration is called population or universe.

Number of questions asked: 1

Apr. 11 – *2M*
Define the term with illustration:
Population

Statistical Population

In the statistical investigation, it is assumed that every object in the population under consideration possesses certain quantitative or qualitative characteristics. The observations on these characteristics collectively called as *statistical population.*

An individual or object belonging to a population is called a member or an elementary unit.

Each elementary unit possesses certain characteristics, which may be qualitative or quantitative. One or more characteristics of elementary units may be of interest in a given population or problem. The result of study of characteristic of elementary unit is called as an observation. Hence, sometimes population may be defined as totality of observations of all elementary units.

For example,

1. In study of performance of students in an examination, marks of the students will form statistical population,
2. In the study of relationship between height and weight of normal adults, the observations on height and weight will form a statistical population.

A population containing finite number of members (elementary units) is called finite population, (No. of students in college). A population containing infinitely many elementary units is called infinite population. In many cases, the number of members in a population is so large as to be practically infinite. e.g., no. of leaves on a tree, no. of screws produced etc.

To study properties of population, data are collected by *census* or may be by *sampling methods*.

4. Census

In this method, each and every elementary unit in the population is examined and observations are recorded. It is also called as complete enumeration. Population census in India is conducted regularly at the interval of 10 years. The necessary information about every individual in the family is collected by enumerators visiting every household.

Limitations of Census Method

i. Census method provides reliable results, being complete enumeration. It is a voluminous work, therefore, it is expensive, time consuming. It is labourious, that is, it requires a large amount of manpower.

ii. In some cases census is not possible. If the objects under study are exhaustive or destructive in nature, then census is impracticable. *For example,* in blood examination, entire blood can not be tested. Similarly, in testing destructive power of explosives/ ammunition, testing life of an electronic component, etc. census method can not be used.

iii. If the population under study is infinite, census method can not be used.

5. Sampling

One of the characteristic features of science of Statistics is that, it is concerned with the study of properties of population and not with the study of properties of individuals. If we are required to study a particular characteristic, say X, of a certain population, it may be difficult to obtain information about X from every member of the population as it may be large and collection of information may mean huge expenditure in terms of time, money and labour. In such cases, a small group of elementary units, which represents the population in all its characteristics, is selected for observation. We collect the required information from such a small group of members of the population and from this information we infer about the properties of the population. This is known as the *sampling method*. A population from which a sample is drawn is called parent population. The purpose of sampling is to give the maximum information about the parent population with minimum efforts.

Definition

Any part of population under study is called sample.

Number of questions asked. **1**

Apr. 11 – *2M*
Define the term with illustration:
Sample

Illustrations

i. In medical diagnosis, blood examination is very common. For the purpose a small quantity of blood is taken from individual. This we call as blood sample. The blood sample represents the whole quantity of blood (blood population) in the body and represents all the characteristics of the population. Note that, here blood census is impracticable.

ii. While purchasing food grains, we inspect a handful of food grains (sample) from the lot of grains in a gunny bag (population). We infer about the quality of grains from this sample. It is not necessary to inspect the whole lot.

iii. In computing BSE Sensex, trading figures of stocks of a small group of listed companies are used. More than 8000 companies are listed in Bombay Stock Exchange.

iv. For market promotion of a newly developed product, sample surveys are conducted by selecting families from a city. These families are generally contacted by telephone. Sample is drawn using telephone directory of the city. Thus the list of telephone holders becomes the population.

v. A manufacturing company accepts lots from its vendors based on sample inspection. The lot submitted for acceptance is population. The rule may be if the sample contains less than or equal to "d" defectives, then accept the whole lot.

Note that, sampling is accepted as a way to collect required information and it is considered to be scientific procedure of selecting units from the population. A sample is a representative of the population in all its characteristics, therefore, it is necessary to draw a sample carefully using proper scientific methods.

6. Advantages of Sampling

In most of the cases the sample survey is beneficial for various reasons over census.

i. **Less time consuming:** A sample is a small part of the population. Thus, the volume of work is very small as compared to census. Thus, considerable time and labour are saved when sample survey is carried out. Time is saved not only in collecting data but also in processing it. Consequently, a sample provides more timely data in practice than census.

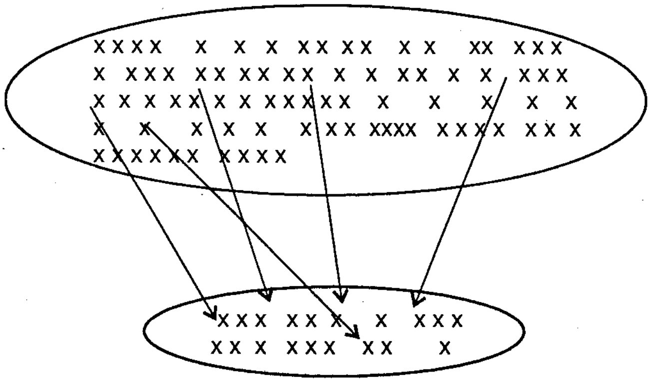
ii. **Less cost:** The cost per unit of collecting information from units in a sample is always more than that in case of complete enumeration. Even then the total financial burden of sample survey is generally very less than that of census. This is because a sample is a part of the population and hence the number of units to be examined is very less. This ensures less expenditure on sample survey.

iii. **More reliable results:** As compared to complete enumeration, a limited number of units are to be examined or processed. Therefore, it is possible of avail services of experts, well trained staff and use of modern techniques of measurements to increase accuracy of results. As volume of work is very less in sample survey, it can be completed efficiently and without fatigue.

iv. **Greater scope:** If the population under study is infinite or too large or cost per unit of examination is very high, census is impracticable. If the units in population are exhaustive or destructive in nature, then sampling is the only alternative to study the population. *For example*, to estimate life of an electric bulb, a sample of bulbs can be tested. Life test of a bulb means burning the bulb until it fuses off. A new remedial drug manufactured on a particular disease can not be tried on all the patients suffering from the disease. Instead a sample of patients can be treated to start with.

However, in certain cases population census is unavoidable. When the sampling errors are not permissible or the characteristic under study is a rare one that may not be properly reflected in the sample, complete enumeration becomes mandatory.

Population



Sample

7. Classification

In previous sections, we have seen process of data collection. In special statistical investigation for the given population, we are interested in some particular characteristic of the individual, that characteristic is referred to as 'characteristic under study'. Then the first step is to collect information on the characteristic under study from the individuals. The information is collected through surveys or experiments. The collected data are usually in a haphazard and unsystematic form. These data are not easy to comprehend, analyze and interpret about the population. These data are known as '**raw data**'. Hence it becomes necessary to arrange or organize data in a form, which is suitable for identifying groups of units, for comparison and for further statistical treatment or analysis of data.

In general, the raw data consist of observations on two types of characteristics, namely, descriptive and numerical.

For example,

i. a person is either male or female,

ii. a system is functioning or non-functioning,

iii. an individual is a senior citizen or not,

iv. a novel is good or bad,

v. income of a family,

vi. wholesale price index,

vii. height (in cm) of a person,

viii. percent marks secured by a student in an examination, etc.

Observe that, the characteristics described in (i) – (iv) will be descriptive and that in (v) – (viii) will be numerical in nature. Therefore, we need to use of proper methods while arranging the data in suitable form. In the following discussion, we learn how to arrange the data for further treatment.

The placement of data in different groups according to resemblance is called **Grouping.** Grouping the data and thus describing measurements and observations is the basis of statistical analysis.

Classification: The process of arranging raw data into homogeneous and non-overlapping groups of observations according to similarities and dissimilarities is called classification.

Classification helps to exhibit a characteristic properly. The collected information, that is, raw data are very complex, thus to make it easy to understand, classification is necessary.

7.1 Objectives of Classification

The principal objectives of classifying the data are:

i. To condense the mass of data in such form that similarities and dissimilarities can be readily observed.

ii. To arrange large number of figures in a few classes according to common features.

iii. To omit unnecessary details.

iv. To facilitate the comparison within and between data.

v. To bring out prominent figures in the data.

vi. To facilitate further analysis of data.

7.2 Basis of Classification

Data are classified on the basis of descriptive and numerical characteristics. Classification based on descriptive characteristic is known as qualitative classification or classification by attributes. While classification based on numerical characteristic is known as quantitative classification or classification by variables. These bases of classification give us different types of classification. They are described below:

7.3 Types of Classification

i. **Geographical Classification**: Grouping the data according to geographical areas, that is, according to cities, districts or states is called geographical classification. *For example*, number of accidents in Maharashtra can be classified with cities.

City	No. of accidents
Mumbai	5402
Nagpur	2371
Nashik	1259
Pune	3267

Geographical classifications are usually listed in alphabetical order. Geographical classification is essentially a type of qualitative classification, but is generally considered as a distinct classification.

ii. **Chronological Classification**: In this classification, data are arranged according to time, that is, weekly, monthly, quarterly, half yearly, annually, quinquennially, etc. Chronological data show figures concerning a particular phenomenon at various specified time points. Such data are also known as time series.

Year	Population (in crores)
1951	35
1961	44
1971	55
1981	68
1991	84
2001	100

iii. **Qualitative Classification**: The characteristic which is qualitative or descriptive in nature and can not be measured, that is, not expressible in numerical figures is called as an **attribute**. It is not possible to measure the attribute but its presence or absence in an individual can be observed. *For example*, Religion, literacy, gender, etc. If the data are classified on the basis of attributes, then the classification is called as qualitative classification.

Number of questions asked. **2**

Oct. 12, Apr. 11 – *2M*
Explain the following terms with illustrations: Attribute.

Qualitative classification is of two types.

a. Simple classification: If the units are classified in to two groups, one possessing a particular attribute and the other not possessing the attribute, then it is called simple classification. *For example*, a group of individuals can be divided into group of males and group of females or group of literates and group of illiterates, it is simple classification. It is also known as dichotomy. In numerical terms a lot of 1000 items has 12 defective items. Defective items – 12, non-defective items – 988.

b. Manifold classification: If the units are divided into more than two subclasses on the basis of presence and absence of the characteristic then it is known as manifold classification. *For example*, a group of literates can be further classified as educated up to primary level, secondary level, graduates and post graduates. Result of an examination of a college can be shown as follows:

Result	No. of students
First class with distinction	37
First class	164
Second class	180
Pass class	111
Fail	84

iv. Quantitative Classification: In this type, data are arranged according to certain characteristic that has been measured, *For example*, according to height, weight, or income of persons, vitamin content in substance, rainfall on a day etc.

Number of questions asked. 1

Oct. 2012 – *2M*
Explain the following terms with illustrations: Variable.

Variable: The characteristic which is quantitative in nature and can be measured directly by instruments or scales, that is, the characteristic which is expressible in numerical terms is called variable. *For example*, height, weight, temperature, income, marks etc.

Number of questions asked. 1

Apr. 11 – *2M*
Explain the following terms with illustrations: Discrete variable.

1. Discrete variable: The variable which takes some specified isolated values in a given range is called discrete variable. *For example*, no. of marks obtained in an examination paper, no. of misprints on a page of a book, no. of defectives in a lot of finished product, no. of heads in tossing a coin, no. of children in a family, etc. The classification of discrete variable can shown as follows:

No. of children in a family	0	1	2	3	4
No. of families	3	18	35	17	2

2. **Continuous variable:** The variable which takes all possible values in a given range is called continuous variable. *For example*, height, weight, temperature, rainfall, etc. Grouping of continuous variable is exhibited in following way:

Weight (kg.)	30 – 40	40 – 50	50 – 60	60 – 70	70 – 80
No. of students	76	94	131	56	18

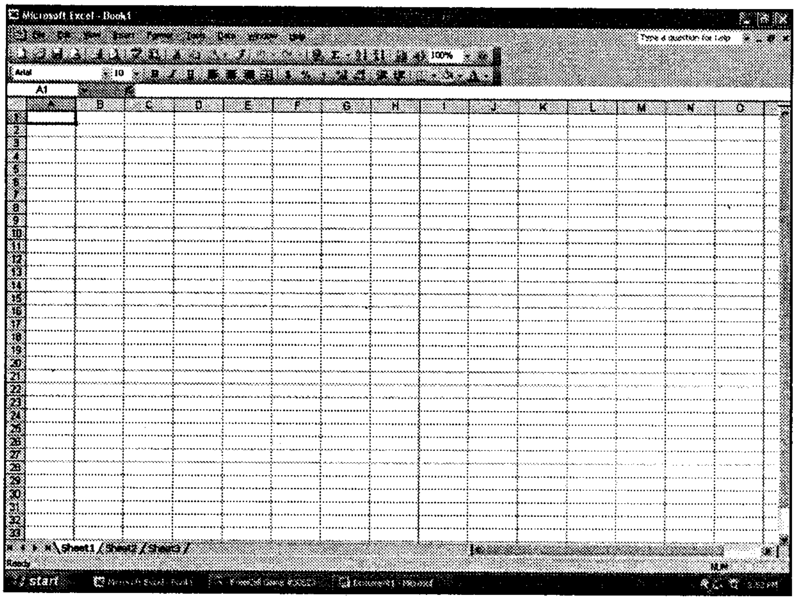
Both types of variables are generally denoted by capital alphabets A, B, C, …, X, Y, Z and their values by small alphabets a, b, c, …, x, y, z.

The classification based on these types of variables is known as Quantitative classification. Such a classification is represented in the form of frequency distribution. The tables showing classification of discrete variable and continuous variable are the corresponding frequency distributions.

8. Spreadsheet

A spreadsheet is an interactive computer application program for organization and analysis of data in tabular form. Initially spreadsheets developed as computerized simulation of paper accounting worksheets. Every sheet has arrays of cells, arranged in number of rows and columns. The program operates on data represented in cells organized in rows and columns.

A view of excel spreadsheet



One sheet of excel has 256 columns named as A, ..., IV and every column has 65536 cells. Hence, a sheet is an array of the size 256×65536. Each cell of the array is model-view-controller element that can contain either numeric or text data or the results of formulas that automatically calculate and display a value based on the contents of other cells.

The user can make changes in any stored value and observe the effects on calculated values. This makes the spreadsheet useful for "what-if" analysis since many cases can be rapidly investigated without tedious manual recalculations. Modern spreadsheets software can have multiple interacting sheets and can display data either as numerals or text or in graphical form.

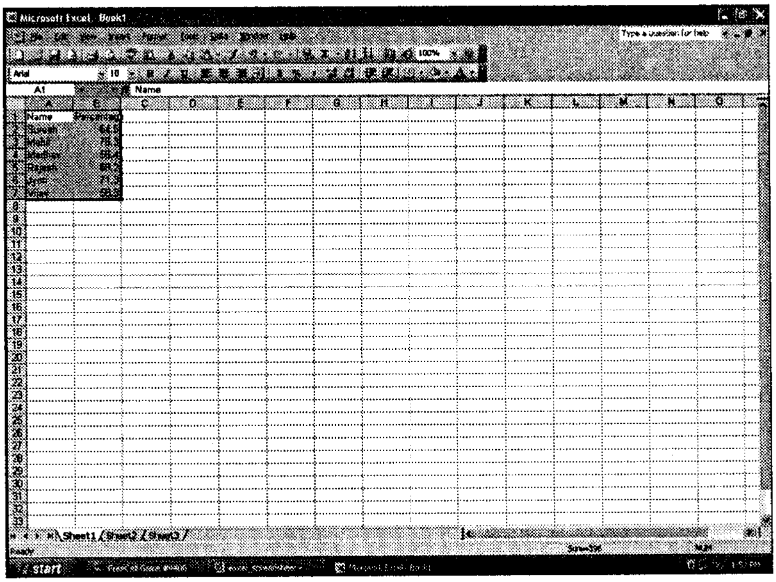
The modern spreadsheets provide built-in functions for common financial and statistical operations. *For example,* net present value or mean, median, mode, standard deviation etc. which can be applied to entered data with pre-programmed function in a necessitated formula. There are programs provided for conditional expressions, functions to convert text and numbers and functions that operate on strings of data.

Data Entry

Earlier it is stated that a cell can contain either numeric or text data. The user can enter the data in the cells as per the requirements. *For example*, consider the following table which gives names of students and their percentage scores in an examination.

Name	Percentage
Suresh	64.5
Mohit	76.3
Madhav	56.4
Rajesh	69.2
Jyoti	71.3
Vijay	58.3

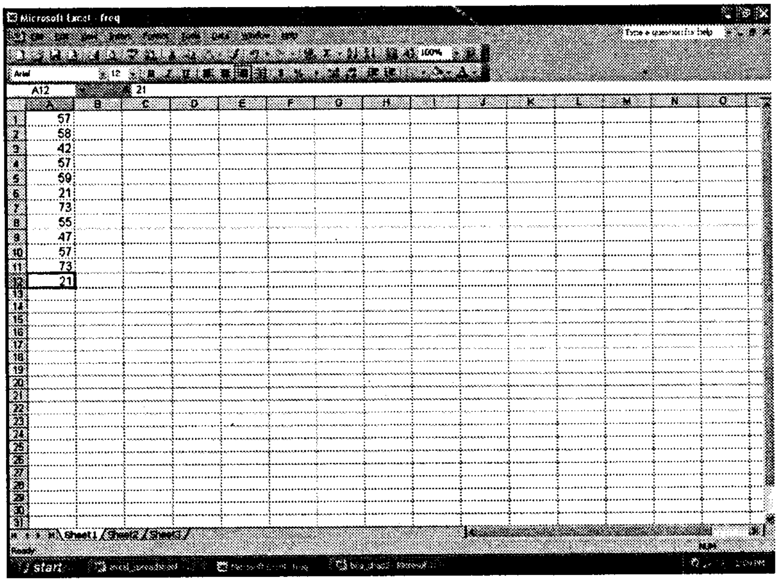
These data can be entered as follows: In the spreadsheet in column A in first row give the title 'Name' and in column B in first row give title 'Percentage'. Below the title 'Name' enter the names of the candidate and below title 'Percentage' enter the percentage scores. The spreadsheet will look as given below:



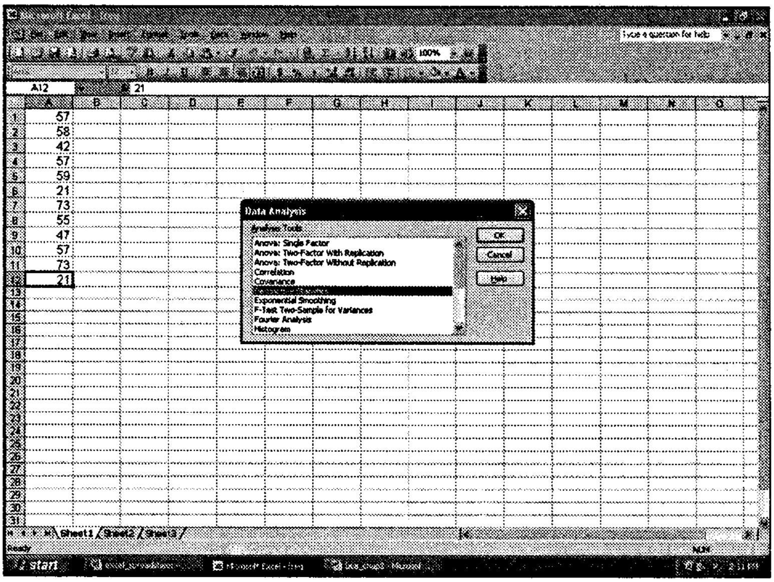
The data are written in 2 columns and in 6 rows. This is an illustration to show that text and numeric data can be entered simultaneously in a spreadsheet. However, it is not necessary that every time w have to enter text and numeric data simultaneously. Data can be entered in any column at any cell.

Summary Statistics

Consider the following data given in spreadsheet.



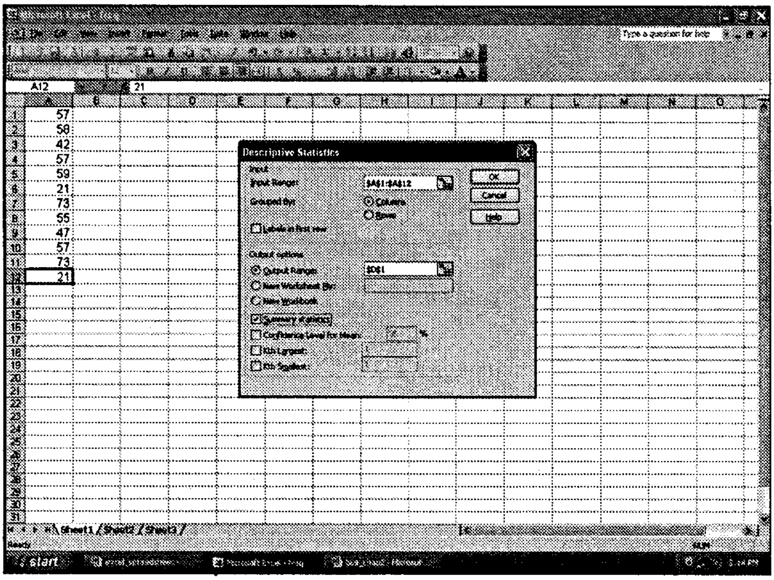
Twelve observations are entered in column A. We wish to obtain the descriptive statistics for these data.



Step 1: Select icon tools on the top of the spreadsheet.

Step 2: Click on 'Data analysis'.

Step 3: Data analysis window will be displayed. Select 'Descriptive Statistics'.



Step 4: In descriptive statistics window enter the required information.

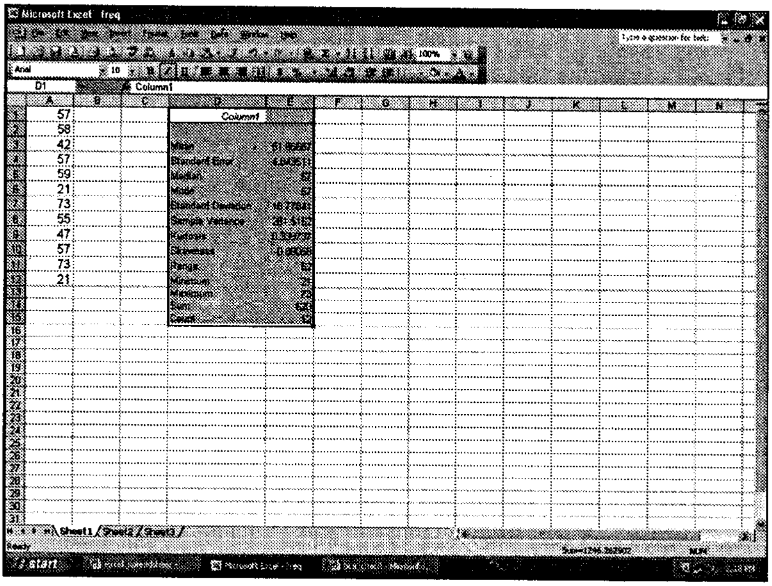
i. Input range. Here it is '$A$1:$A$12'.

ii. Group in column or rows. Here it is column.

iii. Click on output range. Click on a cell where you wish to get the results. Here it is cell '$D$1'.

iv. Then click on 'summary statistics'.

v. Then click on 'OK'. You will get the results in columns D and E. as follows:



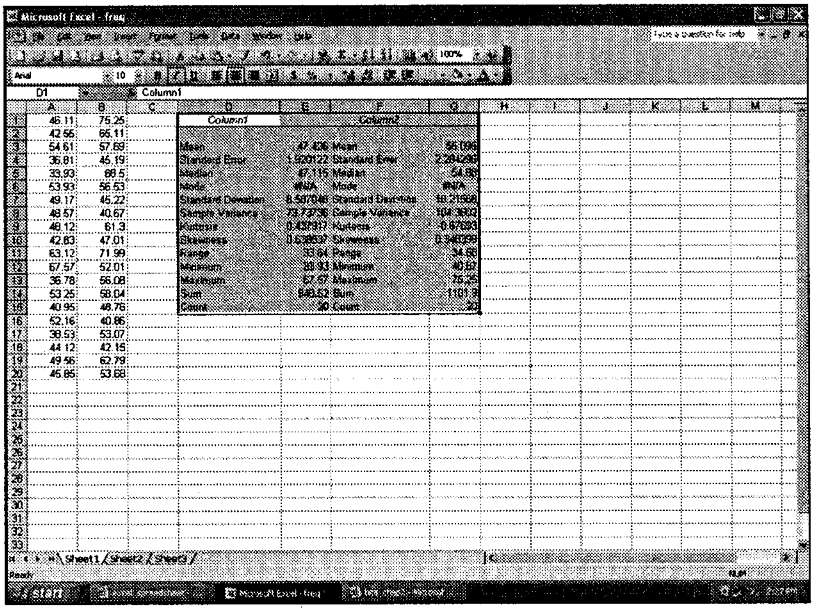
The results are as follows:

Column1	
Mean	51.66667
Standard Error	4.843511
Median	57
Mode	57
Standard Deviation	16.77841
Sample Variance	281.5152
Kurtosis	0.339737
Skewness	–0.88058
Range	52
Minimum	21
Maximum	73
Sum	620
Count	12

It is very clear that the summary statistics gives the results for Mean, Standard Error, Median, Mode, Standard Deviation, Sample Variance, Kurtosis, Skewness, Range, Minimum, Maximum, Sum, count related to given data.

This can be extended to any number of columns. However, the descriptive statistics results will be display for every column separately.

For example, consider



The above display shows the descriptive statistics results are shown separately for column 1 and column 2 and they are as follows:

Column1		Column2	
Mean	47.426	Mean	55.095
Standard Error	1.920122	Standard Error	2.284296
Median	47.115	Median	54.88
Mode	#N/A	Mode	#N/A
Standard Deviation	8.587046	Standard Deviation	10.21568
Sample Variance	73.73736	Sample Variance	104.3602
Kurtosis	0.437917	Kurtosis	-0.67693
Skewness	0.638537	Skewness	0.348399
Range	33.64	Range	34.58
Minimum	33.93	Minimum	40.67
Maximum	67.57	Maximum	75.25
Sum	948.52	Sum	1101.9
Count	20	Count	20

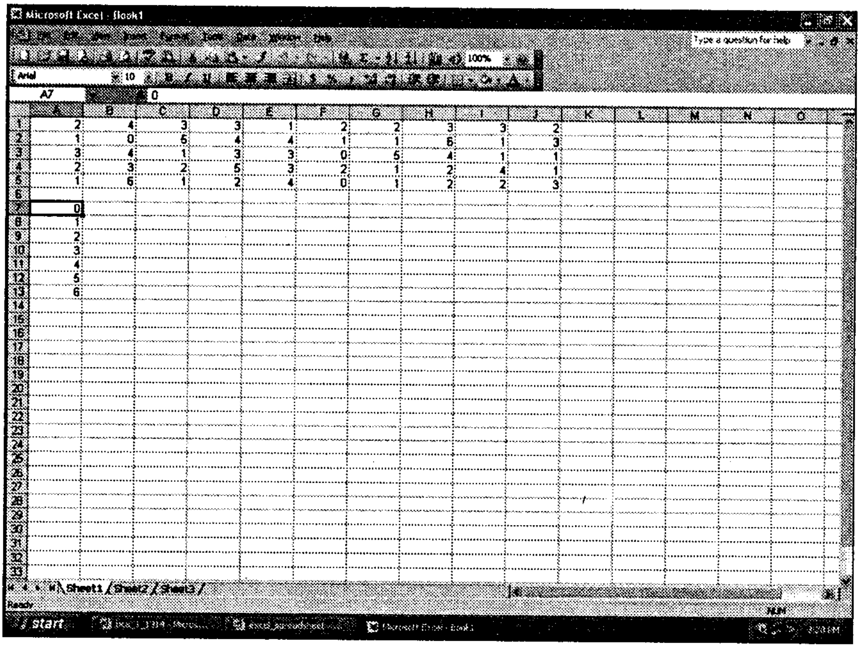
8.1 Preparation of Frequency Distribution

Ungrouped/ Discrete Frequency Distribution

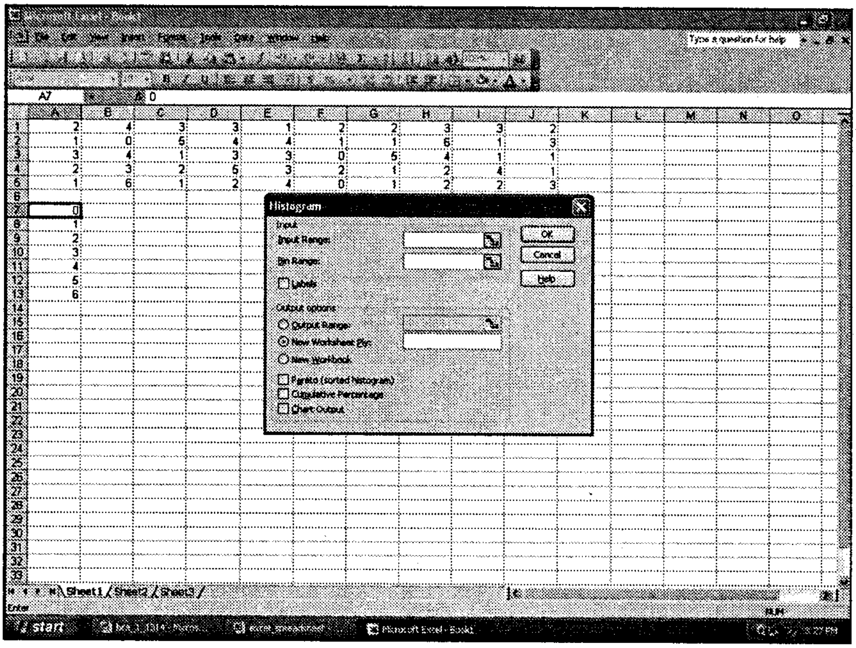
Consider the following data set of number of defectives in 50 lots each of 100 items.

2	4	3	3	1	2	2	3	3	2
1	0	5	4	4	1	1	6	1	3
3	4	1	3	3	0	5	4	1	1
2	3	2	5	3	2	1	2	4	1
1	6	1	2	4	0	1	2	2	3

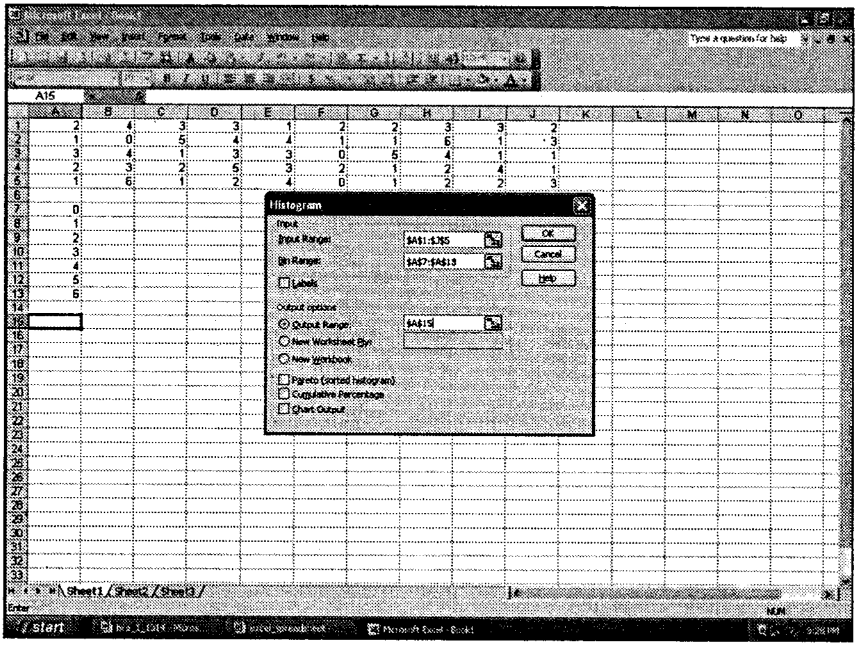
Enter these data as it is in Excel spreadsheet. Find minimum and maximum in the data. Enter the successive values in a column separating them from entered data. Here '0' is minimum and '6' is maximum. Thus, the digits 0, 1, 2, 3, 4, 5 and 6 are entered in column starting from cell A7. This is called as Bin Range using which frequency distribution will be prepared. It will look like picture given below:



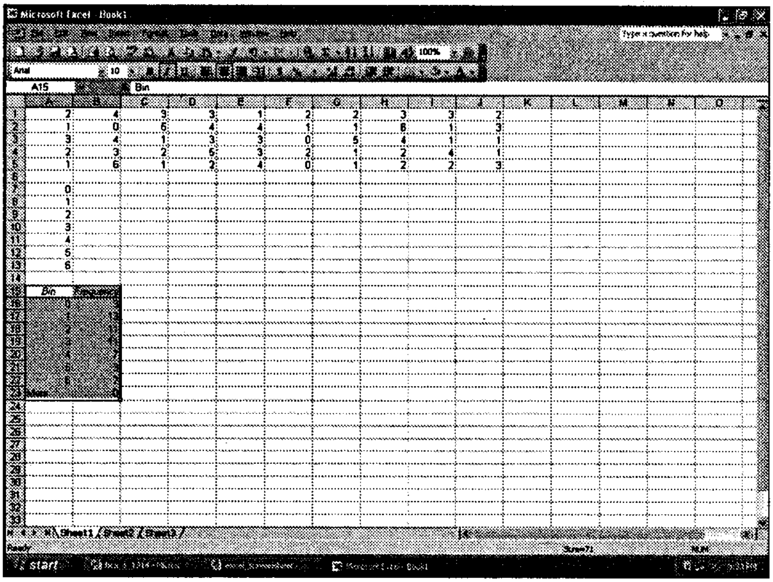
Then go to menu 'Tools', followed by 'Data Analysis'. In data analysis select 'Histogram'. You will get the window as follows:



You have to enter the input range '$A$1:j5' and in bin range the range of seed values 0 – 6, that is '$A$7:$A$13'. Then enter the output range that is indicating the position of the output, any empty cell such as '$A$15'. Then click on 'ok'. You will get discrete frequency distribution as follows:



Then click on 'ok'. You will get discrete frequency distribution as follows:



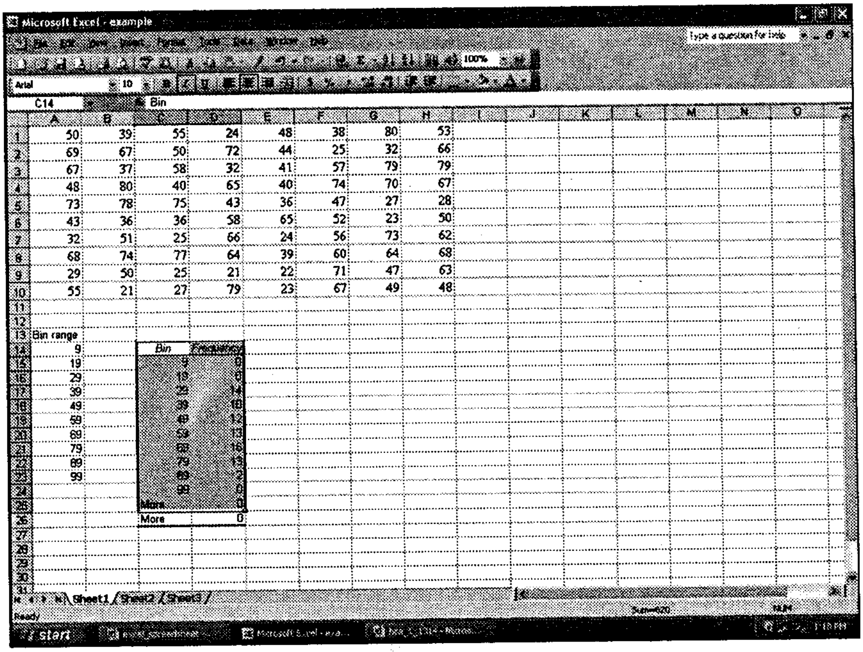
The result is as follows:

Bin	Frequency
0	3
1	13
2	11
3	11
4	7
5	3
6	2
More	0

Grouped or Interval Type Frequency Distribution

Consider following data set entered into a spreadsheet:

50	39	55	24	48	38	80	53
69	67	50	72	44	25	32	66
67	37	58	32	41	57	79	79
48	80	40	65	40	74	70	67
73	78	75	43	36	47	27	28
43	36	36	58	65	52	23	50
32	51	25	66	24	56	73	62
68	74	77	64	39	60	64	68
29	50	25	21	22	71	47	63
55	21	27	79	23	67	49	48



Follow the steps:

Click on to menu Tools.

Select "Data Analysis".

Select "Histogram"

Enter "Input Range".

Here it is "$A$1:$H$10".

Enter "Bin Range".

Here it is "$A$14:$A$23".

Specify "Output Range".

Here it is "$C$14".

Get the resulting frequency distribution.

'or the above data, we get the following result. The meaning of the same is given in adjoining olumns:

Bin	Frequency	Class	Frequency
9	0	0 – 9	0
19	0	10 – 19	0
29	14	20 – 29	14
39	10	30 – 39	10
49	12	40 – 49	12
59	13	50 – 59	13
69	16	60 – 69	16
79	13	70 – 79	13
89	2	80 – 89	2
99	0	90 – 99	0
More	0	100 and above	0

lote that spreadsheet always uses inclusive method of frequency distribution. This values specified n been range are used as upper limits. Any value equal to the bin value is included in the same class. 'hus we get different distributions with bin values 10, 20, 30, ... and 9, 19, 29, 39, ... When data ontain fractional values this problem may not arise.

f we use the bin range as 0, 10, 20, 30,... we get the following frequency distribution:

Bin	Frequency	Class	Frequency
0	0	0 – 10	0
10	0	10 – 20	0
20	0	20 – 30	0
30	14	30 – 40	14
40	12	40 – 50	12
50	14	50 – 60	14
60	10	60 – 70	10
70	16	70 – 80	16
80	14	80 – 90	14
90	0	90 – 100	0
More	0		

'he above frequency distribution is exclusive type, however the spreadsheet prepares it on the rinciples of inclusive method of frequency distribution.

8.2 Bar Chart

Bar chart is known as one dimensional diagram as only one axis is scaled for the values c characteristics under study. Other axis is used to represent qualitative characteristic. Enter the data i two columns, one containing the classes of characteristic and the other with values observed. Follov the steps given below:

i. Select the data that you want to display in the bar chart.

ii. On the Insert menu, click Chart.

iii. In the Chart type box, click Column or Bar. Note that by column we get vertical rectangle where Y axis is scaled for values and by bar we get horizontal rectangles where X axis i scaled for values.

iv. Under Chart sub-type, click any of seven choices of bars. These choices are to get simple ba diagram, multiple bar diagram or divided bar diagram. First choice is preferable as the data ar entered in one column only.

For a quick preview of the chart you are creating, click Press and Hold to View Sample.

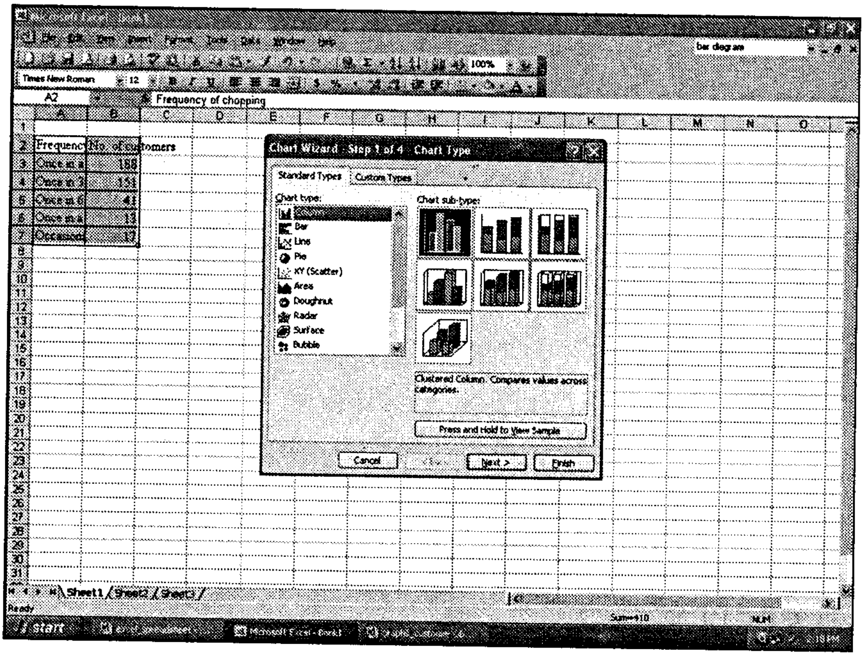
v. Click Next, and continue with Steps 2 through 4 of the Chart Wizard.

For help on any of the steps, click the question mark (?) on the Chart Wizard title bar.

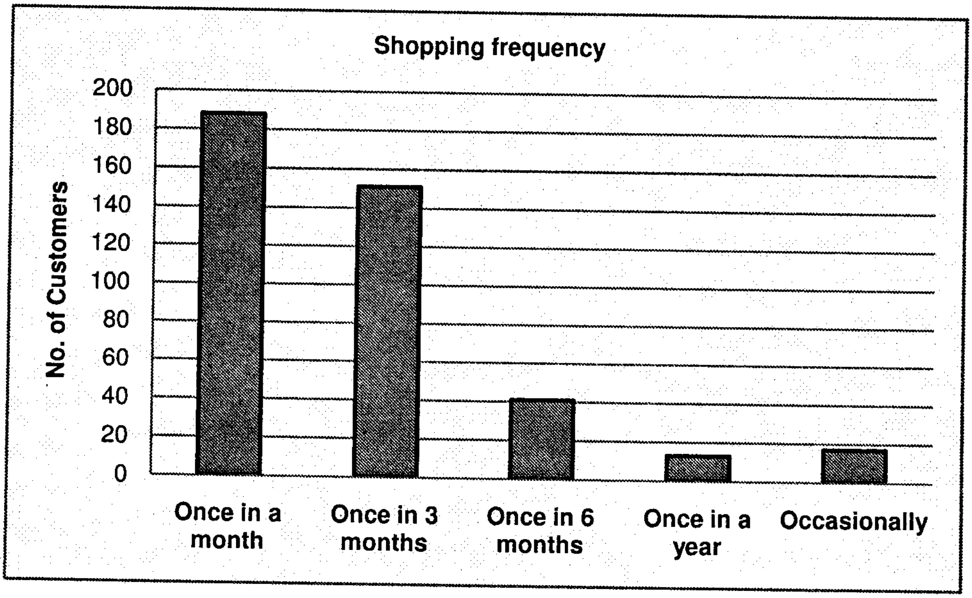
Consider the following data: In a survey related to cloths/ready wares purchasing habits of custome the following data were obtained:

Frequency of chopping	No. of customers
Once in a month	188
Once in 3 months	151
Once in 6 months	41
Once in a year	13
Occasionally	17

Enter these data in a spreadsheet in columns A and B and follows the steps 1 to 5 given above.



After following the given steps you get the following bar diagram:

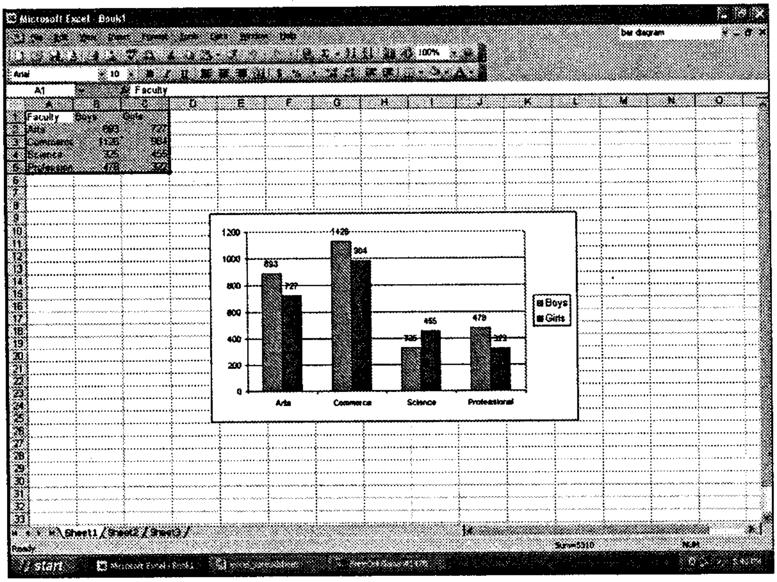


Multiple Bar Diagram

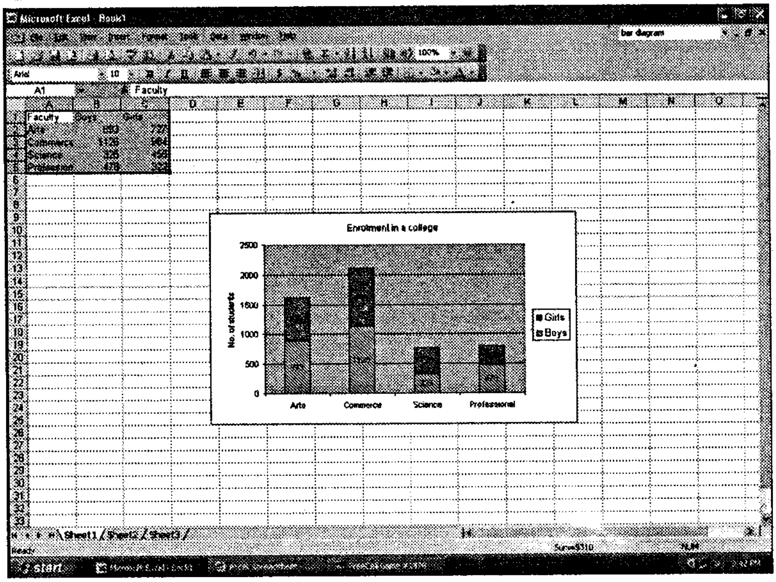
Consider the data set related to enrolment of students in a college. The data are entered into two columns with category name.

Faculty	Boys	Girls
Arts	893	727
Commerce	1126	984
Science	325	455
Professional	478	322

Using three columns we get the bar diagram as follows:



The same can be represented by divided bar diagram using proper choice of diagrams.



8.3 Pie Chart

Pie diagrams are very much popular in practice to show percentage breakdowns with sectors of a circle. The area of a sector is proportional to the percentage of associated category. Pie charts are excellent for displaying data points as a percentage of the whole. However, when several data points each amount to less than 5 percent of the pie, it becomes hard to distinguish the slices.

Create a Pie chart

i. Select the data that you want to display in the Pie chart.

ii. On the **Insert** menu, click **Chart.**

iii. In the **Chart type** box, click **Pie.**

iv. Under **Chart sub-type**, click **Pie of Pie.**

For a quick preview of the chart you are creating, click **Press and Hold to View Sample.**

v. Click **Next** and continue with Steps 2 through 4 of the Chart Wizard.

For help on any of the steps, click the question mark (?) on the **Chart Wizard** title bar.

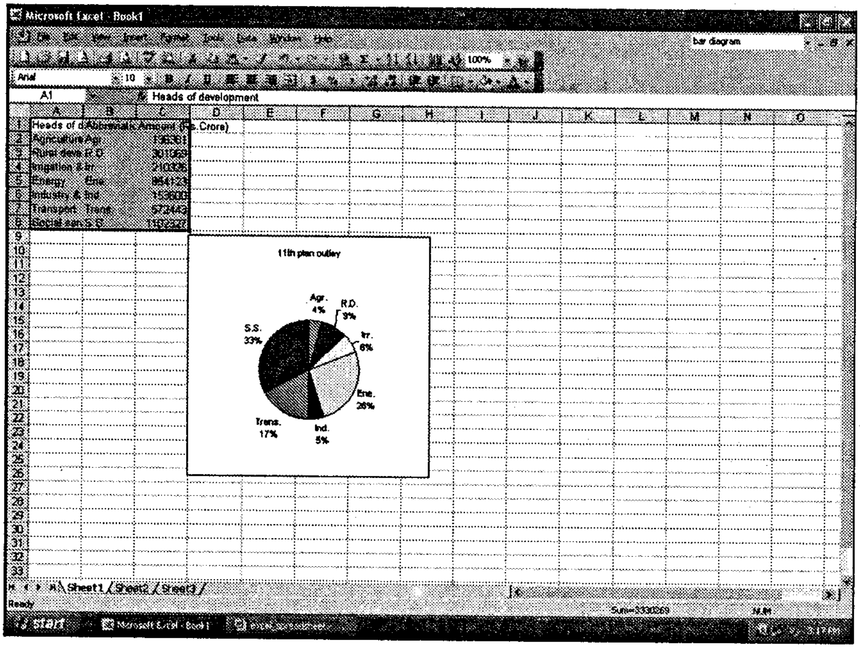
Note: Depending on how many decimal places are specified for percentages on the **Number** tab of the **Format Cells** dialog box (**Format** menu, **Cells** command), percentages that are displayed in data labels may be rounded so that they don't add up correctly.

Consider the following data on outlay on different heads of development in 11^{TH} 5 year plan given by Government of India.

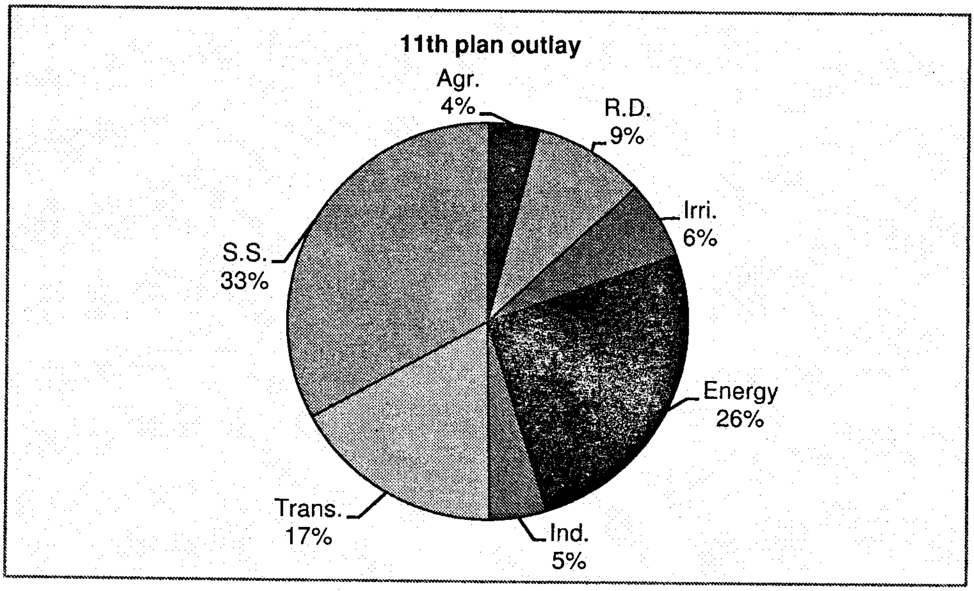
Heads of development	Abbreviations	Amount (Rs.Crore)
Agriculture	Agr.	136381
Rural development	R.D.	301069
Irrigation & Flood control	Irr.	210326
Energy	Ene.	854123
Industry & Minerals	Ind.	153600
Transport	Trans.	572443
Social services	S.S.	1102327

Source: Planning Commission of India

Following the above steps we get pie diagram as follows:



Copied diagram is as follows:



8.4 Frequency Curves and Ogive Curves

Consider a frequency distribution of students according to their heights (in cm). The following table shows frequencies, class marks and cumulative frequencies.

Frequency distribution of heights of students

Class Boundaries	Frequency	Class marks	CFLT/CF↓	CFMT/CF↑
130-135	12	132.5	12	400
135-140	25	137.5	37	388
140-145	37	142.5	74	363
145-150	60	147.5	134	326
150-155	108	152.5	242	266
155-160	80	157.5	322	158
160-165	52	162.5	374	78
165-170	26	167.5	400	26

How we shall see how can we draw frequency curve and ogive curves. To get frequency curve enter the data points class marks and associated frequencies in spreadsheet. Then follow the steps as given below:

Create a Frequency Distribution

i. Select the data that you want to display in the Pie chart.

ii. On the **Insert** menu, click **Chart**.

iii. In the **Chart type** box, click Xy (scatter).

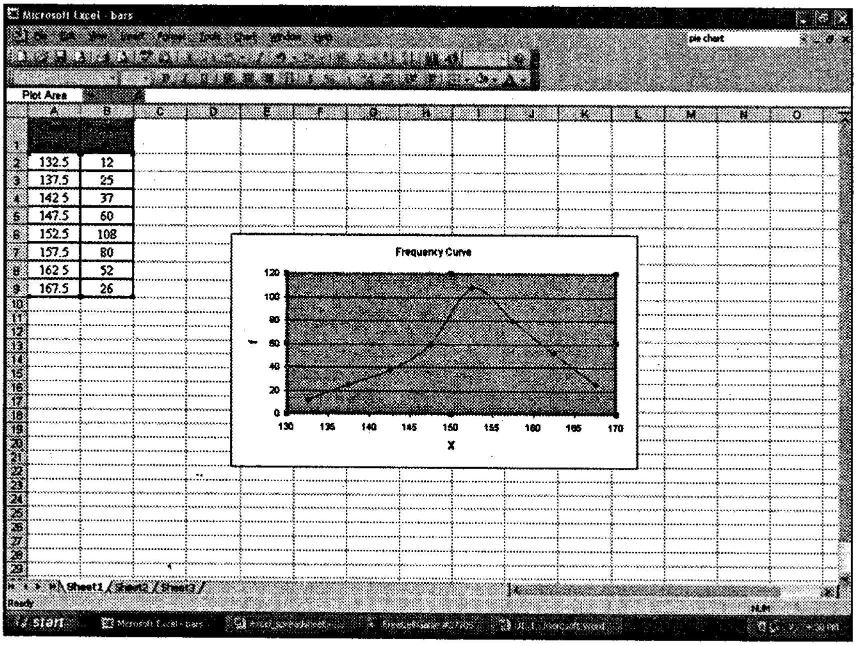
iv. Under **Chart sub-type**, click diagram number 2, indicating scatter with data points connected by smoothed lines.

For a quick preview of the chart you are creating, click **Press and Hold to View Sample**.

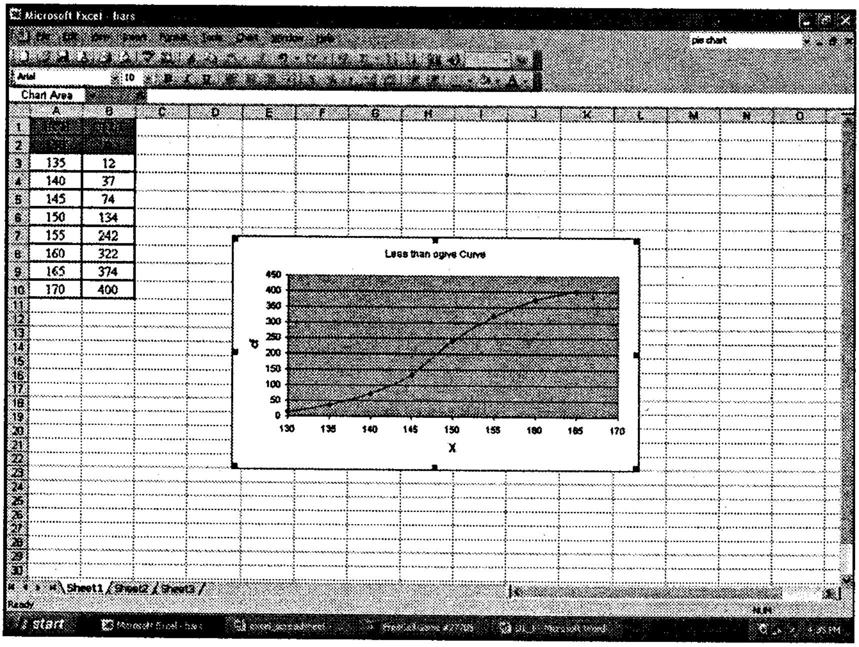
v. Click **Next** and continue with Steps 2 through 4 of the Chart Wizard.

For help on any of the steps, click the question mark (?) on the **Chart Wizard** title bar.

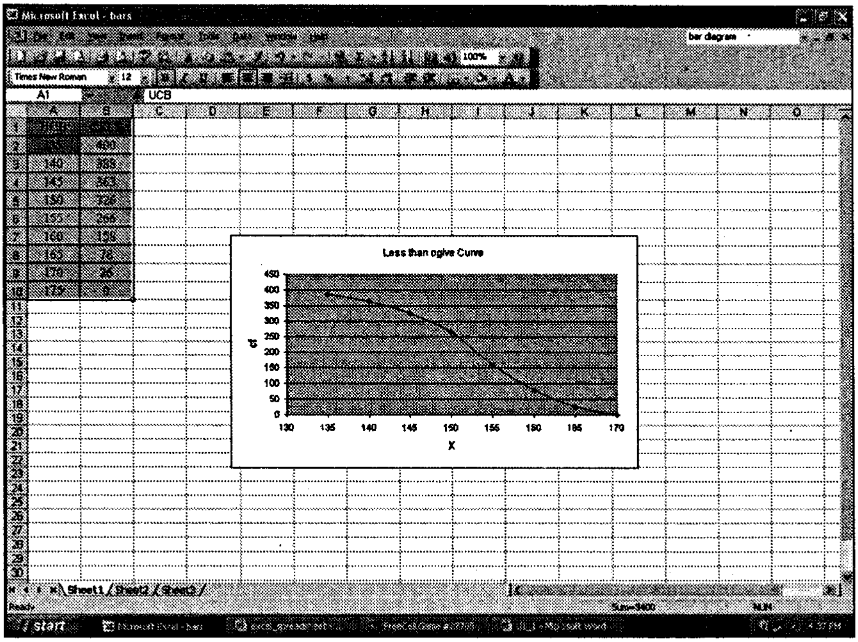
For the above frequency distribution we get the frequency curve as follows:



Similarly for less than ogive curve, enter data points viz. upper class boundary (x coordinate) and corresponding less than cumulative frequency (y coordinate) in two columns. Proceed as in case of frequency curve to get less than ogive. The following picture shows less than ogive curve.



For more than ogive curve, enter data points viz. lower class boundary (x coordinate) and corresponding more than cumulative frequency (y coordinate) in two columns. Proceed as in case of frequency curve to get more than ogive. Following picture shows more than ogive curve:



EXERCISES

A. Fill in the blanks:

1. The word statistics is used in _________ senses namely _________ and _________.
2. The word statistics is derived from the word _________.
3. _________ is the real giant in the development of the theory of statistics.
4. A sample is a study of _________ of the population.
5. A sample can be drawn from the population using either _________ or _________ method.
6. Random sampling is also referred to as _________ sampling.

B. State whether the following statements are true or false:

1. Statistics deals with aggregate of facts.
2. All facts numerically expressed are statistics.
3. Census is always preferable over sample survey.
4. A sample is less expensive than a census.

5. The results obtained in complete enumeration are always more reliable than that obtained in sample survey.
6. Classification is the first step in analysis of data.
7. Classification is the process of arranging data in different rows.
8. In chronological classification data are classified on the basis of time.
9. Number of persons in a family is a discrete variable.
10. Age of a person is a continuous variable.
11. An attribute is a non-measurable characteristic.

C. Answer in brief:

1. Define 'Statistics'.
2. What is probability sampling?
3. What is 'census'?
4. Explain the term 'statistical population'.
5. Explain the terms "finite population" and 'infinite population'.

D. Answer the following:

1. Explain the importance of statistics.
2. Discuss the merits of sampling over census.
3. Explain with illustrations the concept of Statistical Population and sample from population.
4. "Sampling is a necessity under certain conditions". Illustrate giving suitable examples.
5. Explain the terms: classification, qualitative classification, quantitative classification, variable, attribute, discrete variable, continuous variable.
6. Explain the terms: class limits, class boundaries, frequency, class width, frequency density, relative frequencies.
7. Explain the need of classification.
8. Write a note on inclusive and exclusive methods of frequency distribution.
9. What is a spreadsheet?

E. Numerical

1. Following are the data on number of accidents on a highway in a day. Prepare a frequency distribution for the following data taking classes 0, 1, 2, ...

4	2	3	2	3	2	2	6	5	4	3	4	6	4	3	6
3	5	4	4	5	5	3	1	2	6	4	5	1	6	3	2
4	3	5	7	2	4	3	3	1	5	4	3	2	0	4	4
5	2	4	5	4	5	6	3	1	7	4	6	1	3	4	1
3	1	6	1	4	4	5	3	2	4	4	2	3	1	2	4

2. Prepare a frequency distribution for the following data taking classes 0, 1, 2, ...

1	2	5	5	6	6	2	0	4	5
4	2	3	1	4	4	1	2	6	2
3	3	1	3	5	1	5	4	4	5
7	3	2	1	5	1	3	7	3	1
4	1	2	4	4	2	3	1	1	5

3. Prepare interval type frequency distribution for the following data taking classes 0-10, 10-20, 20-30, 30-40 ...

46.11	75.25	63.12	71.99	47.56	42.15	31.78	45.07	65.94	34.00
42.55	65.11	67.57	52.01	53.37	51.73	32.49	37.52	46.19	58.64
54.61	57.69	36.78	56.08	71.08	42.45	62.61	51.52	55.72	46.36
36.81	45.19	53.25	58.04	46.86	41.52	57.50	47.08	62.50	76.85
33.93	68.50	40.95	48.76	41.01	54.92	42.70	64.53	56.49	56.47
53.93	56.53	52.16	40.86	45.14	55.85	40.61	50.70	55.20	41.77
49.17	45.22	38.53	53.07	45.60	40.19	27.88	54.25	49.14	68.47
48.57	40.67	44.12	42.15	6.07	60.53	50.91	58.14	75.06	72.08
48.12	61.30	49.56	62.79	59.21	38.36	43.45	36.62	28.76	43.88
42.83	47.01	45.85	53.68	48.55	53.65	44.02	52.87	45.69	53.60

4. Prepare a frequency distribution for the following data classes 20-29, 30-39, 40-49, 50-59, ...

63	49	69	30	60	48	99	66
86	84	63	90	55	31	40	83
84	46	73	40	51	71	99	99
60	99	50	81	50	93	88	84
91	98	94	54	45	59	34	35
54	45	45	73	81	65	29	63
40	64	31	83	30	70	91	78
85	93	96	80	49	75	80	85
36	63	31	26	28	89	59	79
69	26	34	99	29	84	61	60

5. Draw frequency curve and ogive curves using MS-EXCEL for the following frequency distributions:

a.

Class	10- 20	20- 30	30- 40	40- 50	50- 60	60-70	70- 80
Frequency	1	2	10	39	30	12	6

b.

Class	20-29	30-39	40-49	50-59	60-69	70-79	80-89
Frequency	14	10	12	13	16	13	2

ANSWERS

A.

1. two, singular, plural
2. status
3. Sir R. A. Fisher
4. part
5. non-probability sampling
6. probability sampling
7. probability

B.

1.	True	2.	False	3.	True	4.	False	5.	True	6.	True
7.	True	8.	False	9.	True	5.	True	6.	True	7.	True

E.

5.

Class	0	1	2	3	4	5	6	7
Frequency	1	9	12	16	21	11	8	2

6.

Class	0	1	2	3	4	5	6	7
Frequency	1	11	8	8	9	8	3	2

7.

Class	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
Frequency	1	0	2	10	39	30	12	6

8.

Class	20-29	30-39	40-49	50-59	60-69	70-79	80-89	90-99
Frequency	5	9	10	8	13	7	15	13

PU Questions

[Oct. 2012, 2011 - *4M*] 1. Explain the term Population and Sample with one illustration of each.

[Apr. 2011 - *4M*] 2. Define the following terms with illustration:

i. Population ii. Sample

Chapter 2

METHODS OF COUNTING

1. Counting Techniques

Many of the basic concepts of probability theory may be considered in the context of finite sample description spaces. It in turn requires the study of mathematical technique of counting. In this chapter, we shall be dealing with the elementary ideas of counting of number of possible outcomes of an experiment, arranging and grouping r objects taken at a time out of n objects. Consider the following problems:

i. Two balls are to be drawn from an urn containing six balls of different colours. In how many ways can such a sample be drawn?

ii. A departmental store has twelve different varieties of bags and each variety comes in four sizes and six colours. Find the total number of different bags that the departmental store would have to stock if

   a. all varieties of bags in all the sizes and colours are to be displayed;

   b. only three varieties are to be displayed,

   c. if only three varieties are to be displayed only in two colours and three sizes.

iii. Consider a configuration of ten electrons and four orbits. Enumerate the total possible configurations if

a. no orbit is empty,

b. one, two or three orbits may be empty.

Above types of problems can be answered if we apply the fundamental principle of counting.

2. Fundamental Principle of Counting

In the following section we shall study two basic principles of counting, namely, multiplication principle and addition principle. We will learn how to apply these principles to calculate counting numbers. These principles are helpful in computations of probability when listing elements of sample space becomes very difficult as the number of sample points may be very large.

2.1 Multiplication Principle

If a procedure can be performed in n_1 ways and another procedure can be performed in n_2 ways, then these two procedures can be performed in the same order in $n_1 \times n_2$ ways.

Examples

1. Suppose a student has a choice of five books in the subject Economics and four books in the subject Sociology then he has $5 \times 4 = 20$ different choices of selecting two books of different subjects. It is possible to construct a table showing all possible pairs.

E_1S_1	E_1S_2	E_1S_3	E_1S_4
E_2S_1	E_2S_2	E_2S_3	E_2S_4
E_3S_1	E_3S_2	E_3S_3	E_3S_4
E_4S_1	E_4S_2	E_4S_3	E_4S_4
E_5S_1	E_5S_2	E_5S_3	E_5S_4

In general, if we have several procedures which can be performed in specific order, one followed by the other, we have $n_1 \times n_2 \times n_3 \times n_4 \times \ldots \times n_k$ number of ways to perform the procedures.

2. A departmental store has twelve varieties of bags, in four sizes and in six colours. So if each type of bag is to be displayed, the store must stock $12 \times 4 \times 6 = 288$ different bags.

2.2 Addition Principle

If procedure A can be performed in n_1 ways and procedure B can be performed in n_2 ways, then there are $n_1 + n_2$ ways in which one can perform either procedure A or procedure B. We assume that no two selections can be carried out simultaneously.

Examples

1. Five books are recommended for a course in Statistics and Eight for the course in Mathematics. An undergraduate student can borrow a book from either of these books. A student can choose the book in $5 + 8 = 13$ number of ways.

   In general, if we have several procedures and we can perform only one, in all possible alternatives, then there are $n_1 + n_2 + \ldots + n_k$ ways in which one can perform any one of the procedures. We assume that no two or more selections can be carried out simultaneously.

2. Six different routes are recommended for amateur trekkers in Kulu-Manali Valley; three routes only for driving through the same valley and four "tough" routes for experienced trekkers only. A group of students, therefore, has $6 + 3 + 4 = 13$ choices for selection of a route.

2.3 Factorial Notation

The product of the positive numbers from 1 to n is called as 'n factorial' or 'factorial n'. It is denoted by 'n!'. Thus $n! = 1 \cdot 2 \cdot 3 \ldots (n-2)\,(n-1)\,n$.

Examples

1. $3! = 1 \cdot 2 \cdot 3 = 6,$ $\quad 4! = 1 \cdot 2 \cdot 3 \cdot 4 = 3! \cdot 4 = 24,$

   $5! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 = 4! \cdot 5 = 120,$ $\quad 6! = 5! \cdot 6 = 720$

2. $\dfrac{9!}{6!} = \dfrac{9 \cdot 8 \cdot 7 \cdot 6!}{6!} = 9 \cdot 8 \cdot 7 = 504,$ $\quad 13 \cdot 12 \cdot 11 \cdot 10 = \dfrac{13 \cdot 12 \cdot 11 \cdot 10 \cdot 9!}{9!} = \dfrac{13!}{9!}$

3. Permutations

Number of questions asked. 1

Oct. 2012 – 2M
Define the following term: Permutations

A Permutation of n dissimilar elements (objects) taken r at a time ($r \leq n$): The number of ways in which r dissimilar ('distinct') elements can be arranged from a set of n elements, that is, the number of permutations of n distinct elements taken r at a time, is $n(n-1)(n-2) \ldots [n-(r-1)]$.

The number of permutations of n distinct elements taken r at a time will be denoted by ${}^nP_r$ or P_r^n or $P(n, r)$. We can write

$${}^nP_r = n(n-1)(n-2)\ldots(n-r+1) = \frac{n(n-1)(n-2)\ldots(n-r+1)(n-r)!}{(n-r)!} = \frac{n!}{(n-r)!}$$

As a special case for $r = n$, ${}^nP_n = n!$.

Note that these are the number of distinct ways in which all the elements of a set of n elements can be arranged. In this discussion, it is assumed that the elements are distinct.

Examples

1. Consider a set of four letters {a, b, c, d}. Then we have

   i. abcd, acbd, bdca, cabd, dabc, adcb, etc. are permutation of 4 letter taken all at time.

   ii. abc, acd, bcd, dac, dbc, cab, cdb, bda, etc. are permutations of 4 letter taken 3 at a time.

   iii. ab, ac, bd, cd, ca, bc, db, etc. are permutations of 4 letters taken 2 at a time.

---

2. **How many three digit numbers can be formed from the six digits 2, 4, 5, 6, 8 and 9; if each digit is to be used only once? How many of these are divisible by five?**

Solution

A three digit number can be formed from the given six digits is clearly ${}^6P_3 = 120$. A number is divisible by five only if it is a multiple of five.

It means the digit at units place (out of above six digits) can be selected only in one way, that is choose the digit 5. However first two digits can be written in ${}^5P_2$ ways.

Now using multiplication principle we claim that the total number of 3-digit numbers that are divisible by five will be equal to ${}^5P_2 \times 1 = \left(\frac{5!}{3!}\right) \times 1 = 20$.

A Permutation of Elements with Repetitions

The number of permutations of n elements (objects, things, symbols) of which n_1 are alike, n_2 others are alike, …, and n_k are alike is given by $P(n; n_1, n_2, \ldots, n_k)$, provided $n_1 + n_2 + \ldots + n_k = n$. In notation

$$P(n; n_1, n_2, \ldots, n_k) = \frac{n!}{n_1!\, n_2! \ldots n_k!} \text{ where } \sum_{i=1}^{k} n_i.$$

Examples

1. **Enumerate the set of all possible "words" using all the letters of the word**

   i. RAM ii. ODD.

   Here we consider any combination of given letters as a word.

Solution

i. Word is RAM. It has three letters R, A, M. The set of all possible words is {RAM, RMA, MRA, MAR, ARM, AMR}.

ii. Word is ODD. It is a three letter word. The set of all possible words is {ODD, DDO, DOD}.

2. **How many arrangements can be made using all the letters of the word**

   i. MADAM ii. STATISTICS iii. MISSISSIPPI.

Solution

Here

i. $n = 5, n_1 = 2(M), n_2 = 2(A), n_3 = 1(D)$.

$\therefore$ No. of distinct arrangements $= \dfrac{5!}{2!\,2!\,1!} = \dfrac{120}{4} = 30$

ii. $n = 10, n_1 = 3(S), n_2 = 3(T), n_3 = 2(I), n_4 = 1(A), n_5 = 1(C)$.

No. of distinct arrangements $= \dfrac{10!}{3!\,3!\,2!\,1!\,1!} = \dfrac{3628800}{(6)\,(6)\,(1)\,(1)} = 100{,}800$

iii. $n = 11, n_1 = 1(M), n_2 = 4(I), n_3 = 4(S), n_4 = 2(P)$

No. of distinct arrangements $= \dfrac{11!}{1!\,4!\,4!\,2!} = \dfrac{39916800}{(1)\,(24)\,(24)\,(2)} = 34{,}650$

3. Find the number of ways in which six boys and four girls can be asked to sit in a row so that no two girls may be put together?

Solution

The six boys can be asked to sit in ${}^6P_6 = 6!$ distinct ways. Corresponding to any one arrangement of boys, we have seven positions shown below by * for four girls.

1		2		3		4		5		6		7
*	Boy_1	*	Boy_2	*	Boy_3	*	Boy_4	*	Boy_5	*	Boy_6	*

Out of these seven positions, girls can be placed in any four positions in ${}^7P_4$ ways. Hence the required answer is

$${}^6P_6 \times {}^7P_3 = 6! \times \frac{7!}{3!} = 6{,}04{,}800$$

Asked. Number of Questions 1

Apr. 2010 – 4M

4. A family of 4 brothers and 3 sisters is to be arranged for a photograph in one row. In how many ways can they be seated, if no two sisters sit together?

Solution

The four brothers (B) can be asked to sit in ${}^4P_4 = 4! = 24$ distinct ways. Corresponding to any one arrangement of brothers, we have five positions shown below by * for three sisters.

1		2		3		4		5
*	$Brother_1$	*	$Brother_2$	*	$Brother_3$	*	$Brother_4$	*

Out of these five positions, sisters can be placed in any three positions in ${}^5P_3$ ways. Hence the required answer is ${}^4P_4 \times {}^5P_3 = 4! \times \frac{5!}{2!} = 24 \times 60 = 1440$ ways.

5. A committee consisting of twelve members visits a metropolitan city to investigate the changing scenario with respect to traffic problems. All the members are required to stay in a guest house. The manager of the guest house informs that he has only one triple, three double and three single rooms available. In how many ways all the ten participants can be assigned to the available rooms?

Solution

The total number of ways the twelve participants can be assigned to one triple, three double and three single rooms is given by

$$P(12; 3, 2, 2, 2, 1, 1, 1) = \frac{12!}{3!\,2!\,2!\,1!\,1!\,1!\,1!} = 39{,}916{,}800$$

6. How many four digit numbers can be formed by using digits 1, 3, 5, 6, 7, 8 and 9 (no digit should be repeated)? How many of these will be greater than 3,000?

Number of questions asked 1

Apr. 2010 – *4M*

Solution

We want to form a four digit number. There are in all 6 digits.

Consider

Thousand	Hundred	Ten	Unit

We can place any one of the digit in thousand's place, which is possible in 6 ways. Now for hundred's place we have remaining 5 digits, hence it can be filled in 5 ways, then for ten's place we have 4 digits, therefore it can be filled in 4 ways and for units place there are 3 digits, hence it can be filled in 3 ways.

By multiplication principle, the total number of permutations is $6 \times 5 \times 4 \times 3 = 360$. Hence, we can form 360 different four digit numbers. Alternatively, 4 digit numbers out of 6 digits can be formed in $^6P_2 = \frac{6!}{(6-4)!} = \frac{6!}{2!} = \frac{720}{2} = 360$.

Now, if the digit at thousand's place is greater than or equal to 3 then the number will be greater than 3000. There are 5 digits greater than or equal to 3, thus we can place any one these five digits in thousand's place, which is possible in 5 ways. Now for hundred's place we have remaining 5 digits, hence it can be filled in 5 ways, then for ten's place we have 4 digits, therefore it can be filled in 4 ways and for units place there are 3 digits, hence it can be filled in 3 ways. By multiplication principle, the total number of permutations is $5 \times 5 \times 4 \times 3 = 300$. Hence we can form 300 different four digit numbers which are greater than 3000.

4. Combinations

Number of questions asked 1

Oct. 2012 – *2M*
Define the following term: Combinations

When using formula ^nP_r, we are concerned with the order in which the elements of subsets of r elements are selected from a set of n elements. e.g., Consider permutations of two of the first three letters of the alphabets. These are listed below:

1	2	3	4	5	6
AB	AC	BA	BC	CA	CB

There are 6 different permutations of letters A, B and C taken two at time. However, if we are interested only in the number of distinct subsets and not in order in which these elements are arranged, it should be noted that among the above six permutations there are only three distinct subsets, namely, {A, B}, {A, C}, {B, C}.

In general, there are 'r!' permutations of the elements of a subset r and the ${}^nP_r$ permutations of 'r' elements selected from a set of 'n' contain, therefore, each subset 'r!' times. Therefore to get the number of distinct ways in which subsets of 'r' elements can be selected from a set of 'n' elements, we divide ${}^nP_r$ by r!. This is what we call as combination of the n objects, taken r at a time ($r \leq n$). We denote it by ${}^nC_r$ or C(n, r) or ${}_nC_r$.

Hence $${}^nC_r = \frac{{}^nP_r}{r!} = \frac{n!}{(n-r)!\,r!}$$ where n is a positive integer and r = 0, 1, …, n.

Let us denote by $n_{(r)} = n(n-1)\ldots(n-r+1)$. Then we can express ${}^nC_r$ as follows:

$${}^nC_r = \frac{n(n-1)(n-2)\ldots(n-r+1)(n-r)!}{(n-r)!\,r!} = \frac{n(n-1)(n-2)\ldots(n-r+1)}{r!} = \frac{n_{(r)}}{r!}$$

Note that, in a permutation order is taken into account; in a combination, order is not taken into account. By studying the nature of the problem we will be required to decide whether permutations or combinations are involved.

Examples

1. Interpret $\mathbf{{}^nC_n = 1 = {}^nC_0}$

Solution

It simply means that selection or non-selection of all the n or 0 objects out of n objects can be done in only one way. Recall that $0! \equiv 1$.

2. Find r, if $\mathbf{{}^nP_r = 24 \times {}^nC_r}$

Solution

$${}^nP_r = 24 \times \frac{{}^nP_r}{r!} \Rightarrow r! = 24 = 4 \times 3 \times 2 \times 1 \Rightarrow r = 4$$

3. Interpret $\mathbf{{}^nC_r = {}^nC_{n-r}}$. Hence or otherwise show that if $\mathbf{{}^{27}C_r = {}^{27}C_{r-3}}$ then the value of $\mathbf{{}^rC_{12} = 455}$.

Solution

Selection of r objects out of n, also means non-selection of n – r objects out of n.

Therefore ${}^nC_r = {}^nC_{n-r}$. Observe that $${}^nC_r = \frac{n!}{r!\,(n-r)!} = \frac{n!}{(n-r)!\,r!}$$

Hence, $r + (r-3) = 27 \Rightarrow r = 15$ and $${}^{15}C_{12} = \frac{15!}{12!\,3!} = \frac{15 \times 14 \times 13}{3 \times 2 \times 1} = 455$$

4. (Geometrical Application) There are twelve points on a plane, no three of which are in the same line. How many straight lines can be obtained by joining the points? How many triangles?

Solution

We may select any two points to draw a straight line. This can be done in $^{12}C_2 = \dfrac{12 \times 11}{2 \times 1}$ different ways. That is, 66 straight lines can be drawn.

Similarly $^{12}C_3 = 220$ triangles can be drawn.

5. A committee of 5 persons is to be formed from a group of 9 men and 6 women. Find the number of ways in which the committee can be formed. Also find the number of ways so that the committee contains

i. only one woman, ii. at least three women.

Solution

Total number of persons is 15. A Committee of 5 is to be formed. Thus, we can select any 5 persons out of 15, which is possible in $^{15}C_5 = \dfrac{15!}{5!\,10!} = 3003$.

i. Only one woman in the committee; means there will be 4 men in the committee. 4 men can be selected from 9 men in $^9C_4$ and one woman out of 6 women in $^6C_1$ ways, hence the committee can be formed in $^9C_4 \times {}^6C_1 = 756$ ways.

ii. At least three women are in the committee. The committee can be formed as follows: Select 3 women and 2 men, this can be done in $^6C_3 \times {}^9C_2 = 720$ ways or select 4 women and one man, it can be done in $^6C_4 \times {}^9C_1 = 135$ ways or select all five women, this can be done in $^6C_5 \times {}^9C_0 = 6$ ways. Hence the committee consisting of at least three women can be formed using $720 + 135 + 6 = 861$ ways.

6. Prove that $^nC_r + {}^nC_{r-1} = {}^{n+1}C_r$.

Solution

Consider

$$
\begin{aligned}
{}^nC_r + {}^nC_{r-1} &= \frac{n!}{r!\,(n-r)!} + \frac{n!}{(r-1)!\,(n-r+1)!} \\
&= \frac{n!}{(r-1)!\,(n-r)!}\left[\frac{1}{r} + \frac{1}{n-r+1}\right] \\
&= \frac{n!}{(r-1)!\,(n-r)!}\left[\frac{n-r+1+r}{r\,(n-r+1)}\right]
\end{aligned}
$$

$$= \frac{n!}{(r-1)!\,(n-r)!}\left[\frac{n+1}{r\,(n-r+1)}\right]$$

$$= \frac{(n+1)!}{r!\,(n-r+1)!} = {}^{n+1}C_r$$

5. Some Standard Results

i. We have seen that ${}^nC_r = {}^nC_{n-r}$

ii. ${}^nC_r + {}^nC_{r-1} = {}^{n+1}C_r$

The result is important. It can be used to prepare the famous Pascal's triangle.

i. ${}^mC_n = {}^{m-1}C_{n-1} + {}^{m-1}C_n \;;\quad 0 \le n \le m.$

ii. ${}^{-n}C_r = (-1)^r\, {}^{n+r-1}C_r$

iii. ${}^{-1}C_r = (-1)^r$

iv. $1 \cdot 3 \cdot 5 \,.\, \cdots (2n-1)\,(2n-3) = \frac{(2n)!}{2^n(n!)}.$

v. $\sum_{r=0}^{n} {}^nC_r = 2^n$

vi. $\sum_{r=0}^{n/2} {}^nC_{2r} = \sum_{r=1}^{n/2} {}^nC_{2r-1} = 2^{n-1}$

vii. For any integer $n \ge 2$, $\sum_{r=0}^{n} r \cdot {}^nC_r = n2^{n-1}$

viii. For any positive integers m, n, k, $\sum_{r=0}^{k} {}^mC_r \times {}^nC_{k-r} = {}^{m+n}C_k$

Solved Examples

1 Find ${}^7C_4$

Solution

${}^7C_4 = {}^6C_3 + {}^5C_3 + {}^4C_3 + {}^3C_3 = 20 + 10 + 4 + 1 = 35$

2. Compute values of:

i. $^5P_3$ **ii.** $^9C_3$ **iii.** $^7C_3$ **iv.** $^7C_3 \times {}^8C_2$

Number of questions asked. 1

Apr. 2011 – 4M

Solution

i. $$^5P_3 = \frac{5!}{(5-3)!} = \frac{5!}{2!} = \frac{5 \times 4 \times 3 \times 2!}{2!} = 60.$$

ii. $$^9C_3 = \frac{9!}{3!\,(9-3)!} = \frac{9!}{3!\,6!} = \frac{9 \times 8 \times 7}{3 \times 2 \times 1} = 84.$$

iii. $$^7C_3 = \frac{7!}{3!\,(7-3)!} = \frac{7!}{3!\,4!} = \frac{7 \times 6 \times 5}{3 \times 2 \times 1} = 35.$$

iv. $$^7C_3 \times {}^8C_2 = 35 \times \frac{8!}{2!\,(8-2)!} = 35 \times \frac{8!}{2!\,6!} = 35 \times \frac{8 \times 7}{2 \times 1}$$

$$= 35 \times 28 = 980.$$

3. There are 12 points in a plane of which 5 are collinear. Find number of triangles that can be formed using these points.

Number of questions asked. 1

Apr. 2011 – 4M

Solution

If all 12 points were non-collinear, then we could have triangles by selecting any three points out of 12 points, that is, there would be $^{12}C_3 = \frac{12!}{3!\,9!} = \frac{12 \times 11 \times 10}{3 \times 2 \times 1} = 220$ triangles. Since 5 points are collinear, selecting three points from these 5 would not form triangle. Hence, the number will reduce by

$C_3 = \frac{5!}{3!\,2!} = \frac{5 \times 4 \times 3}{3 \times 2 \times 1} = 10$. Thus, the number of triangles formed will be $220 - 10 = 210$.

4. How many four digit numbers can be formed by using digits 0, 1, 2, 3 and 4 (no digit should be repeated)? How many of these numbers will be divisible by 5?

Number of questions asked. 1

Apr. 2010 – 4M

Solution

To form a four digit number, we cannot have '0' at thousands place, therefore, the number of ways in which 4 digit number can be formed $= 4 \times 4 \times 3 \times 2 = 96$. A number is divisible by 5 if its units place digit is either 5 or 0. Therefore, the number of ways in which a number is divisible by '5' is given by $= 4 \times 3 \times 2 \times 1 = 24$.

Number of questions asked 1

Oct. 2010– *4M*

5. Find values of

i. $^7P_4$ **ii.** $^7C_3$ **iii.** $^8C_5$ **iv.** $^4C_2$

Solution

i. $^7C_4 = \frac{7!}{3!} = 840$ ii. $^7C_3 = \frac{7!}{3!\,4!} = 35$

iii. $^8C_5 = \frac{8!}{5!\,3!} = 56$ iv. $^4C_2 = \frac{4!}{2!\,2!} = 6$

Number of questions asked 1

Oct. 2010 – *4M*

6. There are 7 gents and 4 ladies. A committee of 4 is to be formed. Find number of committees that can be formed so as to include at least one lady.

Solution

There are 11 persons of which 7 are gents and 4 ladies. The number of ways to form a committee of 4 with at least one lady, then the committee would consist 3 gents and 1 lady or 2 gents and 2 ladies or 1 gents and 3 ladies or all 4 ladies. Thus, the committee can be formed in
$^7C_3\,{}^4C_1 + {}^7C_2\,{}^4C_2 + {}^7C_1\,{}^4C_3 + {}^7C_0\,{}^4C_4 = 140 + 126 + 28 + 1 = 295$ ways.

Number of questions asked 1

Apr. 2010 – *4M*

7. Compute values of

i. $^7P_2$ **ii.** $^{10}C_2$ **iii.** $^8C_4 \times {}^5C_2$

Solution

We have $^nP_r = \frac{n!}{(n-r)!}$ and $^nC_r = \frac{n!}{r!\,(n-r)!}$

Therefore,

i. $^7P_2 = \frac{7!}{(7-2)!} = \frac{7!}{5!} = \frac{7 \times 6 \times 5!}{5!} = 7 \times 6 = 42$

ii. $^{10}C_2 = \frac{10!}{2!\,(10-2)!} = \frac{10!}{2!\,8!} = \frac{10 \times 9 \times 8!}{2!\,8!} = \frac{10 \times 9}{2} = 45$

iii. $^8C_4 \times {}^5C_2 = \frac{8!}{4!\,(8-4)!} \times \frac{5!}{2!\,(5-2)!}$

$$= \frac{8!}{4!\,4!} \times \frac{5!}{2!\,3!} = \frac{8 \times 7 \times 6 \times 5}{4 \times 3 \times 2 \times 1} \times \frac{5 \times 4}{2 \times 1} = 70 \times 10 = 700$$

EXERCISES

1. Explain the difference between permutation and combination.
2. Prove that ${}^nC_r + {}^nC_{r-1} = {}^{n+1}C_r$.
3. Mr. Devendra possesses two pairs of shoes, five trousers, and ten shirts. How many combinations can he wear?
4. Three books are recommended for the courses in Statistics, five for the course in Mathematics while four for Ecology. A student can put his demand for a book from the above books. How many choices does he have?
5. How many three letter words can be formed from the letters of the word "COMPUTER"?
6. How many arrangements can be made using all letters of the word
   - i. YIPPEE
   - ii. STATISTICS
   - iii. SOCIOLOGICAL
   - iv. INDEPENDENT
   - v. SYLLOGISM
   - vi. HETEROSCEDASTICITY
   - vii. ASSASSINATION
7. Using 7 consonants and 5 vowels, how many words consisting of 4 consonants and 3 vowels can be formed?
8. Find the number of ways in which six boys and two girls can be asked to sit in a row so that will not sit together.
9. If repetitions are not permitted
   - i. How many three digit numbers can be formed from the six digits 2, 3, 5, 6, 7 and 9?
   - ii. How many of them are less than 400?
   - iii. How many are even?
   - iv. How many are odd?
   - v. How many are multiples of 5?
10. Suppose an urn contains eight balls. Find the number of ordered samples of size 3 when sampling is done
    - i. with replacement
    - ii. without replacement.

11. Find n if

i. ${}^{n}P_2 = 72$ ii. ${}^{n}P_4 = 2\ {}^{n}P_2$ iii. $2\ {}^{n}P_2 + 50 = {}^{2n}P_2$

iv. ${}^{n}P_r = 5040$ and ${}^{n}C_r = 210$ v. ${}^{n}P_3 = {}^{n-1}P_3 + 3\ {}^{7}P_2$

12. A student is to answer eight out of ten questions in an examination.

i. How many choices has he?

ii. How many if he must answer the first three questions?

iii. How many if he must answer at least four of the first five questions?

13. Find the number of four letter words that can be formed from the letters of the word `HISTORY'.

i. How many of them contain only consonants?

ii. How many of them begin and end in consonant?

iii. How many of them begin with vowel?

iv. How many of them begin with T and end in a vowel?

v. How many of them begin with T and contain S?

vi. How many of them contain only vowels?

14. Seven politicians meet at a party. How many handshakes are exchanged if each politician shakes hands with every other politician once and only once?

15. From 8 statisticians, 15 economists and 20 biologists, a committee consisting of 5 members is to be formed. Find the number of ways in which it can be formed, if

i. exactly 2 statisticians are to be included

ii. at most 2 biologists are to be included.

16. An urn contains 4 white and 6 black balls. 4 balls are drawn randomly without replacement from the urn. In how many ways can this be done so as to include

i. exactly 3 white balls? ii. at least 3 white balls?

iii. at most 3 black balls?

17. Write expressions for ${}^{-n}C_r$, ${}^{-1}C_r$.

ANSWERS

3. 100

4. 12

5. ${}^8P_3 = 336$

6. i. 180 ii. 50400 iii. 9979200
   iv. 554400 v. 90720 vi. 18!/(3! 3! 2! 2! 2!)
   vii. 10810800

7. ${}^{21}C_7 \times {}^7C_4 \times {}^5C_3 \times {}^7P_7$

8. 30240

9. i. 120 ii. 40 iii. 40 iv. 80 v. 20

10. i. 512 ii. 336

11. i. 9 ii. 4 iii. 5 iv. 10 v. 8

12. i. 45 ii. 21 iii. 35

13. 840 i. 120 ii. 400 iii. 240 iv. 40 v. 20 vi. 0

14. 21

15. i. 183260 ii. 547239

16. i. 24 ii. 25 iii. 195

PU Questions

[Oct. 2012 - 4M] 1. Define the following terms:
i. Permutations ii. Combinations

[Oct. 2012 - 4M] 2. Find n if: i. ${}^7P_3 = n \cdot {}^7C_3$ ii. ${}^nP_{n-3} = 84$

[Oct. 2011 - 4M] 3. Define Permutation and Combination. Hence show that $\binom{n}{r} = \binom{n}{n-r}$.

[Oct. 2010 - 4M] 4. How many four digit numbers can be formed by using digits 0, 1, 2, 3 and 4 (no digit should be repeated)? How many of these numbers will be divisible by 5?

[Oct. 2010 - 4M] 5. Find values of i. ${}^7P_4$ ii. ${}^7C_3$ iii. ${}^8C_5$ iv. ${}^4C_2$

[Oct. 2010 - 4M] 6. There are 7 gents and 4 ladies. A committee of 4 is to be formed. Find number of committees that can be formed so as to include at least one lady.

[Apr. 2010 - 4M] 7. Compute values of i. ${}^7P_2$ ii. ${}^{10}C_2$ iii. ${}^8C_4$ iv. ${}^5C_2$

[Apr. 2010 - 4M] 8. How many four digit numbers can be formed by using digits 1, 3, 5, 6, 7, 8 and 9 (no digit should be repeated)? How many of these will be greater than 3,000?

[Apr. 2010 - 4M] 9. A family of 4 brothers and 3 sisters is to be arranged for a photograph in one row. In how many ways can they be seated, if no two sisters sit together?

[Oct. 2009 - 4M] 10. A family of 4 brothers and 3 sisters is to be arranged for a photograph in one row. In how many ways can they be seated if
i. All the sisters sit together?
ii. No two sisters sit together?

[Oct. 2009 - 4M] 11. Define Permutations and Combinations. Hence show that ${}^nC_r = \frac{{}^nP_r}{r\,!}$.

[Apr. 2009 - 4M] 12. A person has 12 friends of whom 8 are relatives. In how many ways can he invite 7 guests such that 5 of them are relatives?

[Apr. 2009 - 4M] 13. Define Permutation and Combination. Hence show that ${}^nC_r = \frac{{}^nP_r}{r!}$.

[Apr. 2009 - 4M] 14. There are 12 points in a plane of which 5 are collinear. Find the number of triangles that can be formed using these points.

Chapter 3

ELEMENTS OF PROBABILITY THEORY

1. Introduction

In daily life, we talk about chances, such as chance of rain today, chance of winning a cricket match, chance of passing in an examination, chance of winning a lottery, so on. These type of happenings are associated with uncertainties. The study of such phenomena in a scientific manner is the study of probability. In other words, probability is the study of experiments in which the result cannot be predicted in advance with certainty. These experiments are called as random or *non-deterministic* experiments. In particular, an experiment with more than one possible outcomes and whose result cannot be predicted in advance is called a *random Experiment. For example,* if a coin is tossed in air, then it is certain that the coin will come down, but it is not certain that it will show 'head'. It may show 'tail' as well. Thus, the result of the experiment is not predictable in advance. However, if we repeat this experiment of tossing a coin 'n' number of times and we observe 'm' number of heads, then the ratio m/n, called as relative frequency becomes stables in long run, that is, approaches a limit ½. This stability is the basis of probability theory.

2. Sample Space and Events

The first step in the study of random experiments is to specify the set 'S' of all possible outcomes of the experiment under consideration.

- In the experiment of tossing a coin, we have S = {Head, Tail}.
- In the experiment of rolling a die, S = {1, 2, 3, 4, 5, 6}.
- In the experiment noting suit of a card drawn from a pack of playing cards, S = {Club, Diamond, Heart, Spade}.

Number of questions asked 3

Oct.12,11,Apr.10–*1/2M*
Define Sample space.

Sample space: The set of all possible outcomes of a random experiment E is called *sample space* of the experiment. It is denoted by 'S'.

Each individual outcome of E, is called a *point* or a *sample point*. It is denoted by 'ω'.

The basic object we have is thus a sample point 'ω'. The collection of all sample points is called a sample space S. S may have a finite number of sample points or countably infinite sample points.

For example, in the random experiment of tossing of a coin S = {Head, Tail}. S has only two sample points and thus is a finite sample space.

Now, think of an experiment of tossing a coin till you observe head. In this experiment, you may get head at the first toss or at the second toss or at the third toss or perhaps do not observe it at all. Clearly S = {H, TH, TTH, ...}. The sample points in S are countable, but not finite. Such a sample space is called a countable infinite sample space. We can also think of S which does not have either finite or countable infinite sample points.

Examples

Number of questions asked 1

Oct. 2010 – *8M*

1. Write Sample space for each of the following experiments:

i. A card is drawn from a well shuffled pack of playing cards and the suit is noted.

ii. Number of defectives are noted from a lot of 10 items.

iii. A coin is tossed until a head appears first time.

iv. Two dice are tossed and upper-most faces are noted.

Solution

i. $S = \{\text{Spade, Heart, Diamond, Club}\}$

ii. $S = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$

iii. $S = \{H, TH, TTH, TTTH, TTTTH, \ldots\}$

iv. $S = \{(1, 1), (1, 2), \ldots, (6, 6)\}$

---

2. List elements of sample space for the following experiments:

Number of questions asked: 1
Apr. 2010 – *8M*

i. A coin is tossed 3 times.

ii. A student attempts an examination till he passes.

iii. Ten seeds are planted and total numbers of seeds germinated are recorded.

iv. Two cards are drawn from a pack of playing cards and colour is noted.

Solution

i. Sample space Ω = {HHH, HHT, HTH, THH, HTT, THT, TTH, TTT}

where H: head, T: tail

ii. Let P denote that the student passes and F denote that he fails in the examination, then

Sample space Ω = {P, FP, FFP, FFFP, FFFFP, …}

iii. Sample space $\Omega = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$

iv. There are only two colours red (R) and black (B)

Sample space Ω = {RR, RB, BR, BB}

In above example 1, sample spaces in sub-questions (i), (ii) and (iv) are finite sample spaces and sample space in (iii) is countable infinite sample space. Similarly in example 2, sample space in sub-questions (i), (iii) and (iv) are finite sample spaces and that in (ii) is countable infinite sample space.

Number of questions asked: 1
Oct. 11 – *1M*
Define:
Discrete sample space

2.1 Discrete Sample Space

A sample space containing a finite number of points or a countable infinity of points is called a *discrete sample space*.

The examples of discrete sample space having finite sample points as well as having countable infinity of points are given in the earlier discussion. We can prepare a list of such examples from real life situations. It is not necessary always to prepare or enumerate a list of the S completely. It is enough if we are able to describe the same and can justify it. If necessary, we must be able to spell out these assumptions explicitly. *For example*, when a fair coin is tossed, we visualize only two outcomes, Head (H) or Tail (T). We have ignored the possibility that a coin will stand on its edge. This is an assumption. It is 'reasonable' to make such an assumption. It helps us to have simple mathematical models. When we make use of Newton's laws of motion, we assume that velocities of the objects are far less than that of light. If the assumption is not true, that is, if the velocities of objects are comparable with that of light then the theory of relativity propounded by Einstein explains laws of motion. Many times wherever feasible we must try to perform simple experiments to understand the nature of a sample space. Students may perform simple experiments such as tossing a pair of coins (dice).

For example, consider that two fair coins are tossed to observe the number of heads. The standard approach is to consider S = {HH, HT, TH, TT}. However, if you are noting the number of heads, then the number will be 0, 1 or 2 and S = {0, 1, 2}, as a possible sample space. This is the case if you do not want to distinguish between whether the first was head and the latter was tail (HT) or whether the first was tail and the latter was head (TH).

2.2 Events

When we discussed the concept of random experiment and of the associated sample space S, we understood that S is the collection of all possible outcomes of an experiment for which sample point is just a formal name. Many times we come across a situation where we are not interested only in a sample point but on a subset of a sample space associated with the experiment.

Suppose a fair coin is tossed three times. This is a random experiment, of which the possible outcomes are HHH, HHT, HTH, THH, HTT, THT, TTH, TTT. These, we denote by ω_1, ω_2, ω_3, ω_4, ω_5, ω_6, ω_7, ω_8 respectively. In other words, the sample space $S = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6, \omega_7, \omega_8\}$. Suppose we are interested in those outcomes which correspond to (1) exactly one coin shows head, (2) at least one coin shows head, (3) at most one coin shows head, (4) a coin shows neither head nor tail, and (5) it shows either head or tail.

An event A_1 described in (1) can be written as $A_1 = \{\omega_5, \omega_6, \omega_7\}$. ω_5 = {HTT}, ω_6 = {THT} and ω_7 = {TTH}. We can observe exactly one head if we observe either of ω_5, ω_6, ω_7. This is why we write event A_1 = exactly one coin shows head = $\{\omega_5, \omega_6, \omega_7\}$.

Similarly, we can write $A_2 = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6, \omega_7\}$. A_2 summarizes the event 'at least one coin shows head' in (2). $A_3 = \{\omega_5, \omega_6, \omega_7, \omega_8\}$ describes the event in (3). While (4) is given by $A_4 = \{\ \} = \phi$, the empty set or we may call it as an impossible event. And lastly $A_5 = S$, the sample space itself, this is sure (certain) event.

Thus, a set and an event can be associated with each other. In this example also observe that some of the sample points, *for example,* ω_5, ω_6, ω_7 belong to event A_1, A_2 and A_3. This is possible. Impossible event 'ϕ' and certain event S are usually uninteresting from an experimenter's point of view, however, for the sake of completeness we need them. Also one may construct a singleton set, that is, define an event so that the set contains only one sample point 'ω'. However, there is a difference between a sample point ω and an event $\{\omega\}$. The natural question is, 'How many subsets (events) can be written from a finite sample space?'. Remember that empty set, ϕ, is always a subset of S and S is also a subset of itself. Before we answer this question, we define first the term 'event'.

The set A_1 is said to be a subset of a set A_2 if every point in A_1 is also a point in A_2 *(see figure 3.1).* In notation, $\omega \in A_1 \Rightarrow \omega \in A_2$, we write $A_1 \subset A_2$ (read as A_1 is subset of A_2). The other notation also used is $A_2 \supset A_1$ (A_2 is superset of A_1).

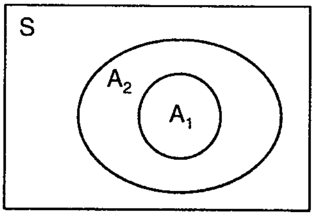


Figure 3.1: $A_1 \subset A_2$ or $A_2 \supset A_1$

If $A_1 \subset A_2$ and $A_2 \subset A_1$, then we say that two sets are equal and we write $A_1 = A_2$.

Definition: Any subset of the sample space S is called as an event.

Thus, when the sample space of an experiment is discrete, any set consisting of some or all points of the sample space, is called an *event.*

The impossible event ϕ is defined as the event that contains no descriptions and therefore cannot occur.

In set theory the impossible event is called the empty set. The study of elementary combinatorial techniques tells us that 2^n subsets are possible if a set has n elements. As an illustration, consider an experiment of tossing of a true coin. Then the sample space $S = \{H, T\}$. The subsets (and therefore events) of S that one may construct are ϕ, $\{H\}$, $\{T\}$, $\{H, T\} = S$; $2^2 = 4$ in total. If the fair coin is tossed three times then the sample space S contains eight sample points, therefore the possible subsets would be $2^8 = 256$. In general, a sample space consisting of 'n' points will have 2^n subsets or events defined in it.

2.3 Occurrence of an event A

When an experiment is performed, we observe only one outcome. However, it could have resulted in any one of the possible outcomes listed in the sample space S. Consider an experiment of tossing a fair die. We know that a die is fair and has six faces labelled as 1, 2, 3, 4, 5, and 6. Now, define an event A in which a number observed is even. Clearly A = {2, 4, 6}. When we say that event A has occurred, we mean that the outcome of tossing of a die has resulted in either observing face labelled as 2 or 4 or 6. That is, if $\omega \in A$, and if $A \subseteq S$, we say that A has occurred. Conversely, if $\omega \notin A$, $A \subseteq S$, we say that event A has not occurred or equivalently A' has occurred. Note that A and A' are complements of each other. If A occurs, then A' cannot occur and vice versa. If we know that A has occurred, we mean that the outcome $\omega \in A$ is observed. It can be any one of the possible points of S which are in A.

3. Algebra of Events

Here after we assume that we are given a random experiment E, associated with it we have a sample space S. S is a discrete sample space and ω stands for a sample point in S. An event always belongs to S. Consider a random experiment E of tossing a six faced fair die.

For the experiment E, S = {1, 2, 3, 4, 5, 6}. Consider the following events:

A: A face showing even number is observed.

B: The number observed is divisible by three.

C: A prime number is observed.

Clearly A = {2, 4, 6}; B = {3, 6} and C = {2, 3, 5}.

Now study the following statements carefully. Here $2 \in A$ means sample point '2' is contained in event A. Therefore, if '2' is the result of an experiment we will claim that event A has occurred. Similarly, we can also claim that event C has occurred. In other words, the outcome of an experiment can belong to more than one event.

Here $1 \notin A$ means sample point 1 does not belong to event A. That is, If '1' is the outcome of the experiment, we can say that A has not occurred. Equivalently one may claim that 'not A has occurred'. It means that if a sample point belongs to an event A, it can not belong to an event 'not A and vice versa. This gives us the definition of 'complementary event'.

3.1 Complementary Event

Number of questions asked. 1

Oct. 2010 – *2M*
Explain the following term:
Complementary event

The event consisting of all points not contained in the event A will be called the *complementary event* of A and will be denoted by A^c or by A'. That is, $A' = \{\omega \mid \omega \notin A\}$.

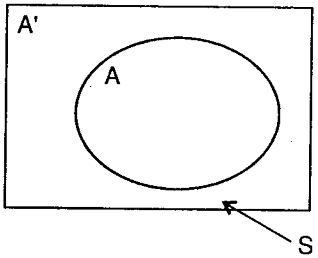


Figure 3.2: A and A' S

Clearly A and A' are complements of each other. Similarly $\phi' = S$ and $S' = \phi$. In the above example, $A = \{2, 4, 6\}$ and $A' = \{1, 3, 5\}$; $B = \{3, 6\}$ and $B' = \{1, 2, 4, 5\}$; $C = \{2, 3, 5\}$ and $C' = \{1, 4, 6\}$.

3.2 Union and Intersection of Two Events

Number of questions asked. 2

Oct. 11, Apr. 10 – *2M*
Define the following terms:
Union of two events.

Let A and B be any two events of S. Now we can construct the events such as $A \cup B$ (read as A union B) and $A \cap B$ (read as A intersection B) using the definitions given below:

Union

The event $A \cup B$ is the set of all sample points which belong to either A or B or to both A and B. That is, $A \cup B = \{\omega \mid \omega \in A \text{ or } \omega \in B\}$.

Let $A = \{2, 4, 6\}$ and $B = \{3, 6\}$ then $A \cup B = \{2, 3, 4, 6\}$.

Note that

- $A \cup A' = S$
- $A \cup A = A$

- $S \cup \phi = S$
- $A \subseteq A \cup B$ and $B \subseteq A \cup B$

Number of questions asked. 3

Oct. 12, 11, 10 – 1/2M
Define the following terms:
Intersection of two events

Intersection

The event $A \cap B$ is the set of all sample points which belong to both A and B, that is, $A \cap B = \{\omega \mid \omega \in A \text{ and } \omega \in B\}$.

Let $A = \{2, 3, 6\}$ and $C = \{2, 3, 5\}$ then $A \cap C = \{2, 3\}$.

Note that

- $A \cap A' = \phi$
- $A \cap A = A$
- $A \cap S = A$
- $A \cap \phi = \phi$
- $A \cap B \subseteq A$ and $A \cap B \subseteq B$.

Figure 3.3 shows $A \cap B$ as well as $A \cup B$.

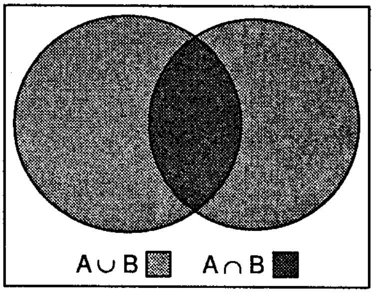


Figure 3.3

It is easy to see that $A \cup B = (A \cap B') \cup (A \cap B) \cup (A' \cap B)$.

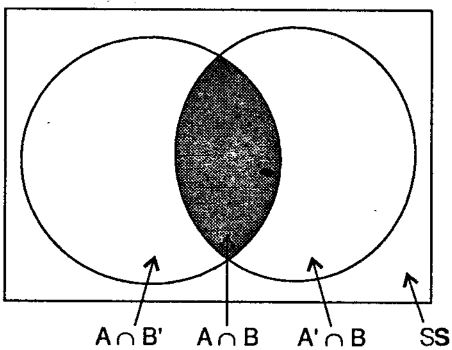


Figure 3.4

Relative complement: Relative complement of A with respect to B is denoted by $A - B$. It is the set of elements of A that are not contained in B. Thus, we can represent $A - B$ by $A \cap B'$. It is easy to see that $A - B \subseteq A$ and $A - B \subseteq B'$.

In the example of tossing a fair die, if we observe the uppermost face showing number 2 (or the number of dots equal to 2), it is certainly an even number and it is also a prime number. It means event A has occurred, event C has occurred and also $A \cap C$ has occurred. Also $A \cup B$ has occurred.

3.3 Union of Three Events

$$A_1 \cup A_2 \cup A_3 = \{\omega | \omega \in A_1 \text{ or } \omega \in A_2 \text{ or } \omega \in A_3\}$$

That is, if point ω belongs to event A_1 or to event A_2 or to event A_3, we say that it belongs to union of these events. In other words, it must belong to at least one of A_i's, if we claim that it belongs to their union.

3.4 Intersection of Three Events

$$A_1 \cap A_2 \cap A_3 = \{\omega | \omega \in A_1 \text{ and } \omega \in A_2 \text{ and } \omega \in A_3\}$$

That is, if point ω belongs to event A_1 and to event A_2 and to event A_3, we say that it belongs to intersection of these events. In other words, the sample point must belong to each and every event A_i, if we claim that it belongs to their intersection.

For example, if $S = \{1, 2, 3, 4, 5, 6\}$, $A_1 = \{1, 3, 5\}$, $A_2 = \{2, 3, 5\}$ and $A_3 = \{3, 6\}$ then

$A_1 \cup A_2 \cup A_3 = \{1, 2, 3, 5, 6\}$ and $A_1 \cap A_2 \cap A_3 = \{3\}$.

With the help of union and intersection of two or more events now we can create several events. Let us understand the meaning of such events with the help of an example.

Examples

Asked Number of questions: 1

Apr. 2010 – 4M

1. **Let the sample space $\Omega = \{1, 2, 3, \ldots, 10\}$**
   $A = \{2, 4, 6, 8, 10\}$, $B = \{6, 7, 8, 10\}$
   List elements of the sets:

   i. $A \cup B$ **ii.** $A \cap B$ **iii.** A' **iv.** $A' \cap B$

Solution

i. $A \cup B = \{2, 4, 6, 7, 8, 10\}$

ii. $A \cap B = \{6, 8, 10\}$

iii. $A' = \{1, 3, 5, 7, 9\}$

iv. $A' \cap B = \{7\}$.

---

2. Consider a family having two children. What is the possible sample space S, if we are to record the gender of a child?

Solution

Clearly, $S = \{BB, BG, GB, GG\}$. Define the following events:

i. A_1: At least one boy;

ii. A_2: At most one boy;

iii. A_3: Children of both the genders.

Now we will describe these and related events.

i. $A_1 = \{BB, BG, GB\}$; $A_1' = \{GG\}$, no boy or both girls;

ii. $A_2 = \{BG, GB, GG\}$; $A_2' = \{BB\}$, no girl or both boys;

iii. $A_3 = \{BG, GB\}$; $A_3' = \{BB, GG\}$, children of same gender only.

Now consider,

- $A_1 \cap A_2 = \{BG, GB\}$, children of both the genders.
- $A_1 \cap A_3 = A_3$, children of both genders. (Note $A_3 \subseteq A_1$).
- $A_2 \cap A_3 = A_3$, children of both genders. (Note $A_3 \subseteq A_2$).
- $A_1 \cup A_2 = S$, but $A_1 \cup A_3 = A_1$ since $A_3 \subseteq A_1$ and $A_2 \cup A_3 = A_2$ since $A_3 \subseteq A_2$

Similar relationships as above for A_i' can be worked out.

3.5 De Morgan's Laws

i. **$A' \cup B' = (A \cap B)'$:** Union of complements of events is same as complement of intersection of events.

ii. **$A' \cap B' = (A \cup B)'$:** Intersection of complements of events is same as complement of union of events.

In our earlier examples, we have come across situations where intersection of two events is a null event (set) ϕ. Using the idea of a null event we now define *mutually exclusive* or *disjoint* events.

Mutually Exclusive or Disjoint Events

Two events A and B are said to be *mutually exclusive or disjoint* if $A \cap B$ is the null event ϕ.

Number of questions asked. 3

Oct.12,Apr.11,10 – *1/2M*
Define the following terms:
Mutually exclusive events.

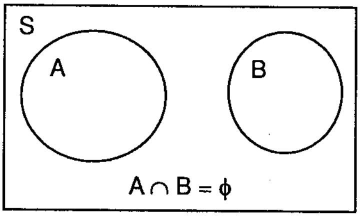


Figure 3.5

We have seen earlier that A and A' are disjoint events. It is easy to prove that

- $A \cap B$ and $A - B$,
- $A \cap B$ and A',
- $A \cup B$ and $A' \cap B'$,
- A and $B - A$ are disjoint events.

Contrary to this we may experience that occurrence (non-occurrence) of A implies occurrence (non-occurrence) of B.

For example, let S = {HH, HT, TH, TT}, A = {HT, TH}, B = {HH, HT, TH}.

Occurrence of event A guarantees occurrence of event B. All the points in A are also the points in B, however the reverse is not true.

Mutually Exclusive (Disjoint) and Exhaustive Events

Number of questions asked. 2

Oct.12, Apr.11 – *1/2M*
Define the following terms:
Exhaustive events

Two events A and B are said to be *mutually exclusive (disjoint) and exhaustive* if and only if $A \cap B = \phi$ and $A \cup B = S$.

(In this case we say that {A, B} forms a partition of the sample space. We need the concept of partition in the study of Bayes' theorem.)

Also note that A = B' and B = A'

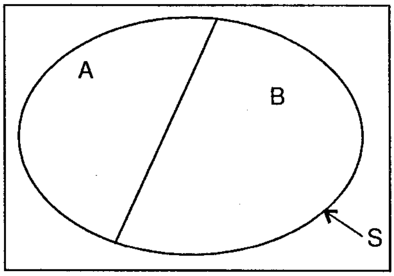


Figure 3.6: Partition of sample space

Certain Event and Impossible Event

Since $S \subset S$, S is itself an event consisting of all outcomes of the random experiment. Since every outcome $\omega \in S$, the event S always occurs and S is called the certain event.

Similarly $\phi \subset S$, it is an event consisting of no outcome. Since every outcome $\omega \notin \phi$, the event ϕ never occurs and it is called the 'impossible event'. In fact there is little point in using the phrase occurrence of an impossible event.

Number of questions asked. 1

Oct. 2010 – 4M

Example

1. If P(A) = 0.5, P(B) = p and $P(A \cup B) = 0.7$, find p, if A and B are mutually exclusive events.

Solution

If A and B are mutually exclusive events then

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

Therefore, we have

$$0.7 = 0.5 + p$$

$$\text{Thus, } p = 0.7 - 0.5$$

$$= 0.2$$

4. Classical Approach of Probability

Number of questions asked: 1

Apr. 2012 – *2M*
Define the following terms:
Classical Definition of Probability.

The classical definition of probability reduces the concept of probability to the concept of equal probability of outcomes. This is a primitive concept. This means every outcome of a random experiment has equal chance of occurrence. That is, in case of coin tossing experiment probability of getting head is same as probability of getting tail and equals half. Consider 'S' as the sample space having n sample points, corresponding to a random experiment E. Let A be any event. We assume that it can be decomposed into 'r' mutually exclusive and equally likely sample points. The ratio $\frac{r}{n}$ is called the probability P(A) of an event A. In other words

$$P(A) = \frac{\text{Number of outcomes belonging to event A}}{\text{Total number of outcomes in the sample space S}} = \frac{r}{n}$$

This statement is also called the frequency interpretation of P(A). Note that for any event A, $0 \le P(A) \le 1$ since $0 \le r \le n$. Thus, the probability of an event is always between 0 and 1.

We will now learn the classical definition of the probability of an event and apply the same to solve the simple problems.

4.1 Probability of an Event: Equally likely Outcomes

Suppose a sample space S has a finite number 'n' of points $\omega_1, \omega_2, \omega_3, \ldots, \omega_n$. We assume that $P(\omega_j) = \frac{1}{n}$; for $j = 1,2,3, \ldots, n$. That is, S is an *equiprobable* sample space.

Suppose now that an event A ($A \subset S$) contains 'r' points. Then under the classical assignment, the probability P(A) of an event A is $\frac{r}{n}$. In other words, 'r' is the number of cases favourable to A and 'n' is the total number of cases. Thus according to classical definition,

$$P(A) = \frac{\text{Number of outcomes belonging to event A}}{\text{Total number of outcomes in the sample space S}} = \frac{r}{n}$$

We can think of an experiment in which a finite number of possible outcomes may occur. It is assumed that each outcome is equally likely. For this, we consider typical problems involving games of chance such as tossing of a coin, rolling a dice, picking a lottery ticket, drawing cards from a pack of cards and so on. We further use the phrase such as the coin is unbiased, the dice is fair, pack of cards is well shuffled and so on. This allows us to assume that all elementary events are equally likely. There are several experiments where S is not finite, moreover outcomes are not equally likely. In such cases, we can not use the classical approach.

4.2 Limitations of the Classical Definition

i. Both 'r' and 'n' being positive integers, P(A) is reduced to a rational number. Irrational numbers are not taken care of by this definition.

ii. The definition is valid when n is finite. For n infinite, we have no satisfactory answer.

iii. In this approach, we consider only equally likely cases. Can we modify it if all the cases are not equally likely? Answer is no.

We will now solve some examples using the classical definition.

Examples

1. A box contains 25 balls of which 15 are white and rest black. A ball is drawn at random from this box. Find the probability of

i. white ball ii. black ball

Solution

Here the number of sample points in S is equal to 25. Let W and B respectively denote the events that white ball is drawn and black ball is drawn. Then, by classical approach,

$$P(W) = \frac{r}{n} = \frac{15}{25} = \frac{3}{5} = 0.60$$

$$P(B) = \frac{r}{n} = \frac{10}{25} = \frac{2}{5} = 0.40$$

2. A card is drawn from a well shuffled pack of playing cards. What is the probability that the card is

i. a king ii. a face card iii. an ace of spade iv. diamond card

Solution

From a well shuffled pack of fifty two playing cards a card can be drawn in C(52, 1) = 52 different ways. Thus, number of sample points in sample space S equals 52.

i. There are 4 kings of which any one can be drawn in C(4, 1) different ways. Hence using the classical approach, P(King) $= \frac{4}{52} = \frac{1}{13}$.

ii. There are 12 face cards of which any one can be drawn in C(12, 1) = 12 different ways. P(a face card) $= \frac{12}{52} = \frac{3}{13}$.

iii. Note that there is only one ace of spade. Thus, P(ace of spade) $= \frac{1}{52}$.

iv. Lastly we want to compute the probability of a diamond card. Number of diamond cards is 13. Hence we get P(a diamond card) $= \frac{13}{52} = \frac{1}{4}$.

---

3. A bag contains six red and four blue balls. Three balls are drawn at random, what are the odds against these being all red?

Solution

Total number of ways in which 3 balls can be drawn out of 10 balls is equal to C(10, 3) = 120. Number of ways in which 3 red balls can be drawn is C(6, 3) = 20. Thus, the probability that all the three balls drawn are red is $\frac{20}{120} = \frac{1}{6} = \frac{1}{1+5}$

Hence, *odds in favour of these balls being red 1:5 and the odds against these being all red 5:1.*

---

4. From a pack of 52 cards, 2 are drawn at random. What is the probability that one is a king and the other is a queen?

Solution

Here 2 cards can be drawn in C(52, 2) = 1326 ways. There are 4 kings and 4 queens in the pack. Thus the number of ways in which a king and a queen can be drawn simultaneously equals C(4, 1) × C(4, 1) = 16. Therefore, assuming 'S' to be equiprobable, the required probability is $\frac{16}{1326} = \frac{8}{663}$.

---

5. If n distinct balls are placed in n cells what is the probability that each cell will be occupied?

Solution

Total number of possible arrangements is n^n because each ball can go to any cell and thus each ball has n choices. All these cases are equally likely and are mutually exclusive. The number of ways each cell may be kept occupied equals the number of ways in which 'n' balls may be arranged among themselves. This number is clearly n!. Hence required probability is $\frac{n!}{n^n}$.

Number of questions asked.
2
Apr. 11, Oct. 10 – 4M

6. **Two cards are drawn at random from a well shuffled pack of playing cards. Find probability that:**

   i. **Both king cards are drawn,**

   ii. **One king and one queen card is drawn.**

Solution

Two cards can be drawn from a well shuffled pack of cards in

$$^{52}C_2 = \frac{52!}{2!50!} = 1326 \text{ ways}$$

Let A denote the event that both king cards are drawn. There are 4 kings in the pack, hence 2 kings can be drawn in $^4C_2 = \frac{4!}{2!2!} = 6$ ways. The required probability is

$$P(A) = \frac{6}{1326} = \frac{1}{221} = 0.004525$$

Let B denote the event that one king and one queen are drawn. There are 4 kings and 4 queens in a pack, hence one king can be drawn in 4 ways and one queen can be drawn in 4 ways. Therefore, one king and one queen can be drawn in $4 \times 4 = 16$ ways. The required probability is

$$P(B) = \frac{16}{1326} = \frac{8}{663} = 0.012066$$

Number of questions asked.
1
Apr. 2010 – 4M

7. **If two dices are rolled, find probability that the sum of the integers on the uppermost faces is almost 4.**

Solution

Two dice are rolled hence the sample space is

$$S = \left\{ \begin{matrix} (1, 1) & (1, 2) & (1, 3) & (1, 4) & (1, 5) & (1, 6) \\ (2, 1) & (2, 2) & (2, 3) & (2, 4) & (2, 5) & (2, 6) \\ (3, 1) & (3, 2) & (3, 3) & (3, 4) & (3, 5) & (3, 6) \\ (4, 1) & (4, 2) & (4, 3) & (4, 4) & (4, 5) & (4, 6) \\ (5, 1) & (5, 2) & (5, 3) & (5, 4) & (5, 5) & (5, 6) \\ (6, 1) & (6, 2) & (6, 3) & (6, 4) & (6, 5) & (6, 6) \end{matrix} \right\}$$

Let A be the event that the sum of the integers on the uppermost faces is at most 4, then we have A = {(1, 1), (1, 2), (1, 3), (2, 1), (2, 2), (3, 1)}. Number of sample points in S is 36 and that in A is 6, hence the required probability is

$$P(A) = \frac{\text{Number of outcomes belonging to event A}}{\text{Total number of outcomes in the sample space S}} = \frac{r}{n} = \frac{6}{36} = \frac{1}{6}$$

5. Axiomatic Approach

Let 'S' be a discrete sample space associated with a random experiment E. Let A be any event in S. 'P' is called the probability function or probability measure on S if the following axiom is satisfied.

Axioms

Given $S = \{\omega_1, \omega_2, \omega_3, \ldots, \omega_n\}$ is a discrete sample space. We can assign to each sample point ω_j of S a number $P\{\omega_j\}$, such that

i. $0 \le P\{\omega_j\} \le 1$ and

ii. $P\{\omega_1\} + P\{\omega_2\} + \ldots + P\{\omega_j\} = \sum_j P(\omega_j) = 1$

We can have two or more valid assignments of probabilities to the same sample space. *For example,* in the example of tossing a dice, the sample space S has six sample points, viz. $\omega_1 = 1$, $\omega_2 = 2$, $\omega_3 = 3$, $\omega_4 = 4$, $\omega_5 = 5$, $\omega_6 = 6$. Following table shows probability assignments under different models.

Probability assignment	Sample point						Total
	ω_1	ω_2	ω_3	ω_4	ω_5	ω_6	
Model 1	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{6}{6} = 1$
Model 2	$\frac{1}{21}$	$\frac{2}{21}$	$\frac{3}{21}$	$\frac{4}{21}$	$\frac{5}{21}$	$\frac{6}{21}$	$\frac{21}{21} = 1$
Model 3	$\frac{1}{4}$	0	$\frac{1}{4}$	$\frac{1}{4}$	0	$\frac{1}{4}$	$\frac{4}{4} = 1$
Model 4	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{9}{9} = 1$

In model 1, it is assumed that the dice is fair where as in model 2, the probability of observing ω_j is assumed to be proportional to j for j = 1, ..., 6. In model 3, $P\{\omega_j\} = \frac{1}{4}$ for j = 1, 3, 4, 6 and $P\{\omega_2\} = P\{\omega_5\} = 0$. In model 4, $P\{\omega_1\} = P\{\omega_3\} = P\{\omega_5\} = \frac{1}{9}$ and $P\{\omega_2\} = P\{\omega_4\} = P\{\omega_6\} = \frac{2}{9}$.

Observe that, in all these models $P(\omega_j) \ge 0$ and $\sum_j P(\omega_j) = 1$ We can construct such any number of probability models. Further, if number of sample points in S equals n, a finite number, $\sum_{j=1}^{n} P(\omega_j) = 1$ can easily be checked.

With the help of above axioms, we define probability of an event A.

Probability of an event

The Probability P(A) of an event A is, the sum of the probabilities of the sample points in A. In symbols, $P(A) = \sum_{\omega \in A} P(\omega_j)$; where A can be any event. It may be an impossible event, or an elementary event or S itself. Recall that we have defined an impossible event ϕ. It does not contain a sample point. Hence there is little point in talking about occurrence of ϕ. Hence by convention $P(\phi) = 0$. Also S is a sure event. It always occurs.

Hence, we must have $P(S) = 1$. Note clearly that $P(A) = 0$ does not imply that A is an impossible event and also $P(A) = 1$ does not imply that A is sample space.

Consider model 3. Let A = {2, 5,}, then $P(A) = 0$. But A has two points. It is not an empty set. Similarly, let B = {1, 3, 4, 6}, $B \subset S$ and $P(B) = 1$; but $B \neq S$. We, therefore, define *null event* as an event with probability zero and *almost sure event}* is an event with probability one. This should make it clear that the impossible event is a null event and the sure event is almost sure event, however, a null event is not the impossible event and an almost sure event is not necessarily the sure event.

6. Probability of an Event: Properties

We will see some important properties of the probability of an event. We assume that S is a sample space associated with a random experiment E and A is any event defined on S. P is called the probability function or probability measure on S if the following axioms are satisfied:

- **Axiom 1.** P(A) is a real number such that $P(A) \geq 0$ for any event A on S.
- **Axiom 2.** $P(S) = 1$.
- **Axiom 3.** If A and B are any two events defined on S such that $A \cap B = \phi$, that is, A and B are disjoint event, then $P(A \cup B) = P(A) + P(B)$.

In general, if $A_1, A_2, \ldots, A_n$ are events on S such that for any two events $A_i \cap A_j = \phi$, for $i \neq j$ then

$$P\left(\bigcup_{i=1}^{n} A_j\right) = \sum_{j=1}^{n} P(A_i)$$

Property 1: Let A be any event. A' denotes the complement of A. Then $P(A') = 1 - P(A)$.

Property 2: $P(\phi) = 0$. That is, the probability of an impossible event is zero.

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

Property 4: If $A \subset B$ and $A, B \subseteq S$, then $P(A) \leq P(B)$.

Property 5: If A and B are any two events, then $P(A - B) = P(A) - P(A \cap B)$, that is,

$P(A \cap B') = P(A) - P(A \cap B)$.

Property 6: **(Addition theorem of probability)** For any two events A and B defined on S, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. It can be seen from following figure.

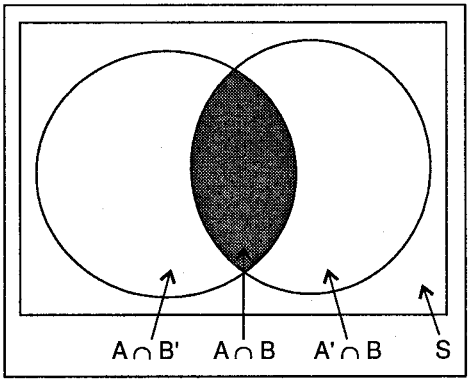


Figure 3.7

Also observe that, $P(A \cup B) \leq P(A) + P(B)$, since $P(A \cap B) \geq 0$. It is known as *Boole's inequality*. The equality holds if and only if $P(A \cap B) = 0$.
These are some important properties of the probability of an event. With the help of these properties, one can investigate the relationships between the probabilities of an event. We have introduced the axiomatic approach of the probability of an event.

Property 7: Let A, B, C be any three events defined on S, then

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

This is the probability that at least one of three events A, B or C occurs.

Examples

1. **Let A and B be events with $P(A) = \frac{3}{8}$, $P(B) = \frac{1}{2}$ and $P(A \cap B) = \frac{5}{16}$.**

Find

i.	**$P(A \cup B)$**	**ii.**	**$P(A'), P(B')$**	**iii.**	**$P(A' \cap B')$**
iv.	**$P(A' \cup B')$**	**v.**	**$P(A \cap B')$**	**vi.**	**$P(A' \cap B)$.**

Solution

i. $P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{3}{8} + \frac{1}{2} - \frac{5}{16} = \frac{9}{16}$

ii. $P(A') = 1 - P(A) = 1 - \frac{3}{8} = \frac{5}{8}$, $P(B') = 1 - P(B) = 1 - \frac{1}{2} = \frac{1}{2}$

iii. Using De Morgan's law, $(A' \cap B') = (A \cup B)'$, we have

$P(A' \cap B') = P(A \cup B)' = 1 - P(A \cup B) = 1 - \frac{9}{16} = \frac{7}{16}$

iv. By De Morgan's law we have $(A' \cup B') = (A \cap B)'$, we have

$P(A' \cup B') = P(A \cap B)' = 1 - P(A \cap B) = 1 - \frac{5}{16} = \frac{11}{16}$

Equivalently,

$P(A' \cup B') = P(A') + P(B') - P(A' \cap B') = \frac{5}{8} + \frac{1}{2} - \frac{7}{16} = \frac{11}{16}$

v. $P(A \cap B') = P(A) - P(A \cap B) = \frac{3}{8} - \frac{5}{16} = \frac{1}{16}$

vi. $P(A' \cap B) = P(B) - P(A \cap B) = \frac{1}{2} - \frac{5}{16} = \frac{3}{16}$

Number of questions asked. 1

Oct. 2010 – 4*M*

2. If $P(A) = 0.6$, $P(B) = 0.3$ and $P(A \cap B) = 0.2$,

Find:

i. $P(A \cup B)$ ii. $P(B')$ iii. $P(A' \cap B)$

Solution

i. $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.6 + 0.3 - 0.2 = 0.70.$

ii. $P(B') = 1 - 0.3 = 0.7$

iii. $P(A' \cap B) = 0.3 - 0.2 = 0.1$

Number of questions asked. 1

Apr. 2010 – 4*M*

3. If $P(A) = 0.6$, $P(B) = 0.4$, $P(A \cap B) = 0.2$

Compute:

i. $P(A \cup B)$ ii. $P(A')$ iii. $P(A' \cap B)$

Solution

i. $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.6 + 0.4 - 0.2 = 0.8$

ii. $P(A') = 1 - P(A) = 1 - 0.6 = 0.4.$

iii. $P(A' \cap B) = P(B) - P(A \cap B) = 0.4 - 0.2 = 0.2$

4. Let A, B, C be three events defined on sample space Ω. If $P(A) = 0.5$, $P(B) = 0.4$, $P(C) = 0.6$, $P(A \cap B) = P(B \cap C) = 0.2$, $P(A \cap C) = 0.3$, $P(A \cap B \cap C) = 0.1$.

Number of questions asked. 1

Apr. 2012 – 8*M*

Compute:

i. $P(A \cup B \cup C)$ **ii.** $P(A' \cap B' \cap C')$

iii. $P(A \cap B')$ **iv.** $P(A/B)$

Solution

i. $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(B \cap C) - P(A \cap C) + P(A \cap B \cap C).$

$$= 0.5 + 0.4 + 0.6 - 0.2 - 0.2 - 0.3 + 0.1 = 0.9$$

ii. $P(A' \cap B' \cap C') = 1 - P(A \cup B \cup C)' = 1 - 0.9 = 0.1$ (By De Morgan's law)

iii. $P(A \cap B') = P(A) - P(A \cap B) = 0.5 - 0.2 = 0.3$

iv. $P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{0.2}{0.4} = 0.5.$

5. A and B stand in a queue at random with seven other persons. What is the probability that there will be two persons between A and B. Assume that all the positions are occupied randomly.

Solution

Eight positions including A and B are:

Position	1	2	3	4	5	6	7	8	9
	*	+	#	*	+	#	*	+	#

The table shows possible arrangement of positions. If A is positioned at 1, B must be at 4 in order that there are two persons between A and B, if A is positioned at 2, B must be at 5, if A is positioned at 3, B must be at 6, if A is positioned at 4, B must be at 7, if A is positioned at 5, B must be at 8 and if A is positioned at 6, B must be at 9. Thus there are 6 positions in all while A and B can interchange their positions in 2! ways. Hence A and B can take described positions in $2! \times 6 = 12$ ways. Since out of 9 places 2 positions can be selected in $C(9, 2) = 36$ different ways, the required probability equals $\frac{12}{36} = \frac{1}{3}$.

6. A personal assistant randomly puts 8 letters in 8 addressed envelopes. What is th chance that all letters are wrongly placed?

Solution

8 letters can be placed on 8! different ways, whereas there is only one way of placing all of them i their right envelopes. Hence the required probability $= 1 - \frac{1}{8!}$.

7. Three swimmers A, B and C are in a competition. A is twice as likely to win as B and B i twice as likely to win as C. What are their respective probabilities of winning? What i the probability that either B or C wins? Assume that there are no ties.

Solution

Assume that only one competitor can win in such a race. Further, let us assume that the probabilit that C wins is p. Thus, $P(C) = p$, where $\{(0 < p < 1)\}$. It gives $P(B) = 2p$ and $P(A) = 4p$. Further

$$P(A) + P(B) + P(C) = 1 \Rightarrow 4p + 2p + p = 1 \Rightarrow 7p = 1 \Rightarrow p = \frac{1}{7}$$

$$\therefore P(A) = \frac{4}{7}, P(B) = \frac{2}{7}, P(C) = \frac{1}{7}$$

Also $P(B \cup C) = P(B) + P(C) - P(B \cap C) = \frac{2}{7} + \frac{1}{7} = \frac{3}{7} \because P(B \cap C) = 0$

8. A fair coin is tossed three times. Define event A as observing at least one head and a least one tail and event B as observing at most one head. Find P(A), P(B), and P(A ∩ B).

Solution

Here S = {HHH, HHT, HTH, HTT, THH, THT, TTH, TTT} and it has 8 sample points, whe A = {HHT, HTH, HTT, THH, THT, TTH} $\Rightarrow P(A) = \frac{6}{8} = 0.75$ and B = {HTT, THT, TTH, TTT $\Rightarrow P(B) = \frac{4}{8} = 0.50$.

Also $A \cap B$ = {HTT, THT, TTH} $\Rightarrow P(A \cap B) = \frac{3}{8} = 0.375$. Note that $P(A \cap B) = P(A) \cdot P(B)$.

9. A box contains 20 light bulbs of which 6 are defective. Four light bulbs are chosen a random from the box. Find the probability that

i. none is defective, **ii. exactly one is defective,**

iii. at least one is defective.

Solution

Four light bulbs can be chosen from the box in C(20, 4) = 4845 (number of points in sample space). We assume that the sample space is equiprobable.

i. There are 6 defective bulbs, that is, there are 20 – 6 =14 non-defective bulbs. We want P(none is defective). There are C(14, 4) = 1001 ways to chose 4 non-defective bulbs. Thus required probability = $\frac{1001}{4845}$ =0.206605.

ii. We want the probability that there is exactly one defective bulb in four chosen bulbs. There are C(6, 1) ways to choose one defective bulb and C(14, 3) to choose 3 non-defective bulbs. Thus, the number of ways to have exactly one defective bulb is C(6, 1) × C(14, 3) = 2184. Therefore, the required probability is $\frac{2184}{4845}$ = 0.450774.

iii. The event that at least one defective bulb is the complement of the event that none is defective, therefore the required probability = 1 – 0.206605 = 0.793395.

10. Two amateur bird-watchers Shirish and Nitin visit a habitat. The probability that Shirish will identify Black-winged Kite is 0.68 and that of Nitin is 0.54. The probability that both will identify the Black-winged Kite is 0.32. What is the probability that only one of them will correctly identify the bird species?

Solution

Let A be the event that Shirish identifies Black-winged Kite and B be the event that Nitin identifies Black-winged Kite. We want probability that only one of them will correctly identify the bird species, that is, $P(A \cap B') + P(A' \cap B)$. We have

$$P(A \cap B') + P(A' \cap B) = P(A) - P(A \cap B) + P(B) - P(A \cap B) = P(A) + P(B) - 2P(A \cap B)$$

Therefore, the required probability = 0.68 + 0.54 – 2 × 0.32 = 0.58.

11. A problem in Statistics is given to three students A, B, and C, whose chances of solving it are respectively 0.3, 0.4 and 0.6. A and B together can solve the problem with probability 0.15, A and C together can solve the problem with probability 0.10 and B and C can solve it with probability 0.25. If A, B and C sit together and solve the problem jointly, the chances of solving it are 0.05. What is the probability that the problem is solved?

Solution

Let A, B and C denote the event that problem is solved by the students A, B and C respectively. We need $P(A \cup B \cup C)$. We have

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

$$P(A \cup B \cup C) = 0.30 + 0.40 + 0.60 - 0.15 - 0.25 - 0.10 + 0.05 = 0.85$$

Number of questions asked. 1

Apr. 2011 – 4M

12. A coin is tossed three times, find the probability that:

i. two heads appear

ii. at least two heads appear.

Solution

The sample space in this experiment is

$$\Omega = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}.$$

Here $n(\Omega) = 8$

i. Let A be the event that 'two heads appear'. Then A = {HHT, HTH, THH}. n(A) = 3.

Thus, $P(A) = \frac{3}{8}$.

ii. Let B be the event that 'at least two heads appear'.

Then B = {HHH, HHT, HTH, THH}. n(B) = 4. Thus, $P(B) = \frac{4}{8} = \frac{1}{2}$.

Number of questions asked. 1

Apr. 2012 – 4M

13. If two dice are rolled, find the probability that sum of the points appears on the uppermost faces is more than 4.

Solution

The sample space is given by

$$S = \left\{\begin{matrix} (1,1) & (1,2) & (1,3) & (1,4) & (1,5) & (1,6) \\ (2,1) & (2,2) & (2,3) & (2,4) & (2,5) & (2,6) \\ (3,1) & (3,2) & (3,3) & (3,4) & (3,5) & (3,6) \\ (4,1) & (4,2) & (4,3) & (4,4) & (4,5) & (4,6) \\ (5,1) & (5,2) & (5,3) & (5,4) & (5,5) & (5,6) \\ (6,1) & (6,2) & (6,3) & (6,4) & (6,5) & (6,6) \end{matrix}\right\}$$

There are 36 sample points in S, that is, n(S) = 36. Let A be the event that sum of the points appears on the uppermost faces is more than 4. Thus A', the complement of event A is sum of the points appears on the uppermost faces is less than or equal to 4.

Hence A'= {(1, 1), (1, 2), (1, 3), (2, 1), (2, 2), (3, 1)}. There are 6 points in A'. Therefore,

$P(A') = \frac{6}{36} = \frac{1}{6}$. Now $P(A) = 1 - P(A') = 1 - \frac{1}{6} = \frac{5}{6}$.

14. Explain Non-deterministic experiment with two illustrations.

Number of questions asked. 1

Apr. 2012 – 4M

Solution

In daily life, we talk about chances, such as chance of rain today, chance of winning a cricket match, chance of passing in an examination, chance of winning a lottery, so on. These types of happenings are associated with uncertainties. These are the experiments in which the result cannot be predicted in advance with certainty. These experiments are called as random or non-deterministic experiments. In particular, an experiment with more than one possible outcome and whose result cannot be predicted in advance is called a *random Experiment.*

For example,

i. Tossing a coin,

ii. Rolling a dice,

iii. Whether item produced by a machine is defective, etc.

15. Distinguish between Deterministic and Non-deterministic experiments.

Number of questions asked. 1

Apr. 2011 – 4M

Solution

	Deterministic experiments	Non-Deterministic experiments
i.	Only one outcome of the experiment.	More than one outcomes of the experiment.
ii.	Outcome can be predicted in advance.	Outcome cannot be predicted in advance, however, list of outcomes can be prepared.
iii.	The outcome is sure.	Occurrence of an outcome is based on chance factor.
iv.	No chance factor involved.	Based on past experience and statistical regularity probability of an outcome can be determined.
v.	*Example:* Newton's laws, Ohm's law, Boyle's law.	*Example:* Tossing a coin, Rolling a die, Germination of a seed.

7. Independence of events

Two or more events defined on sample space S are said to be independent when the outcome of one does not affect and is not affected by the other. For example, if a coin is tossed by right hand and left hand, then the results obtained are not affected by hand used. Similarly, if a dice is tossed twice, the result of the second toss would in no way affected by the result of the first toss.

Number of questions asked. 2

Oct.12, Apr. 10 – *2M*
Define the following terms:
Independence of two events.

Definition: Let A and B be any two events of sample space S associated with random experiment E. They are said to be *independent* or *stochastically independent* if

$P(A \cap B) = P(A) \cdot P(B)$ .... **(1)**

Remarks

- Relation equation (1) does not need P(A) or P(B) be positive. That is, even if $P(A) = 0$ or $P(B) = 0$ then relation equation (1) can hold good.
- If $A \cap B = \phi$, then $P(A \cap B) = P(\phi) = 0 \Rightarrow P(A) = 0$ or $P(B) = 0$. This means that if A and B are mutually exclusive events then either $P(A) = 0$ or $P(B) = 0$ (or both). If we insist that P(A) and P(B) must be positive, then A and B can not be mutually exclusive and independent.
- If A is an almost sure event, then A and any other event B are independent. This is really trivial.
- If A is a null event then A and any other event B are also independent.
- If A occurs, A' does not occur and vice versa. Stochastic independence implies that no inference can be drawn from the occurrence of B to that of A; and therefore stochastic independence of A and B should mean the same as independence of A and B'. Of course, we must also have then independence of A' and B and also of A' and B'.
- The concept of independence of two events can be generalized to three or more events.

Examples

1. **A card is drawn at random from a well-shuffled deck of playing cards. Define event A, 'an ace card' and event B, "a club card". Are A and B independent?**

Solution

$P(A) = P(\text{ace card}) = \frac{4}{52} = \frac{1}{13}$ and $P(B) = P(\text{club card}) = \frac{13}{52} = \frac{1}{4}$. $A \cap B$ denotes the event that an ace of club is drawn, hence $P(A \cap B) = \frac{1}{52} = P(A) \times P(B)$. This shows that A and B are independent.

2. Let A and B be two independent events defined on a sample space Ω. If $P(A) = 0.6$, $P(B) = K$ and $P(A \cup B) = 0.8$, find value of K.

Number of questions asked 1

Apr. 2010 – *4M*

Solution

Given: $P(A) = 0.6, P(B) = K$ and $P(A \cup B) = 0.8$.

Since events A and B are independent $P(A \cap B) = P(A) \times P(B) = 0.6 \times K = 0.6K$. We know that $P(A \cup B) = P(A) + P(B) - P(A \cap B) \Rightarrow 0.8 = 0.6 + K - 0.6K$

$$\therefore\ 0.2 = 0.4K \quad \text{or} \quad K = \frac{0.2}{0.4} = 0.5$$

3. Sam tosses a fair coin two times and Dave tosses the same coin three times. Let A denote the event "at most one head" and B denote the event "at least one head and at least one tail". Examine the independence of A and B and comment.

Solution

Here we have two random experiments, say E_1 and E_2. Let S_1 and S_2 denote, respectively, the sample spaces associated with them. Thus we have

Player	Sam	Dave
Experiment	E_1	E_2
Sample space	S_1 = {HH, HT, TH, TT}	S_2 = {HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}
Event A	{HT, TH, TT}	{HTT, THT, TTH, TTT}
Event B	{HT, TH}	{HHT, HTH, THH, HTT, THT, TTH}
$A \cap B$	{HT, TH}	{HTT, THT, TTH}
Probabilities	$P(A) = \frac{3}{4}$ $P(B) = \frac{2}{4} = \frac{1}{2}$ $P(A \cap B) = \frac{2}{4} = \frac{1}{2}$	$P(A) = 4/8 = \frac{1}{2}$ $P(B) = 6/8 = \frac{3}{4}$ $P(A \cap B) = \frac{3}{8}$

In experiment E_1, $P(A \cap B) = \frac{1}{2}$, $P(A) = \frac{3}{4}$ and $P(B) = \frac{1}{2} \Rightarrow P(A \cap B) \neq P(A) \times P(B)$, hence A and B are dependent events.

In experiment E_2, $P(A \cap B) = \frac{3}{8}$, $P(A) = \frac{1}{2}$ and $P(B) = \frac{3}{4} \Rightarrow P(A \cap B) = P(A) \times P(B)$, hence A and B are independent events.

This example demonstrates that when we speak about independence (or dependence) of events we have events $A \subseteq S$ and $B \subseteq S$, that is, these events must belong to the same sample space S.

Complete Independence of Three Events

Let A, B and C be any three events defined on S of a random experiment E. Events A, B and C are independent if

$$\begin{aligned} P(A \cap B) &= P(A) \times P(B); \\ P(B \cap C) &= P(B) \times P(C); \\ P(C \cap A) &= P(C) \times P(A); \\ P(A \cap B \cap C) &= P(A) \times P(B) \times P(C) \quad \ldots\ldots (2) \end{aligned}$$

For complete independence, we need all these relations to hold good. In case of pairwise independence, first three relations in equation (2) must hold good. *Example 1* demonstrates that *pairwise independence does not guarantee complete independence (or mutual independence).*

Solved Examples

1. A fair coin is tossed two times. Let A_1 denote the event that first toss is head, A_2 denote the event that second toss is head and A_3 denote the event that outcomes of these two tosses are the same. Show that conditions of pair-wise independence are satisfied, but not of complete independence.

Solution

Here $S = \{HH, HT, TH, TT\}$, $A_1 = \{HH, HT\}$, $A_2 = \{HH, TH\}$ and $A_3 = \{HH, TT\}$.

Hence $P(A_1) = P(A_2) = P(A_3) = 0.5$. Also,

$A_1 \cap A_2 = \{HH\} \Rightarrow P(A_1 \cap A_2) = 0.25 = P(A_1)\,P(A_2)$;

$A_1 \cap A_3 = \{HH\} \Rightarrow P(A_1 \cap A_3) = 0.25 = P(A_1)\,P(A_3)$; and

$A_2 \cap A_3 = \{HH\} \Rightarrow P(A_2 \cap A_3) = 0.25 = P(A_2)\,P(A_3)$.

Thus, all the three conditions for pair-wise independence of three events defined on S are satisfied. Now for complete independence, in addition to above three we require fourth condition, viz.,

$$P(A_1 \cap A_2 \cap A_3) = P(A_1) \times P(A_2) \times P(A_3) \text{ to hold good.}$$

Here, $A_1 \cap A_2 \cap A_3 = \{HH\} \Rightarrow P(A_1 \cap A_2 \cap A_3) = 0.25 \neq P(A_1) \times P(A_2) \times P(A_3) = 0.125$.

Thus, pair-wise independence does not guarantee complete independence.

Remark: The concepts of dependence and independence are very important in probability theory. Incomplete understanding of these concepts will invariably lead to mistakes. e.g., If a coin is tossed ten times and it has shown head every time, what is most probable in the eleventh trial? Heads or tails? If you feel that head has already occurred ten times and if the coin is to be fair it must now start showing tail (in order to compensate!), you are then absolutely wrong. If the coin is fair, then on the eleventh toss, just as on the first one, the probability that a head appears is 0.5. However, you have right to verify the statement that the coin is fair? But how do you verify? For doing this, you must study statistical inference and testing of hypothesis.

2. The probabilities that three students A, B and C will solve a problem in Statistics independent of each other are $\frac{1}{3}, \frac{3}{8}$ and $\frac{1}{4}$. Find the probability that

i. at least one of them will solve the problem; ii. only A will solve the problem;
iii. B and C will solve the problem.

Solution

Let us denote the probabilities as $P(A) = \frac{1}{3}$, $P(B) = \frac{3}{8}$ and $P(C) = \frac{1}{4}$. It is given that the students solve the problem independent of each other.

i. *To find*: $P(A \cup B \cup C)$.

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

By independence of the events we have

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

$$= \frac{1}{3} + \frac{3}{8} + \frac{1}{4} - \frac{1}{3} \times \frac{3}{8} - \frac{3}{8} \times \frac{1}{4} - \frac{1}{3} \times \frac{1}{4} + \frac{1}{3} \times \frac{3}{8} \times \frac{1}{4} = \frac{11}{16}$$

ii. *To find*: $P(A \cap B' \cap C')$.

$$P(A \cap B' \cap C') = P(A)\,P(B')\,P(C') = \frac{1}{3} \times \left(1 - \frac{3}{8}\right) \times \left(1 - \frac{1}{4}\right) = \frac{1}{3} \times \frac{5}{8} \times \frac{3}{4} = \frac{5}{32}$$

iii. *To find*: $P(B \cap C)$.

$$P(B \cap C) = \frac{3}{8} \times \frac{1}{4} = \frac{3}{32}$$

3. A coin is tossed two times. Let event A denote head occurs on the first toss and B denote head occurs on second toss. Verify whether A and B are independent events.

Number of asked questions 1

Oct. 2010 – 4M

Solution

The sample space is given by S = {HH, HT, TH, TT}. Therefore, A = {HH, HT} and B = {HH, TH}.
Hence $A \cap B$ = {HH}

Now, $P(A) = \frac{1}{2}$, $P(B) = \frac{1}{2}$, $P(A \cap B) = \frac{1}{4}$. Here $P(A \cap B) = P(A)P(B)$.

Therefore, A and B are independent events.

EXERCISES

1. Distinguish between deterministic and non-deterministic experiments.
2. A list of experiments is given below. Classify them as random or non-random experiments giving reasons.
   i. A light bulb is placed in service and the time of operation until it burns out is measured.
   ii. A coin is tossed repeatedly until a head occurs.
   iii. The velocity of a falling body after a certain length of time t (in seconds) is observed.
   iv. Water in a pot is heated for a sufficiently long time to a temperature in excess of 100°C.
   v. A radio-active substance emits α particles. The number of such particles reaching an observation screen during t seconds is noted.
   vi. Cancer patients are divided into two groups, smokers and non-smokers. Ten patients are put under experimentation and their smoking habit is noted.
   vii. A card is drawn from a pack of well shuffled pack of playing cards and its suit is noted.
   viii. In an experiment in fermentation studies, a liquid containing yeast cells was thoroughly mixed and poured in a flat dish.
   ix. Number of accidents on Mumbai-Agra highway is recorded.
   x. Number of arches on the fingertips of an individual is noted.
   xi. A company produces items. A quality control engineer examines a batch of n items and records the number of defectives in it.
   xii. A textile manufacturer observes the number of faults in a meter of cloth chosen from the daily production.
   xiii. A biased coin is tossed.
3. Specify the sample space 'S' for the following experiments.
   i. A coin is tossed.
   ii. A die is tossed.
   iii. A pair of two coins is tossed.
   iv. A card is drawn from a pack of fifty two well shuffled playing cards and its suit is noted.
   v. Two cards are drawn from 5 cards numbered 1 to 5.
   vi. A coin and a die is tossed together.
   vii. Two dice are rolled and the sum of points on their upper most faces is noted.
   viii. A student attempts an examination till he passes.
   ix. A problem is given to a group of 20 students. Number of students correctly solving the problem is noted.
   x. A family has two children, elder child is a girl.

4. A coin is tossed. If head occurs, it is tossed again only once, otherwise it is tossed two times. Describe

   i. the sample space S,

   ii. event A that at least one head is observed,

   iii. the event B that exactly 2 tails occur.

5. Let A_i (i = 1, 2, 3) be any three arbitrary events. Find the expressions for the events that correspond to the occurrence of

   i. only A_1,
   ii. only A_1 and A_2,
   iii. none of A_i's,
   iv. only one of A_i's,
   v. at least one of A_i's.

6. Give the alternate forms of the events:

   i. $(A \cup B) \cap (A \cup C)$,
   ii. $(A \cup B) \cap (A' \cup B)$,
   iii. $(A \cup B) \cap (A' \cup B) \cap (A' \cup B')$.

7. A family has 4 children. Define the events: A: boys and girls alternate, B: the first and fourth child are boys, C: as many boys as girls, and D: three successive children of the same sex. Write S, A, B, C, D.

8. Explain the classical approach to probability.

9. Explain modern axiomatic approach in probability theory.

10. Discuss classical definition of the probability of an event.

11. Explain the terms

    i. equiprobable sample space,
    ii. unequiprobable sample space.

12. State the axioms of probability.

13. Explain the meaning of

    i. odds in favour of an event are a : b,
    ii. odds against of an event are a : b.

14. State the limitations of classical approach of probability.

15. State addition theorem of probability.

16. State giving reasons whether following statements are true or false.

    i. If P(A) is zero, then $A = \phi$.

    ii. If P(A) is unity, then A = S.

    iii. If A, B and C are any three events such that P(A) = 0.5, P(B)= 0.6 and P(C) = 0.25, then $P(A \cap B) = 0.35$, $P(A \cap C) = 0.40$ and $P(B \cap C) = 0.45$.

    iv. If $A \subset B$, then P(A) < P(B) always.

17. Verify whether the following are valid probability models ($\omega_j \in S$, for all j).

i. $P\{\omega_j\} = \frac{j^2}{30}$; for j = 1, 2, 3, 4 ii. $P\{\omega_j\} = \frac{1}{j(j+1)}$; for j = 1, 2, 3, 4, 5, 6, 7, 8.

18. Evaluate unknown constant k in the following models so that the model can be viewed as probability model.

i. $P\{\omega_j\} = kj$, for j = 1, 2, 3, 4, 5, 6 ii. $P\{\omega_j\} = kj^3$, for j = 1, 2, 3, 4.

19. Miss Anandita owns two television sets, one colour and one black-and-white set. Let A be the event that the colour set is on and B the event the black and white set is on. If P(A) = 0.4, and P(B) = 0.3, and $P(A \cup B) = 0.5$, find the probability of each event:

i. both are on, ii. the colour set is on and the other is off,

iii. exactly one set is on, iv. neither set is on.

20. Three coins are tossed together. Let A be the event that exactly 2 coins show heads and B be the event that at least 2 coins show heads. List the elements of A, B, A', B'. Verify whether A and B are mutually exclusive. Obtain the probabilities of these events.

21. A fair coin has numbers 3 and 5 on its sides. The coin is tossed four times. Find the probability that the sum of numbers observed is less than 16.

22. Two dice are rolled. Let A be the event that the sum of the points on the upper most faces is odd and B be the event that there is at least one '3' shown. Describe the following events. $A \cup B$, $A \cap B$ and $(A \cap B') \cup A'$ and list the elements contained in it.

23. J, U, N are in a race; J is thrice as likely to win as U and U is twice as likely to win as N. Find P(J), P(U) and P(N). Also find

i. $P(J \cup U)$, ii. $P(J' \cup N)$.

24. Two women W_1 and W_2 and three men M_1, M_2 and M_3 are in a chess tournament. Those of the same gender have equal probabilities of winning, but each man is twice as likely to win as any woman.

i. Find probability that a woman wins the tournament.

ii. If M_1 and W_2 are married, find probability that one of them wins the tournament.

25. A class consists of 6 girls and 10 boys. If a committee of 4 is chosen at random from the class, find the probability that

i. 4 boys are selected, ii. exactly 2 boys are selected,

iii. at least one boy is selected, iv. at most 3 girls are selected.

26. Given $P(A) = \frac{1}{3}$, $P(B) = \frac{1}{4}$, $P(A \cap B) = \frac{1}{6}$. Find $P(A')$, $P(A' \cup B)$, $P(A \cup B')$, $P(A \cap B')$, $P(A' \cup B')$.

27. If $P(A) = \frac{3}{4}$ and $P(B) = \frac{3}{8}$, show that $P(A \cup B) \geq \frac{3}{4}$ and $\frac{1}{8} \leq P(A \cup B) \leq \frac{3}{8}$.

28. Three winning tickets are drawn from a box of 100 tickets. What is the probability of winning for a person who buys:

    i. 4 tickets? ii. only one ticket?

29. Let A, B, C be three events defined on S with probabilities $P(A) = 0.35$, $P(B) = 0.25$, $P(C) = 0.45$, $P(A \cap B) = 0.18$, $P(A \cap C) = 0.22$, $P(B \cap C) = 0.16$, $P(A \cap B \cap C) = 0.11$. Compute

    i. $P(A \cup B \cup C)$, ii. $P(A \cup C)$, iii. $P(A' \cap B \cap C)$,

    iv. $P(A' \cap B' \cap C)$, v. $P(A' \cap B' \cap C')$.

30. Let n be the number of elements in the sample space S and N(E) the number of elements in the event E. Verify whether the following data are consistent:

    $n = 1000$, $N(A) = 525$, $N(B) = 312$, $N(C) = 470$, $N(A \cap B) = 42$, $N(A \cap C) = 147$, $N(B \cap C) = 86$ and $N(A \cap B \cap C) = 25$.

31. Define independence of two events. State axioms, if any.

32. What is complete independence?

33. If A and B are independent, $P(A) = 0.20$ and $P(B) = 0.45$, find

    i. $P(A \cap B)$, ii. $P(A \cup B)$, iii. $P(A' \cap B')$.

34. If possible, give two examples each of two events which are

    i. mutually exclusive and independent,

    ii. mutually exclusive but dependent,

    iii. not mutually exclusive but independent,

    iv. not mutually exclusive and dependent.

35. Three independent components are hooked in series. Each component may fail with probability p. What is the probability that the system does not fail? Also evaluate the probability that the system does not fail if these three components are hooked in parallel.

36. Mark the correct answer:

i. Assume that a dice is rolled twice in succession. How many sample points are in S?

a. 6 b. 12
c. 36 d. 42

ii. Ten numbered balls are placed in an urn. Numbers 1-4 are red and numbers 5-10 are blue. What is the probability that a ball drawn at random is blue?

a. 0.1 b. 0.4
c. 0.6 d. 1.0

iii. Assume that A_i (i = 1, 2, 3) are non-null events such that $P(A_i \cap A_j) = 0$ for $i \neq j$. Then $P(A_1 \cup A_2 \cup A_3)$ is

a. 1 b. $P(A_1) + P(A_2) + P(A_3)$
c. less than $P(A_1) + P(A_2) + P(A_3)$ d. $\max\{P(A_1), P(A_2), P(A_3)\}$

37. The events A, B and C are independent with P(A) = 0.3, P(B) = 0.2 and P(C) = 0.4. What is the probability that at least two will occur among the three events?

38. In Science class of a certain college 25% of the students failed Mathematics, 15% of the students failed Physics and 10 % failed both Mathematics and Physics. A student is selected at random.

i. What is the probability that he failed Mathematics or Physics?

ii. What is the probability he failed Mathematics only?

iii. If he failed Physics, what is the probability that he failed Mathematics?

39. A certain football team wins (W) with probability 0.55, loses (L) with probability 0.35 and ties (T) with probability 0.10. The team plays three games in league. Determine the probability that

i. team wins at least twice and doesn't lose;

ii. team wins, loses and ties.

ANSWERS

2.	i.	Random (R)	ii.	R	iii.	Non-random (NR)	iv.	NR
	v.	R	vi.	R	vii.	R	viii.	R
	ix.	R	x.	R	xi.	R	xii.	R
	xiii. R							

3. i. {H, T},

ii. {1, 2, 3, 4, 5, 6}

iii. {HH, HT, TH, TT}

iv. {Spade, Club, Heart, Diamond}

v. {11, 12, 13, 14, 15, 22, 23, 24, 25, 33, 34, 35, 44, 45, 55}

vi. {H1, H2, H3, H4, H5, H6, T1, T2, T3, T4, T5, T6}

vii. {2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12}

viii. {P, FP, FFP, FFFP, …}

ix. {0, 1, 2, 3, …, 20}

x. {GB, GG}

4. S = {HH, HT, THH, THT, TTH, TTT}, A = {HH, HT, THH, THT, TTH},

B = {THT, TTH},

5. i. $A_1 \cap A_2' \cap A_3'$

ii. $A_1 \cap A_2 \cap A_3'$

iii. $A_1' \cap A_2' \cap A_3' = (A_1 \cup A_2 \cup A_3)'$

iv. $(A_1 \cap A_2' \cap A_3') \cup (A_1' \cap A_2 \cap A_3') \cup (A_1' \cap A_2' \cap A_3)$

v. $A_1 \cup A_2 \cup A_3$

6. i. $(A \cap B) \cup (A \cap C) \cup (B \cap C)$, ii. B, iii. S,

7. S = {BBBB, BBBG, BBGB, BGBB, GBBB, BBGG, BGBG, BGGB, GBBG, GBGB, GGBB, BGGG, GBGG, GGBG, GGGB, GGGG},

A = {BGBG, GBGB},

B = {BBBB, BBGB, BGBB, BGGB},

C = {BBGG, BGBG, BGGB, GGBB, GBGB, GBBG},

D = {BBBB, BBBG, GBBB, BGGG, GGGG},

16. i. False ii. False iii. False iv. True

17. i. True ii. False $\sum P(j) < 1$

18. i. $k = \frac{1}{21}$ ii. $k = \frac{1}{100}$

19. i. 0.2 ii. 0.2 iii. 0.3 iv. 0.5

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

A = {HHT, HTH, THH},
B = {HHH, HHT, HTH, THH},

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

B' = {HTT, THT, TTH, TTT},

A and B are not mutually exclusive. $P(A) = \frac{3}{8}$, $P(B) = \frac{1}{2}$, $P(A') = \frac{5}{8}$, $P(B') = \frac{1}{2}$

21. $\frac{5}{16}$

22. $A \cup B$ = {12, 13, 14, 16, 21, 23, 25, 31, 32, 33, 34, 35, 36, 41, 43, 45, 52, 53, 54, 56, 61, 63, 65},

$A \cap B$ = {23, 32, 34, 36, 43, 63},

$(A \cap B') \cup A'$ = {11, 12, 14, 15, 16, 21, 22, 24, 25, 26, 41, 42, 44, 45, 46, 51, 52, 54, 55, 56, 61, 62, 64, 65, 66} = S – B,

23. $\frac{6}{9}, \frac{2}{9}, \frac{1}{9}$ i. $\frac{8}{9}$ ii. $\frac{4}{9}$

24. i. $\frac{2}{8}$, ii. $\frac{3}{8}$

25. i. 0.115385 ii. 0.370879 iii. 0.991758 iv. 0.991758

26. $\frac{2}{3}, \frac{5}{6}, \frac{11}{12}, \frac{1}{6}, \frac{5}{6}$

27. i. 0.000105 ii. 0.03

29. i. 0.60 ii. 0.58 iii. 0.05 iv. 0.18 v. 0.40

30. $P(A \cup B \cup C) = 1.057$, therefore the data are inconsistent.

33. i. 0.09 ii. 0.56 iii. 0.44

35. $(1 - p)^3$, $1 - p^3$

36. i. c ii. c iii. b

37. 0.212

38. 0.30, 0.15, 0.666667

39. 0.331375, 0.1155

PU Questions

1. Define the following terms: **[Oct. 2012 - *4M*]**
   i. Mutually Exclusive events
   ii. Exhaustive events
   iii. Sample space
   iv. Intersection of two events
2. Define the following terms: **[Oct. 2012 - *4M*]**
   i. Independence of two events
   ii. Conditional probability
3. In a group of 10 men, 6 are graduates. If 3 men are selected at random, what is the probability that consists of **[Oct. 2012 - *4M*]**
   i. All graduate? ii. At least one graduate?
4. A card is drawn from a pack of 52 cards. What is the probability that **[Oct. 2012 - *4M*]**
   i. Card is either red or black?
   ii. Card is either red or face card?
5. Explain Non-deterministic experiment with two illustrations. **[Apr. 2012 - *4M*]**
6. If two dice are rolled, find the probability that sum of the points appears on the uppermost faces is more than 4. **[Apr. 2012 - *4M*]**
7. Define the following terms: **[Apr. 2012 - *4M*]**
   i. Classical Definition of Probability
   ii. Conditional Probability
8. Find the probability that a leap year selected at random will contain 53 Sundays. **[Apr. 2012 - *4M*]**
9. The probability that a contractor will get a plumbing contract is 0.4 and the probability that he will not get an electric contract is 0.7. If the probability of getting at least one contract is 0.6, what it's the probability that he will get: **[Oct. 2011 - *4M*]**
   i. both the contracts? ii. Exact one contract?

[Oct. 2011 - 4M] 10. Define

i. Sample space ii. Discrete sample space

iii. Union of two events iv. Intersection of two events

[Oct. 2011 - 4M] 11. Consider the following sample space $\Omega = \{1, 2, 3, 4, 5, 6, 7, 8\}$. Write down the following events:

i. A: An odd number appears

ii. B: Number is greater than 4

iii. At least one of A and B occurs

iv. None of A and B occurs

[Oct. 2011 - 4M] 12. Suppose A and B are two events defined on sample space Ω. If $P(A) = 0.8$, $P(A \cup B) = 0.9$ and $P(B) = x$, find the value of x if,

i. A and B are independent

ii. A and B are mutually exclusive.

[Oct. 2011 - 4M] 13. Define classical definition of probability. State the addition theorem for three events.

[Oct. 2011 - 4M] 14. Consider an experiment of rolling of a fair die with sample space $\Omega = \{1, 2, 3, 4, 5, 6\}$ with the event A = occurrence of an even number and B = occurrence of a number greater than 4. Check whether A and B are independent or not.

[Apr. 2011 - 4M] 15. Define the following terms with illustration:

i. Mutually Exclusive Events,

ii. Exhaustive events.

[Apr. 2011 - 4M] 16. If $P(A) = 0.6$, $P(B) = 0.4$, $P(A \cap B) = 0.2$,

find: i. $P(A \cup B)$ ii. $P(A')$ iii. $P(A' \cap B)$

[Apr. 2011 - 4M] 17. Distinguish between Deterministic and Non-deterministic experiments.

[Apr. 2011 - 4M] 18. A coin is tossed three times, find the probability that:

i. two heads appear

ii. at least two heads appear.

19. Two cards are drawn from a pack of 52 cards. Find probability that: **[Apr. 2011 - *4M*]**

    i. both are kings, ii. one king and one queen.

20. Two cards are drawn at random from a well shuffled pack of playing cards. Find probability that: **[Oct. 2010 - *4M*]**

    i. Both king cards are drawn,

    ii. One king and one queen card is drawn.

21. What is the probability that a leap year selected at random will contain 53 Sundays? **[Oct. 2010 - *4M*]**

22. Explain the following terms with one illustration of each: **[Oct. 2010 - *4M*]**

    i. Complementary event,

    ii. Intersection of two events.

23. If $P(A) = 0.6$, $P(B) = 0.3$ and $P(A \cap B) = 0.2$, **[Oct. 2010 - *4M*]**

    find: i. $P(A \cup B)$, ii. $P(B')$,

    iii. $P(A' \cap B)$.

24. A coin is tossed two times. Let event A denote head occurs on the first toss and B denote head occurs on second toss. Verify whether A and B are independent events. **[Oct. 2010 - *4M*]**

25. If $P(A) = 0.5$, $P(B) = p$ and $P(A \cup B) = 0.7$. Find p, if A and B are mutually exclusive events. **[Oct. 2010 - *4M*]**

26. Define: **[Apr. 2010 - *4M*]**

    i. Sample Space ii. Union of Two Events

27. An urn contains 6 white and 8 red balls. Two balls are drawn at random. Find probability that: **[Apr. 2010 - *4M*]**

    i. both balls are of red colour.

    ii. both balls are of the same colour.

28. If $P(A) = 0.6$, $P(B) = 0.4$, $P(A \cap B) = 0.2$ **[Apr. 2010 - *4M*]**

    Compute:

    i. $P(A \cup B)$ ii. $P(A')$ iii. $P(A' \cap B)$

[Apr. 2010 - 4M] 29. Explain the following terms with one illustration of each:

i. Mutually Exclusive Events

ii. Independence of Two Events

[Apr. 2010 - 4M] 30. If two dices are rolled, find probability that the sum of the integers on the uppermost faces is almost 4.

[Apr. 2010 - 4M] 31. Let A and B be two independent events defined on a sample space Ω. If $P(A) = 0.6$, $P(B) = K$ and $P(A \cup B) = 0.8$, find value of K.

[Apr. 2010 - 4M] 32. Let the sample space $\Omega = \{1, 2, 3, \ldots, 10\}$

$A = \{2, 4, 6, 8, 10\}$, $B = \{6, 7, 8, 10\}$

List elements of the sets:

i. $A \cup B$ ii. $A \cap B$

iii. A' iv. $A' \cap B$

[Oct. 2009 - 4M] 33. Distinguish between Deterministic and Non-deterministic Experiments.

[Oct. 2009 - 4M] 34. A box contains 6 red and 8 black balls. Of two balls are drawn at random one by one, find probability of getting

i. both the balls are of different colour.

ii. both the balls of same colour.

[Oct. 2009 - 4M] 35. Find $P(A \cup B)$, if $P(A) = 0.2$ and $P(B) = 0.5$, given that:

i. A and B are independent events, and

ii. A and B are mutually exclusive events.

[Oct. 2009 - 4M] 36. Define the following terms: Sample Space and Event with illustrations

[Apr. 2009 - 4M] 37. Define: Sample Space and Event with one illustration of each.

[Apr. 2009 - 4M] 38. Explain: Population and Sample with an illustration.

39. A husband and wife appear in an interview for two vacancies in the same post. The probability of husband's selection is $\frac{1}{7}$ and that of wife is $\frac{1}{5}$. What is the probability that: [Apr. 2009 - 4M]

i. both of them will be selected?

ii. none of them will be selected?

40. Suppose A and B are any two events defined on the sample space, if $P(A) = 0.8$, $P(B) = K$, $P(A \cup B) = 0.9$. Find the value of K, if A and B is: [Apr. 2009 - 4M]

i. Independent ii. Mutually exclusive

41. Define Mutually Exclusive Events and Mutually Exhaustive Events with one illustration each. [Apr. 2009 - 4M]

42. Suppose A and B are any two events defined on the sample space, if $P(A) = 0.6$, $P(B) = 0.3$, $P(A \cap B)= 0.2$, find the probability that: [Apr. 2009 - 4M]

i. at least one of the two events A and B will occur.

ii. exactly one of the two events A and B will occur.

43. Let A, B, C be three events defined on sample space Ω. If $P(A) = P(B) = P(C) = \frac{1}{4}$, $P(A \cap B) = 0$, $P(B \ \Omega \ C) = 0$, $P(A \cap C) = \frac{1}{8}$, [Oct. 2012 - 8M]

Calculate:

i. $P(A \cup B \cup C)$ ii. $P(A \cap C')$

iii. $P(A' \cap B' \cap C')$ iv. $P(A/C)$

44. Let A, B, C be three events defined on sample space Ω. If $P(A) = 0.5$, $P(B) = 0.4$, $P(C) = 0.6$, $P(A \cap B) = P(B \cap C) = 0.2$, $P(A \cap C) = 0.3$, $P(A \cap B \cap C) = 0.1$. Compute: [Apr. 2012 - 8M]

i. $P(A \cup B \cup C)$ ii. $P(A' \cap B' \cap C')$

iii. $P(A \cap B')$ iv. $P(A/B)$

[Oct. 2011 - *8M*]

45. Let A, B, C be three events defined on Ω with the following probabilities:

$P(A) = 0.3$, $P(B) = 0.2$, $P(C) = 0.5$,

$P(A \cap C) = 0.25$, $P(B \cap C) = 0.15$, $P(A \cap B) = 0.17$ and $P(A \cap B \cap C) = 0.1$

Compute:

i. $P(A \cup B)$ ii. $P(A' \cap B' \cap C)$

iii. $P(A' \cap B' \cap C)$ iv. $P(A' \cap B' \cap C')$

[Oct. 2010 - *8M*]

46. Write Sample space for each of the following experiments:

i. A card is drawn from a well shuffled pack of playing cards and the suit is noted.

ii. Number of defectives are noted from a lot of 10 items.

iii. A coin is tossed until a head appears first time.

iv. Two dice are tossed and upper-most faces are noted.

[Apr. 2010 - *8M*]

47. List elements of sample space for the following experiments:

i. A coin is tossed 3 times.

ii. A student attempts an examination till he passes.

iii. Ten seeds are planted and total numbers of seeds germinated are recorded.

iv. Two cards are drawn from a pack of playing cards and colour is noted.

Chapter 4

STANDARD DISCRETE DISTRIBUTIONS

1. Introduction

We have considered probability evaluation based on the definition of sample space and events. In this chapter, we extend the idea to obtain the probabilities using the relationship between the outcomes in the sample space and a set of real numbers. This relationship in general is a function from sample space to set of real numbers. The function under consideration is known as 'random variable'. We shall now proceed to study the concept of random variable and its probability distribution.

2. Concept of a Random Variable

We have earlier understood the difference between a deterministic model and a non-deterministic model. In the study of probability, we mainly deal with non-deterministic models. We can associate a sample space, S with every random experiment. S may or may not be equiprobable. However, we can specify the probability of a sample point or the probability of an event. This enables us to study the uncertainties associated with the outcome(s) of random experiments.

A list of simple experiments that are observable in our daily life given below:

- Consider a family having n children. Observe the number of male children, say X, in it. Possible values of X are 0, 1, 2, …, n.
- Suppose you bet an amount of Rs.100 in favour of India's win in a match. Your net gains (X) are -100, 100.
- In a road traffic study, number of accidents taking place is a variable (X) of interest. The possible values of X are 0, 1, 2, ….
- Let X denote the number of telephone channels in use during a specified time interval. The possible values of X are 0, 1, 2, …, n, where n is the capacity of the exchange.

Note that, the experiments listed above are random experiments. A real number is associated with each outcome of the experiment. Since the outcome itself depends upon chance factor so also must be the real number associated with outcome. Thus, we have a quantity which varies. Hence we call it a variable. Moreover it must be called as a random variable (r.v.) because the values that it assumes depend upon chance factor. If the possible values of r.v. are 'isolated' (different) and are countable then such a r.v. is called a discrete r.v. There are other types of random variables, however, in this book we will be dealing only with discrete r.v.s.

Definition: Let S be sample space corresponding to a random experiment E. Let X be a real valued function defined from S to set of real numbers R. That is $X: S \to R$, then X is called a random variable on S.

X is real valued function from S to R means $X(\omega) \in R$, $\omega \in S$. The r.v.s are usually denoted by capital letters such as X, Y, Z, … with or without suffixes. The value of a r.v. X at the point $\omega \in S$, is denoted by $X(\omega)$. Customarily we make use of corresponding small letters to denote these values, that is, $X(\omega) = x$.

Consider an experiment of tossing a fair coin two times. Here, S = {HH, HT, TH, TT}. It consists of four points. viz., $\omega_1 = \{HH\}$, $\omega_2 = \{HT\}$, $\omega_3 = \{TH\}$, $\omega_4 = \{TT\}$.

Let X be the number of heads observed in two successive tosses of the coin. So $X(\omega_1) = X(HH) = 2$. Similarly, $X(\omega_2) = X(HT) = 1$, $X(\omega_3) = X(TH) = 1$ and $X(\omega_4) = X(TT) = 0$.Thus a r.v. X is a function defined on S. The domain set is S itself and co-domain is the set of real numbers R, in particular the range set is {0, 1, 2}. Note that $X: S \to R$. The correspondence from S to R is of many-one type.

In the above example, we are dealing with a fair coin. Observe that

$$X = \begin{cases} 0, & \text{if and only if the outcome is } \omega_4 \\ 1, & \text{if and only if the outcome is } \omega_2 \text{ or } \omega_3 \\ 2, & \text{if and only if the outcome is } \omega_1 \end{cases}$$

The above approach helps us to identify the events under consideration. Then the probabilities of these events can be obtained. Suppose, we want to find the probabilities of events:

i. no head,

ii. exactly one head,

iii. exactly two heads.

It is easy to see that

$$P\{\text{No head}\} = P\{X = 0\} = P\{\omega_4\} = \frac{1}{4}.$$

$$P\{\text{exactly one head}\} = P\{X = 1\} = P\{\omega_2 \text{ or } \omega_3\} = P\{\omega_2\} + P\{\omega_3\} = \frac{1}{4} + \frac{1}{4} = \frac{2}{4}.$$

$$P\{\text{exactly two heads}\} = P\{X = 2\} = P\{\omega_1\} = \frac{1}{4}.$$

We could obtain probabilities of above-mentioned events, because we have identified the possible values of the r.v. with the subsets of S. Since S is assumed to be an equiprobable sample space. We have $P\{\omega_j\} = \frac{1}{4}$ for j = 1, 2, 3 and 4. Note that $P\{X = x\} \geq 0$ and $\sum_x P\{X = x\} = 1$.

Now we give formal definition of a discrete r.v.

3. Discrete Random Variable

A discrete r.v. can take countable values. It means either its values are finite in number or its values can be arranged in a sequence. Thus, a discrete r.v. can assume only isolated values. Note that, these values need to be real numbers. These values need not necessarily be integers. In this chapter, we consider only univariate situations. Further, we will be dealing with a discrete sample space and specified probabilities of its sample points.

Definition: Let $S = \{\omega_1, \omega_2, \ldots, \omega_n\}$ be a finite sample space. A r.v. X, is a function which assigns to every sample point $\omega_i \in S$, a real number x_i, such that $X(\omega_i) = x_i$, for i = 1, 2, 3, …, n. Then X is called a discrete random variable.

In simple words, X is a real valued function defined on S. It takes finite values corresponding to finite sample space. Note that, the domain set is always S and the range set is some subset of the set of real numbers.

The definition can be extended to countable sample space $S = \{\omega_1, \omega_2, \ldots, \omega_i, \ldots\}$. The corresponding random variable will take countable values.

A list of some discrete r.v.s. are given below:

i. Number of female children in a family having n children.

ii. Number of peas in a pod.

iii. Number of seedling survived in a plantation.

iv. Number of telephone calls arrived at an electronic switch-board in a given time interval.

v. Number of days required to give a hearing in a lawsuit in the court of a law since the case is filed.

vi. Number of accidents per day on a Mumbai - Bangalore national high way.

vii. Number of defective items in a lot of size N.

viii. Number of misprints on a page of a book.

Suppose that the r.v. X takes different values x_j, for $j = 1, 2, \ldots, n$. We are interested in finding out the probabilities with which these values are taken by X. So we must obtain $P[X = x_j] = f(x_j)$, for $j = 1, 2, \ldots, n$. This leads us to a definition of probability mass function (pmf) of a r.v. X.

4. Probability Mass Function (PMF)

Definition: The function $f(x_j) = P[X = x_j]$, $j = 1, 2, \ldots, n$ defined for the values $x_1, x_2, \ldots, x_n$ assumed by X satisfying

$$f(x_j) = P[X = x_j] = p_j \geq 0; \;\; \forall j = 1,2,\ldots,n$$

$$\sum_{j=1}^{n} f(x_j) = \sum_{j=1}^{n} P[X = x_j] = \sum_{j=1}^{n} p_j = 1. \quad \ldots\ldots (1)$$

is called the probability mass function (pmf) of X.

Remark: Any function satisfying Equation (1) qualifies to be a pmf.

Example

Let X denote the number of heads in two successive tosses of a fair coin. Obtain pmf of X.

Solution

Consider a random experiment of tossing a fair coin two times. We then will have S = {HH, HT, TH, TT}= $\{\omega_1, \omega_2, \omega_3, \omega_4\}$. It is an equiprobable sample space.

Clearly, $X(HH) = X(\omega_1) = 2$, $(HT) = X(\omega_2) = 1$, $X(TH) = X(\omega_3) = 1$ and $X(TT) = X(\omega_4) = 0$.

Thus, the possible values of X are 0, 1 and 2. The pmf of X is given by $f(x) = P[X = x]$ for x = 0, 1, 2.

$$f(0) = P[X = x] = P[TT] = \frac{1}{4}$$

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

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

We need not use the notation f(x) always. Some authors prefer to use the notation P[X =x] or P(x) to denote f(x).The pmf of X can also be written in a tabular form as follows:

Pmf of X

X	0	1	2	Total
P[X = x]	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	1

5. Cumulative Distribution Function (CDF)

Definition: Let X be a discrete r.v. having pmf P(x). The cdf of X is $F: R \to [0, 1]$ such that for each $x \in R$, $F_X(x) = P[X \leq x]$

Note that, the domain of a cdf is the entire real line and not just the set of values of the r.v. X.

For discrete r.v.s we consider the function $F_X(x) = P[X \leq X] = \sum_{t \in A} P(t)$, where we consider the set A of all points in R that are less than or equal to x; that is, $A = \{t \mid t \leq x \text{ and } t \in R\}$. The function $F_X(x)$ is called *distribution function* (or *cumulative distribution function*) of the discrete r.v. X. The cdf of a r.v. X is denoted either by $F_X(x)$ or by F(x) or by $F(\cdot)$.

Properties of CDF

i. $F(x)$ is defined for every $x \in R$.

ii. $0 \leq F(x) \leq 1$, since $F(x)$ is a probability.

iii. $F(x)$ is a non-decreasing function of x. That is if $x_1 \leq x_2$, then $F(x_1) \leq F(x_2)$.

iv. $\lim\limits_{x \to -\infty} F(x) = 0 = f(-\infty)$ and $\lim\limits_{x \to \infty} F(x) = 0 = F(\infty)$.

v. $F(x)$ is right continuous everywhere.

Note that,

- Any function $F: R \to [0, 1]$ which satisfies above properties is a cdf of some r.v.
- Several random variables can have the same cdf. If two random variables have the same cdf then we say that they have the same probability law.
- Suppose $\lim\limits_{x \to \infty} F(x) = c > 0$ $(1 \neq c < \infty)$, then properly scaling this function we can obtain

  $H(x) = \dfrac{F(x)}{c}$ as a cdf of $x \in R$.

- If $a, b \in R$ and $a \leq b$, then

  i. $P[a < X \leq b] = F(b) - F(a)$.

  ii. $P[a \leq X \leq b] = F(b) - F(a) + P[X = a]$.

  iii. $P[a \leq X < b] = F(b) - F(a) - P[X = b] + P[X = a]$.

  iv. $P[a < X < b] = F(b) - F(a) - P[X = b]$.

  v. $P[X > a] = 1 - P[X \leq a] = 1 - F(a)$.

- **Relation to pmf:** $P[X = x_j] = F(x_j) - F(x_{j-1})$, $(j = 1, 2, \ldots)$.Using cdf (pmf), we can obtain pmf (cdf).
- Cdf of a discrete r.v. is usually, though not always, a step function. This can be seen by plotting a graph of $F(x)$ Vs X. It is easy to see that $F(x)$ is the theoretical counterpart of 'less than cumulative frequency curve'. Also draw the graph of $H(x) = 1 - F(x)$. Suppose X denotes the life-time of a component. Then $H(x)$ will denote the probability that the component will survive beyond the time x. It is referred as the *survival function*.

Example

Consider the following pmf and corresponding cdf

x	2	3	4	5	6	7	8	9	10	11	12
P[X = x]	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$
F(x)	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{6}{36}$	$\frac{10}{36}$	$\frac{15}{36}$	$\frac{21}{36}$	$\frac{26}{36}$	$\frac{30}{36}$	$\frac{33}{36}$	$\frac{35}{36}$	$\frac{36}{36} = 1$

6. Mathematical Expectation

Mathematical expectation is very important concept in probability distributions. It is useful in locating the point of equilibrium of the distribution.

Definition: The expectation E(X) of a r.v. X assuming values $x_1, x_2, \ldots$ with probabilities $p(x_1)$, $p(x_2), \ldots$ is given by $E(X) = \sum_j x_j\, p(x_j)$ provided $\sum_j |x_j|\, p(x_j)$ is finite.

Remarks:

- Many times the suffixes in the expression are suppressed and the formula is written as

  $E(X) = \sum xp(x)$

- Recall that a discrete sample space has either a finite number of points or its points can be arranged in a sequence. Since a r.v. is a function on the sample space, it can have either a finite number of values or its values can be arranged in a sequence. If a r.v. X takes finite values, we can use the phrase that 'X is a r.v. having finite sample space'. In such a case E(X) is simply the usual sum, viz., $E(X) = \sum_j x_j\, p(x_j)$. It always exists.

- The most common r.v.s have finite expectations. Hence the concept. In this text we will study the r.v.s having finite expectation only. Interested students may verify the existence of absolute convergence.

- If X assumes a countable number of values then it is necessary to show that the series $E(X) = \sum_j x_j\, p(x_j)$ converges absolutely; that is $\sum_j |x_j|\, p(x_j)$ must be a convergent series.

- The terms mean, average and mathematical expectation are synonymous. E(X) is referred as the mean of the distribution.

- *The centre-of-mass interpretation of E(X)*: E(X) is a 'weighted average'. It can be considered as a measure of the 'center' of the associated probability distribution. Think of E(X) as the centre of gravity of the probability distribution. If a mass P[X = x] is placed at the point x for each x on a beam then the balance point of the beam is at E(X). Alternatively, if the random experiment E is repeated large number of times and the average of the observed values of X is computed then with a high probability, this average will be close to E(X).

Examples

1. Consider a r.v. X having pmf as given below:

X	–1	0	1	2	3	4
P[X = x]	0.07	0.18	0.24	0.20	0.24	0.07

Find E(X).

Solution

By definition, $E(X) = \sum_j x_j\, p(x_j)$.

$$\therefore\ E(X) = (-1)(0.07) + (0)(0.18) + (1)(0.24) + (2)(0.20) + (3)(0.24) + (4)(0.07)$$

$$= 1.57$$

2. Consider cdf of r.v. X,

X	0	1	2	3	4	5	6
P[X ≤ x]	0.08	0.24	0.50	0.50	0.72	0.93	1.00

Find E(X).

Solution

We are given cdf of X. Therefore, first step is to obtain pmf of X. The table below gives pmf of X and $x_j p(x_j)$.

X	0	1	2	3	4	5	6	Total
P[X ≤ x]	0.08	0.24	0.50	0.50	0.72	0.93	1.00	–
P[X = x]	0.08	0.16	0.26	0.00	0.22	0.21	0.07	1.00
xP[X = x]	0	0.16	0.52	0.00	0.88	1.05	0.42	3.03

Thus, $E(X) = \sum_j x_j\, p(x_j) = 3.03$.

6.1 Properties of Expectation

We are given a r.v. X and its pmf. Many times our interest is not in E(X) but in E[g(X)], where g is any real-valued function of X. *For example*, g(X) may be X^2 or $(X - E(X))^2$, or e^{tx}, where t is any real number and so on. Since X is a r.v. so also will be g(X). We, therefore, can certainly use the definition of E(X) by obtaining the probability distribution of Y = g(X) and find E(Y). However it may not be easy to obtain pmf of g(X). Moreover, it is not needed also. To know more about it study the following example:

Examples

1. Suppose the r.v. X has following probability distribution. The pmf of X is

X	–2	–1	0	1	2
P(x)	0.10	0.20	0.30	0.25	0.15

Find E(X), E(X^2).

Solution

By definition,

$$E(X) = \Sigma\, xp(x) = (-2)(0.1) + (-1)(0.2) + (0)(0.3) + (1)(0.25) + (2)(0.15) = 0.15$$

Let $g(X) = X^2$.

Note the possible values of g(X) are 0, 1 and 4. These occur respectively with probabilities 0.30, 0.45 and 0.25.

Thus, the pmf of Y = g(X) is given by

y	0	1	4
P[Y = y]	0.30	0.45	0.25

Thus, $E(Y) = \Sigma\, yp(y) = (0)(0.30) + (1)(0.45) + (4)(0.25) = 1.45$

We shall now compute $E[g(X)] = E(X^2)$ without using pmf of g(X).

$$E(X^2) = \Sigma\, x^2p(x) = (4)(0.10) + (1)(0.20) + (0)(0.30) + (1)(0.25) + (4)(0.15) = 1.45$$

Note that $E(X^2)$ obtained by both ways is the same. It happens because of the property 1, stated below.

Property 1 (Without Proof): Let X be a discrete r.v. with pmf P(x) with range space in R.

Let g(X) be any real valued function of X, then E[g(X)] is given by

$$E[g(X)] = \Sigma\, g(x)\, p(x) \text{ provided the sum converges absolutely.}$$

Since X is a r.v. the function g(X) of the r.v. X itself, is also a r.v.

Let us denote g(X) by Y, that is, when $X = x_j$, $Y = g(x_j)$. By definition,

$E(Y) = \sum_j y\, P_Y(y)$, where $P_Y(y)$ is the pmf of Y.

Property 2: Let X be a discrete r.v. with pmf P(x) then $E[aX + b] = a\,E[X] + b$, where a and b are constants.

Property 3 (The linearity property of expectation): Let X and Y be two r.v.s having finite expectations, then $E[aX + bY]$ exists and $E[aX + bY] = a\,E[X] + b\,E[Y]$, where a and b are constants.

If we put $a = b = 1$ in $E[a X + b Y]$. We then have $E[X + Y] = E[X] + E[Y]$. Similarly, if we put $a = 1$ and $b = -1$ to get $E[X - Y] = E[X] - E[Y]$.

---

2. A biased die is rolled. Let X denote the score on the die. Let P[X = x] be proportional to x. Find E[X].

Solution

Let X denote the number on the die. Since $P[X = x]$ is proportional to x, we write $P(x) = kx$, where $k > 0$. Then we must have $\sum P(x) = 1 \Rightarrow k + 2k + 3k + 4k + 5k + 6k = 1 \Rightarrow 21k = 1$ or $k = \frac{1}{21}$. Thus, the pmf of X is

X	1	2	3	4	5	6
P(x)	$\frac{1}{21}$	$\frac{2}{21}$	$\frac{3}{21}$	$\frac{4}{21}$	$\frac{5}{21}$	$\frac{6}{21}$

Clearly

$E(X) = \sum_{x=1}^{6} xP(x) = \frac{(1 + 4 + 9 + 16 + 25 + 36)}{21} = \frac{91}{21}$. Note that $E[X] = \frac{91}{21}$ is not one of the possible values of X. In fact it need not be so.

6.2 Variance

Let X be a discrete r.v. with pmf P(x), having finite mean 'm', the variance of X denoted by Var(X) or σ_x^2 is given by Var(X): $\sigma_x^2 = E[X - E(X)]^2 = \Sigma_x [x - E(X)]^2\, P(x)$.

Simplified form: Var(X) : $\sigma_x^2 = E(X^2) - \{E(X)\}^2 = E(X^2) - m^2$.

Remarks

- Variance is always a non-negative number.
- The positive square root of the variance is called as the standard deviation (s.d.) and is usually denoted by σ.
- The variance and s.d. are the measures of the 'spread' or 'dispersion' of a distribution. If a r.v. X takes values near to E(X) with a large probability, X will have a "concentrated" distribution, otherwise it will have a large variance.
- The s.d. must be expressed in the same units as the individual values of the r.v.
- Variance is zero if and only if the r.v. X assumes only one constant value. Such a r.v. is called a **degenerate** r.v.
- $Var(aX + b) = a^2 Var(X)$, variance is invariant to change of origin but not of scale.

Examples

1. **Consider pmf of a discrete r.v. X**

x	–1	0	1	2	3	4	5
P[X = x]	0.05	0.18	0.25	0.22	0.15	0.09	0.06

Find Var(X) and also Var(4X – 3).

Solution

First we compute E(X).

$$E(X) = (-1)(0.05) + (0)(0.18) + (1)(0.25) + (2)(0.22) + (3)(0.15) + (4)(0.09) + (5)(0.06)$$
$$= 1.75$$

Now we compute Var(X) using its definition.

X	–1	0	1	2	3	4	5
P[X = x]	0.05	0.18	0.25	0.22	0.15	0.09	0.06
X – E(X)	–2.75	–1.75	–0.75	0.25	1.25	2.25	3.25
$(X - E(X))^2$	7.5625	3.0625	0.5625	0.0625	1.5625	5.0625	10.5625
$(X - E(X))^2 P(x)$	0.378125	0.551250	0.140625	0.013750	0.234375	0.455625	0.633750

Thus, we have Var(X): $\sigma_x^2 = E[X - E(X)]^2 = \Sigma_x [x - E(X)]^2 P(x) = 2.4075$.

$Var(4X - 3) = 4^2 Var(X) = 16 \times 2.4075 = 38.52$. (–3 is change of origin factor, which has no effect on variance.)

2. In example 1, in section 6.1, we have $E(X) = 0.15$ and $E(X^2) = 1.45$ thus $Var(X) = 1.45 - (0.15)^2 = 1.4275$ and standard deviation $\sigma = 1.1948$.

7. Standard Probability Distributions

In this section, we shall introduce two important standard probability distribution, viz., binomial distribution and Poisson distribution.

7.1 Discrete Uniform Distribution

In some real life situations involving classical assignment of probability can be modelled by a discrete r.v. that assumes all of its values with the same probability. *For example*, a fair dice is rolled; a lucky ticket is drawn out of N lottery tickets. In these experiments, the use of discrete uniform distribution is appropriate. The possible outcomes of these experiments can be related to a set of consecutive natural numbers.

Definition: A r.v. is said to have a discrete uniform distribution on integers 1, 2, ..., N if its pmf is given by

$$P[X = x] = \begin{cases} \frac{1}{N}, & \text{for } x = 1, 2, 3, \ldots, N \\ 0, & \text{otherwise} \end{cases}$$

This can be abbreviated as $X \rightarrow DU$

The pmf can be represented in table form as follows:

X	1	2	3	......	N	Total
P[X = x]	$\frac{1}{N}$	$\frac{1}{N}$	$\frac{1}{N}$	......	$\frac{1}{N}$	1

Example

Suppose a fair dice is rolled. Let X denote the outcome on the dice. Clearly X takes values 1, 2, 3, 4, 5 and 6, each with probability $\frac{1}{6}$. The pmf of X in table form is as follows:

$$X \rightarrow DU\ (6)$$

X	1	2	3	4	5	6	Total
P[X = x]	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	1

7.2 Cumulative Distribution Function (CDF)

By definition, cdf $F_X(x)$ is given by

$$F_x(x) = P[X \le x] = \sum_{k=1}^{[x]} P[X = k]$$

Since $X \to DU(N)$, we have $F_X(x) = \underbrace{\frac{1}{N} + \frac{1}{N} + \frac{1}{N} + \ldots + \frac{1}{N}}_{[x] \text{ times}} = \frac{[X]}{N}$

where [x] is the largest integer smaller than x, that is $[x] \le x < [x] + 1$. Hence $F_X(x) = F_X([x])$.

We can represent pmf and cdf in the tabular form as follows:

x	1	2	3	……	N − 1	N
P[X = x]	$\frac{1}{N}$	$\frac{1}{N}$	$\frac{1}{N}$	……	$\frac{1}{N}$	$\frac{1}{N}$
FX(x)	$\frac{1}{N}$	$\frac{2}{N}$	$\frac{3}{N}$	……	$\frac{N-1}{N}$	$\frac{N}{N} = 1$

Mean

The mean of discrete uniform r.v. X is given by

$$E(X) = \sum_{x=1}^{N} xP[X = x] = \sum_{x=1}^{N} x \cdot \frac{1}{N} = \frac{1}{N} \sum_{x=1}^{N} x = \frac{1}{N}(1 + 2 + 3 + \ldots + N) = \frac{1}{N} \times \frac{N(N+1)}{2} = \frac{N+1}{2}$$

Variance

The variance of X is given by $\sigma^2 = E(X^2) - [E(X)]^2$.

Consider

$$E(X^2) = \sum_{x=1}^{N} x^2 P[X = x]$$

$$= \sum_{x=1}^{N} x^2 \cdot \frac{1}{N}$$

$$= \frac{1}{N}\sum_{x=1}^{N} x^2 = \frac{1}{N}(1^2 + 2^2 + 3^2 + \ldots + N^2)$$

$$= \frac{1}{N} \times \frac{N(N+1)(2N+1)}{6} = \frac{(N+1)(2N+1)}{6}$$

$$\therefore \sigma^2 = \frac{(N+1)(2N+1)}{6} - \left[\frac{(N+1)}{2}\right]^2 = \frac{N^2-1}{12}$$

Examples

1. A r.v. $X \to DU(N)$ such that mean = variance, find N.

Solution

Given: $X \to DU(N)$ such that mean = variance, therefore

$$\frac{N+1}{2} = \frac{N^2-1}{12} \Rightarrow N+1 = \frac{(N+1)(N-1)}{6} \Rightarrow N-1 = 6 \Rightarrow N = 7$$

2. A fair die is tossed twice. Let X denote the outcome in first toss and Y denote the outcome in second toss. Obtain the pmf of X + Y.

Solution

Clearly $X \to DU(6)$ and $Y \to DU(6)$ and X and Y are independent r.v.s.

The pmf of X is given by

x	1	2	3	4	5	6	Total
P[X = x]	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	1

The pmf of Y is given by

y	1	2	3	4	5	6	Total
P[Y = y]	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	1

Let Z = X + Y. The possible values of Z are 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 and 12. The table below gives the pairs (x, y) associated with the possible values of Z and corresponding probabilities.

Probability Mass Function of Z = X + Y

Z	2	3	4	5	6	7	8	9	10	11	12	
Sample Points	(1,1)	(1,2) (2,1)	(1,3) (2,2) (3,1)	(1,4) (2,3) (3,2) (4,1)	(1,5) (2,4) (3,3) (4,2) (5,1)	(1,6) (2,5) (3,4) (4,3) (5,2) (6,1)	(2,6) (3,5) (4,4) (5,3) (6,2)	(3,6) (4,5) (5,4) (6,3)	(4,6) (5,5) (6,4)	(5,6) (6,5)	(6,6)	Total
P[Z = z]	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	1

The above pmf can also be written in the following form:

$$P(x) = P[X = x] = \frac{6 - |x - 7|}{36}; \; x = 2, 3, 4, \ldots, 12$$

Observe that the properties given in equation (1) are satisfied by above two functions P(x), therefore they are proper pmfs.

Graphical representation of pmf: The pmf of any discrete r.v. can be represented by line diagram. The line diagram is drawn by plotting on scaled graph with co-ordinates (x, p(x)) and then perpendiculars are drawn from these points on to X-axis. We illustrate this with the help of above pmf.

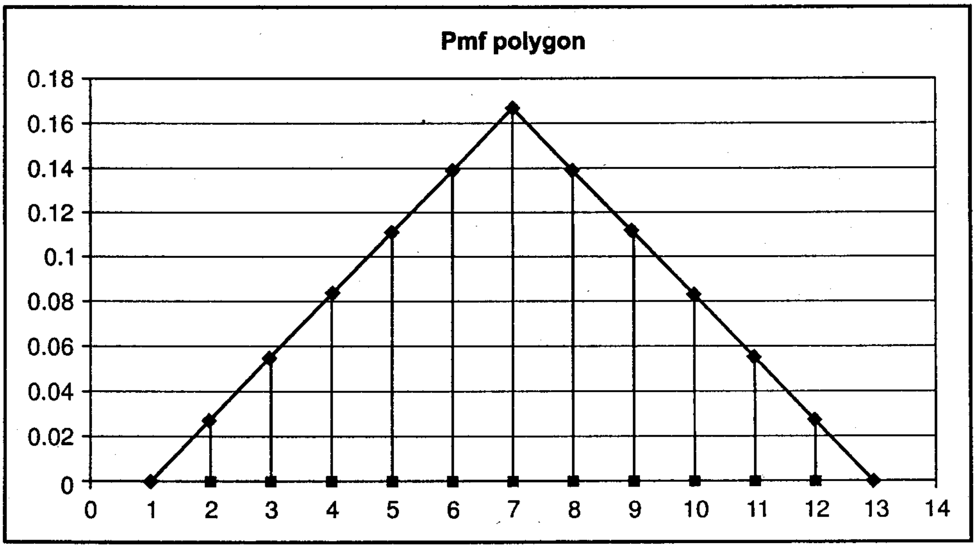


Figure 4.1: Graphical representation of pmf

From the above graph it can be seen that the pmf polygon is a triangle. It can be shown in general also, if X and Y are independent identically distributed DU(N) r.v.s then X + Y has triangular distribution taking values 2, 3, 4, ..., 2N.

7.3 Bernoulli Distribution

Consider an experiment which results in either success or failure in one trial. *For example*, tossing a coin, result is either head or tail, student appearing for an examination either passes or fails, an item inspected is either good or bad, so on. Such an experiment is called as Bernoulli experiment or Bernoulli trial. Usually one of the outcomes of the trial is labelled as 'success' and the other as 'failure'. We assign value '1' to success and '0' to failure. Let 'p' is the probability of success. Then we can write

Outcome	X	P[X = x]
Failure	0	$q = 1 - p$
Success	1	$p, 0 < p < 1$

A random variable (r.v.) X which assumes only two values 0 or 1 is called Bernoulli variable.

Definition: A r.v. X is said to have Bernoulli distribution with parameter 'p' if its pmf is given by $P[X = 0] = q = 1 - p$ and $P[X = 1] = p; 0 < p < 1$.

It also can be written as

$$P[X = x] = \begin{cases} p^x (1-p)^{1-x}; & \text{for } x = 0,1; 0 < p < 1 \\ 0; & \text{otherwise} \end{cases}$$

CDF

The cumulative distribution function is given by $F(x) = P[X \le x]$. Thus we have

X	0	1
P[X = x]	q	p
F(x)	q	q + p = 1

Mean

Mean is given by

$$E(x) = \sum_{x=0}^{1} x\,P(x) = 0 \times q + 1 \times p = p$$

Variance

First we shall compute $E(X^2)$.

$$E(X^2) = \sum_{x=0}^{1} x^2 P(x) = 0^2 \times q + 1^2 \times p = p$$

Variance 'σ^2' is given by $\sigma^2 = E(X^2) - [E(X)]^2 = p - p^2 = p(1-p) = pq$.

7.4 Binomial Distribution

Consider an experiment which results in either success or failure in one trial. *For example*, tossing a coin, result is either head or tail, student appearing for an examination either passes or fails, an item inspected is either good or bad, so on. If 'p' is the probability of success and the experiment is conducted under identical conditions for a fixed number of time, say n, then the number of successes is a random variable (X) which possesses binomial distribution with parameters 'n' and 'p'.

Definition: A discrete random variable is said to have binomial distribution with parameters n and p if its pmf is given by

$P[X = x] = {}^nC_x\, p^x (1-p)^{n-x}$; $x = 0, 1, 2, \ldots, n$; $0 < p < 1$.

Sometimes $1 - p$ is replaced by q, then the pmf is written as

$P[X = x] = {}^nC_x\, p^x q^{n-x}$; $x = 0, 1, 2, \ldots, n$; $0 < p < 1$; $q = 1 - p$.

Remark

- Notation $X \to B(n, p)$.
- Mean: $E(X) = np$, $Var(X) = npq$. Mean > Variance.
- Binomial distribution is the basis for developing control charts for fraction defectives, that is, p-charts.

Solved Examples

1. Let $X \to B(8, 0.4)$. Find $P[X = 3 \text{ or } X = 4]$.

Solution

$X \to B(8, 0.4)$.

$P[X = 3 \text{ or } X = 4] = P[X = 3] + P[X = 4]$

$= 0.2787 + 0.2322 = 0.5109$ (From Statistical tables).

2. A fair coin is tossed 6 times, if getting head is a success find the probability that

i. exactly 2 heads occur, ii. at least four heads occur.

Solution

Let X denote the number of heads occurred in 6 tosses. Clearly $X \to B(6, 0.5)$.

The probability that exactly 2 heads occur is

$P(X = 2) = {}^6C_2\,(0.5)^2\,(0.5)^4 = 0.234375.$

The probability that at least 4 heads occur is

$$\begin{aligned} P(X \geq 4) &= P(X = 4) + P(X = 5) + P(X = 6) \\ &= {}^6C_4\,(0.5)^4\,(0.5)^2 + {}^6C_5\,(0.5)^5\,(0.5)^1 + {}^6C_6\,(0.5)^6\,(0.5)^0 \\ &= 0.234375 + 0.093750 + 0.015625 = 0.343750. \end{aligned}$$

3. A machine produces 15% defective items of its total output. A sample of 10 items is drawn randomly, what is the probability that the sample will contain

i. at most 2 defective items, ii. at least 8 defective items?

Solution

Let X denote the number of defective items in the sample of size 10. Here success is to get an defective item, hence the probability of success is 0.15. Therefore, $X \to B(10, 0.15)$.

The probability that at most 2 defective items is

$$\begin{aligned} P(X \leq 2) &= P(X = 0) + P(X = 1) + P(X = 2) \\ &= {}^{10}C_0\,(0.15)^0\,(0.85)^{10} + {}^{10}C_1\,(0.15)^1\,(0.85)^9 + {}^{10}C_2\,(0.15)^2\,(0.85)^8 \\ &= 0.196874 + 0.347425 + 0.275897 = 0.820196 \end{aligned}$$

The probability that at least 8 defective items is

$$\begin{aligned} P(X \geq 8) &= P(X = 8) + P(X = 9) + P(X = 10) \\ &= {}^{10}C_8\,(0.15)^8\,(0.85)^2 + {}^{10}C_9\,(0.15)^9\,(0.85)^1 + {}^{10}C_{10}\,(0.15)^{10}\,(0.85)^0 \\ &= 8.6651332 \times 10^{-6} = 0.0000086651332. \end{aligned}$$

4. The mean and variance of a binomial distribution are 5.4 and 2.97 respectively, find $P[X = 3]$.

Solution

We have $X \to B(n, p)$, where n and p are unknown. We know that $E(X) = 5.4$ and $Var(X) = 2.97$.

That is np = 5.4 and npq = 2.97. Therefore, $\frac{npq}{np} = q = \frac{2.97}{5.4} = 0.55.$

Thus, $p = 1 - q = 0.45$. This gives us n = 12.

Hence $X \to B(12, 0.45)$. Then $P[X = 3] = {}^{12}C_3\,(0.45)^3\,(0.55)^9 = 0.092326.$

---

5. The incidence of a certain disease is such that on the average 20% of workers suffer from it. If 12 workers are selected at random, find the probability that

i. exactly 3 workers suffer from it, **ii. at most 2 workers suffer from it.**

Solution

Let X denote the number of workers suffering from a certain disease out of 12. The probability that worker suffers from disease is $\frac{20}{100} = 0.20$. Therefore, $X \to B(12, 0.20)$.

i. The probability that exactly 3 workers suffer from it is

$$P(X = 3) = {}^{12}C_3\,(0.20)^3\,(0.80)^9 = 0.236223$$

ii. The probability that at most 2 workers suffer from it is

$$\begin{aligned} P(X \le 2) &= P(X = 0) + P(X = 1) + P(X = 2) \\ &= {}^{12}C_0\,(0.20)^0\,(0.80)^{12} + {}^{12}C_1\,(0.20)^1\,(0.80)^{11} + {}^{12}C_2\,(0.20)^2\,(0.80)^{10} \\ &= 0.068719 + 0.206158 + 0.283468 \quad = 0.558365 \end{aligned}$$

---

6. In a pot 10 seeds are planted. The probability of germination of a seed is 0.90. Assuming independence of seed germination, obtain the probability that

i. all seeds germinate, **ii. no seed germinates,**

iii. exactly 5 seeds germinate.

Solution

Let X denote the number of seeds germinate out of 10 seeds. The probability that a seed germinates is 0.90. Therefore, $X \to B(10, 0.90)$.

i. The probability that all seeds germinate

$$P[X = 10] = {}^{10}C_{10}\,(0.90)^{10}(0.10)^0 = 0.348678.$$

ii. The probability that no seed germinates

$$P[X = 0] = {}^{10}C_0\,(0.90)^0(0.10)^{10} = \frac{1}{10} \approx 0.$$

iii. The probability that exactly 5 seeds germinate

$$P[X = 5] = {}^{10}C_5\,(0.90)^5(0.10)^5 = 0.001635.$$

7. The probability that a machine produces a defective item is 0.05. Find the probability that in a lot of 20 items

i. there is no defective item, **ii. exactly 2 defective items,**

iii. at most three defective items.

Solution

Let X denote the number of defective items in a lot of 20 items. The probability that an item is defective equals 0.05. Therefore, $X \rightarrow B(20, 0.05)$.

i. The probability that there is no defective item

$$P[X = 0] = {}^{20}C_0\,(0.05)^{20}\,(0.95)^0 = 0.358486.$$

ii. The probability that exactly 2 defective items.

$$P[X = 2] = {}^{20}C_2\,(0.05)^2\,(0.95)^{18} = 0.188677.$$

iii. The probability that at most 3 defectives.

$$\begin{aligned} P(X \leq 3) &= P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) \\ &= {}^{20}C_0\,(0.05)^0\,(0.95)^{20} + {}^{20}C_1\,(0.05)^1\,(0.95)^{19} + {}^{20}C_2\,(0.05)^2\,(0.95)^{18} \\ &\quad + {}^{20}C_3\,(0.05)^3\,(0.95)^{17} \\ &= 0.358486 + 0.377354 + 0.188677 + 0.059582 = 0.984098 \end{aligned}$$

8. How many independent trail each with p = 0.10 must be performed to ensure that the probability of at least one success is 0.70 or more?

Solution

The probability of at least one success in 'n' independent Bernoulli trails is

$$P(X \geq 1) = 1 - P(X = 0) = 1 - (0.90)^n.$$

We must find 'n' such that $1 - (0.90)^n \geq 0.70$, that is, $0.30 \geq (0.90)^n$.

Consider the following table where we calculate $(0.90)^n$.

n	$(0.90)^n$
1	0.9000
2	0.8100
3	0.7290
4	0.6561
5	0.5905
6	0.5314
7	0.4783
8	0.4305
9	0.3874
10	0.3487
11	0.3138
12	0.2824*

Here we find that for $n = 12$, $(0.90)^n < 0.30$. Thus 12 independent trails are required.

(Alternately, consider the equation

$$\log_{10}(0.30) = n \log_{10}(0.90) \Rightarrow -0.5229 \geq -0.0458n.$$

This gives us $n = \dfrac{-0.5229}{-0.0458} = 11.43$. That is $0.30 = (0.90)^{11.43}$. So as to have $0.30 \geq (0.90)^n$, we must have $n > 11.43$, hence we take $n = 12$.)

---

9. A manufacturer of regulator valves claims that only 2% of his production is defective. He supplies valves in a shipment containing 1000 boxes of 10 regulator valves. Find the expected number of boxes having at most 1 defective valve.

Solution

Here we have $X \to B(10, 0.02)$. First we shall obtain $P(X \leq 1)$.

$$\begin{aligned} P(X \leq 1) &= P(X = 0) + P(X = 1) \\ &= {}^{10}C_0\,(0.02)^0\,(0.98)^{10} + {}^{10}C_1\,(0.02)^1\,(0.98)^9 \\ &= 0.817073 + 0.166750 = 0.983822. \end{aligned}$$

Expected number of boxes having at most 1 defective valve

$$= 1000 \times P(X \leq 1) = 1000 \times 0.983822 = 983.822 \approx 984.$$

EXERCISES

1. Explain the following terms:
   i. Discrete sample space
   ii. Discrete random variable
   iii. Probability mass function
   iv. Cumulative distribution function
2. Check whether the functions given below are valid pmf. Obtain unknown constants, if any, in P(x) so that P(x) becomes a pmf.
   i. $P(x) = k(x + 2)$, for $x = 1, 2, 3, 4$
   ii. $P(x) = \frac{k}{1 + x^2}$; for $x = 0, 1, 2$
   iv. $P(x) = kx^2$; for $x = 1, 2, 3, 4, 5$
   v. $P(x) = k(x + 1)^2$; for $x = -3, -2, -1, 1, 2, 3$
3. A r.v. X has following probability distribution:

0	2	4	6	8	10	12
c	2c	2c	3c	c^2	$2c^2$	$6c^2$

   i. Find c,
   ii. Obtain $P[X \le 4]$ and $P[(0 \le X \le 6) \cup (8 < X \le 12)]$.
4. A r.v. X has following probability distribution.

X	1	2	3	4	5	6
P[X = x]	$\frac{2}{5}$	p	$\frac{p}{2}$	$\frac{p}{4}$	$\frac{p}{8}$	$\frac{5p}{8}$

   Find p, $P[X \le 5]$, $P[X \ge 3]$, pmf of $4X - 3$.
5. The probability distribution of a r.v. is
   $P[X = x] = kx$; for $x = 6, 7, \ldots, 20$
   $= 0$; otherwise
   Find k, $P[X \ge 8]$, $P[6 \le X \le 13]$, $P[(7 < X \le 12) \cup (10 < X \le 15)]$.
6. Find the probability distribution of the absolute difference of the numbers when a pair of fair dice is rolled.
7. From an urn containing 8 red and 5 white balls a man is to draw three balls at random with replacement, find the probability distribution of number of white balls drawn.
8. An urn contains 8 white balls and 4 black balls. From it five balls are selected by the procedure of
   i. simple random sampling with replacement, and
   ii. simple random sampling without replacement.
   For each of these cases find the distribution of the number of white balls in the sample.
9. A discrete r.v. takes values 0, 1, 2. If for some constant k, $P[X = i] = kP[X = i - 1]$, for $i = 1, 2$. Find E(X) and V(X).
10. Suppose the r.v. X has following pmf: $P[X = x] = \frac{1}{6}$ for $x = 1, 2, 3, 4, 5, 6$.
    Find mean, median, mode and V(X).

11. If $X \to DU(10)$; which of the following statements is correct.

a. $E(X) = 5.5, Var(X) = \frac{100}{12}$ b. $E(X) = 5.5, Var(X) = \frac{101}{12}$

c. $E(X) = 5.5, Var(X) = \frac{99}{12}$ d. $E(X) = 5.0, Var(X) = \frac{99}{12}$

12. A one digit random number (X) is drawn, assuming that X takes values 0, 1, 2, ..., 9 with equal probability.

Obtain $P[X \leq 5]$, $P[X > 4]$, $P[2 < X \leq 7]$, $P[X \geq 3 \mid X < 6]$.

13. Let $X \to DU(3)$ and $Y \to DU(5)$, assuming independence of X and Y, find the probability distribution of

i. $U = X + Y$, ii. $V = |X - Y|$.

14. A person lands at an international airport. He has the option to select any one of the four travel routes, say A_i for i = 1, 2, 3 and 4, to reach his destination. The route A_1 costs Rs. 500/- and requires two hours of travel time. The route A_2 costs Rs. 350/- and requires three hours. The routes A_3 and A_4 both cost Rs. 300 but respectively require four and five hours of journey time. Assume that the person chooses his options randomly each time, find his expected journey time, his expected cost to reach the destination and also the average cost per hour.

15. The number of hardware failures of a computer system in a fortnight of operation has the following pmf:

No. of failures	0	1	2	3	4	5	6
Probability	0.20	0.30	0.25	0.10	0.08	0.05	0.02

Find average number of failures by

i. mean, ii. mode,

iii. median and iv. the standard deviation of number of failures.

16. Suppose X takes values 0.5 and 2 with equal probability. Find E(X), E(1/X) and comment on the results.

17. Let $X \to B(10, 0.45)$. Find

i. $P[X = 4]$, ii. $P[X < 3]$, iii. $P[X > 7]$.

18. Let $X \to B(2, p)$; $Y \to B(4, p)$. If $P[X \geq 1] = \frac{21}{25}$; find $P[Y \geq 2]$.

19. Let $X \to B(n, p)$; $E(X) = 6$ and $Var(X) = 4.2$, find n and p.

20. Three out of every hundred income-tax returns are found to be illegitimate. If an office chooses 12 returns at random find the probability that

i. none return is illegitimate, ii. at most 3 returns are illegitimate.

21. The probability that the battery cell will last for at least 180 days is 0.70. Find the probability that among 15 such cells at least 12 cells will last for 180 or more days.

ANSWERS

2. i. $\frac{1}{18}$ ii. $\frac{10}{17}$ iii. $\frac{1}{55}$ iv. $\frac{1}{34}$

3. i. $\frac{1}{9}$, ii. $\frac{5}{9}$ and $\frac{80}{81}$

4. $\frac{6}{25}, \frac{17}{20}, \frac{9}{25}$, Probability distribution of 4X – 3

4X – 3	1	5	10	13	17	21
P[X = x]	$\frac{2}{5}$	$\frac{6}{25}$	$\frac{3}{25}$	$\frac{3}{50}$	$\frac{3}{100}$	$\frac{3}{20}$

5. $k = \frac{1}{195}, \frac{21}{65}, \frac{92}{195}$

6. $X = |a - b|$

X	0	1	2	3	4	5
P[X = x]	$\frac{3}{18}$	$\frac{5}{18}$	$\frac{4}{18}$	$\frac{3}{18}$	$\frac{2}{18}$	$\frac{1}{18}$

7. $P[0] = \left(\frac{8}{13}\right)^3$, $P[1] = 3\left(\frac{8}{13}\right)^2\left(\frac{5}{13}\right)$, $P[2] = 3\left(\frac{8}{13}\right)\left(\frac{5}{13}\right)^2$, $P[3] = \left(\frac{5}{13}\right)^3$

8. Probability distribution X: No. of white balls:

X	0	1	2	3	4	5
P[X = x]	0	0.041152	0.164609	0.329218	0.329218	0.135803

9. Let P(0) = p, then P(1) = kp, $P(2) = k^2p$. E(X) = kp(1 + 2k), $Var(X) = kp(1 + 8p) - kp^2(1 + 2k)^2$

10. 3.5, 3, mode does not exist, Var = 2.916667

11. c

12. $\frac{3}{5}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}$

13. Pmf of X + Y

Z = X + Y	2	3	4	5	6	7	8
P[Z = z]	$\frac{1}{15}$	$\frac{2}{15}$	$\frac{3}{15}$	$\frac{3}{15}$	$\frac{3}{15}$	$\frac{2}{15}$	$\frac{1}{15}$

Pmf of |X – Y|

Z = |X – Y|	0	1	2	3	4
P[Z = z]	$\frac{3}{15}$	$\frac{5}{15}$	$\frac{4}{15}$	$\frac{2}{15}$	$\frac{1}{15}$

14. Rs. 362.5, 3.5 hr, Rs.103.57/hr

15. 1.79, 1, 1, 1.498633

16. E(X) = E(1/X) = 1.25. This happens because $x_1 = \frac{1}{x_1}$ with equal probabilities.

17. 0.2060, 0.0652, 0.0610

18. p = 0.6, 0.8208

19. n = 20, p = 0.3

20. 0.693842, 0.99967

Chapter 5

SIMULATION TECHNIQUES

1. Introduction

In decision making we come across different types of models. However, every situation can not be represented in a perfect mathematical or stochastic model. The situation or system may be too complex to be represented by a model. There are situations in businesses, commerce, economics, social, medical and physical sciences it is not possible to build an appropriate model. This drawback can be overcome by making use of virtual or imitative models. Such models are generally called as simulation models.

In this chapter we study some simple simulation techniques.

2. Definition

R. E. Shannon defines simulation as 'Simulation is the process of designing a model of a real system and conducting experiments with this model for the purpose either of understanding the behavior of the system or of evaluating various strategies (within the limits imposed by a criterion or set of criteria) for the operation of the system'.

This can be further described as the solution of any problem in a system becomes complicated if the system is not adequately represented by theoretical models. Then by the knowledge of the important characteristics and rules of operation of a system, we can visualize the behavior of the system. It is possible by applying random sampling to the components of the objects or events of a population to find the corresponding probability and then the expressions in the model are evaluated. In certain models, there are terms and variables which can not be evaluated exactly. However, by assuming some theoretical probability distribution of various components we can find the solution. The procedure is generally called as simulation.

Simulations models are abstractions of reality. Thus, simulation is the imitation of the operation of a real-world process or system over time. In the act of simulating something we first require to develop a model which represents the key characteristics or behaviors/functions of the selected physical or abstract system or process. The model represents the system itself, whereas the simulation represents the operation of the system over time.

In day to day life, we come across events like paying bills, withdrawing money from bank or ATM, filling petrol in vehicle; where we need to wait in queues for some time. How much do we need to wait? This can not be answered. Taking into account the arrival rate and departure rate of customers we can build a waiting line model. Simulating the model the service providers can take proper decisions so that the customers need to wait in queue for shortest time on an average.

Simulation techniques are used because

i. the system as yet does not exist,

ii. experimentation with the system is expensive, too time consuming, too dangerous,

iii. experimentation with the system is inappropriate, *for example*, disaster planning.

However, if the system can be modeled analytically simulation is not carried out.

3. Random Number Generator

The procedure exercised for simulating a model is based on the use of set of random numbers. Random numbers are the numbers selected in such a way that every number has an equal chance of selection.

In 1930, *Kendall and Bobington_Smith* published random number tables which were generated with the help of a spinning disc illuminated by a flash lamp. The simplest random number generators are coins, dice and bags of colored balls, playing card. Coins and Dice are impracticable for all simulations. Many times simulations are conducted with aid of readily available random number tables. In recent years, computers are used to generate random numbers.

Pseudo-Random Number Generator

When a computer is used in simulation then it is not advisable to fill computer memory with a large number of random digits. Programs have been developed to generate predictable and reproducible random numbers. A **Pseudorandom Number Generator** (**PRNG**), also known as a **Deterministic Random Bit Generator** (**DRBG**) is an algorithm for generating a sequence of numbers that approximates the properties of random numbers. The sequence is not truly random because it is completely determined by a relatively small set of initial values, called the PRNG's state, which includes a truly random seed.

In MS-EXCEL, the standard function **RAND()** generates uniform random numbers between 0 and 1. Also, there are many methods to generate pseudo-random numbers. The commonly used methods are the mid-square and congruential techniques.

4. Monte Carlo Simulation

Monte Carlo technique is a part of simulation procedure where random observations are selected within simulated model. The samples of random variables associated either with inputs of the processing operations or both are drawn.

Monte Carlo simulation was named after the city in Monaco (famous for its casino) where games of chance (e.g., roulette) involve repetitive events with known probabilities. The technique came into existence in early 1900's and the main contributors are Von-Neumann, *Ulam and Fermi*. The method has found useful in the field of social science, military, industrial and business operations. *Dr. A. S. Householder* defines *Monte Carlo* Techniques as the device for studying an artificial stochastic model of a physical or mathematical process. It is also opined that these methods are a combination of probability methods and sampling techniques providing solutions to complicated partial or integral differential equations.

Monte Carlo methods are useful in following situations:

i. The cases where mathematical and statistical problems are very complex to deal with.

ii. Situations where it is not possible to have practical experience since the physical system may not exist.

iii. To study transitional process.

iv. To estimate parameters of the model.

v. To construct the functional form of a model.

vi. To create course of action that cannot be formulated into a model.

General procedure of Monte Carlo Method

The method uses random numbers for generating some data by which a problem can be solved. These random numbers are used in creating a new set of hypothetical data of a problem. This is based on past experience. If nothing of the past is known randomness can be assumed for the problem. The Cumulative Distribution Function (CDF) is the basic and most important instrument in the use of Monte Carlo methods. This probability function gives the probability of falling within a given interval and the next higher one. Monte Carlo technique involves the assumption of probability distribution for variables at different stages of the process. Using random number tables representative samples are drawn from the distribution and numerical realization of the process is built.

To simulate the process we use following steps:

i. Prepare CDF for the process using past experience or knowledge about the process.

ii. Select a sequence of random numbers from random number tables.

iii. The random numbers so obtained are taken as cumulative probabilities.

iv. These cumulative probabilities are compared with CDF.

v. The range in which the probability falls, corresponding X value is the desired sampling value.

Note that in Monte Carlo simulation, the entire system is simulated a large number of times. Each simulation is equally likely, referred to as a realization of the system. For each realization, all of the uncertain parameters are sampled. The system is then simulated through time using given set of input parameters such that the performance of the system can be computed. This results in a large number of separate and independent results, each representing a possible "future" for the system. The results of the independent system realizations are assembled into probability distributions of possible outcomes. As a result, the outputs are not single values, but probability distributions.

5. Model Sampling from Discrete Distributions

Consider a discrete probability distribution such that the random variable X assumes values $x_1, x_2, \ldots, x_k$ with probabilities $p_1, p_2, \ldots, p_k$. To draw model sample of size 'n' follow the steps given below:

Obtain the cumulative distribution of X, F(x).

i. Use random number tables and select a set of 'n' random numbers having 4 or 6 digits.

ii. Represent these random numbers as random probabilities, say U_i, $i = 1, 2, \ldots, n$.

iii. Compare U_i with cumulative probabilities in the given distribution.

iv. Find two successive F(.) such that $F(x_j - 1) < U_i \leq F(x_j)$, then select x_j as sample value.

v. Continue the procedure till you get a sample of size 'n'.

Examples

1. Draw model sample of size 8 from the distribution of X given below:

X	1	2	3	4	5	Total
P(x)	0.12	0.25	0.39	0.16	0.08	1.00

Solution

We proceed as follows:

x	P(x)	F(x)	U	Comparison	Sample value
1	0.12	0.12	0.0347	U < F(1)	1
2	0.25	0.37	0.9774	F(4) < U ≤ F(5)	5
3	0.39	0.76	0.1676	F(1) < U ≤ F(2)	2
4	0.16	0.92	0.1256	F(1) < U ≤ F(2)	2
5	0.08	1.00	0.5559	F(2) < U ≤ F(3)	3
			0.1622	F(1) < U ≤ F(2)	2
			0.8442	F(3) < U ≤ F(4)	4
			0.6301	F(2) < U ≤ F(3)	3

Hence the model sample is 1, 5, 2, 2, 3, 2, 4, 3.

(Table XXXIII. Random Numbers (I), Column – 1, Row – 1, Selection columnwise)

This procedure can be used to draw model samples from standard discrete distributions such as Discrete Uniform and Binomial Distribution.

2. Let X → DU (8). Draw model sample of size 10. Also find mean of the sample drawn.

Solution

Here the probability distribution of X is given by

X	1	2	3	4	5	6	7	8	Total
P[X=x]	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125	1.000

Now consider

X	P[X = x]	F(x)	U	Comparison	Sample value
1	0.125	0.125	0.5374	$F(4) < U \le F(5)$	5
2	0.125	0.250	0.6338	$F(5) < U \le F(6)$	6
3	0.125	0.375	0.3530	$F(2) < U \le F(3)$	3
4	0.125	0.500	0.6343	$F(5) < U \le F(6)$	6
5	0.125	0.625	0.9825	$F(7) < U \le F(8)$	8
6	0.125	0.750	0.0263	$U < F(1)$	1
7	0.125	0.875	0.6455	$F(5) < U \le F(6)$	6
8	0.125	1.000	0.8507	$F(6) < U \le F(7)$	7
			0.5854	$F(4) < U \le F(5)$	5
			0.3485	$F(2) < U \le F(3)$	3
				Total	50

(Table XXXIII. Random Numbers (II), Column – 1, Row – 1, Selection columnwise)

The mean $\bar{x} = \frac{50}{10} = 5$.

3. Draw model sample of size 8 from binomial distribution with n = 6 and p = 0.40. Also compute mean and variance for the model sample.

Solution

Here the probability distribution of X is given by

X	0	1	2	3	4	5	6	Total
P[X=x]	0.0467	0.1866	0.3110	0.2765	0.1382	0.0369	0.0041	1.000

Now consider

X	P[X = x]	F(x)	U	Comparison	Sample value (x)	x^2
0	0.0467	0.0467	0.1027	F(0) < U ≤ F(1)	1	1
1	0.1866	0.2333	0.2841	F(1) < U ≤ F(2)	2	4
2	0.3110	0.5443	0.3421	F(1) < U ≤ F(2)	2	4
3	0.2765	0.8208	0.6181	F(2) < U ≤ F(3)	3	9
4	0.1382	0.9590	0.6115	F(2) < U ≤ F(3)	3	9
5	0.0369	0.9959	0.9176	F(3) < U ≤ F(4)	4	16
6	0.0041	1	0.0097	U < F(0)	0	0
			0.3646	F(1) < U ≤ F(2)	2	4
				Total	17	47

(Table XXXIII. Random Numbers (IV), Column – 1, Row – 1, Selection columnwise)

The mean, $\bar{x} = \frac{17}{8} = 2.125.$

The variance is given by

$$\sigma^2 = \frac{\sum_{i=1}^{n} x^2}{n} - \bar{x}^2$$

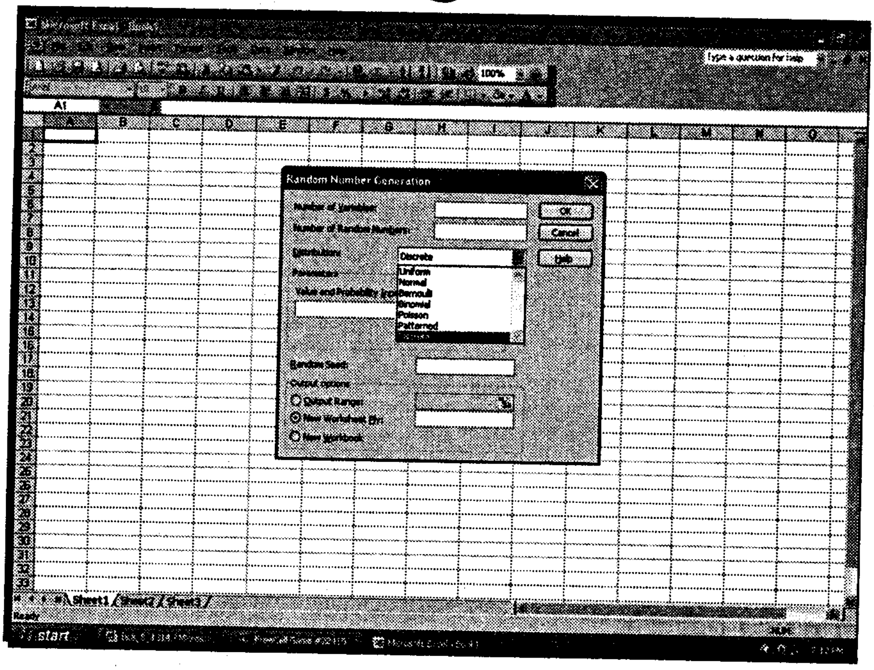
$$= \frac{47}{8} - (2.125)^2$$

$$= 5.8750 - 4.515625$$

$$= 1.359375.$$

6. Computer Aided Simulation

Following picture of MS-EXCEL shows options for simulation of data from Uniform, Normal, Bernoulli, Binomial, Poisson, Patterned and Discrete distributions.



Examples

1. To generate sample from a discrete distribution.

i. Enter the values of random variable and corresponding probabilities in two columns of spreadsheet.

ii. Select tools menu. In which select Random Number Generation.

iii. Get view of the window.

iv. Enter number of variables: say 1.

v. Enter Number of Random Numbers, size of sample.

vi. Select Distribution: Discrete.

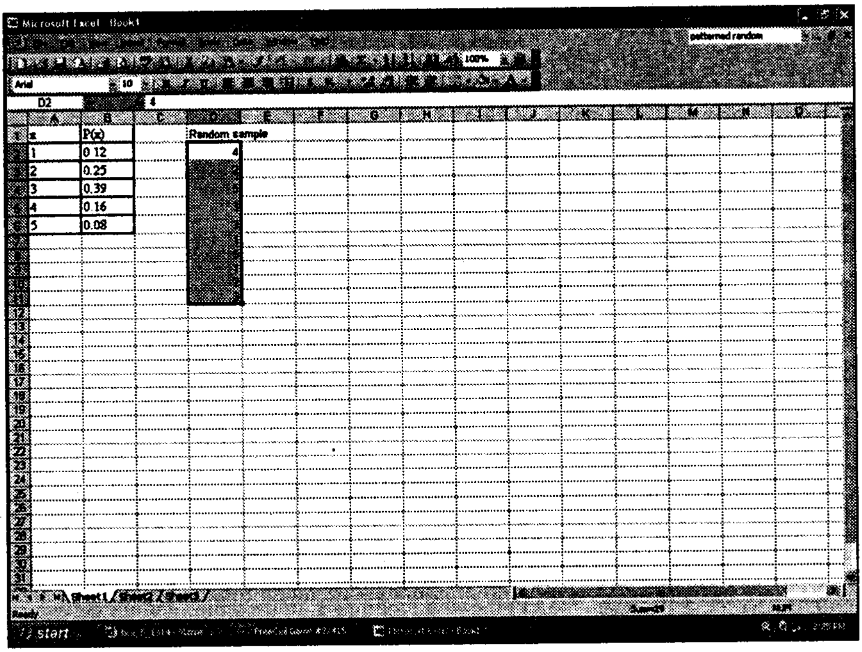
vii. Enter parameters: Value and Probability input range. Enter the starting and end cells of probability distribution.

viii. Enter random seed.

ix. Specify output range: an empty cell where output is expected.

x. Get the random sample.

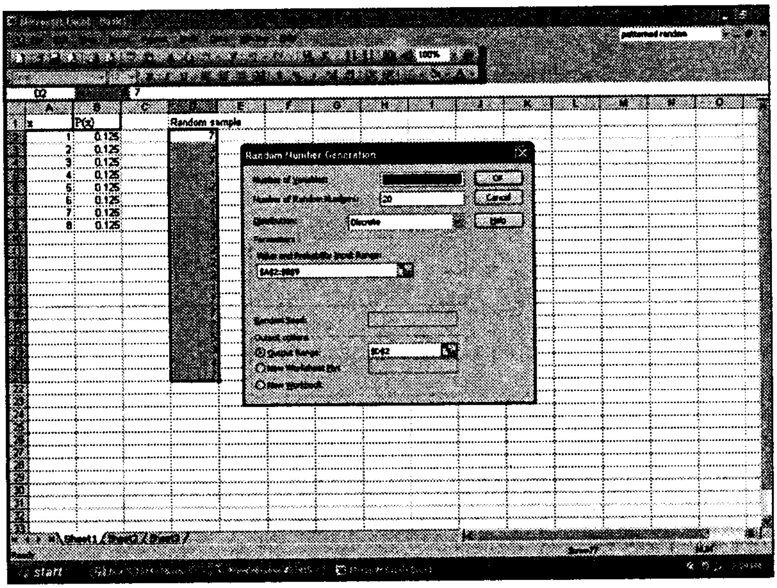
Following is output of a simulation from discrete distribution.



2. Draw a sample of size 20 from DU (8).

In this case we use the above procedure with P(X = x) = 1/8; x = 1, 2, 3, …, 8.

Following is the result.



3. Draw a model sample of size 25 from B(6, 0.40).

i. In this case we don't need to specify the possible values of the random variable and corresponding probabilities.

i. Select menu Tools.

ii. Select Random Number Generation.

iii. Enter number of variables: say 1.

iv. Enter Number of Random Numbers, size of sample.

v. In Distributions select Binomial.

vi. Enter the value of 'p'. Here p = 0.40.

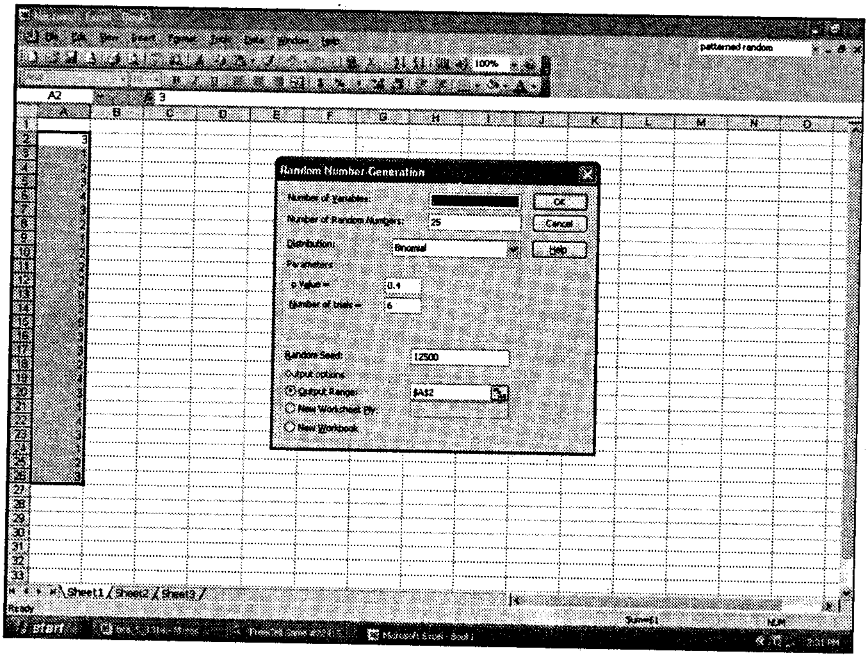
vii. Enter 'Number of trails', the value of 'n'. Here n = 6.

viii. Enter random seed.

ix. Specify output range: an empty cell where output is expected.

x. Get the random sample.

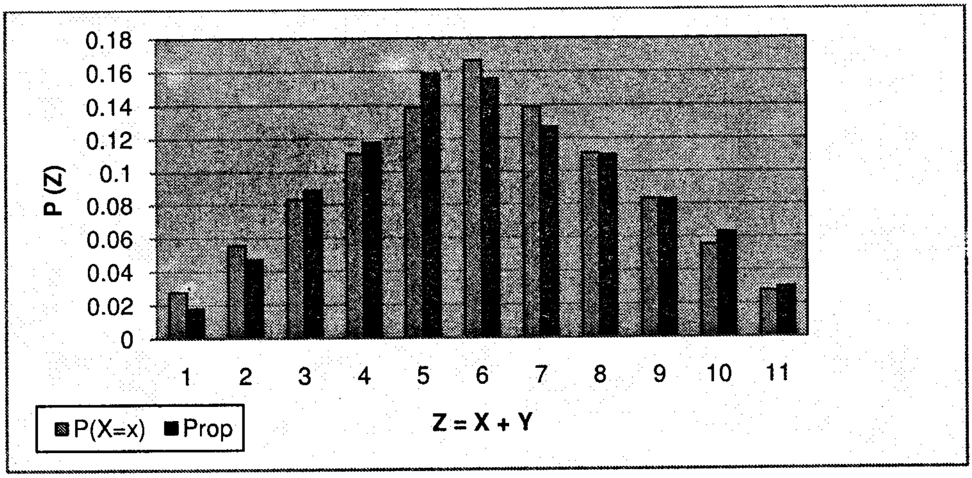
Following picture shows the result.



4. **Following is an output of simulation related to rolling of two dice 1000 times. The table below gives the exact probabilities of $Z = X + Y$, sum of numbers on upper faces and corresponding simulated proportions.**

x	P(X=x)	Prop
2	0.027778	0.018
3	0.055556	0.047
4	0.083333	0.089
5	0.111111	0.118
6	0.138889	0.159
7	0.166667	0.156
8	0.138889	0.127
9	0.111111	0.110
10	0.083333	0.083
11	0.055556	0.063
12	0.027778	0.030

The diagrammatic representation of the same is given below:



7. Merits and Demerits of Simulation

Merits

i. Simulation techniques are easier than mathematical models.

ii. Simulation techniques provide solutions to complex process which otherwise may not be represented mathematically.

iii. It is less time consuming, inexpensive and gives better results as compared to real experimentation.

iv. It can be used to verify the accuracy of an approximate solution to the given system.

v. It is widely accepted and understood by a common man because rigorous mathematics is not involved.

vi. With the aid of computers in simulation it has become very easy to study the results for huge data set in very short time.

vii. Simulation study is repeatable hence one can get complete control over the development of the model.

Demerits

i. It does not produce optimum solution.

ii. Each simulation runs like a single experiment conducted under a given set of condition therefore it is unable to describe each and every characteristic of real-life experiment.

iii. Simulation involves repetition of an experiment hence it becomes time consuming when conducted manually.

iv. Whenever the situation can not be quantified, simulation is not possible.

v. One may unnecessarily rely more often on the simulation technique as it is simple in adaptation although the mathematical model is more suitable in the situation.

EXERCISES

1. What do you mean by 'simulation'?
2. Explain simulation. State its merits and demerits.
3. Explain the concept of pseudo-random numbers. How they are generated?
4. Explain the use of computers in simulation.
5. Explain the procedure of drawing model sample from discrete uniform distribution.
6. Explain the procedure of drawing model sample from binomial distribution.
7. Draw a model sample of size 10 from DU(8). Find mean of the sample.
8. Draw a model sample of size 15 from B(10, 0.60). Find mean and variance of the sample.
9. Draw a model sample of size 50 from DU(12) using MS-EXCEL.
10. Draw a model sample of size 100 from B(8, 0.45) using MS-EXCEL.

Suggestive Reading

Books

1. R. A. Becker, J. M. Chambers, and A. R. Wilks. The New S Language. Chapman & Hall, Boca Raton, FL, 1988.
2. W. John Braun and Duncan J. Murdoch. A First Course in Statistical Programming with R. Cambridge University Press, Cambridge, 2007.
3. Cynthia A. Brewer. Designing Better Maps: A Guide for GIS Users. ESRI Press, Redlands, CA, 2005.
4. John M. Chambers. Data Management in S. Technical report, AT&T Bell Laboratories Statistics Research Report No. 99, 1991.
5. John M. Chambers. Programming with Data: A Guide to the S Language. Springer, New York, 1998.
6. John M. Chambers and Trevor Hastie, editors. Statistical Models in S. Chapman & Hall, Boca Raton, FL, 1992.
7. John M. Chambers, David A. James, Diane Lambert, and Scott Vander Wiel. Monitoring networked applications using incremental quantiles. Statistical Science, 21:463–475, 2006.
8. William S. Cleveland. Elements of Graphing Data. Hobart Press, Summit, NJ, 1985.
9. William S. Cleveland. Visualizing Data. Hobart Press, Summit, NJ, 1993.
10. D. Cook and D. F. Swayne. Interactive and Dynamic Graphics for Data Analysis With R and GGobi. Springer, New York, 2007.
11. Michael J. Crawley. Statistics: An Introduction Using R. John Wiley & Sons, Hoboken, NJ, 2005. 479480
12. Peter Dalgaard. Introductory Statistics With R. Springer, New York, 2002.
13. Editors of the American Heritage Dictionaries. The American Heritage Dictionary of the English Language. Houghton Mifflin, Boston, 4th edition, 2000.
14. Brian S. Everitt and Torsten Hothorn. A Handbook of Statistical Analyses Using R. Chapman & Hall, Boca Raton, FL, 2006.
15. Trevor Hastie, Robert Tibshirani, and Jerome Friedman. The Elements of Statistical Learning. Springer, New York, 2001.

16. W. Kahan. Lecture Notes on the Status of IEEE Standard 754 for Binary Floating-Point Arithmetic. Technical report, Elect. Eng. & Computer Science University of California, Berkley, October 1997.
17. Paul Murrell. R Graphics. Chapman & Hall, Boca Raton, FL, 2005.
18. John North. Stonehenge, Neolithic Man and the Cosmos. Harper & Collins, New York, 1996.
19. Deepayan Sarkar. Lattice: Multivariate Data Visualization with R. Springer, New York, 2008.
20. G. W. Stewart. Matrix Algorithms, Vol. I, II. SIAM, 1998, 2001.
21. W. N. Venables and B. D. Ripley. Modern Applied Statistics with S. Springer, New York, 4th edition, 2002.
22. John Verzani. Using R for Introductory Statistics. Chapman & Hall, Boca Raton, FL, 2005.
23. Larry Wall, Tom Christiansen, and Jon Orwant. Programming Perl. O'Reilly, Sebastopol, CA, 3rd edition, 2000.