

# **FUNDAMENTALS OF MATHEMATICAL STATISTICS**

**(A Modern Approach)**

*A Textbook written completely on modern lines for Degree, Honours, Post-graduate Students of all Indian Universities and Indian Civil Services, Indian Statistical Service Examinations.*

**(Contains, besides complete theory, more than 650 fully solved examples and more than 1,500 thought-provoking Problems with Answers, and Objective Type Questions)**

**S.C. GUPTA**

Reader in Statistics  
Hindu College,  
University of Delhi  
Delhi

**V.K. KAPOOR**

Reader in Mathematics  
Shri Ram College of Commerce  
University of Delhi  
Delhi

**Tenth Revised Edition  
(Greatly Improved)**



**SULTAN CHAND & SONS**

*Educational Publishers  
New Delhi*

- \* First Edition : Sept. 1970
- Tenth Revised Edition : August 2000
- Reprint: 2002

\* Price : **Rs. 210.00**

ISBN 81-7014-791-3

- \* Exclusive publication, distribution and promotion rights reserved with the Publishers.
- \* *Published by :*  
**Sultan Chand & Sons**  
23, Darya Ganj, New Delhi-110002  
Phones : 3277843, 3266105, 3281876
- \* *Laser typeset by :* T.P.  
Printed at:- New A.S. Offset Press Laxmi Nagar Delhi-92

DEDICATED TO  
OUR TEACHER  
**PROFESSOR H.C. GUPTA**  
WHO INITIATED  
THE TEACHING OF  
MATHEMATICAL STATISTICS  
AT THE  
*UNIVERSITY OF DELHI*



# PREFACE

## TO THE TENTH EDITION

The book has been revised keeping in mind the comments and suggestions received from the readers. An attempt is made to eliminate the misprints/errors in the last edition. Further suggestions and criticism for the improvement of the book will be most welcome and thankfully acknowledged.

August 2000

S.C. GUPTA  
V.K. KAPOOR

## TO THE NINTH EDITION

The book originally written twenty-four years ago has, during the intervening period, been revised and reprinted several times. The authors have, however, been thinking, for the last few years that the book needed not only a thorough revision but rather a complete rewriting. They now take great pleasure in presenting to the readers the ninth completely revised and enlarged edition of the book. The subject-matter in the whole book has been rewritten in the light of numerous criticisms and suggestions received from the users of the previous editions in India and abroad.

### Some salient features of the new edition are:

- The entire text, especially Chapter 5 (Random Variables), Chapter 6 (Mathematical Expectation), Chapters 7 and 8 (Theoretical Discrete and Continuous Distributions), Chapter 10 (Correlation and Regression), Chapter 15 (Theory of Estimation), has been restructured, rewritten and updated to cater to the revised syllabi of Indian universities, Indian Civil Services and various other competitive examinations.
- During the course of rewriting, it has been specially borne in mind to retain all the basic features of the previous editions especially the simplicity of presentation, lucidity of style and analytical approach which have been appreciated by teachers and students all over India and abroad.
- A number of typical problems have been added as solved examples in each chapter. These will enable the reader to have a better and thoughtful understanding of the basic concepts of the theory and its various applications.
- Several new topics have been added at appropriate places to make the treatment more comprehensive and complete. Some of the obvious ADDITIONS are:

§ 8.1.5 Triangular Distribution p. 8.10 to 8.12

§ 8.8.3 Logistic Distribution p. 8.92 to 8.95

§ 8.10 Remarks 2, Convergence in Distribution of Law p. 8.106

§ 8.10.3. Remark 3, Relation between Central Limit Theorem and Weak Law of Large Numbers p. 8.110

§ 8.10.4 Cramer's Theorem p. 8.111-8.112, 8.114-8.115 — Example 8.46

§ 8.14 to	<i>Order Statistics — Theory, Illustrations and</i>
§ 8.14.6	<i>Exercise Set p. 8.136 to 8.151</i>
§ 8.15	<i>Truncated Distributions—with Illustrations</i> p. 8.151 to 8.156
§ 10.6.1	<i>Derivation of Rank Correlation Formula for Tied Ranks</i> p. 10.40-10.41
§ 10.7.1	<i>Lines of Regression—Derivation (Aliter)</i> p. 10.50-10.51. Example 10.21 p. 10.55
§ 10.10.2	<i>Remark to § 10.10.2 — Marginal Distributions of Bivariate Normal Distribution p. 10.88-10.90</i> <i>Theorem 10.5, p. 10.86, and Theorem 10.6, p. 10.87 on Bivariate Normal Distribution.</i> <i>Solved Examples 10.31, 10.32, pages 10.96-10.97 on BVN Distribution.</i> <i>Theorem 13.5 Alternative Proof of Distribution of <math>(X, s^2)</math> using m.g.f. p.13.19 to 13.21</i>
§ 13.11	$\chi^2$ -Test for pooling of Probabilities ( $P_\lambda$ Test) p. 13.69
§ 15.4.1	Invariance property of Consistent Estimators—Theorem 15.1, pp 15.3
§ 15.4.2	Sufficient Conditions for Consistency—Theorem 15.2, p. 15.3
§ 15.5.5	MVUE : Theorem 15.4, p. 15.12-15.13
§ 15.7	Remark 1. Minimum Variance Bound (MVB), Estimator, p.15.24
§ 15.7.1	Conditions for the equality sign in Cramer-Rao (CR) inequality, p. 15.25 to 15.27
§ 15.8	Complete family of Distributions (with illustrations), p. 15.31 to 15.34 Theorem 15.10 (Blackwellisation), p. 15.36. Theorems 15.16 and 15.17 on MLE, p. 15.55.
§ 16.5.1	Unbiased Test and Unbiased Critical Region. Theorem 16.2-pages 16.9-16.10
§ 16.5.2	Optimum Regions and Sufficient Statistics, p.16.10-16.11 Remark to Example 16.6, p. 16.17-16.18 and Remarks 1, 2 to Example 16.7, p. 16.20 to 16.22; Graphical Representation of Critical Regions.

- Exercise sets at the end of each chapter are substantially reorganised. Many new problems are included in the exercise sets. Repetition of questions of the same type (more than what is necessary) has been avoided. Further in the set of exercises, the problems have been carefully arranged and properly graded. More difficult problems are put in the miscellaneous exercise at the end of each chapter.

- Solved examples and unsolved problems in the exercise sets have been drawn from the latest examination papers of various Indian Universities, Indian Civil Services, etc.

- An attempt has been made to rectify the errors in the previous editions.

- The present edition incorporates modern viewpoints. In fact with the addition of new topics, rewriting and revision of many others and restructuring of exercise sets, altogether a new book, covering the revised syllabi of almost all the Indian universities, is being presented to the reader. It is earnestly hoped that, in the new form, the book will prove of much greater utility to the students as well as teachers of the subject.

We express our deep sense of gratitude to our Publishers M/s Sultan Chand & Sons and printers DRO Phototypesetter for their untiring efforts, unfailing courtesy, and co-operation in bringing out the book, in such an elegant form. We are also thankful to our several colleagues, friends and students for their suggestions and encouragement during the preparing of this revised edition.

Suggestions and criticism for further improvement of the book as well as intimation of errors and misprints will be most gratefully received and duly acknowledged.

S C. GUPTA & V.K. KAPOOR

#### **TO THE FIRST EDITION**

Although there are a large number of books available covering various aspects in the field of Mathematical Statistics, there is no comprehensive book dealing with the various topics on Mathematical Statistics for the students. The present book is a modest though determined bid to meet the requirements of the students of Mathematical Statistics at Degree, Honours and Post-graduate levels. The book will also be found of use by the students preparing for various competitive examinations. While writing this book our goal has been to present a clear, interesting, systematic and thoroughly teachable treatment of Mathematical Statistics and to provide a textbook which should not only serve as an introduction to the study of Mathematical Statistics but also carry the student on to such a level that he can read with profit the numerous special monographs which are available on the subject. In any branch of Mathematics, it is certainly the teacher who holds the key to successful learning. Our aim in writing this book has been simply to assist the teacher in conveying to the students more effectively a thorough understanding of Mathematical Statistics.

The book contains sixteen chapters (equally divided between two volumes). The first chapter is devoted to a concise and logical development of the subject. The second and third chapters deal with the frequency distributions, and measures of average and dispersion. Mathematical treatment has been given to the proofs of various articles included in these chapters in a very logical and simple manner. The theory of probability which has been developed by the application of the set theory has been discussed quite in detail. A large number of theorems have been deduced using the simple tools of set theory. The

*simple applications of probability are also given. The chapters on mathematical expectation and theoretical distributions (discrete as well as continuous) have been written keeping the latest ideas in mind. A new treatment has been given to the chapters on correlation, regression and bivariate normal distribution using the concepts of mathematical expectation. The thirteenth and fourteenth chapters deal mainly with the various sampling distributions and the various tests of significance which can be derived from them. In chapter 15, we have discussed concisely statistical inference (estimation and testing of hypothesis). Abundant material is given in the chapter on finite differences and numerical integration. The whole of the relevant theory is arranged in the form of serialised articles which are concise and to the point without being insufficient. The more difficult sections will, in general, be found towards the end of each chapter. We have tried our best to present the subject so as to be within the easy grasp of students with varying degrees of intellectual attainment.*

*Due care has been taken of the examination needs of the students and, wherever possible, indication of the year, when the articles and problems were set in the examination as been given. While writing this text, we have gone through the syllabi and examination papers of almost all Indian universities where the subject is taught so as to make it as comprehensive as possible. Each chapter contains a large number of carefully graded worked problems mostly drawn from university papers with a view to acquainting the student with the typical questions pertaining to each topic. Furthermore, to assist the student to gain proficiency in the subject, a large number of properly graded problems mainly drawn from examination papers of various universities are given at the end of each chapter. The questions and problems given at the end of each chapter usually require for their solution a thoughtful use of concepts. During the preparation of the text we have gone through a vast body of literature available on the subject, a list of which is given at the end of the book. It is expected that the bibliography given at the end of the book will considerably help those who want to make a detailed study of the subject*

*The lucidity of style and simplicity of expression have been our twin objects to remove the awe which is usually associated with most mathematical and statistical textbooks.*

*While every effort has been made to avoid printing and other mistakes, we crave for the indulgence of the readers for the errors that might have inadvertently crept in. We shall consider our efforts amply rewarded if those for whom the book is intended are benefited by it. Suggestions for the improvement of the book will be highly appreciated and will be duly incorporated.*

# **Contents**

<i>Art</i>	<i>Pages</i>
<i>Chapter 1</i>	
<b>Introduction — Meaning and Scope</b>	<b>1·1 — 1·8</b>
1·1 Origin and Development of Statistics 1·1	
1·2 Definition of Statistics 1·2	
1·3 Importance and Scope of Statistics 1·4	
1·4 Limitations of Statistics 1·5	
1·5 Distrust of Statistics 1·6	
<i>Chapter 2</i>	
<b>Frequency Distributions and Measures of Central Tendency</b>	<b>2·1 — 2·44</b>
2·1 Frequency Distributions 2·1	
2·1·1 Continuous Frequency Distribution 2·4	
2·2 Graphic Representation of a Frequency Distribution 2·4	
2·2·1 Histogram 2·4	
2·2·2 Frequency Polygon 2·5	
2·3 Averages or Measures of Central Tendency or Measures of Location 2·6	
2·4 Requisites for an Ideal Measure of Central Tendency 2·6	
2·5 Arithmetic Mean 2·6	
2·5·1 Properties of Arithmetic Mean 2·8	
2·5·2 Merits and Demerits of Arithmetic Mean 2·10	
2·5·3 Weighted Mean 2·11	
2·6 Median 2·13	
2·6·1 Derivation of Median Formula 2·19	
2·6·2 Merits and Demerits of Median 2·16	
2·7 Mode 2·17	
2·7·1 Derivation of Mode Formula 2·19	
2·7·2 Merits and Demerits of Mode 2·22	
2·8 Geometric Mean 2·22	
2·8·1 Merits and Demerits of Geometric Mean 2·23	
2·9 Harmonic Mean 2·25	
2·9·1 Merits and Demerits of Harmonic Mean 2·25	
2·10 Selection of an Average 2·26	
2·11 Partition Values 2·26	
2·11·1 Graphical Location of the Partition Values 2·27	

**Chapter — 3****Measures of Dispersion, Skewness and Kurtosis 3·1 — 3·40**

3·1	Dispersion	3·1
3·2	Characteristics for an Ideal Measure of Dispersion	3·1
3·3	Measures of Dispersion	3·1
3·4	Range	3·1
3·5	Quartile Deviation	3·1
3·6	Mean Deviation	3·2
3·7	Standard Deviation ( $\sigma$ ) and Root Mean Square Deviation (s)	3·2
3·7·1	Relation between $\sigma$ and s	3·3
3·7·2	Different Formulae for Calculating Variance	3·3
3·7·3	Variance of the Combined Series	3·10
3·8	Coefficient of Dispersion	3·12
3·8·1	Coefficient of Variation	3·12
3·9	Moments	3·21
3·9·1	Relation Between Moments About Mean in Terms of Moments About Any Point and Vice Versa	3·22
3·9·2	Effect of Change of Origin and Scale on Moments	3·23
3·9·3	Sheppard's Correction for Moments	3·23
3·9·4	Charlier's Checks	3·24
3·10	Pearson's $\beta$ and $\gamma$ Coefficients	3·24
3·11	Factorial Moments	3·24
3·12	Absolute Moments	3·25
3·13	Skewness	3·32
3·14	Kurtosis	3·35

**Chapter — 4****Theory of Probability**

4·1 — 4·116

4·1	Introduction	4·1
4·2	Short History	4·1
4·3	Definitions of Various Terms	4·2
4·3·1	Mathematical or Classical Probability	4·3
4·3·2	Statistical or Empirical Probability	4·4
4·4	Mathematical Tools : Preliminary Notions of Sets	4·14
4·4·1	Sets and Elements of Sets	4·14
4·4·2	Operations on Sets	4·15
4·4·3	Algebra of Sets	4·15
4·4·4	Limit of Sequence of Sets	4·16
4·4·5	Classes of Sets	4·17
4·5	Axiomatic Approach to Probability	4·17

4.5.1	Random Experiment (Sample space)	4.18
4.5.2	Event	4.19
4.5.3	Some Illustrations	4.19
4.5.4	Algebra of Events	4.21
4.6	Probability - Mathematical Notion	4.25
4.6.1	Probability Function	4.25
4.6.2	Law of Addition of Probabilities	4.30
4.6.3	Extension of General Law of Addition of Probabilities	4.31
4.7	Multiplication Law of Probability and Conditional Probability	4.35
4.7.1	Extension of Multiplication Law of Probability	4.36
4.7.2	Probability of Occurrence of At Least One of the $n$ Independent Events	4.37
4.7.3	Independent Events	4.39
4.7.4	Pairwise Independent Events	4.39
4.7.5	Conditions for Mutual Independence of $n$ Events	4.40
4.8	Bayes Theorem	4.69
4.9	Geometric Probability	4.80

**Chapter 5****Random Variables and Distribution Functions** 5.1 — 5.82

5.1	Random Variable	5.1
5.2	Distribution Function	5.4
5.2.1	Properties of Distribution Function	5.6
5.3	Discrete Random Variable	5.6
5.3.1	Probability Mass Function	5.6
5.3.2	Discrete Distribution Function	5.7
5.4	Continuous Random Variable	5.13
5.4.1	Probability Density Function	5.13
5.4.2	Various Measures of Central Tendency, Dispersion, Skewness and Kurtosis for Continuous Distribution	5.15
5.4.3	Continuous Distribution Function	5.32
5.5	Joint Probability Law	5.41
5.5.1	Joint Probability Mass Function	5.41
5.5.2	Joint Probability Distribution Function	5.42
5.5.3	Marginal Distribution Function	5.43
5.5.4	Joint Density Function	5.44
5.5.5	The Conditional Distribution Function	5.46
5.5.6	Stochastic Independence	5.47
5.6	Transformation of One-dimensional Random Variable	5.70
5.7	Transformation of Two-dimensional Random Variable	5.73

**Chapter — 6****Mathematical Expectation and Generating Functions**

6·1 — 6·138

6·1	Mathematical Expectation	6·1
6·2	Expectation of a Function of a Random Variable	6·3
6·3	Addition Theorem of Expectation	6·4
6·4	Multiplication Theorem of Expectation	6·6
6·5	Expectation of a Linear Combination of Random Variables	6·8
6·6	Covariance	6·11
6·7	Variance of a Linear Combination of Random Variables	6·11
6·8	Moments of Bivariate Probability Distributions	6·54
6·9	Conditional Expectation and Conditional Variance	6·54
6·10	Moment Generating Function	6·67
6·10·1	Some Limitations of Moment Generating Functions	6·68
6·10·2	Theorems on Moment Generating Functions	6·71
6·10·3	Uniqueness Theorem of Moment Generating Function	6·72
6·11	Cumulants	6·72
6·11·1	Additive Property of Cumulants	6·73
6·11·2	Effect of Change of Origin and Scale on Cumulants	6·73
6·12	Characteristic Function	6·77
6·12·1	Properties of Characteristic Function	6·78
6·12·2	Theorems on Characteristic Functions	6·79
6·12·3	Necessary and Sufficient Conditions for a Function $\phi(t)$ to be a Characteristic Function	6·83
6·12·4	Multivariate Characteristic Function	6·84
6·13	Chebychev's Inequality	6·97
6·13·1	Generalised Form of Bienayme-Chebychev Inequality	6·98
6·14	Convergence in Probability	6·100
6·15	Weak Law of Large Numbers	6·101
6·15·1	Bernoulli's Law of Large Numbers	6·103
6·15·2	Markoff's Theorem	6·104
6·15·3	Khintchin's Theorem	6·104
6·16	Borel Cantelli Lemma	6·115
6·17	Probability Generating Function	6·123
6·17·1	Convolutions	6·126

7·0	Introduction	7·1
7·1	Bernoulli Distribution	7·1
7·1·1	Moments of Bernoulli Distribution	7·1
7·2	Binomial Distribution	7·1
7·2·1	Moments	7·6
7·2·2	Recurrence Relation for the Moments of Binomial Distribution	7·9
7·2·3	Factorial Moments of Binomial Distribution	7·11
7·2·4	Mean Deviation about Mean of Binomial Distribution	7·11
7·2·5	Mode of Binomial Distribution	7·12
7·2·6	Moment Generating Function of Binomial Distribution	7·14
7·2·7	Additive Property of Binomial Distribution	7·15
7·2·8	Characteristic Function of Binomial Distribution	7·16
7·2·9	Cumulants of Binomial Distribution	7·16
7·2·10	Recurrence Relation for Cumulants of Binomial Distribution	7·17
7·2·11	Probability Generating Function of Binomial Distribution	7·18
7·2·12	Fitting of Binomial Distribution	7·19
7·3	Poisson Distribution	7·40
7·3·1	The Poisson Process	7·42
7·3·2	Moments of Poisson Distribution	7·44
7·3·3	Mode of Poisson Distribution	7·45
7·3·4	Recurrence Relation for the Moments of Poisson Distribution	7·46
7·3·5	Moment Generating Function of Poisson Distribution	7·47
7·3·6	Characteristic Function of Poisson Distribution	7·47
7·3·7	Cumulants of Poisson Distribution	7·47
7·3·8	Additive or Reproductive Property of Independent Poisson Variates	7·47
7·3·9	Probability Generating Function of Poisson Distribution	7·49
7·3·10	Fitting of Poisson Distribution	7·61
7·4	Negative Binomial Distribution	7·72
7·4·1	Moment Generating Function of Negative Binomial Distribution	7·74
7·4·2	Cumulants of Negative Binomial Distribution	7·74
7·4·3	Poisson Distribution as Limiting Case of Negative Binomial Distribution	7·75

<b>7·4·4</b>	<b>Probability Generating Function of Negative Binomial Distribution</b>	<b>7·76</b>
<b>7·4·5</b>	<b>Deduction of Moments of Negative Binomial Distribution From Binomial Distribution</b>	<b>7·79</b>
<b>7·5</b>	<b>Geometric Distribution</b>	<b>7·83</b>
<b>7·5·1</b>	<b>Lack of Memory</b>	<b>7·84</b>
<b>7·5·2</b>	<b>Moments of Geometric Distribution</b>	<b>7·84</b>
<b>7·5·3</b>	<b>Moment Generating Function of Geometric Distribution</b>	<b>7·85</b>
<b>7·6</b>	<b>Hypergeometric Distribution</b>	<b>7·88</b>
<b>7·6·1</b>	<b>Mean and Variance of Hypergeometric Distribution</b>	<b>7·89</b>
<b>7·6·2</b>	<b>Factorial Moments of Hypergeometric Distribution</b>	<b>7·90</b>
<b>7·6·3</b>	<b>Approximation to the Binomial Distribution</b>	<b>7·91</b>
<b>7·6·4</b>	<b>Recurrence Relation for Hypergeometric Distribution</b>	<b>7·91</b>
<b>7·7</b>	<b>Multinomial Distribution</b>	<b>7·95</b>
<b>7·7·1</b>	<b>Moments of Multinomial Distribution</b>	<b>7·96</b>
<b>7·8</b>	<b>Discrete Uniform Distribution</b>	<b>7·101</b>
<b>7·9</b>	<b>Power Series Distribution</b>	<b>7·101</b>
<b>7·9·1</b>	<b>Moment Generating Function of p-s-d</b>	<b>7·102</b>
<b>7·9·2</b>	<b>Recurrence Relation for Cumulants of p-s-d</b>	<b>7·102</b>
<b>7·9·3</b>	<b>Particular Cases of q-p-s-d</b>	<b>7·103</b>

**Chapter — 8**

## Theoretical Continuous Distributions

$$8.1 = 8.166$$

8·1	Rectangular or Uniform Distribution	8·1
8·1·1	Moments of Rectangular Distribution	8·2
8·1·2	M·G·F· of Rectangular Distribution	8·2
8·1·3	Characteristic Function	8·2
8·1·4	Mean Deviation about Mean	8·2
8·1·5	Triangular Distribution	8·10
8·2	Normal Distribution	8·17
8·2·1	Normal Distribution as a Limiting form of Binomial Distribution	8·18
8·2·2	Chief Characteristics of the Normal Distribution and Normal Probability Curve	8·20
8·2·3	Mode of Normal distribution	8·22
8·2·4	Median of Normal Distribution	8·23
8·2·5	M·G·F· of Normal Distribution	8·23
8·2·6	Cumulant Generating Function ( <i>c·g·f</i> ) of Normal Distribution	8·24
8·2·7	Moments of Normal Distribution	8·24

8·2·8	A Linear Combination of Independent Normal Variates is also a Normal Variate	8·26
8·2·9	Points of Inflection of Normal Curve	8·28
8·2·10	Mean Deviation from the Mean for Normal Distribution	8·28
8·2·11	Area Property : Normal Probability Integral	8·29
8·2·12	Error Function	8·30
8·2·13	Importance of Normal Distribution	8·31
8·2·14	Fitting of Normal Distribution	8·32
8·2·15	Log-Normal Distribution	8·35
8·3	Gamma Distribution	8·68
8·3·1	M·G·F of Gamma Distribution	8·68
8·3·2	Cumulant Generating Function of Gamma Distribution	8·68
8·3·3	Additive Property of Gamma Distribution	8·70
8·4	Beta Distribution of First Kind	8·70
8·4·1	Constants of Beta Distribution of First Kind	8·71
8·5	Beta Distribution of Second Kind	8·72
8·5·1	Constants of Beta Distribution of Second Kind	8·72
8·6	The Exponential Distribution	8·85
8·6·1	M·G·F of Exponential Distribution	8·86
8·7	Laplace Double Exponential Distribution	8·89
8·8	Weibul Distribution	8·90
8·8·1	Moments of Standard Weibul Distribution	8·91
8·8·2	Characterisation of Weibul Distribution	8·91
8·8·3	Logistic Distribution	8·92
8·9	Cauchy Distribution	8·98
8·9·1	Characteristic Function of Cauchy Distribution	8·99
8·9·2	Moments of Cauchy Distribution	8·100
8·10	Central Limit Theorem	8·105
8·10·1	Lindeberg-Levy Theorem	8·107
8·10·2	Applications of Central Limit Theorem	8·108
8·10·3	Liapounoff's Central Limit Theorem	8·109
8·10·4	Cramer's Theorem	8·111
8·11	Compound Distributions	8·116
8·11·1	Compound Binomial Distribution	8·116
8·11·2	Compound Poisson distribution	8·117
8·12	Pearson's Distributions	8·120
8·12·1	Determination of the Constants of the Equation in Terms of Moments	8·121
8·12·2	Pearson Measure of Skewness	8·121
8·12·3	Criterion 'k'	8·121
8·12·4	Pearson's Main Type I	8·122
8·12·5	Pearson Type IV	8·124
8·12·6	Pearson Type VI	8·125

8·12·7	Type III	8·125
8·12·8	Type V	8·126
8·12·9	Type II	8·126'
8·12·10	Type VIII	8·127
8·12·11	Zero Type (Normal Curve)	8·127
8·12·12	Type VIII to XII	8·127
8·13	Variate Transformations	8·132
8·13·1	Uses of Variate Transformations	8·132
8·13·2	Square Root Transformation	8·133
8·13·3	Sine Inverse or $\sin^{-1}$ Transformation	8·133
8·13·4	Logarithmic Transformation	8·134
8·13·5	Fisher's Z-Transformation	8·135
8·14	Order Statistics	8·136
8·14·1	Cumulative Distribution Function of a Single Order Statistic	8·136
8·14·2	Probability Density Function ( <i>p.d.f.</i> ) of a Single Order Statistic	8·137
8·14·3	Joint <i>p.d.f.</i> of two Order Statistics	8·138
8·14·4	Joint <i>p.d.f.</i> of $k$ -Order Statistics	8·139
8·14·5	Joint <i>p.d.f.</i> of all $n$ -Order Statistics	8·140
8·14·6	Distribution of Range and Other Systematic Statistics	8·140
8·15	Truncated Distributions	8·151

#### *Chapter — 9*

#### **Curve Fitting and Principle of Least Squares**      9·1 — 9·24

9·1	Curve Fitting	9·1
9·1·1	Fitting of a Straight Line	9·1
9·1·2	Fitting of a Second Degree Parabola	9·2
9·1·3	Fitting of a Polynomial of $k$ th Degree	9·3
9·1·4	Change of Origin	9·5
9·2	Most Plausible Solution of a System of Linear Equations	9·8
9·3	Conversion of Data to Linear Form	9·9
9·4	Selection of Type of Curve to be Fitted	9·13
9·5	Curve Fitting by Orthogonal Polynomials	9·15
9·5·1	Orthogonal Polynomials	9·17
9·5·2	Fitting of Orthogonal Polynomials	9·18
9·5·3	Finding the Orthogonal Polynomial $P_p$	9·18
9·5·4	Determination of Coefficients	9·21

#### *Chapter — 10*

#### **Correlation and Regression**      10·1 — 10·128

10·1	Bivariate Distribution, Correlation	10·1
10·2	Scatter Diagram	10·1

10·3	Karl Pearson Coefficient of Correlation	10·1
10·3·1	Limits for Correlation Coefficient	10·2
10·3·2	Assumptions Underlying Karl Pearson's Correlation Coefficient	10·5
10·4	Calculation of the Correlation Coefficient for a Bivariate Frequency Distribution	10·32
10·5	Probable Error of Correlation Coefficient	10·38
10·6	Rank Correlation	10·39
10·6·1	Tied Ranks	10·40
10·6·2	Repeated Ranks (Continued)	10·43
10·6·3	Limits for Rank Correlation Coefficient	10·44
10·7	Regression	10·49
10·7·1	Lines of Regression	10·49
10·7·2	Regression Curves	10·52
10·7·3	Regression Coefficients	10·58
10·7·4	Properties of Regression Coefficients	10·58
10·7·5	Angle Between Two Lines of Regression	10·59
10·7·6	Standard Error of Estimate	10·60
10·7·7	Correlation Coefficient Between Observed and Estimated Value	10·61
10·8	Correlation Ratio	10·76
10·8·1	Measures of Correlation Ratio	10·76
10·9	Intra-class Correlation	10·81
10·10	Bivariate Normal Distribution	10·84
10·10·1	Moment Generating Function of Bivariate Normal Distribution	10·86
10·10·2	Marginal Distributions of Bivariate Normal Distribution	10·88
10·10·3	Conditional Distributions	10·90
10·11	Multiple and Partial Correlation	10·103
10·11·1	Yule's Notation	10·104
10·12	Plane of Regression	10·105
10·12·1	Generalisation	10·106
10·13	Properties of Residuals	10·109
10·13·1	Variance of the Residuals	10·110
10·14	Coefficient of Multiple Correlation	10·111
10·14·1	Properties of Multiple Correlation Coefficient	10·113
10·15	Coefficient of Partial Correlation	10·114
10·15·1	Generalisation	10·116
10·16	Multiple Correlation in Terms of Total and Partial Correlations	10·116
10·17	Expression for Regression Coefficient in Terms of Regression Coefficients of Lower Order	10·118
10·18	Expression for Partial Correlation Coefficient in Terms of Correlation Coefficients of Lower Order	10·118

**Chapter—11**  
**Theory of Attributes**

11·1 — 11·22

11·1	Introduction	11·1
11·2	Notations	11·1
11·3	Dichotomy	11·1
11·4	Classes and Class Frequencies	11·1
11·4·1	Order of Classes and Class Frequencies	11·1
11·4·2	Relation between Class Frequencies	11·2
11·5	Class Symbols as Operators	11·3
11·6	Consistency of Data	11·8
11·6·1	Conditions for Consistency of Data	11·8
11·7	Independence of Attributes	11·12
11·7·1	Criterion of Independence	11·12
11·7·2	Symbols $(AB)_o$ and $\delta$	11·14
11·8	Association of Attributes	11·15
11·8·1	Yule's Coefficient of Association	11·16
11·8·2	Coefficient of Colligation	11·16

**Chapter—12**

**Sampling and Large Sample Tests**

12·1 — 12·50

12·1	Sampling - Introduction	12·1
12·2	Types of Sampling	12·1
12·2·1	Purposive Sampling	12·2
12·2·2	Random Sampling	12·2
12·2·3	Simple Sampling	12·2
12·2·4	Stratified Sampling	12·3
12·3	Parameter and Statistic	12·3
12·3·1	Sampling Distribution	12·3
12·3·2	Standard Error	12·4
12·4	Tests of Significance	12·6
12·5	Null Hypothesis	12·6
12·5·1	Alternative Hypothesis	12·6
12·6	Errors in Sampling	12·7
12·7	Critical Region and Level of Significance	12·7
12·7·1	One Tailed and Two Tailed Tests	12·7
12·7·2	Critical or Significant Values	12·8
12·7·3	Procedure for Testing of Hypothesis	12·10
12·8	Test of Significance for Large Samples	12·10
12·9	Sampling of Attributes	12·11
12·9·1	Test for Single Proportion	12·12

12·9·2	Test of Significance for Difference of Proportions	12·15
12·10	Sampling of Variables	12·28
12·11	Unbiased Estimates for Population Mean ( $\mu$ ) and Variance ( $\sigma^2$ )	12·29
12·12	Standard Error of Sample Mean	12·31
12·13	Test of Significance for Single Mean	12·31
12·14	Test of Significance for Difference of Means	12·37
12·15	Test of Significance for Difference of Standard Deviations	12·42

*Chapter — 13***Exact Sampling Distributions (Chi-Square Distribution)**

13·1 — 13·72

13·1	Chi-square Variate	13·1
13·2	Derivation of the Chi-square Distribution: First Method – Method of M·G·F.	13·1
	Second Method – Method of Induction	13·2
13·3	M·G·F. of $\chi^2$ Distribution	13·5
13·3·1	Cumulant Generating Function of $\chi^2$ Distribution	13·5
13·3·2	Limiting Form of $\chi^2$ Distribution	13·6
13·3·3	Characteristic Function of $\chi^2$ Distribution	13·7
13·3·4	Mode and Skewness of $\chi^2$ Distribution	13·7
13·3·5	Additive Property of Chi-square Variates	13·7
13·4	Chi-Square Probability Curve	13·9
13·5	Conditions for the Validity of $\chi^2$ test	13·15
13·6	Linear Transformation	13·16
13·7	Applications of Chi-Square Distribution	13·37
13·7·1	Chi-square Test for Population Variance	13·38
13·7·2	Chi-square Test of Goodness of Fit	13·39
13·7·3	Independence of Attributes	13·49
13·8	Yates Correction	13·57
13·9	Brandt and Snedecor Formula for $2 \times k$ Contingency Table	13·57
13·9·1	Chi-square Test of Homogeneity of Correlation Coefficients	13·66
13·10	Bartlett's Test for Homogeneity of Several Independent Estimates of the same Population Variance	13·68
13·11	$\chi^2$ Test for Pooling the Probabilities (P $_{\lambda}$ Test.)	13·69
13·12	Non-central $\chi^2$ Distribution	13·69
13·12·1	Non-central $\chi^2$ Distribution with Non-Centrality Parameter $\lambda$	13·70
13·12·2	Moment Generating Function of Non-central $\chi^2$ Distribution	13·70

13·12·3	Additive Property of Non-central Chi-square Distribution	13·72
13·12·4	Cumulants of Non-central Chi-square Distribution	13·72

***Chapter — 14*****Exact Sampling Distributions (Continued)**

(*t*, *F* and *z* distributions) 14·1 — 14·74

14·1	Introduction	14·1
14·2	Student's "t"	14·1
14·2·1	Derivation of Student's <i>t</i> -distribution	14·2
14·2·2	Fisher's "t"	14·3
14·2·3	Distribution of Fisher's "t"	14·4
14·2·4	Constants of <i>t</i> -distribution	14·5
14·2·5	Limiting form of <i>t</i> -distribution	14·14
14·2·6	Graph of <i>t</i> -distribution	14·15
14·2·7	Critical Values of 't'	14·15
14·2·8	Applications of <i>t</i> -distribution	14·16
14·2·9	<i>t</i> -Test for Single Mean	14·16
14·2·10	<i>t</i> -test for Difference of Means	14·24
14·2·11	<i>t</i> -test for Testing Significance of an Observed Sample Correlation Coefficient	14·37
14·2·12	<i>t</i> -test for Testing Significance of an Observed Regression Coefficient	14·39
14·2·13	<i>t</i> -test for Testing Significance of an Observed Partial Correlation Coefficient	14·39
14·3	Distribution of Sample Correlation Coefficient when Population Correlation Coefficient $\rho = 0$	14·39
14·4	Non-central <i>t</i> -distribution	14·43
14·5	<i>F</i> -statistic (Definition)	14·44
14·5·1	Derivation of Snedecor's <i>F</i> -Distribution	14·45
14·5·2	Constants of <i>F</i> -distribution	14·46
14·5·3	Mode and Points of Inflexion of <i>F</i> -distribution	14·48
14·5·4	Applications of <i>F</i> -distribution	14·57
14·5·5	<i>F</i> -test for Equality of Population Variances	14·57
14·5·6	Relation Between <i>t</i> and <i>F</i> -distributions	14·64
14·5·7	Relation Between <i>F</i> and $\chi^2$	14·65
14·5·8	<i>F</i> -test for Testing the Significance of an Observed Multiple Correlation Coefficient	14·66
14·5·9	<i>F</i> -test for Testing the Significance of an Observed Sample Correlation Ratio	14·66
14·5·10	<i>F</i> -test for Testing the Linearity of Regression	14·66

14·5·11	F-test for Equality of Several Means	14·67
14·6	Non-Central 'F'-Distribution	14·67
14·7	Fisher's Z-Distribution	14·69
14·7·1	M·G·F· of Z-Distribution	14·70
14·8	Fisher's Z-Transformation	14·71

***Chapter — 15*****Statistical Inference — I**(Theory of Estimation) 15·1 — 15·92

15·1	Introduction	15·1
15·2	Characteristics of Estimators	15·1
15·3	Consistency	15·2
15·4	Unbiasedness	15·2
15·4·1	Invariance Property of Consistent Estimators	15·3
15·4·2	Sufficient Conditions for Consistency	15·3
15·5	Efficient Estimators	15·7
15·5·1	Most Efficient Estimator	15·8
15·6	Sufficiency	15·18
15·7	Cramer-Rao Inequality	15·22
15·7·1	Conditions for the Equality Sign in Cramer-Rao (C·R·) Inequality	15·25
15·8	Complete Family of Distributions	15·31
15·9	MVUE and Blackwellisation	15·34
15·10	Methods of Estimation	15·52
15·11	Method of Maximum Likelihood Estimation	15·52
15·12	Method of Minimum Variance	15·69
15·13	Method of Moments	15·69
15·14	Method of Least Squares	15·73
15·15	Confidence Intervals and Confidence Limits	15·82
15·15·1	Confidence Intervals for Large Samples	15·87

***Chapter — 16*****Statistical Inference — II**Testing of Hypothesis, Non-parametric Methods  
and Sequential Analysis 16·1 — 16·80

16·1	Introduction	16·1
16·2	Statistical Hypothesis (Simple and Composite)	16·1
16·2·1	Test of a Statistical Hypothesis	16·2
16·2·2	Null Hypothesis	16·2
16·2·3	Alternative Hypothesis	16·2
16·2·4	Critical Region	16·3

16·2·5	Two Types of Errors	16·4
16·2·6	Level of Significance	16·5
16·2·7	Power of the Test	16·5,
16·3	Steps in Solving Testing of Hypothesis Problem	16·6
16·4	Optimum Tests Under Different Situations	16·6
16·4·1	Most Powerful Test (MP Test.)	16·6
16·4·2	Uniformly Most Powerful Test	16·7
16·5	Neyman-Pearson Lemma	16·7
16·5·1	Unbiased Test and Unbiased Critical Region	16·9
16·5·2	Optimum Regions and Sufficient Statistics	16·10
16·6	Likelihood Ratio Test	16·34
16·6·1	Properties of Likelihood Ratio Test	16·37
16·7·1	Test for the Mean of a Normal Population	16·37
16·7·2	Test for the Equality of Means of Two Normal Populations	16·42
16·7·3	Test for the Equality of Means of Several Normal Populations	16·47
16·7·4	Test for the Variance of a Normal Population	16·50
16·7·5	Test for Equality of Variances of two Normal Populations	16·53
16·7·6	Test for the Equality of Variances of several Normal Populations	16·55
16·8	Non-parametric Methods	16·59
16·8·1	Advantages and Disadvantages of N·P· Methods over Parametric Methods	16·59
16·8·2	Basic Distribution	16·60
16·8·3	Wald-Wolfowitz Run Test	16·61
16·8·4	Test for Randomness	16·63
16·8·5	Median Test	16·64
16·8·6	Sign Test	16·65
16·8·7	Mann-Whitney-Wilcoxon U-test	16·66
16·9	Sequential Analysis	16·69
16·9·1	Sequential Probability Ratio Test (SPRT)	16·69
16·9·2	Operating Characteristic (O.C.) Function of S.P.R.T	16·71
16·9·3	Average Sample Number (A.S.N.)	16·71

**APPENDIX****Numerical Tables (I to VIII)**

1·1 — 1·11

**Index**

1—5

# Fundamentals of Mathematical Statistics

S.C. GUPTA      V.K. KAPOOR

Hindu College,      Shri Ram College of Commerce  
University of Delhi, Delhi      University of Delhi, Delhi

Tenth Edition 2002      Pages xx + 1284      22 x 14 cm      ISBN 81-7014-791-3      Rs 210.00

## Special Features

- Comprehensive and analytical treatment is given of all the topics.
- Difficult mathematical deductions have been treated logically and in a very simple manner.
- It conforms to the latest syllabi of the Degree and Post graduate examinations in Mathematics, Statistics and Economics.

## Contents

Introduction • Frequency Distribution and Measures of Central Tendency • Measures of Dispersion, Skewness and Kurtosis • Theory of Probability • Random Variables—Distribution Function • Mathematical Expectation, Generating Functions and Law of Large Numbers • Theoretical Discrete Distributions • Theoretical Continuous Distributions • Curve Fitting and Principle of Least Squares • Correlation, Regression, Bivariate Normal Distribution and Partial & Multiple Correlation • Theory of Attributes • Sampling and Large Sample Tests of Mean and Proportion • Sampling Distribution Exact (Chi-square Distribution) • Exact Sampling Distributions (I, F and Z Distributions) • Theory of Estimation • Testing of Hypothesis, Sequential Analysis and Non-parametric Methods.

# Elements of Mathematical Statistics

S.C. GUPTA      V.K. KAPOOR

Third Edition 2001      Pages xiv + 489      ISBN 81-7014-291-1      Rs 70.00

- Prepared specially for B.Sc. students, studying Statistics as subsidiary or ancillary subject.

## Contents

Introduction—Meaning and Scope • Frequency Distributions and Measures of Central Tendency • Measures of Dispersion, Skewness and Kurtosis • Theory of Probability • Random Variables—Distribution Functions • Mathematical Expectation, Generating Functions and Law of Large Numbers • Theoretical Discrete Distributions • Theoretical Continuous Distributions • Curve Fitting and Principle of Least Squares • Correlation and Regression • Theory of Attributes • Sampling and Large Sample Tests • Chi-square Distribution • Exact, Sampling Distribution • Theory of Estimation • Testing of Hypothesis • Analysis of Variance • Design of Experiments • Design of Sample Surveys • Tables.

# Fundamentals of Applied Statistics

S.C. GUPTA      V.K. KAPOOR

Third Edition 2001      Pages xvi + 628      ISBN 81-7014-151-6      Rs 110.00

## Special Features

- The book provides comprehensive and exhaustive theoretical discussion.
- All basic concepts have been explained in an easy and understandable manner.
- 125 stimulating problems selected from various university examinations have been solved.
- It conforms to the latest syllabi of B.Sc. (Hons.) and post-graduate examination in Statistics, Agriculture and Economics.

## Contents

Statistical Quality Control • Analysis of Time Series (Mathematical Treatment) • Index Number • Demand Analysis, Price and Income Elasticity • Analysis of Variance  
Design of Experiments, Completely Randomised Design • Randomised Block Design, Latin Square Design • Factorial Designs and Confounding.  
Design of Sample Surveys (Mathematical Treatment) • Sample Random Sampling, Stratified Sampling, Systematic Sampling, Multi-stage Sampling • Educational and Psychological Statistics • Vital Statistical Methods.

# Operations Research

for Managerial Decision-making

V. K. KAPOOR

*Co-author of Fundamentals of Mathematical Statistics*

Sixth Revised Edition 2002

Pages xviii + 904

ISBN 81-7014-130-3

Rs 225.00

This well-organised and profusely illustrated book presents updated account of the Operations Research Techniques.

## Special Features

- It is lucid and practical in approach.
- Wide variety of carefully selected, adapted and specially designed problems with complete solutions and detailed workings.
- 221 Worked examples are expertly woven into the text.
- Useful sets of 740 problems as exercises are given.
- The book completely covers the syllabi of M.B.A., M.M.S. and M.Com. courses of all Indian Universities.

## Contents

Introduction to Operations Research • Linear Programming : Graphic Method • Linear Programming : Simplex Method • Linear Programming : Duality • Transportation Problems • Assignment Problems • Sequencing Problems • Replacement Decisions • Queuing Theory • Decision Theory • Game Theory • Inventory Management • Statistical Quality Control • Investment Analysis • PERT & CPM • Simulation • Work Study Value Analysis • Markov Analysis • Goal, Integer and Dynamic Programming

# Problems and Solutions in Operations Research

V.K. KAPOOR

Fourth Rev. Edition 2004

Pages xii + 835

ISBN 81-7014-605-4

Rs 235.00

## Salient Features

- The book fully meets the course requirements of management and commerce students. It would also be extremely useful for students of professional courses like ICA, ICWA.
- Working rules, aid to memory, short-cuts, alternative methods are special attractions of the book.
- Ideal book for the students involved in independent study.

## Contents

Meaning & Scope • Linear Programming : Graphic Method • Linear Programming : Simplex Method • Linear Programming : Duality • Transportation Problems • Assignment Problems • Replacement Decisions • Queuing Theory • Decision Theory • Inventory Management • Sequencing Problems • Pert & CPM • Cost Consideration in Pert • Game Theory • Statistical Quality Control • Investment Decision Analysis • Simulation.

**Sultan Chand & Sons**

*Providing books — the never failing friends*

23, Daryaganj, New Delhi-110 002

Phones : 3266105, 3277843, 3281876, 3286788; Fax : 011-326-6357

# *Introduction – Meaning and Scope*

**1.1. Origin and Development of Statistics.** Statistics, in a sense, is as old as the human society itself. Its origin can be traced to the old days when it was regarded as the 'science of State-craft' and was the by-product of the administrative activity of the State. The word 'Statistics' seems to have been derived from the Latin word '*status*' or the Italian word '*statista*' or the German word '*statistik*' each of which means a 'political state'. In ancient times, the government used to collect the information regarding the population and 'property or wealth' of the country – the former enabling the government to have an idea of the manpower of the country (to safeguard itself against external aggression, if any), and the latter providing it a basis for introducing new taxes and levies.

In India, an efficient system of collecting official and administrative statistics existed even more than 2,000 years ago, in particular, during the reign of Chandragupta Maurya (324 - 300 B.C.). From Kautilya's *Arthashastra* it is known that even before 300 B.C. a very good system of collecting 'Vital Statistics' and registration of births and deaths was in vogue. During Akbar's reign (1556 - 1605 A.D.), Raja Todarmal, the then land and revenue minister, maintained good records of land and agricultural statistics. In *Aina-e-Akbari* written by Abul Fazl (in 1596 - 97), one of the nine gems of Akbar, we find detailed accounts of the administrative and statistical surveys conducted during Akbar's reign.

In Germany, the systematic collection of official statistics originated towards the end of the 18th century when, in order to have an idea of the relative strength of different German States, information regarding population and output – industrial and agricultural – was collected. In England, statistics were the outcome of Napoleonic Wars. The Wars necessitated the systematic collection of numerical data to enable the government to assess the revenues and expenditure with greater precision and then to levy new taxes in order to meet the cost of war.

Seventeenth century saw the origin of the 'Vital Statistics.' Captain John Grant of London (1620 - 1674), known as the 'father' of Vital Statistics, was the first man to study the statistics of births and deaths. Computation of mortality tables and the calculation of expectation of life at different ages by a number of persons, viz., Casper Newman, Sir William Petty (1623 - 1687), James Dodson, Dr. Price, to mention only a few, led to the idea of 'life insurance' and the first life insurance institution was founded in London in 1698.

The theoretical development of the so-called modern statistics came during the mid-seventeenth century with the introduction of 'Theory of Probability' and 'Theory of Games and Chance', the chief contributors being mathematicians and gamblers of France, Germany and England. The French mathematician Pascal (1623 - 1662), after lengthy correspondence with another French mathematician

P. Fermat (1601 - 1665) solved the famous 'Problem of Points' posed by the gambler Chevalier de Mere. His study of the problem laid the foundation of the theory of probability which is the backbone of the modern theory of statistics. Pascal also investigated the properties of the co-efficients of binomial expansions and also invented mechanical computation machine. Other notable contributors in this field are : James Bernoulli ( 1654 - 1705 ), who wrote the first treatise on the 'Theory of Probability'; De-Moivre (1667 - 1754 ) who also worked on probabilities and annuities and published his important work "The Doctrine of Chances" in 1718, Laplace (1749 - 1827 ) who published in 1782 his monumental work on the theory of probability, and Gauss (1777 - 1855 ), perhaps the most original of all writers on statistical subjects, who gave the principle of least squares and the normal law of errors. Later on, most of the prominent mathematicians of 18th, 19th and 20th centuries, viz., Euler, Lagrange, Bayes, A. Markoff, Khintchin, Kolmogoroff, to mention only a few, added to the contributions in the field of probability.

Modern veterans in the development of the subject are Englishmen. Francis Galton (1822-1921), with his works on 'regression', pioneered the use of statistical methods in the field of Biometry. Karl Pearson (1857-1936), the founder of the greatest statistical laboratory in England (1911), is the pioneer in correlational analysis. His discovery of the 'chi square test', the first and the most important of modern tests of significance, won for Statistics a place as a science. In 1908 the discovery of Student's ' $t$ ' distribution by W.S. Gosset who wrote under the pseudonym of 'Student' ushered in an era of exact sample tests (small samples).

Sir Ronald A. Fisher (1890 - 1962), known as the 'Father of Statistics', placed Statistics on a very sound footing by applying it to various diversified fields, such as genetics; biometry, education, agriculture, etc. Apart from enlarging the existing theory, he is the pioneer in introducing the concepts of 'Point Estimation' (efficiency, sufficiency, principle of maximum likelihood, etc.), 'Fiducial Inference' and 'Exact Sampling Distributions.' He also pioneered the study of 'Analysis of Variance' and 'Design of Experiments.' His contributions won for Statistics a very responsible position among sciences.

**1.2. Definition of Statistics.** Statistics has been defined differently by different authors from time to time. The reasons for a variety of definitions are primarily two. *First*, in modern times the field of utility of Statistics has widened considerably. In ancient times Statistics was confined only to the affairs of State but now it embraces almost every sphere of human activity. Hence a number of old definitions which were confined to a very narrow field of enquiry were replaced by new definitions which are much more comprehensive and exhaustive. *Secondly*, Statistics has been defined in two ways. Some writers define it as '*statistical data*', i.e., numerical statement of facts, while others define it as '*statistical methods*', i.e., complete body of the principles and techniques used in collecting and analysing such data. Some of the important definitions are given below.

### **Statistics as 'Statistical Data'**

Webster defines Statistics as "*classified facts representing the conditions of the people in a State ... especially those facts which can be stated in numbers or in any other tabular or classified arrangement.*" This definition, since it confines Statistics only to the data pertaining to State, is inadequate as the domain of Statistics is much wider.

Bowley defines Statistics as "*numerical statements of facts in any department of enquiry placed in relation to each other.*"

A more exhaustive definition is given by Prof. Horace Secrist as follows :

"*By Statistics we mean aggregates of facts affected to a marked extent by multiplicity of causes numerically expressed, enumerated or estimated according to reasonable standards of accuracy, collected in a systematic manner for a pre-determined purpose and placed in relation to each other.*"

### **Statistics as Statistical Methods**

Bowley himself defines Statistics in the following three different ways :

- (i) Statistics may be called the science of counting.
- (ii) Statistics may rightly be called the science of averages.
- (iii) Statistics is the science of the measurement of social organism, regarded as a whole in all its manifestations.

But none of the above definitions is adequate. The first because statistics is not merely confined to the collection of data as other aspects like presentation, analysis and interpretation, etc., are also covered by it. The second, because averages are only a part of the statistical tools used in the analysis of the data, others being dispersion, skewness, kurtosis, correlation, regression, etc. The third, because it restricts the application of Statistics to sociology alone while in modern days Statistics is used in almost all sciences – social as well as physical.

According to Boddington, "Statistics is the science of estimates and probabilities." This also is an inadequate definition since probabilities and estimates constitute only a part of the statistical methods.

Some other definitions are :

"The science of Statistics is the method of judging collective, natural or social phenomenon from the results obtained from the analysis or enumeration or collection of estimates." – King.

"Statistics is the science which deals with collection, classification and tabulation of numerical facts as the basis for explanation, description and comparison of phenomenon." – Lovitt.

Perhaps the best definition seems to be one given by Croxton and Cowden, according to whom Statistics may be defined as "the science which deals with the collection, analysis and interpretation of numerical data."

**1.3. Importance and Scope of Statistics.** In modern times, Statistics is viewed not as a mere device for collecting numerical data but as a means of developing sound techniques for their handling and analysis and drawing valid inferences from them. As such it is not confined to the affairs of the State but is intruding constantly into various diversified spheres of life – social, economic and political. It is now finding wide applications in almost all sciences – social as well as physical – such as biology, psychology, education, economics, business management, etc. It is hardly possible to enumerate even a single department of human activity where statistics does not creep in. It has rather become indispensable in all phases of human endeavour.

**Statistics and Planning.** Statistics is indispensable to planning. In the modern age which is termed as 'the age of planning', almost all over the world, governments, particularly of the budding economies, are resorting to planning for the economic development. In order that planning is successful, it must be based soundly on the correct analysis of complex statistical data.

**Statistics and Economics.** Statistical data and technique of statistical analysis have proved immensely useful in solving a variety of economic problems, such as wages, prices, analysis of time series and demand analysis. It has also facilitated the development of economic theory. Wide applications of mathematics and statistics in the study of economics have led to the development of new disciplines called Economic Statistics and Econometrics.

**Statistics and Business.** Statistics is an indispensable tool of production control also. Business executives are relying more and more on statistical techniques for studying the needs and the desires of the consumers and for many other purposes. The success of a businessman more or less depends upon the accuracy and precision of his statistical forecasting. Wrong expectations, which may be the result of faulty and inaccurate analysis of various causes affecting a particular phenomenon, might lead to his disaster. Suppose a businessman wants to manufacture readymade garments. Before starting with the production process he must have an overall idea as to 'how many garments are to be manufactured', 'how much raw material and labour is needed for that', and 'what is the quality, shape, colour, size, etc., of the garments to be manufactured'. Thus the formulation of a production plan in advance is a must which cannot be done without having quantitative facts about the details mentioned above. As such most of the large industrial and commercial enterprises are employing trained and efficient statisticians.

**Statistics and Industry.** In industry, Statistics is very widely used in 'Quality Control'. In production engineering, to find whether the product is conforming to specifications or not, statistical tools, viz., inspection plans, control charts, etc., are of extreme importance. In inspection plans we have to resort to some kind of sampling – a very important aspect of Statistics.

**Statistics and Mathematics.** Statistics and mathematics are very intimately related. Recent advancements in statistical techniques are the outcome of wide applications of advanced mathematics. Main contributors to statistics, namely, Bernouli, Pascal, Laplace, De-Moivre, Gauss, R. A. Fisher, to mention only a few, were primarily talented and skilled mathematicians. Statistics may be regarded as that branch of mathematics which provided us with systematic methods of analysing a large number of related numerical facts. According to Connor, " Statistics is a branch of Applied Mathematics which specialises in data." Increasing role of mathematics in statistical analysis has resulted in a new branch of Statistics called Mathematical Statistics.

**Statistics and Biology, Astronomy and Medical Science.** The association between statistical methods and biological theories was first studied by Francis Galton in his work in 'Regression'. According to Prof. Karl Pearson, the whole 'theory of heredity' rests on statistical basis. He says, "*The whole problem of evolution is a problem of vital statistics, a problem of longevity, of fertility, of health, of disease and it is impossible for the Registrar General to discuss the national mortality without an enumeration of the population, a classification of deaths and knowledge of statistical theory.*"

In astronomy, the theory of Gaussian 'Normal Law of Errors' for the study of the movement of stars and planets is developed by using the 'Principle of Least Squares'.

In medical science also, the statistical tools for the collection, presentation and analysis of observed facts relating to the causes and incidence of diseases and the results obtained from the use of various drugs and medicines, are of great importance. Moreover, the efficacy of a manufactured drug or injection or medicine is tested by using the 'tests of significance' – (*t*-test).

**Statistics and Psychology and Education.** In education and psychology, too, Statistics has found wide applications, e.g., to determine the reliability and validity of a test, 'Factor Analysis', etc., so much so that a new subject called 'Psychometry' has come into existence.

**Statistics and War.** In war, the theory of 'Decision Functions' can be of great assistance to military and technical personnel to plan 'maximum destruction with minimum effort'.

Thus, we see that the science of Statistics is associated with almost all the sciences – social as well as physical. Bowley has rightly said, "*A knowledge of Statistics is like a knowledge of foreign language or of algebra; it may prove of use at any time under any circumstance.*"

**1-4. Limitations of Statistics.** Statistics, with its wide applications in almost every sphere of human activity; is not without limitations. The following are some of its important limitations :

(i) *Statistics is not suited to the study of qualitative phenomenon.* Statistics, being a science dealing with a set of numerical data, is applicable to the study of only those subjects of enquiry which are capable of quantitative measurement. As such, qualitative phenomena like honesty, poverty, culture, etc., which cannot be expressed numerically, are not capable of direct statistical analysis. However, statistical techniques may be applied indirectly by first reducing the qualitative expressions to precise quantitative terms. For example, the intelligence of a group of candidates can be studied on the basis of their scores in a certain test.

(ii) *Statistics does not study individuals.* Statistics deals with an aggregate of objects and does not give any specific recognition to the individual items of a series. Individual items, taken separately, do not constitute statistical data and are meaningless for any statistical enquiry. For example, the individual figures of agricultural production, industrial output or national income of any country for a particular year are meaningless unless, to facilitate comparison, similar figures of other countries or of the same country for different years are given. Hence, statistical analysis is suited to only those problems where group characteristics are to be studied.

(iii) *Statistical laws are not exact.* Unlike the laws of physical and natural sciences, statistical laws are only approximations and not exact. On the basis of statistical analysis we can talk only in terms of probability and chance and not in terms of certainty. Statistical conclusions are not universally true – they are true only on an average. For example, let us consider the statement: "It has been found that 20 % of a certain surgical operations by a particular doctor are successful." The statement does not imply that if the doctor is to operate on 5 persons on any day and four of the operations have proved fatal, the fifth must be a success. It may happen that fifth man also dies of the operation or it may also happen that of the five operations on any day, 2 or 3 or even more may be successful. By the statement we mean that as number of operations becomes larger and larger we should expect, on the average, 20 % operations to be successful.

(iv) *Statistics is liable to be misused.* Perhaps the most important limitation of Statistics is that it must be used by experts. As the saying goes, " Statistical methods are the most dangerous tools in the hands of the in experts. Statistics is one of those sciences whose adepts must exercise the self-restraint of an artist." The use of statistical tools by inexperienced and untrained persons might lead to very fallacious conclusions. One of the greatest shortcomings of Statistics is that they do not bear on their face the label of their quality and as such can be moulded and manipulated in any manner to support one's way of argument and reasoning. As King says, "*Statistics are like clay of which one can make a god or devil as one pleases.*" The requirement of experience and skill for judicious use of statistical methods restricts their use to experts only and limits the chances of the mass popularity of this useful and important science.

**1.5. Distrust of Statistics.** We often hear the following interesting comments on Statistics :

(i) 'An ounce of truth will produce tons of Statistics',

(ii) 'Statistics can prove anything',

(iii) 'Figures do not lie. Liars figure',

(iv) 'If figures say so it can't be otherwise',

(v) 'There are three type of lies – lies, damn lies, and Statistics – wicked in the order of their naming, and so on.'

Some of the reasons for the existence of such divergent views regarding the nature and function of Statistics are as follows :

(i) Figures are innocent, easily believable and more convincing. The facts supported by figures are psychologically more appealing.

(ii) Figures put forward for arguments may be inaccurate or incomplete and thus might lead to wrong inferences.

(iii) Figures, though accurate, might be moulded and manipulated by selfish persons to conceal the truth and present a distorted picture of facts to the public to meet their selfish motives. When the skilled talkers, writers or politicians through their forceful writings and speeches or the business and commercial enterprises through advertisements in the press mislead the public or belie their expectations by quoting wrong statistical statements or manipulating statistical data for personal motives, the public loses its faith and belief in the science of Statistics and starts condemning it. We cannot blame the layman for his distrust of Statistics, as he, unlike statistician, is not in a position to distinguish between valid and invalid conclusions from statistical statements and analysis.

It may be pointed out that Statistics neither proves anything nor disproves anything. It is only a tool which if rightly used may prove extremely useful and if misused, might be disastrous. According to Bowley, "*Statistics only furnishes a tool, necessary though imperfect, which is dangerous in the hands of those who do not know its use and its deficiencies.*" It is not the subject of Statistics that is to be blamed but those people who twist the numerical data and misuse them either due to ignorance or deliberately for personal selfish motives. As King points out, "*Science of Statistics is the most useful servant but only of great value to those who understand its proper use.*"

We discuss below a few interesting examples of misrepresentation of statistical data.

(i) A statistical report : "The number of accidents taking place in the middle of the road is much less than the number of accidents taking place on its side. Hence it is safer to walk in the middle of the road." This conclusion is obviously wrong since we are not given the proportion of the number of accidents to the number of persons walking in the two cases.

(ii) "The number of students taking up Mathematics Honours in a University has increased 5 times during the last 3 years. Thus, Mathematics is gaining popularity among the students of the university." Again, the conclusion is faulty since we are not given any such details about the other subjects and hence comparative study is not possible.

(iii) "99% of the people who drink alcohol die before attaining the age of 100 years. Hence drinking is harmful for longevity of life." This statement, too, is incorrect since nothing is mentioned about the number of persons who do not drink alcohol and die before attaining the age of 100 years.

Thus, statistical arguments based on incomplete data often lead to fallacious conclusions.

# **FREQUENCY DISTRIBUTIONS AND MEASURES OF CENTRAL TENDENCY**

**2-1. Frequency Distributions.** When observations, discrete or continuous, are available on a single characteristic of a large number of individuals, often it becomes necessary to condense the data as far as possible without losing any information of interest. Let us consider the marks in Statistics obtained by 250 candidates selected at random from among those appearing in a certain examination:

**TABLE 1 : MARKS IN STATISTICS OF 250 CANDIDATES**

32	47	41	51	41	30	39	18	48	53
54	32	31	46	15	37	32	56	42	48
38	26	50	40	38	42	35	22	62	51
44	21	45	31	37	41	44	18	37	47
68	41	30	52	52	60	42	38	38	34
41	53	48	21	28	49	42	36	41	29
30	33	37	35	29	37	38	40	32	49
43	32	24	38	38	22	41	50	17	46
46	50	26	15	23	42	25	52	38	46
41	38	40	37	40	48	45	30	28	31
40	33	42	36	51	42	56	44	35	38
31	51	45	41	50	53	50	32	45	48
40	43	40	34	34	44	38	58	49	28
40	45	19	24	34	47	37	33	37	36
36	32	61	30	44	43	50	31	38	45
46	40	32	34	44	54	35	39	31	48
48	50	43	55	43	39	41	48	53	34
32	31	42	34	34	32	33	24	43	39
40	50	27	47	34	44	34	33	47	42
17	42	57	35	38	17	33	46	36	23
48	50	31	58	33	44	26	29	31	37
47	55	57	37	41	54	42	45	47	43
37	52	47	46	44	50	44	38	42	19
52	45	23	41	47	33	42	24	48	39
48	44	60	38	38	44	38	43	40	48

This representation of the data does not furnish any useful information and is rather confusing to mind. A better way may be to express the figures in an ascending or descending order of magnitude, commonly termed as *array*. But this does not reduce the bulk of the data. A much better representation is given on the next page.

A bar (|) called *tally mark* is put against the number when it occurs. Having occurred four times, the fifth occurrence is represented by putting a cross tally (/) on the first four tallies. This technique facilitates the counting of the tally marks at the end.

The representation of the data as above is known as *frequency distribution*. Marks are called the *variable* ( $x$ ) and the '*number of students*' against the marks is known as the *frequency* ( $f$ ) of the variable. The word '*frequency*' is derived from '*how frequently*' a variable occurs. For example, in the above case the frequency of 31 is 10 as there are ten students getting 31 marks. This representation, though better than an array', does not condense the data much and it is quite cumbersome to go through this huge mass of data.

TABLE 2

Marks	No. of Students Tally Marks	Total Frequency	Marks	No. of Students Tally Marks	Total Frequency
15		= 2	40		= 11
17		= 3	41		= 10
18		= 2	42		= 13
19		= 2	43		= 8
21		= 2	44		= 12
22		= 2	45		= 7
23		= 3	46		= 7
24		= 4	47		= 8
25		= 1	48		= 12
26		= 3	49		= 3
27		= 1	50		= 10
28		= 3	51		= 4
29		= 2	52		= 5
30		= 5	53		= 4
31		= 10	54		= 3
32		= 10	55		= 2
33		= 8	56		= 2
34		= 11	57		= 2
35		= 5	58		= 2
36		= 5	60		= 3
37		= 12	61	-	= 1
38		= 17	62	-	= 1
39		= 6	68	-	= 1

If the identity of the individuals about whom a particular information is taken is not relevant, nor the order in which the observations arise, then the first real step of condensation is to divide the observed range of variable into a suitable number of *class-intervals* and to record the number of observations in each class. For example, in the above case, the data may be expressed as shown in Table 3.

Such a table showing the distribution of the frequencies in the different classes is called a *frequency table* and the manner in which the class frequencies are distributed over the class intervals is called the *grouped frequency distribution* of the variable.

**Remark.** The classes of the type 15—19, 20—24, 25—29 etc., in which both the upper and lower limits are included are called '*inclusive classes*'. For example the class 20—24, includes

TABLE 3 : FREQUENCY TABLE

Marks ( $x$ )	No. of students ( $f$ )
15—19	9
20—24	11
25—29	10
30—34	44
35—39	45
40—44	54
45—49	37
50—54	26
55—59	8
60—64	5
65—69	1
Total	250

all the values from 20 to 24, both inclusive and the classification is termed as *inclusive type classification*.

In spite of great importance of classification in statistical analysis, no hard and fast rules can be laid down for it. The following points may be kept in mind for classification:

(i) The classes should be clearly defined and should not lead to any ambiguity.

(ii) The classes should be exhaustive, i.e., each of the given values should be included in one of the classes.

(iii) The classes should be mutually exclusive and non-overlapping.

(iv) The classes should be of equal width. The principle, however, cannot be rigidly followed. If the classes are of varying width, the different class frequencies will not be comparable. Comparable figures can be obtained by dividing the value of the frequencies by the corresponding widths of the class intervals. The ratios thus obtained are called '*frequency densities*'.

(v) Indeterminate classes, e.g., the open-end classes, less than 'a' or greater than 'b' should be avoided as far as possible since they create difficulty in analysis and interpretation.

(vi) The number of classes should neither be too large nor too small. It should preferably lie between 5 and 15. However, the number of classes may be more than 15 depending upon the total frequency and the details required, but it is desirable that it is not less than 5 since in that case the classification may not reveal the essential characteristics of the population. The following formula due to Struges may be used to determine an approximate number  $k$  of classes :

$$k = 1 + 3.322 \log_{10} N,$$

where  $N$  is the total frequency.

### The Magnitude of the Class Interval

Having fixed the number of classes, divide the range (the difference between the greatest and the smallest observation) by it and the nearest integer to this value gives the magnitude of the class interval. Broad class intervals (i.e., less number of classes) will yield only rough estimates while for high degree of accuracy small class intervals (i.e., large number of classes) are desirable.

### Class Limits

The class limits should be chosen in such a way that the mid-value of the class interval and actual average of the observations in that class interval are as near to each other as possible. If this is not the case then the classification gives a distorted picture of the characteristics of the data. If possible, class limits should be located at the points which are multiple of 0, 2, 5, 10,... etc., so that the midpoints of the classes are the common figures, viz., 0, 2, 5, 10,... etc., the figures capable of easy and simple analysis.

**2.1.1. Continuous Frequency Distribution.** If we deal with a continuous variable, it is not possible to arrange the data in the class intervals of above type. Let us consider the distribution of age in years. If class intervals are 15—19, 20—24 then the persons with ages between 19 and 20 years are not taken into consideration. In such a case we form the class intervals as shown below.

<i>Age in years</i>
Below 5
5 or more but less than 10
10 or more but less than 15
15 or more but less than 20
20 or more but less than 25
and so on.

Here all the persons with any fraction of age are included in one group or the other. For practical purpose we re-write the above classes as

0 — 5
5 — 10
10 — 15
15 — 20
20 — 25

This form of frequency distribution is known as *continuous frequency distribution*.

It should be clearly understood that in the above classes, the upper limits of each class are excluded from the respective classes. Such classes in which the upper limits are excluded from the respective classes and are included in the immediate next class are known as '*exclusive classes*' and the classification is termed as '*exclusive type classification*'.

**2.2. Graphic Representation of a Frequency Distribution.** It is often useful to represent a frequency distribution by means of a diagram which makes the unwieldy data intelligible and conveys to the eye the general run of the observations. Diagrammatic representation also facilitates the comparison of two or more frequency distributions. We consider below some important types of graphic representation.

**2.2.1. Histogram.** In drawing the histogram of a given continuous frequency distribution we first mark off along the  $x$ -axis all the class intervals on a suitable scale. On each class interval erect rectangles with heights proportional to the frequency of the corresponding class interval so that the area of the rectangle is proportional to the frequency of the class. If, however, the classes are of unequal width then the height of the rectangle will be proportional to the ratio of the frequencies to the width of the classes. The diagram of continuous rectangles so obtained is called *histogram*.

**Remarks.** 1. To draw the histogram for an ungrouped frequency distribution of a variable we shall have to assume that the frequency corresponding to the variate value  $x$  is spread over the interval  $x - h/2$  to  $x + h/2$ , where  $h$  is the jump from one value to the next.

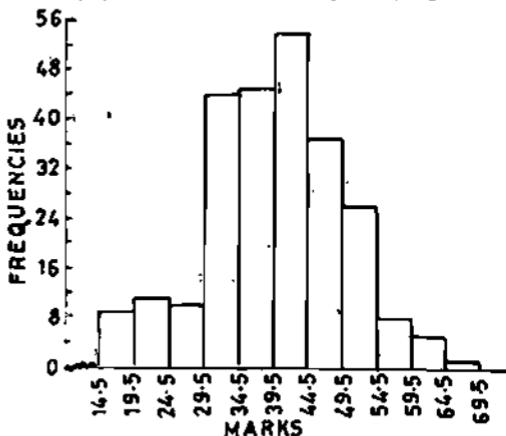
2. If the grouped frequency distribution is not continuous, first it is to be converted into continuous distribution and then the histogram is drawn.

3. Although the height of each rectangle is proportional to the frequency of the corresponding class, the height of a fraction of the rectangle is not proportional to the frequency of the corresponding fraction of the class, so that histogram cannot be directly used to read frequency over a fraction of a class interval.

4. The histogram of the distribution of marks of 250 students in Table 3 (page 2-2) is obtained as follows.

Since the grouped frequency distribution is not continuous, we first convert it into a continuous distribution as follows: **HISTOGRAM FOR FREQ. DISTRIBUTION**

Marks	No. of Students
14.5-19.5	9
19.5-24.5	11
24.5-29.5	10
29.5-34.5	44
34.5-39.5	45
39.5-44.5	54
44.5-49.5	37
49.5-54.5	26
54.5-59.5	8
59.5-64.5	5
64.5-69.5	1



**Remark.** The upper and lower class limits of the new exclusive type classes are known as *class boundaries*.

If  $d$  is the gap between the upper limit of any class and the lower limit of the succeeding class, the class boundaries for any class are then given by :

$$\text{Upper class boundary} = \text{Upper class limit} + \frac{d}{2}$$

$$\text{Lower class boundary} = \text{Lower class limit} - \frac{d}{2}$$

**2-2-2. Frequency Polygon.** For an ungrouped distribution, the frequency polygon is obtained by plotting points with abscissa as the variate values and the ordinate as the corresponding frequencies and joining the plotted points by means of straight lines. For a grouped frequency distribution, the abscissa of points are mid-values of the class intervals. For equal class intervals the frequency polygon can be obtained by joining the middle points of the upper sides of the adjacent rectangles of the histogram by means of straight lines. If the class intervals are of small width the polygon can be approximated by a smooth curve. The frequency curve can be obtained by drawing a smooth freehand curve through the vertices of the frequency polygon.

**2-3. Averages or Measures of Central Tendency or Measures of Location.** According to Professor Bowley, averages are "statistical constants which enable us to comprehend in a single effort the significance of the whole." They give us an idea about the concentration of the values in the central part of the distribution. Plainly speaking, an average of a statistical series is the value of the variable which is representative of the entire distribution. The following are the five measures of central tendency that are in common use:

- (i) *Arithmetic Mean or simply Mean*, (ii) *Median*,
- (iii) *Mode*, (iv) *Geometric Mean*, and (v) *Harmonic Mean*.

**2-4. Requisites for an Ideal Measure of Central Tendency.** According to Professor Yule, the following are the characteristics to be satisfied by an ideal measure of central tendency :

- (i) It should be rigidly defined.
- (ii) It should be readily comprehensible and easy to calculate.
- (iii) It should be based on all the observations.
- (iv) It should be suitable for further mathematical treatment. By this we mean that if we are given the averages and sizes of a number of series, we should be able to calculate the average of the composite series obtained on combining the given series.
- (v) It should be affected as little as possible by fluctuations of sampling.

In addition to the above criteria, we may add the following (which is not due to Prof. Yule) :

- (vi) It should not be affected much by extreme values.

**2-5. Arithmetic Mean.** Arithmetic mean of a set of observations is their sum divided by the number of observations, e.g., the arithmetic mean  $\bar{x}$  of  $n$  observations  $x_1, x_2, \dots, x_n$  is given by

$$\bar{x} = \frac{1}{n} (x_1 + x_2 + \dots + x_n) = \frac{1}{n} \sum_{i=1}^n x_i$$

In case of frequency distribution  $x_i | f_i$ ,  $i = 1, 2, \dots, n$ , where  $f_i$  is the frequency of the variable  $x_i$ ;

$$\bar{x} = \frac{f_1 x_1 + f_2 x_2 + \dots + f_n x_n}{f_1 + f_2 + \dots + f_n} = \frac{\sum_{i=1}^n f_i x_i}{\sum_{i=1}^n f_i} = \frac{1}{N} \sum_{i=1}^n f_i x_i, \left[ \sum_{i=1}^n f_i = N \right] \quad \dots(2-1)$$

In case of grouped or continuous frequency distribution,  $x$  is taken as the mid-value of the corresponding class.

**Remark.** The symbol  $\Sigma$  is the letter capital sigma of the Greek alphabet and is used in mathematics to denote the sum of values.

**Example 2.1.** (a) Find the arithmetic mean of the following frequency distribution:

$x:$	1	2	3	4	5	6	7
$f:$	5	9	12	17	14	10	6

(b) Calculate the arithmetic mean of the marks from the following table :

Marks	: 0-10	10-20	20-30	30-40	40-50	50-60
No. of students	: 12	18	27	20	17	6

**Solution.** (a)

$x$	$f$	$fx$
1	5	5
2	9	18
3	12	36
4	17	68
5	14	70
6	10	60
7	6	42
	73	299

$$\therefore \bar{x} = \frac{1}{N} \sum f_i x_i = \frac{299}{73} = 4.09$$

(b)

Marks	No. of students ( $f$ )	Mid-point ( $x$ )	$fx$
0-10	12	5	60
10-20	18	15	270
20-30	27	25	675
30-40	20	35	700
40-50	17	45	765
50-60	6	55	330
Total	100		2,800

$$\text{Arithmetic mean or } \bar{x} = \frac{1}{N} \sum f_i x_i = \frac{1}{100} \times 2,800 = 28$$

It may be noted that if the values of  $x$  or (and)  $f$  are large, the calculation of mean by formula (2.1) is quite time-consuming and tedious. The arithmetic is reduced to a great extent by taking the deviations of the given values from any arbitrary point 'A', as explained below.

Let  $d_i = x_i - A$ , then  $f_i d_i = f_i (x_i - A) = f_i x_i - A f_i$

Summing both sides over  $i$  from 1 to  $n$ , we get

$$\sum_{i=1}^n f_i d_i = \sum_{i=1}^n f_i x_i - A \sum_{i=1}^n f_i = \sum_{i=1}^n f_i x_i - A \cdot N.$$

$$\Rightarrow \frac{1}{N} \sum_{i=1}^n f_i d_i = \frac{1}{N} \sum_{i=1}^n f_i x_i - A = \bar{x} - A ,$$

where  $\bar{x}$  is the arithmetic mean of the distribution.

$$\therefore \bar{x} = A + \frac{1}{N} \sum_{i=1}^n f_i d_i \quad \dots(2-2)$$

This formula is much more convenient to apply than formula (2-1).

Any number can serve the purpose of arbitrary point 'A' but, usually, the value of  $x$  corresponding to the middle part of the distribution will be much more convenient.

In case of grouped or continuous frequency distribution, the arithmetic is reduced to a still greater extent by taking

$$d_i = \frac{x_i - A}{h} ,$$

where  $A$  is an arbitrary point and  $h$  is the common magnitude of class interval. In this case, we have

$$h d_i = x_i - A ,$$

and proceeding exactly similarly as above, we get

$$\bar{x} = A + \frac{h}{N} \sum_{i=1}^n f_i d_i \quad \dots(2-3)$$

**Example 2-2.** Calculate the mean for the following frequency distribution.

**Class-interval :** 0-8 8-16 16-24 24-32 32-40 40-48

**Frequency :** 8 7 16 24 15 7

**Solution.**

Class-interval	mid-value (x)	Frequency (f)	$d = (x - A) / h$	$fd$
0-8	4	8	-3	-24
8-16	12	7	-2	-14
16-24	20	16	-1	-16
24-32	28	24	0	0
32-40	36	15	1	15
40-48	44	7	2	14
		77		-25

Here we take  $A = 28$  and  $h = 8$ .

$$\therefore \bar{x} = A + \frac{h \sum f d}{N} = 28 + \frac{8 \times (-25)}{77} = 28 - \frac{200}{77} = 25.404$$

### 2-5-1. Properties of Arithmetic Mean

**Property 1.** Algebraic sum of the deviations of a set of values from their arithmetic mean is zero. If  $x_i | f_i$ ,  $i = 1, 2, \dots, n$  is the frequency distribution, then

$$\sum_{i=1}^n f_i (x_i - \bar{x}) = 0, \bar{x} \text{ being the mean of distribution.}$$

**Proof.**  $\sum_i f_i (x_i - \bar{x}) = \sum_i f_i x_i - \bar{x} \sum f_i = \sum_i f_i x_i - \bar{x} N$

Also  $\bar{x} = \frac{\sum f_i x_i}{N} \Rightarrow \sum f_i x_i = N \bar{x}$

Hence  $\sum_{i=1}^n f_i (x_i - \bar{x}) = N \bar{x} - \bar{x} N = 0$

**Property 2.** The sum of the squares of the deviations of a set of values is minimum when taken about mean.

**Proof.** For the frequency distribution  $x_i | f_i$ ,  $i = 1, 2, \dots, n$ , let

$$Z = \sum_{i=1}^n f_i (x_i - A)^2,$$

be the sum of the squares of the deviations of given values from any arbitrary point 'A'. We have to prove that  $Z$  is minimum when  $A = \bar{x}$ .

Applying the principle of maxima and minima from differential calculus,  $Z$  will be minimum for variations in  $A$  if

$$\frac{\partial Z}{\partial A} = 0 \text{ and } \frac{\partial^2 Z}{\partial A^2} > 0$$

Now  $\frac{\partial Z}{\partial A} = -2 \sum_i f_i (x_i - A) = 0 \Rightarrow \sum_i f_i (x_i - A) = 0$

$\Rightarrow \sum f_i x_i - A \sum f_i = 0 \text{ or } A = \frac{\sum f_i x_i}{N} = \bar{x}$

Again  $\frac{\partial^2 Z}{\partial A^2} = -2 \sum_i f_i (-1) = 2 \sum f_i = 2N > 0$

Hence  $Z$  is minimum at the point  $A = \bar{x}$ . This establishes the result.

**Property 3.** (Mean of the composite series). If  $\bar{x}_i$ , ( $i = 1, 2, \dots, k$ ) are the means of  $k$ -component series of sizes  $n_i$ , ( $i = 1, 2, \dots, k$ ) respectively, then the mean  $\bar{x}$  of the composite series obtained on combining the component series is given by the formula:

$$\bar{x} = \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2 + \dots + n_k \bar{x}_k}{n_1 + n_2 + \dots + n_k} = \frac{\sum n_i \bar{x}_i}{\sum n_i} \quad \dots(2.4)$$

**Proof.** Let  $x_{11}, x_{12}, \dots, x_{1n_1}$  be  $n_1$  members of the first series ;  $x_{21}, x_{22}, \dots, x_{2n_2}$  be  $n_2$  members of the second series,  $x_{k1}, x_{k2}, \dots, x_{kn_k}$  be  $n_k$  members of the  $k$ th series. Then, by def.,

$$\left. \begin{array}{l} \bar{x}_1 = \frac{1}{n_1} (x_{11} + x_{12} + \dots + x_{1n_1}) \\ \bar{x}_2 = \frac{1}{n_2} (x_{21} + x_{22} + \dots + x_{2n_2}) \\ \vdots \quad \vdots \quad \vdots \\ \bar{x}_k = \frac{1}{n_k} (x_{k1} + x_{k2} + \dots + x_{kn_k}) \end{array} \right\} \dots (*)$$

The mean  $\bar{x}$  of composite series of size  $n_1 + n_2 + \dots + n_k$  is given by

$$\begin{aligned} \bar{x} &= \frac{(x_{11} + x_{12} + \dots + x_{1n_1}) + (x_{21} + x_{22} + \dots + x_{2n_2}) + \dots + (x_{k1} + x_{k2} + \dots + x_{kn_k})}{n_1 + n_2 + \dots + n_k} \\ &= \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2 + \dots + n_k \bar{x}_k}{n_1 + n_2 + \dots + n_k}, \end{aligned} \quad (\text{From } *)$$

Thus,  $\bar{x} = \sum_i n_i \bar{x}_i / (\sum_i n_i)$

**Example 2-3.** The average salary of male employees in a firm was Rs.520 and that of females was Rs.420. The mean salary of all the employees was Rs.500. Find the percentage of male and female employees.

**Solution.** Let  $n_1$  and  $n_2$  denote respectively the number of male and female employees in the concern and  $\bar{x}_1$  and  $\bar{x}_2$  denote respectively their average salary (in rupees). Let  $\bar{x}$  denote the average salary of all the workers in the firm.

We are given that :

$$\bar{x}_1 = 520, \bar{x}_2 = 420 \text{ and } \bar{x} = 500$$

Also we know

$$\begin{aligned} \bar{x} &= \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2}{n_1 + n_2} \\ \Rightarrow 500(n_1 + n_2) &= 520n_1 + 420n_2 \\ \Rightarrow (520 - 500)n_1 &= (500 - 420)n_2 \\ \Rightarrow 20n_1 &= 80n_2 \\ \Rightarrow \frac{n_1}{n_2} &= \frac{4}{1} \end{aligned}$$

Hence the percentage of male employees in the firm

$$= \frac{4}{4+1} \times 100 = 80$$

and percentage of female employees in the firm

$$= \frac{1}{4+1} \times 100 = 20$$

## 2.5.2. Merits and Demerits of Arithmetic Mean

**Merits.** (i) It is rigidly defined.

(ii) It is easy to understand and easy to calculate.

(iii) It is based upon all the observations.

(iv) It is amenable to algebraic treatment. The mean of the composite series in terms of the means and sizes of the component series is given by

$$\bar{x} = \frac{\sum_{i=1}^k n_i \bar{x}_i}{\sum_{i=1}^k n_i}$$

(v) Of all the averages, arithmetic mean is affected least by fluctuations of sampling. This property is sometimes described by saying that arithmetic mean is a *stable average*.

Thus, we see that arithmetic mean satisfies all the properties laid down by Prof. Yule for an ideal average.

**Demerits.** (i) It cannot be determined by inspection nor it can be located graphically.

(ii) Arithmetic mean cannot be used if we are dealing with qualitative characteristics which cannot be measured quantitatively; such as, intelligence, honesty, beauty, etc. In such cases median (discussed later) is the only average to be used.

(iii) Arithmetic mean cannot be obtained if a single observation is missing or lost or is illegible unless we drop it out and compute the arithmetic mean of the remaining values.

(iv) Arithmetic mean is affected very much by extreme values. In case of extreme items, arithmetic mean gives a distorted picture of the distribution and no longer remains representative of the distribution.

(v) Arithmetic mean may lead to wrong conclusions if the details of the data from which it is computed are not given. Let us consider the following marks obtained by two students A and B in three tests, viz., terminal test, half-yearly examination and annual examination respectively.

Marks in : →	I Test	II Test	III Test	Average marks
A	50%	60%	70%	60%
B	70%	60%	50%	60%

Thus average marks obtained by each of the two students at the end of the year are 60%. If we are given the average marks alone we conclude that the level of intelligence of both the students at the end of the year is same. This is a fallacious conclusion since we find from the data that student A has improved consistently while student B has deteriorated consistently.

(vi) Arithmetic mean cannot be calculated if the extreme class is open, e.g., below 10 or above 90. Moreover, even if a single observation is missing mean cannot be calculated.

(vii) In extremely asymmetrical (skewed) distribution, usually arithmetic mean is not a suitable measure of location.

**2-5-3. Weighted Mean.** In calculating arithmetic mean we suppose that all the items in the distribution have equal importance. But in practice this may not be so. If some items in a distribution are more important than others, then this

point must be borne in mind, in order that average computed is representative of the distribution. In such cases, proper weightage is to be given to various items — the weights attached to each item being proportional to the importance of the item in the distribution. For example, if we want to have an idea of the change in cost of living of a certain group of people, then the simple mean of the prices of the commodities consumed by them will not do, since all the commodities are not equally important, e.g., wheat, rice and pulses are more important than cigarettes, tea, confectionery, etc.

Let  $w_i$  be the weight attached to the item  $x_i$ ,  $i = 1, 2, \dots, n$ . Then we define :

$$\text{Weighted arithmetic mean or weighted mean} = \frac{\sum w_i x_i}{\sum w_i} \quad \dots (2.5)$$

It may be observed that the formula for weighted mean is the same as the formula for simple mean with  $f_i$ , ( $i = 1, 2, \dots, n$ ), the frequencies replaced by  $w_i$ , ( $i = 1, 2, \dots, n$ ), the weights.

Weighted mean gives the result equal to the simple mean if the weights assigned to each of the variate values are equal. It results in higher value than the simple mean if smaller weights are given to smaller items and larger weights to larger items. If the weights attached to larger items are smaller and those attached to smaller items are larger, then the weighted mean results in smaller value than the simple mean.

**Example 2.4.** Find the simple and weighted arithmetic mean of the first  $n$  natural numbers, the weights being the corresponding numbers.

**Solution.** The first natural numbers are  $1, 2$

,  $3, \dots, n$ .

We know that

$$1 + 2 + 3 + \dots + n = \frac{n(n+1)}{2}$$

$$1^2 + 2^2 + 3^2 + \dots + n^2 = \frac{n(n+1)(2n+1)}{6}$$

Simple A.M. is

$$\bar{X} = \frac{\sum X}{n} = \frac{1 + 2 + 3 + \dots + n}{n} = \frac{n+1}{2}$$

Weighted A.M. is

$$\begin{aligned}\bar{X}_w &= \frac{\sum w X}{\sum w} = \frac{1^2 + 2^2 + \dots + n^2}{1 + 2 + \dots + n} \\ &= \frac{n(n+1)(2n+1)}{6} \cdot \frac{2}{n(n+1)}\end{aligned}$$

X	w	wX
1	1	1 <sup>2</sup>
2	2	2 <sup>2</sup>
3	3	3 <sup>2</sup>
:	:	:
n	n	n <sup>2</sup>

**2.6. Median.** Median of a distribution is the value of the variable which divides it into two equal parts. It is the value which exceeds and is exceeded by the same number of observations, i.e., it is the value such that the number of observations above it is equal to the number of observations below it. The median is thus a *positional average*.

In case of ungrouped data, if the number of observations is odd then median is the middle value after the values have been arranged in ascending or descending order of magnitude. In case of even number of observations, there are two middle terms and median is obtained by taking the arithmetic mean of the middle terms. For example, the median of the values 25, 20, 15, 35, 18, i.e., 15, 18, 20, 25, 35 is 20 and the median of 8, 20, 50, 25, 15, 30, i.e., of 8, 15, 20, 25, 30, 50 is  $\frac{1}{2}(20 + 25) = 22.5$ .

**Remark.** In case of even number of observations, in fact any value lying between the two middle values can be taken as median but conventionally we take it to be the mean of the middle terms.

In case of discrete frequency distribution median is obtained by considering the cumulative frequencies. The steps for calculating median are given below:

(i) Find  $N/2$ , where  $N = \sum_i f_i$ .

(ii) See the (less than) cumulative frequency (c.f.) just greater than  $N/2$ .

(iii) The corresponding value of  $x$  is median.

**Example 2.5.** Obtain the median for the following frequency distribution:

$x :$	1	2	3	4	5	6	7	8	9
$f :$	8	10	11	16	20	25	15	9	6

**Solution.**

$x$	$f$	c.f.
1	8	8
2	10	18
3	11	29
4	16	45
5	20	65
6	25	90
7	15	105
8	9	114
9	6	120
		120

$$\text{Hence } N = 120 \Rightarrow N/2 = 60$$

Cumulative frequency (c.f.) just greater than  $N/2$ , is 65 and the value of  $x$  corresponding to 65 is 5. Therefore, median is 5.

In the case of continuous frequency distribution, the class corresponding to the c.f. just greater than  $N/2$  is called the *median class* and the value of median is obtained by the following formula :

$$\text{Median} = l + \frac{h}{f} \left( \frac{N}{2} - c \right) \quad \dots(2.6)$$

where  $l$  is the lower limit of the median class,

$f$  is the frequency of the median class,

$h$  is the magnitude of the median class,

' $c$ ' is the c.f. of the class preceding the median class,

and  $N = \sum f$ .

**Derivation of the Median Formula (2.6).** Let us consider the following continuous frequency distribution, ( $x_1 < x_2 < \dots < x_{n+1}$ ):

Class interval :  $x_1 - x_2, x_2 - x_3, \dots, x_k - x_{k+1}, \dots, x_n - x_{n+1}$

Frequency :  $f_1, f_2, \dots, f_k, \dots, f_n$

The cumulative frequency distribution is given by :

Class interval :  $x_1 - x_2, x_2 - x_3, \dots, x_k - x_{k+1}, \dots, x_n - x_{n+1}$

Frequency :  $F_1, F_2, \dots, F_k, \dots, F_n$

where  $F_i = f_1 + f_2 + \dots + f_i$ . The class  $x_k - x_{k+1}$  is the median class if and only if  $F_{k-1} < N/2 \leq F_k$ .

Now, if we assume that the variate values are uniformly distributed over the median-class which implies that the ogive is a straight line in the median-class, then

we get from the Fig. 1,

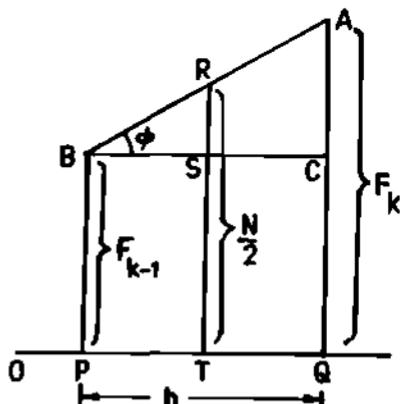
$$\tan \phi = \frac{RS}{BS} = \frac{AC}{BC}$$

$$\text{i.e. } \frac{RT - TS}{BS} = \frac{AQ - CQ}{BC}$$

$$\text{or } \frac{RT - BP}{BS} = \frac{AQ - BP}{PQ}$$

$$\text{or } \frac{N/2 - F_{k-1}}{BS} = \frac{F_k - F_{k-1}}{PQ}$$

$$= \frac{f_k}{h}$$



where  $f_k$  is the frequency and  $h$  the magnitude of the median class.

$$\therefore BS = \frac{h}{f_k} \left( \frac{N}{2} - F_{k-1} \right)$$

Hence

$$\text{Median} = OT = OP + PT = OP + BS$$

$$= l + \frac{h}{f_k} \left( \frac{N}{2} - F_{k-1} \right)$$

which is the required formula.

**Remark.** The median formula (2.6) can be used only for continuous classes without any gaps, i.e., for 'exclusive type' classification. If we are given a frequency

distribution in which classes are of '*inclusive type*' with gaps, then it must be converted into a continuous '*exclusive type*' frequency distribution without any gaps before applying (2-6). This will affect the value of  $l$  in (2-6). As an illustration see Example 2-7.

**Example 2-6.** Find the median wage of the following distribution :

Wages (in Rs.)	20—30	30—40	40—50	50—60	60—70
No. of labourers	3	5	20	10	5

[Gorakhpur Univ. B. Sc. 1989]

**Solution.**

Wages (in Rs.)	No. of labourers	c.f.
20—30	3	3
30—40	5	8
40—50	20	28
50—60	10	38
60—70	5	43

Here  $N/2 = 43/2 = 21.5$ . Cumulative frequency just greater than 21.5 is 28 and the corresponding class is 40–50. Thus median class is 40–50. Hence using (2-6), we get

$$\text{Median} = 40 + \frac{10}{20} (21.5 - 8) = 40 + 6.75 = 46.75$$

Thus median wage is Rs. 46.75.

**Example 2-7.** In a factory employing 3,000 persons, 5 per cent earn less than Rs. 3 per hour, 580 earn from Rs. 3.01 to Rs. 4.50 per hour, 30 percent earn from Rs. 4.51 to Rs. 6.00 per hour, 500 earn from Rs. 6.01 to Rs. 7.50 per hour, 20 percent earn from Rs. 7.51 to Rs. 9.00 per hour, and the rest earn Rs. 9.01 or more per hour. What is the median wage?

[Utkal Univ. B.Sc. 1992]

**Solution.** The given information can be expressed in tabular form as follows.

#### CALCULATIONS FOR MEDIAN WAGE

Earnings (in Rs.)	Percentage of workers	No. of workers (f)	Less than c.f.	Class boundaries
less than 3	5%	$\frac{5}{100} \times 3000 = 150$	150	Below 3.005
3.01—4.50	—	580	730	3.005—4.505
4.51—6.00	30 %	$\frac{30}{100} \times 3000 = 900$	1630	4.505—6.005
6.01—7.50	—	500	2130	6.005—7.505
7.51—9.00	20 %	$\frac{20}{100} \times 3000 = 600$	2730	7.505—9.005
9.01 and above		$3000 - 2730 = 270$	$3000 = N$	9.005 and above

$N/2 = 1500$ . The c.f. just greater than 1500 is 1630. The corresponding class 4.51–6.00, whose class boundaries are 4.505–6.005, is the median class. Using the median formula, we get :

$$\text{Median} = l + \frac{h}{f} \left( \frac{N}{2} - C \right) = 4.505 + \frac{1.5}{900} (1500 - 730) \\ = 4.505 + 1.283 \approx 5.79$$

Hence median wage is Rs. 5.79.

**Example 2-8.** An incomplete frequency distribution is given as follows:

Variable	Frequency	Variable	Frequency
10–20	12	50–60	?
20–30	30	60–70	25
30–40	?	70–80	18
40–50	65	Total	229

Given that the median value is 46, determine the missing frequencies using the median formula.

[Delhi Univ. B. Sc., Oct. 1992]

**Solution.** Let the frequency of the class 30–40 be  $f_1$  and that of 50–60 be  $f_2$ .

$$\text{Then } f_1 + f_2 = 229 - (12 + 30 + 65 + 25 + 18) = 79.$$

Since median is given to be 46, the class 40–50 is the median class.

Hence using median formula (2-6), we get

$$46 = 40 + \frac{114.5 - (12 + 30 + f_1)}{65} \times 10$$

$$46 - 40 = \frac{72.5 - f_1}{65} \times 10 \text{ or } 6 = \frac{72.5 - f_1}{6.5}$$

$$f_1 = 72.5 - 39 = 33.5 \approx 34$$

[Since frequency is never fractional]

$$\therefore f_2 = 79 - 34 = 45$$

[Since  $f_1 + f_2 = 79$ ]

### 2-6-1. Merits and Demerits of Median

**Merits.** (i) It is rigidly defined.

(ii) It is easily understood and is easy to calculate. In some cases it can be located merely by inspection.

(iii) It is not at all affected by extreme values.

(iv) It can be calculated for distributions with open-end classes.

**Demerits.** (i) In case of even number of observations median cannot be determined exactly. We merely estimate it by taking the mean of two middle terms.

(ii) It is not based on all the observations. For example, the median of 10, 25, 50, 60 and 65 is 50. We can replace the observations 10 and 25 by any two values which are smaller than 50 and the observations 60 and 65 by any two values greater than 50 without affecting the value of median. This property is sometimes described

by saying that median is *insensitive*.

(iii) It is not amenable to algebraic treatment.

(iv) As compared with mean, it is affected much by fluctuations of sampling.

**Uses.** (i) Median is the only average to be used while dealing with qualitative data which cannot be measured quantitatively but still can be arranged in ascending or descending order of magnitude, e.g., to find the average intelligence or average honesty among a group of people.

(ii) It is to be used for determining the typical value in problems concerning wages, distribution of wealth, etc.

**2-7. Mode.** Let us consider the following statements :

(i) The average height of an Indian (male) is 5'-6".

(ii) The average size of the shoes sold in a shop is 7.

(iii) An average student in a hostel spends Rs.150 p.m.

In all the above cases, the average referred to is mode. Mode is the value which occurs most frequently in a set of observations and around which the other items of the set cluster densely. In other words, mode is the value of the variable which is predominant in the series. Thus in the case of discrete frequency distribution mode is the value of  $x$  corresponding to maximum frequency. For example, in the following frequency distribution :

$x$	:	1	2	3	4	5	6	7	8
$f$	:	4	9	16	25	22	15	7	3

the value of  $x$  corresponding to the maximum frequency, viz., 25 is 4. Hence mode is 4.

But in any one (or more) of the following cases :

(i) if the maximum frequency is repeated,

(ii) if the maximum frequency occurs in the very beginning or at the end of the distribution, and

(iii) if there are irregularities in the distribution,

the value of mode is determined by the *method of grouping*, which is illustrated below by an example.

**Example 2-9.** Find the mode of the following frequency distribution :

Size ( $x$ ) : 1 2 3 4 5 6 7 8 9 10 11 12

Frequency ( $f$ ) : 3 8 15 23 35 40 32 28 20 45 14 6

**Solution.** Here we see that the distribution is not regular since the frequencies are increasing steadily up to 40 and then decrease but the frequency 45 after 20 does not seem to be consistent with the distribution. Here we cannot say that since maximum frequency is 45, mode is 10. Here we shall locate mode by the method of grouping as explained below :

The frequencies in column (i) are the original frequencies. Column (ii) is obtained by combining the frequencies two by two. If we leave the first frequency and combine the remaining frequencies two by two we get column (iii). Combining

Size (x)	Frequency -					
	(i)	(ii)	(iii)	(iv)	(v)	(vi)
1	3					
2	8	11	23	26	46	73
3	15					
4	23	38	58	98	107	100
5	35					
6	40	75	72			
7	32					
8	28	60	48	80	93	79
9	20					
10	45	65	59	65		
11	14					
12	6	20				

the frequencies two by two after leaving the first two frequencies results in a repetition of column (ii). Hence, we proceed to combine the frequencies three by three, thus getting column (iv). The combination of frequencies three by three after leaving the first frequency results in column (v) and after leaving the first two frequencies results in column (vi).

The maximum frequency in each column is given in black type. To find mode we form the following table :

#### ANALYSIS TABLE

Column Number (1)	Maximum Frequency (2)	Value or combination of values of x giving max. frequency in (2) (3)
(i)	45	10
(ii)	75	5, 6
(iii)	72	6, 7
(iv)	98	4, 5, 6,
(v)	107	5, 6, 7
(vi)	100	6, 7, 8

On examining the values in column (3) above, we find that the value 6 is repeated the maximum number of times and hence the value of mode is 6 and not 10 which is an irregular item.

In case of continuous frequency distribution, mode is given by the formula :

$$\text{Mode} = l + \frac{h(f_1 - f_0)}{(f_1 - f_0) - (f_2 - f_1)} = l + \frac{h(f_1 - f_0)}{2f_1 - f_0 - f_2} \quad \dots(2.7)$$

where  $l$  is the lower limit,  $h$  the magnitude and  $f_1$  the frequency of the modal class,  $f_0$  and  $f_2$  are the frequencies of the classes preceding and succeeding the modal class respectively.

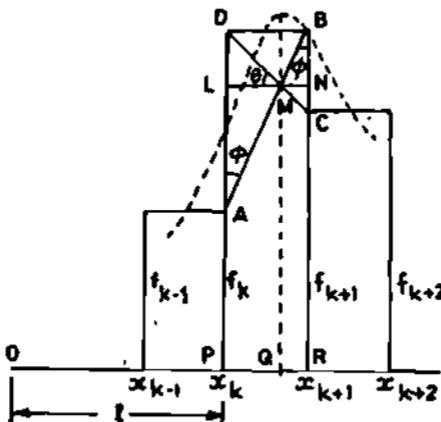
**Derivation of the Mode Formula (2.7).** Let us consider the continuous frequency distribution :

Class :  $x_1 - x_2, x_2 - x_3, \dots, x_k - x_{k+1}, \dots, x_n - x_{n+1}$

Frequency :  $f_1, f_2, \dots, f_k, \dots, f_n$ .

If  $f_1$  is the maximum of all the frequencies, then the modal class is  $(x_k - x_{k+1})$ .

Let us further consider a portion of the histogram, namely, the rectangles erected on the modal class and the two adjacent classes. The mode is the value of  $x$  for which the frequency curve has a maxima. Let the modal point be Q.



From the figure, we have

$$\tan \theta = \frac{LD}{LM} = \frac{NC}{MN}$$

and  $\tan \phi = \frac{LM}{AL} = \frac{MN}{NB}$

$$\therefore \frac{LM}{MN} = \frac{LD}{NC} = \frac{AL}{NB} = \frac{AL + LD}{NB + NC} = \frac{AD}{BC}$$

i.e.,  $\frac{LM}{LN - LM} = \frac{PD - AP}{BR - CR}$

or  $\frac{LM}{h - LM} = \frac{f_1 - f_{k-1}}{f_1 - f_{k+1}}$ , where 'h' is the magnitude of the modal class. Thus solving for LM, we get

$$LM = \frac{h(f_k - f_{k-1})}{(f_k - f_{k+1}) + (f_k - f_{k-1})} = \frac{h(f_k - f_{k-1})}{2f_k - f_{k-1} - f_{k+1}}$$

Hence Mode =  $OQ = OP + PQ = OP + LM$

$$= l + \frac{h(f_k - f_{k-1})}{2f_k - f_{k-1} - f_{k+1}}$$

**Example 2-10.** Find the mode for the following distribution :

Class - interval : 0-10 10-20 20-30 30-40 40-50 50-60 60-70 70-80

Frequency : 5 8 7 12 28 20 10 10

**Solution.** Here maximum frequency is 28. Thus the class 40-50 is the modal class. Using (2-7), the value of mode is given by

$$\text{Mode} = 40 + \frac{10(28 - 12)}{(2 \times 28 - 12 - 20)} = 40 + 6.666 = 46.67 \text{ (approx.)}$$

**Example 2-11.** The Median and Mode of the following wage distribution are known to be Rs. 33.50 and Rs. 34 respectively. Find the values of  $f_3$ ,  $f_4$  and  $f_5$ .

Wages : (in Rs.)	0-10	10-20	20-30	30-40	40-50
Frequency :	4	16	$f_3$	$f_4$	$f_5$
Wages :	50-60	60-70	Total		
Frequency :	6	4	230		

[Gujarat Univ. B.Sc., 1991]

**Solution.**

#### CALCULATIONS FOR MODE AND MEDIAN

Wages (in Rs.)	Frequency ( $f$ )	Less than c.f.
0-10	4	4
10-20	16	20
20-30	$f_3$	$20 + f_3$
30-40	$f_4$	$20 + f_3 + f_4$
40-50	$f_5$	$20 + f_3 + f_4 + f_5$
50-60	6	$26 + f_3 + f_4 + f_5$
60-70	4	$30 + f_3 + f_4 + f_5$
<i>Total</i>	$230 = 30 + f_3 + f_4 + f_5$	

From the above table, we get

$$\Sigma f = 30 + f_3 + f_4 + f_5 = 230$$

$$\Rightarrow f_3 + f_4 + f_5 = 230 - 30 = 200 \quad \dots(i)$$

Since median is 33.5, which lies in the class 30-40, 30-40 is the median class. Using the median formula, we get

$$Md = l + \frac{h}{f} \left( \frac{N}{2} - C \right)$$

$$\Rightarrow 33.5 = 30 + \frac{10}{f_4} [115 - (20 + f_3)]$$

$$\Rightarrow \frac{33.5 - 30}{10} = \frac{95 - f_3}{f_4}$$

$$\Rightarrow 0.35 f_4 = 95 - f_3 \Rightarrow f_3 = 95 - 0.35 f_4 \quad \dots(ii)$$

Mode being 34, the modal class is also 30-40. Using mode formula we get :

$$34 = 30 + \frac{10(f_4 - f_3)}{2f_4 - f_3 - f_5}$$

$$\Rightarrow \frac{34 - 30}{10} = \frac{f_4 + 0.35 f_4 - 95}{2f_4 - (200 - f_4)} \quad [\text{Using (i) and (ii)}]$$

$$\Rightarrow 0.4 = \frac{1.35 f_4 - 95}{3f_4 - 200}$$

$$\Rightarrow 1.2f_4 - 80 = 1.35f_4 - 95$$

$$\Rightarrow f_4 = \frac{95 - 80}{1.35 - 1.20} = \frac{15}{0.15} = 100 \quad \dots(iii)$$

Substituting in (ii) we get :

$$f_3 = 95 - 0.35 \times 100 = 60$$

Substituting the values of  $f_3$  and  $f_4$  in (i) we get :

$$f_5 = 200 - f_3 - f_4 = 200 - 60 - 100 = 40$$

Hence  $f_3 = 60, f_4 = 100$  and  $f_5 = 40$ .

**Remarks.** 1. In case of irregularities in the distribution, or the maximum frequency being repeated or the maximum frequency occurring in the very beginning or at the end of the distribution, the modal class is determined by the method of grouping and the mode is obtained by using (2.7).

Sometimes mode is estimated from the mean and the median. For a symmetrical distribution, mean, median and mode coincide. If the distribution is moderately asymmetrical, the mean, median and mode obey the following empirical relationship (due to Karl Pearson) :

$$\text{Mean} - \text{Median} = \frac{1}{3} (\text{Mean} - \text{Mode})$$

$$\Rightarrow \text{Mode} = 3 \text{Median} - 2 \text{Mean} \quad \dots(2.8)$$

2. If the method of grouping gives the modal class which does not correspond to the maximum frequency, i.e., the frequency of modal class is not the maximum frequency, then in some situations we may get,  $2f_k - f_{k-1} - f_{k+1} = 0$ . In such cases, the value of mode can be obtained by the formula :

$$\text{Mode} = l + \frac{h(f_k - f_{k-1})}{|f_k - f_{k-1}| + |f_k - f_{k+1}|}$$

### 2.7.1. Merits and Demerits of Mode

**Merits.** (i) Mode is readily comprehensible and easy to calculate. Like median, mode can be located in some cases merely by inspection.

(ii) Mode is not at all affected by extreme values.

(iii) Mode can be conveniently located even if the frequency distribution has class-intervals of unequal magnitude provided the modal class and the classes preceding and succeeding it are of the same magnitude. Open-end classes also do not pose any problem in the location of mode.

**Demerits.** (i) Mode is ill-defined. It is not always possible to find a clearly defined mode. In some cases, we may come across distributions with two modes. Such distributions are called *bi-modal*. If a distribution has more than two modes, it is said to be *multimodal*.

(ii) It is not based upon all the observations.

(iii) It is not capable of further mathematical treatment.

(iv) As compared with mean, mode is affected to a greater extent by fluctuations of sampling.

**Uses.** Mode is the average to be used to find the ideal size, e.g., in business forecasting, in the manufacture of ready-made garments, shoes, etc.

**2.8. Geometric Mean.** Geometric mean of a set of  $n$  observations is the  $n$ th root of their product. Thus the geometric mean  $G$ , of  $n$  observations  $x_i, i = 1, 2, \dots, n$  is

$$G = (x_1 \cdot x_2 \cdot \dots \cdot x_n)^{1/n} \quad \dots(2.9)$$

The computation is facilitated by the use of logarithms. Taking logarithm of both sides, we get

$$\begin{aligned} \log G &= \frac{1}{n} (\log x_1 + \log x_2 + \dots + \log x_n) = \frac{1}{n} \sum_{i=1}^n \log x_i \\ \therefore G &= \text{Antilog} \left[ \frac{1}{n} \sum_{i=1}^n \log x_i \right] \end{aligned} \quad \dots(2.9a)$$

In case of frequency distribution  $x_i | f_i, (i = 1, 2, \dots, n)$  geometric mean,  $G$  is given by

$$G = \left[ x_1^{f_1} \cdot x_2^{f_2} \cdot \dots \cdot x_n^{f_n} \right]^{\frac{1}{N}}, \text{ where } N = \sum_{i=1}^n f_i \quad \dots(2.10)$$

Taking logarithms of both sides, we get

$$\begin{aligned} \log G &= \frac{1}{N} (f_1 \log x_1 + f_2 \log x_2 + \dots + f_n \log x_n) \\ &= \frac{1}{N} \sum_{i=1}^n f_i \log x_i \end{aligned} \quad \dots(2.10a)$$

Thus we see that logarithm of  $G$  is the arithmetic mean of the logarithms of the given values. From (2-10a), we get

$$G = \text{Antilog} \left( \frac{1}{N} \sum_{i=1}^n f_i \log x_i \right) \quad \dots(2-10b)$$

In the case of grouped or continuous frequency distribution,  $x$  is taken to be the value corresponding to the mid-point of the class-intervals.

### 2-8-1. Merits and Demerits of Geometric Mean

**Merits.** (i) It is rigidly defined.

(ii) It is based upon all the observations.

(iii) It is suitable for further mathematical treatment. If  $n_1$  and  $n_2$  are the sizes,  $G_1$  and  $G_2$  the geometric means of two series respectively, the geometric mean  $G$ , of the combined series is given by

$$\log G = \frac{n_1 \log G_1 + n_2 \log G_2}{n_1 + n_2} \quad \dots(2-11)$$

**Proof.** Let  $x_{1i}$  ( $i = 1, 2, \dots, n_1$ ) and  $x_{2j}$  ( $j = 1, 2, \dots, n_2$ ) be  $n_1$  and  $n_2$  items of two series respectively. Then by def.,

$$G_1 = (x_{11} \cdot x_{12} \dots x_{1n_1})^{1/n_1} \Rightarrow \log G_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} \log x_{1i}$$

$$G_2 = (x_{21} \cdot x_{22} \dots x_{2n_2})^{1/n_2} \Rightarrow \log G_2 = \frac{1}{n_2} \sum_{j=1}^{n_2} \log x_{2j}$$

The geometric mean  $G$  of the combined series is given by

$$\begin{aligned} G &= (x_{11} \cdot x_{12} \dots x_{1n_1} \cdot x_{21} \cdot x_{22} \dots x_{2n_2})^{1/(n_1+n_2)} \\ \therefore \log G &= \frac{1}{n_1+n_2} \left[ \sum_{i=1}^{n_1} \log x_{1i} + \sum_{j=1}^{n_2} \log x_{2j} \right] \\ &= \frac{1}{n_1+n_2} [n_1 \log G_1 + n_2 \log G_2] \end{aligned}$$

The result can be easily generalised to more than two series.

(iv) It is not affected much by fluctuations of sampling.

(v) It gives comparatively more weight to small items.

**Demerits.** (i) Because of its abstract mathematical character, geometric mean is not easy to understand and to calculate for a non-mathematics person.

(ii) If any one of the observations is zero, geometric mean becomes zero and if any one of the observations is negative, geometric mean becomes imaginary regardless of the magnitude of the other items.

**Uses.** Geometric mean is used –

(i) To find the rate of population growth and the rate of interest.

(ii) In the construction of index numbers.

**Example 2·12.** Show that in finding the arithmetic mean of a set of readings on thermometer it does not matter whether we measure temperature in Centigrade or Fahrenheit, but that in finding the geometric mean it does matter which scale we use. [Patna Univ. B.Sc., 1991]

**Solution.** Let  $C_1, C_2, \dots, C_n$  be the  $n$  readings on the Centigrade thermometer. Then their arithmetic mean  $\bar{C}$  is given by :

$$\bar{C} = \frac{1}{n} (C_1 + C_2 + \dots + C_n)$$

If  $F$  and  $C$  be the readings in Fahrenheit and Centigrade respectively then we have the relation :

$$\frac{F - 32}{180} = \frac{C}{100} \Rightarrow F = 32 + \frac{9}{5} C.$$

Thus the Fahrenheit equivalents of  $C_1, C_2, \dots, C_n$  are

$$32 + \frac{9}{5} C_1, 32 + \frac{9}{5} C_2, \dots, 32 + \frac{9}{5} C_n,$$

respectively.

Hence the arithmetic mean of the readings in Fahrenheit is

$$\begin{aligned}\bar{F} &= \frac{1}{n} \left\{ \left(32 + \frac{9}{5} C_1\right) + \left(32 + \frac{9}{5} C_2\right) + \dots + \left(32 + \frac{9}{5} C_n\right) \right\} \\ &= \frac{1}{n} \left\{ 32n + \frac{9}{5} (C_1 + C_2 + \dots + C_n) \right\} \\ &= 32 + \frac{9}{5} \left( \frac{C_1 + C_2 + \dots + C_n}{n} \right) \\ &= 32 + \frac{9}{5} \bar{C},\end{aligned}$$

which is the Fahrenheit equivalent of  $\bar{C}$ .

Hence in finding the arithmetic mean of a set of  $n$  readings on a thermometer, it is immaterial whether we measure temperature in Centigrade or Fahrenheit.

Geometric mean  $G$ , of  $n$  readings in Centigrade is

$$G = (C_1 \cdot C_2 \cdots C_n)^{1/n}$$

Geometric mean  $G_1$ , (say), of Fahrenheit equivalents of  $C_1, C_2, \dots, C_n$  is

$$G_1 = \left\{ \left(32 + \frac{9}{5} C_1\right) \left(32 + \frac{9}{5} C_2\right) \cdots \left(32 + \frac{9}{5} C_n\right) \right\}^{1/n}$$

which is not equal to Fahrenheit equivalent of  $G$ , viz.,

$$\left\{ \frac{9}{5} (C_1 \cdot C_2 \cdots C_n)^{1/n} + 32 \right\}$$

Hence in finding the geometric mean of the  $n$  readings on a thermometer, the scale, (Centigrade or Fahrenheit) is important.

**2.9. Harmonic Mean.** Harmonic mean of a number of observations is the reciprocal of the arithmetic mean of the reciprocals of the given values. Thus, harmonic mean  $H$ , of  $n$  observations  $x_i$ ,  $i = 1, 2, \dots, n$  is

$$H = \frac{1}{\frac{1}{n} \sum_{i=1}^n \left( \frac{1}{x_i} \right)} \quad \dots(2.12)$$

In case of frequency distribution  $x_i | f_i$ , ( $i = 1, 2, \dots, n$ ),

$$H = \frac{1}{\frac{1}{N} \sum_{i=1}^n \left( \frac{f_i}{x_i} \right)}, \quad \left[ N = \sum_{i=1}^n f_i \right] \quad \dots(2.12a)$$

### 2.9.1. Merits and Demerits of Harmonic Mean

**Merits.** Harmonic mean is rigidly defined, based upon all the observations and is suitable for further mathematical treatment. Like geometric mean, it is not affected much by fluctuations of sampling. It gives greater importance to small items and is useful only when small items have to be given a greater weightage.

**Demerits.** Harmonic mean is not easily understood and is difficult to compute.

**Example 2.13.** A cyclist pedals from his house to his college at a speed of 10 m.p.h. and back from the college to his house at 15 m.p.h. Find the average speed.

**Solution.** Let the distance from the house to the college be  $x$  miles. In going from house to college, the distance ( $x$  miles) is covered in  $\frac{x}{10}$  hours, while in coming from college to house, the distance is covered in  $\frac{x}{15}$  hours. Thus a total distance of  $2x$  miles is covered in  $\left( \frac{x}{10} + \frac{x}{15} \right)$  hours.

$$\begin{aligned} \text{Hence average speed} &= \frac{\text{Total distance travelled}}{\text{Total time taken}} = \frac{2x}{\left( \frac{x}{10} + \frac{x}{15} \right)} \\ &= \frac{2}{\left( \frac{1}{10} + \frac{1}{15} \right)} = 12 \text{ m.p.h.} \end{aligned}$$

**Remark.** 1. In this case the average speed is given by the harmonic mean of 10 and 15 and not by the arithmetic mean.

Rather, we have the following general result :

If equal distances are covered (travelled) per unit of time with speeds equal to  $V_1, V_2, \dots, V_n$ , say, then the average speed is given by the harmonic mean of  $V_1, V_2, \dots, V_n$ , i.e.,

$$\text{Average speed} = \frac{n}{\left( \frac{1}{V_1} + \frac{1}{V_2} + \dots + \frac{1}{V_n} \right)} = \frac{n}{\sum \left( \frac{1}{V} \right)}$$

Proof is left as an exercise to the reader.

$$\text{Hint. Speed} = \frac{\text{Distance}}{\text{Time}} \Rightarrow \text{Time} = \frac{\text{Distance}}{\text{Speed}}$$

$$\text{Average Speed} = \frac{\text{Total distance travelled}}{\text{Total time taken}}$$

**2. Weighted Harmonic Mean.** Instead of fixed (constant) distance being travelled with varying speed, let us now suppose that different distances, say,  $S_1, S_2, \dots, S_n$ , are travelled with different speeds, say,  $V_1, V_2, \dots, V_n$  respectively. In that case, the average speed is given by the weighted harmonic mean of the speeds, the weights being the corresponding distances travelled, i.e.,

$$\text{Average speed} = \frac{S_1 + S_2 + \dots + S_n}{\left( \frac{S_1}{V_1} + \frac{S_2}{V_2} + \dots + \frac{S_n}{V_n} \right)} = \frac{\Sigma S}{\Sigma \left( \frac{S}{V} \right)}$$

**Example 2-14.** You can take a trip which entails travelling 900 km. by train at an average speed of 60 km. per hour, 3000 km. by boat at an average of 25 km. p.h., 400 km. by plane at 350 km. per hour and finally 15 km. by taxi at 25 km. per hour. What is your average speed for the entire distance ?

**Solution.** Since different distances are covered with varying speeds, the required average speed for the entire distance is given by the weighted harmonic mean of the speeds (in km.p.h.), the weights being the corresponding distances covered (in kms.).

COMPUTATION OF WEIGHTED H. M.		
Speed (km. / hr.) $X$	Distance (in km.) $W$	$W/X$
60	900	15.00
25	3000	120.00
350	400	1.43
25	15	0.60
Total	$\Sigma W = 4315$	$\Sigma (W/X) = 137.03$

$$\begin{aligned}\text{Average speed} &= \frac{\Sigma W}{\Sigma (W/X)} \\ &= \frac{4315}{137.03} \\ &= 31.489 \text{ km.p.h.}\end{aligned}$$

**2-10. Selection of an Average.** From the preceding discussion it is evident that no single average is suitable for all practical purposes. Each one of the average has its own merits and demerits and thus its own particular field of importance and utility. We cannot use the averages indiscriminately. A judicious selection of the average depending on the nature of the data and the purpose of the enquiry is essential for sound statistical analysis. Since arithmetic mean satisfies all the properties of an ideal average as laid down by Prof. Yule, is familiar to a layman and further has wide applications in statistical theory at large, it may be regarded as the best of all the averages.

**2-11. Partition Values.** These are the values which divide the series into a number of equal parts.

The three points which divide the series into four equal parts are called *quartiles*. The first, second and third points are known as the first, second and third quartiles respectively. The first quartile,  $Q_1$ , is the value which exceed 25% of the observations and is exceeded by 75% of the observations. The second quartile,  $Q_2$ , coincides with median. The third quartile,  $Q_3$ , is the point which has 75% observations before it and 25% observations after it.

The nine points which divide the series into ten equal parts are called *deciles* whereas *percentiles* are the ninety-nine points which divide the series into hundred equal parts. For example,  $D_1$ , the seventh decile, has 70% observations before it and  $P_{47}$ , the forty-seventh percentile, is the point which exceed 47% of the observations. The methods of computing the partition values are the same as those of locating the median in the case of both discrete and continuous distributions.

**Example 2.15.** Eight coins were tossed together and the number of heads resulting was noted. The operation was repeated 256 times and the frequencies ( $f$ ) that were obtained for different values of  $x$ , the number of heads, are shown in the following table. Calculate median, quartiles, 4th decile and 27th percentile.

$x :$	0	1	2	3	4	5	6	7	8
$f :$	1	9	26	59	72	52	29	7	1

**Solution.**

$x :$	0	1	2	3	4	5	6	7	8
$f :$	1	9	26	59	72	52	29	7	1
$c.f. :$	1	10	36	95	167	219	248	255	256

Median : Here  $N/2 = 256/2 = 128$ . Cumulative frequency ( $c.f.$ ) just greater than 128 is 167. Thus, median = 4.

$Q_1$  : Here  $N/4 = 64$ .  $c.f.$  just greater than 64 is 95. Hence,  $Q_1 = 3$ .

$Q_3$  : Here  $3N/4 = 192$  and  $c.f.$  just greater than 192 is 219. Thus  $Q_3 = 5$ .

$D_4$  :  $\frac{4N}{10} = 4 \times 25.6 = 102.4$  and  $c.f.$  just greater than 102.4 is 167. Hence  $D_4 = 4$ .

$P_{27} : \frac{27N}{100} = 27 \times 2.56 = 69.12$  and  $c.f.$  just greater than 69.12 is 95. Hence  $P_{27} = 3$ .

**2.11.1. Graphical Location of the Partition Values.** The partition values, viz., quartiles, deciles and percentiles, can be conveniently located with the help of a curve called the 'cumulative frequency curve' or 'Ogive'. The procedure is illustrated below.

First form the cumulative frequency table. Take the class intervals (or the variate values) along the  $x$ -axis and plot the corresponding cumulative frequencies along the  $y$ -axis against the *upper limit* of the class interval (or against the variate value in the case of discrete frequency distribution). The curve obtained on joining

the points so obtained by means of free hand drawing is called the *cumulative frequency curve or ogive*. The graphical location of partition values from this curve is explained below by means of an example.

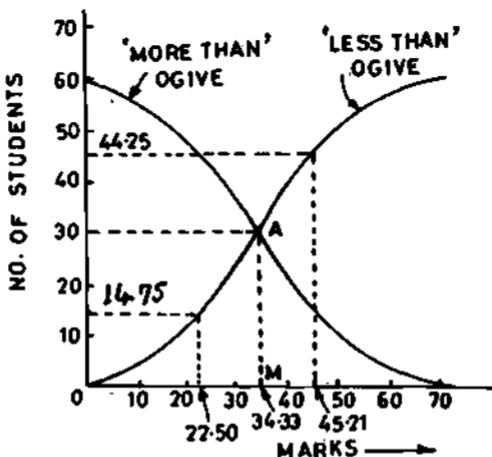
**Example 2.16.** Draw the cumulative frequency curve for the following distribution showing the number of marks of 59 students in Statistics.

Marks-group	0—10	10—20	20—30	30—40	40—50	50—60	60—70
No. of Students	4	8	11	15	12	6	3

**Solution.**

Marks-group	No. of Students	Less than c.f.	More than c.f.
0—10	4	4	59
10—20	8	12	55
20—30	11	23	47
30—40	15	38	36
40—50	12	50	21
50—60	6	56	9
60—70	3	59	3

Taking the marks-group along  $x$ -axis and c.f. along  $y$ -axis, we plot the cumulative frequencies, viz., 4, 12, 23, ..., 59 against the upper limits of the corresponding classes, viz., 10, 20, ..., 70 respectively. The smooth curve obtained on joining these points is called *ogive* or more particularly '*less than*' *ogive*.



If we plot the 'more than' cumulative frequencies, viz., 59, 55, ..., 3 against the lower limits of the corresponding classes, viz., 0, 10, ..., 60 and join the points by a smooth curve, we get cumulative frequency curve which is also known as *ogive* or more particularly '*more than*' *ogive*.

To locate graphically the value of median, mark a point corresponding to  $N/2$  along  $y$ -axis. At this point draw a line parallel to  $x$ -axis meeting the ogive at the point 'A' (say). From 'A' draw a line perpendicular to  $x$ -axis meeting it in 'M' (say). Then abscissa of 'M' gives the value of median.

To locate the values of  $Q_1$  (or  $Q_3$ ), we mark the points along  $y$ -axis corresponding to  $N/4$  (or  $3N/4$ ) and proceed exactly similarly.

In the above example, we get from ogive

$$\text{Median} = 34.33, Q_1 = 22.50, \text{ and } Q_3 = 45.21.$$

**Remarks.** 1. The median can also be located as follows :

From the point of intersection of 'less than' ogive and 'more than' ogive, draw perpendicular to  $OX$ . The abscissa of the point so obtained gives median.

2. Other partition values, viz., deciles and percentiles, can be similarly located from 'ogive'.

### EXERCISE

1. (a) What are grouped and ungrouped frequency distributions? What are their uses? What are the considerations that one has to bear in mind while forming the frequency distribution?

(b) Explain the method of constructing Histogram and Frequency Polygon. Which, out of these two, is better representative of frequencies of (i) a particular group, and (ii) whole group.

2. What are the principles governing the choice of :

- (i) Number of class intervals,
- (ii) The length of the class interval,
- (iii) The mid-point of the class interval.

3. Write short notes on :

- (i) Frequency distribution,
- (ii) Histogram, frequency polygon and frequency curve,
- (iii) Ogive.

4. (a) What are the properties of a good average? Examine these properties with reference to the Arithmetic Mean, the Geometric Mean and the Harmonic Mean, and give an example of situations in which each of them can be the appropriate measure for the average.

(b) Compare mean, median and mode as measures of location of a distribution.

(c) The mean is the most common measure of central tendency of the data. It satisfies almost all the requirements of a good average. The median is also an average, but it does not satisfy all the requirements of a good average. However, it carries certain merits and hence is useful in particular fields. Critically examine both the averages.

(d) Describe the different measures of central tendency of a frequency distribution, mentioning their merits and demerits.

5. Define (i) arithmetic mean, (ii) geometric mean and (iii) harmonic mean of grouped and ungrouped data. Compare and contrast the merits and demerits of them. Show that the geometric mean is capable of further mathematical treatment.

6. (a) When is an average a meaningful statistics? What are the requisites of a satisfactory average? In this light compare the relative merits and demerits of three well-known averages.

(b) What are the chief measures of central tendency? Discuss their merits.

7. Show that (i) Sum of deviations about arithmetic mean is zero.

(ii) Sum of absolute deviations about median is least.

(iii) Sum of the squares of deviations about arithmetic mean is least.

8. The following numbers give the weights of 55 students of a class. Prepare a suitable frequency table.

42	74	40	60	82	115	41	61	75	83	63
53	110	76	84	50	67	65	78	77	56	95
68	69	104	80	79	79	54	73	59	81	100
66	49	77	90	84	76	42	64	69	70	80
72	50	79	52	103	96	51	86	78	94	71

(i) Draw the histogram and frequency polygon of the above data.

(ii) For the above weights, prepare a cumulative frequency table and draw the less than ogive.

9. (a) What are the points to be borne in mind in the formation of frequency table?

Choosing appropriate class-intervals, form a frequency table for the following data:

10.2	0.5	5.2	6.1	3.1	6.7	8.9	7.2	8.9
5.4	3.6	9.2	6.1	7.3	2.0	1.3	6.4	8.0
4.3	4.7	12.4	8.6	13.1	3.2	9.5	7.6	4.0
5.1	8.1	1.1	11.5	3.1	6.8	7.0	8.2	2.0
3.1	6.5	11.2	12.0	5.1	10.9	11.2	8.5	2.3
3.4	5.2	10.7	4.9	6.2				

(b) What are the considerations one has to bear in mind while forming a frequency distribution?

A sample consists of 34 observations recorded correct to the nearest integer, ranging in value from 201 to 337. If it is decided to use seven classes of width 20 integers and to begin the first class at 199.5, find the class limits and class marks of the seven classes.

(c) The class marks in a frequency table (of whole numbers) are given to be 5, 10, 15, 20, 25, 30, 35, 40, 45 and 50. Find out the following :

(i) the true classes.

(ii) the true class limits.

(iii) the true upper class limits.

10. (a) The following table shows the distribution of the number of students per teacher in 750 colleges :-

Students : 1	4	7	10	13	16	19	22	25	28
Frequency : 7	46	165	195	189	89	28	19	9	3

Draw the histogram for the data and superimpose on it the frequency polygon.

(b) Draw the histogram and frequency curve for the following data.

Monthly wages

in Rs.	10-13	13-15	15-17	17-19	19-21	21-23	23-25
No. of workers	6	53	85	56	21	16	8

(c) Draw a histogram for the following data :

Age (in years) : 2-5	5-11	11-12	12-14	14-15	15-16
No. of boys : 6	6	2	5	1	3

11. (a) Three people A, B, C were given the job of finding the average of 5000 numbers. Each one did his own simplification. A's method : Divide the sets into sets of 1000 each, calculate the average in each set and then calculate the average of these averages. B's method : Divide the set into 2,000 and 3,000 numbers, take average in each set and then take the average of the averages. C's method : 500 numbers were unities. He averaged all other numbers and then added one. Are these methods correct?

Ans. Correct, not correct, not correct.

(b) The total sale (in '000 rupees) of a particular item in a shop, on 10 consecutive days, is reported by a clerk as, 35.00, 29.60, 38.00, 30.00, 40.00, 41.00, 42.00, 45.00, 3.60, 3.80. Calculate the average. Later it was found that there was a number 10.00 in the machine and the reports of 4th to 8th days were 10.00 more than the true values and in the last 2 days he put a decimal in the wrong place thus for example 3.60 was really 36.0. Calculate the true mean value.

Ans. 30.8, 32.46.

12. (a) Given below is the distribution of 140 candidates obtaining marks X or higher in a certain examination (all marks are given in whole numbers) :

X : 10	20	30	40	50	60	70	80	90	100
c.f. : 140	133	118	100	75	45	25	9	2	0

Calculate the mean, median and mode of the distribution.

Hint.

Class	Frequency (f)	Class boundaries	Mid value	c.f. (less than)
10-19	$140 - 133 = 7$	9.5-19.5	14.5	7
20-29	$133 - 118 = 15$	19.5-29.5	24.5	22
30-39	$118 - 100 = 18$	29.5-39.5	34.5	40
40-49	$100 - 75 = 25$	39.5-49.5	44.5	65
50-59	$75 - 45 = 30$	49.5-59.5	54.5	95
60-69	$45 - 25 = 20$	59.5-69.5	64.5	115
70-79	$25 - 9 = 16$	69.5-79.5	74.5	131
80-89	$9 - 2 = 7$	79.5-89.5	84.5	138
90-99	$2 - 0 = 2$	89.5-99.5	94.5	140

$$\text{Mean} = 54.5 + \frac{10 \times (-53)}{140} = 50.714$$

$$\text{Median} = 49.5 + \frac{10}{30} \left( \frac{140}{2} - 65 \right) = 51.167$$

(b) The four parts of a distribution are as follows :

Part	Frequency	Mean
1	50	61
2	100	70
3	120	80
4	30	83

Find the mean of the distribution.

(Madurai Univ. B.Sc., 1988)

13. (a) Define a 'weighted mean'. If several sets of observations are combined into a single set, show that the mean of the combined set is the weighted mean of several sets.

(b) The weighted geometric mean of three numbers 229, 275 and 125 is 203. The weights for the first and second numbers are 2 and 4 respectively. Find the weight of third. Ans. 3.

14. Define the weighted arithmetic mean of a set of numbers. Show that it is unaffected if all weights are multiplied by some common factor.

The following table shows some data collected for the regions of a country:

Region	Number of inhabitants (million)	Percentage of literates	Average annual income per person (Rs.)
A	10	52	850
B	5	68	620
C	18	39	730

Obtain the overall figures for the three regions taken together. Prove the formulae you use.

(Calcutta Univ. B.A.(Hons.), 1991)

15. Draw the Ogives and hence estimate the median.

Class	0-9	10-19	20-29	30-39	40-49	50-59	60-69	70-79
Frequency	8	32	142	216	240	206	143	13

16. The following data relate to the ages of a group of workers in a factory.

Ages	No. of workers	Ages	No. of workers
20-25	35	40-45	90
25-30	45	45-50	74
30-35	70	50-55	51
35-40	105	55-60	30

Draw the percentage cumulative curve and find from the graph the number of workers between the ages 28-48.

17. (a) The mean of marks obtained in an examination by a group of 100 students was found to be 49.96. The mean of the marks obtained in the same examination by another group of 200 students was 52.32. Find the mean of the marks obtained by both the groups of students taken together.

(b) A distribution consists of three components with frequencies 300, 200 and 600 having their means 16, 8 and 4 respectively. Find the mean of the combined distribution.

(c) The mean marks got by 300 students in the subject of Statistics are 45. The mean of the top 100 of them was found to be 70 and the mean of the last 100 was known to be 20. What is the mean of the remaining 100 students?

(d) The mean weight of 150 students in a certain class is 60 kilograms. The mean weight of boys in the class is 70 kilograms and that of the girls is 55 kilograms.

Find the number of boys and number of girls in the class.

Ans. (a) 51.53, (b) 8, (c) 45, (d) Boys = 50, Girls = 100.

18. From the following data, calculate the percentage of workers getting wages

(a) more than Rs. 44, (b) between Rs. 22 and Rs. 58, (c) Find  $Q_1$  and  $Q_3$ .

Wages (Rs.)      0–10    10–20    20–30    30–40    40–50    50–60    60–70    70–80

No. of workers    20       45       85       160      70       55       35       30

**Hint.** Assuming that frequencies are uniformly distributed over the entire interval,

(a) Number of persons with wages more than Rs. 44 is

$$\left( \frac{50 - 44}{10} \times 70 \right) + 55 + 35 + 30 = 162$$

Hence the percentage of workers getting over Rs. 44 is

$$= \frac{162}{500} \times 100 = 32.4\%$$

(b) Percentage of workers getting wages between Rs. 22 and Rs. 58 is

$$\left[ \left( \frac{30 - 22}{10} \times 85 \right) + 160 + 70 + \left( \frac{58 - 50}{10} \times 55 \right) \right] \times 100 = 68.4\%$$

19. For the two frequency distributions give below the mean calculated from the first was 25.4 and that from the second was 32.5. Find the values of  $x$  and  $y$ .

Class	Distribution I Frequency	Distribution II Frequency
10–20	20	4
20–30	15	8
30–40	10	4
40–50	$x$	$2x$
50–60	$y$	$y$

Ans.  $x = 3$ ,  $y = 2$

20. A number of particular articles has been classified according to their weights. After drying for two weeks the same articles have again been weighted and similarly classified. It is known that the median weight in the first weighing was 20.83 oz. while in the second weighing it was 17.35 oz. Some frequencies  $a$  and  $b$  in the first weighing and  $x$  and  $y$  in the second are missing. It is known that  $a = \frac{1}{3}x$  and  $b = \frac{1}{2}y$ . Find out the values of the missing frequencies.

Class		Frequencies		Class		Frequencies	
1st weighing		2nd weighing		1st weighing		2nd weighing	
0—5	$a$	$x$		15—20	52		50
5—10	$b$	$y$		20—25	75		30
10—15	11	40		25—30	22		28

Hint. We have  $x = 3a$ ,  $y = 2b$ ,

$$N_1 = \text{Total frequency in 1st weighing} = 160 + a + b.$$

$$N_2 = \text{Total frequency in 2nd weighing} = 148 + x + y = 148 + 3a + 2b.$$

Using Median formula, we shall get

$$\begin{aligned} 20.83 &= 20 + \frac{5}{75} \left[ \frac{N_1}{2} - (63 + a + b) \right] \\ \Rightarrow 15(20.83 - 20) &= \frac{160 + a + b}{2} - (63 + a + b) \\ \Rightarrow 12.45 &= 17 - \frac{a + b}{2} \\ \Rightarrow a + b &= 2(17 - 12.45) = 9.10 \approx 9 \quad \dots(*) \end{aligned}$$

Since  $a$  and  $b$ , being frequencies are integral valued,  $a + b$  is also integral valued. Now the median of 2nd weighing gives :

$$\begin{aligned} 17.35 &= 15 + \frac{5}{50} \left[ \frac{148 + 3a + 2b}{2} - (40 + x + y) \right] \\ \Rightarrow 10 \times 2.35 &= 74 + \frac{3a + 2b}{2} - 40 - 3a - 2b \\ \Rightarrow \frac{3a + 2b}{2} &= 34 - 23.5 = 10.5 \\ \Rightarrow 3a + 2b &= 21 \quad \dots(**) \end{aligned}$$

Multiplying (\*) by 3, we get

$$3a + 3b = 27 \quad \dots(***)$$

Subtracting (\*\*) from (\*\*), we get  $b = 6$ . Substituting in (\*), we get  $a = 9 - 6 = 3$ .

$$\therefore a = 3, b = 6; x = 3a = 9, y = 2b = 12.$$

21. From the following table showing the wage distribution in a certain factory, determine :

- the mean wage,
- the median wage,
- the modal wage,
- the wage limits for the middle 50% of the wage earners,
- the percentages of workers who earned between Rs. 75 and Rs. 125.
- the percentage who earned more than Rs. 150 per week, and
- the percentage who earned less than Rs. 100 per week.

<i>Weekly wages (Rs.)</i>	<i>No. of employees</i>	<i>Weekly wages (Rs.)</i>	<i>No. of employees</i>
20-40	8	120-140	35
40-60	12	140-160	18
60-80	20	160-180	7
80-100	30	180-200	5
100-120	40		

Ans. (a)  $\bar{X} = 108.5$ , (b) Med. = 108.75, (c) Mo = 118.3, (d) 81.25, 129.3  
(e) 48, (f) 12, (g) 40.

22. (a) Explain how the ogives are drawn for any frequency distribution. Point out the method of finding out the values of median, mode, quartiles, deciles and percentiles graphically. Also, write down the formula for the computation of each of them for any frequency distribution.

(b) The following table gives the frequency distribution of marks in a class of 65 students.

<i>Marks</i>	<i>No. of Students</i>	<i>Marks</i>	<i>No. of students</i>
0-4	10	14-18	5
4-8	12	18-20	3
8-12	18	20-25	4
12-14	7	25 and over	6
<i>Total</i>			65

Calculate : (i) Upper and lower quartiles.

(ii) No. of students who secured marks more than 17.

(iii) No. of students who secured marks between 10 and 15.

(c) The following table shows the age distribution of heads of families in a certain country during the year 1957. Find the median, the third quartile and the second decile of the distribution. Check your results by the graphical method.

*Age of head of family*

<i>years</i>	<i>Under 25</i>	<i>25-29</i>	<i>30-34</i>	<i>35-44</i>	<i>45-54</i>	<i>55-64</i>	<i>65-74</i>	<i>above 74</i>
<i>Number</i>	2.3	4.1	5.3	10.6	9.7	6.8	4.4	1.8
<i>(million)</i>								Total 45

Ans. Md = 45.2 yrs.; Q<sub>3</sub> = 57.5 yrs.; D<sub>2</sub> = 32.5 yrs.

23. The following data represent travel expenses (other than transportation) for 7 trips made during November by a salesman for a small firm :

<i>Trip</i>	<i>Days</i>	<i>Expense (Rs.)</i>	<i>Expense per day (Rs.)</i>
1	0.5	13.50	27
2	2.0	12.00	6
3	3.5	17.50	5
4	1.0	9.00	9
5	9.0	27.00	3
6	0.5	9.00	18
7	8.5	17.00	2
<i>Total</i>	25.0	105.00	70

An auditor criticised these expenses as excessive, asserting that the average expense per day is Rs. 10 (Rs. 70 divided by 7). The salesman replied that the average is only Rs. 4.20 (Rs. 105 divided by 25) and that in any event the median is the appropriate measure and is only Rs. 3. The auditor rejoined that the arithmetic mean is the appropriate measure, but that the median is Rs. 6.

You are required to :

- (a) Explain the proper interpretation of each of the four averages mentioned.
- (b) Which average seems appropriate to you ?

24. (a) Define Geometric and Harmonic means and explain their uses in statistical analysis.

You take a trip which entails travelling 900 miles by train at an average speed of 60 m.p.h., 300 miles by boat at an average of 25 m.p.h., 400 miles by plane at 350 m.p.h. and finally 15 miles by taxi at 25 m.p.h. What is your speed for the entire distance?

(b) A train runs 25 miles at a speed of 30 m.p.h., another 50 miles at a speed of 40 m.p.h., then due to repairs of the track travels for 6 minutes at a speed of 10 m.p.h. and finally covers the remaining distance of 24 miles at a speed of 24 m.p.h. What is the average speed in m.p.h.?

(c) A man motors from *A* to *B*. A large part of the distance is uphill and he gets a mileage of only 10 per gallon of gasoline. On the return trip he makes 15 miles per gallon. Find the harmonic mean of his mileage. Verify the fact that this is the proper average to be used by assuming that the distance from *A* to *B* is 60 miles.

(d) Calculate the average speed of a car running at the rate of 15 km.p.h. during the first 30 kms., at 20 km.p.h. during the second 30 kms. and at 25 km.p.h. during the third 30 kms.

25. The following table shows the distribution of 100 families according to their expenditure per week. Number of families corresponding to expenditure groups Rs. (10—20) and Rs.(30—40) are missing from the table. The median and

mode are given to be Rs.25 and 24 Calculate the missing frequencies and then arithmetic mean of the data :

<i>Expenditure</i>	0—10	10—20	20—30	30—40	40—50
<i>No. of families</i>	14	?	27	?	15

**Hint.**

<i>Expenditure</i>	<i>No. of Families</i>	<i>Cumulative frequencies</i>
0—10	14	14
10—20	$f_1$	$14 + f_1$
20—30	27	$41 + f_1$
30—40	$f_2$	$41 + f_1 + f_2$
40—50	15	$56 + f_1 + f_2$

$$\therefore 25 = 20 + \frac{\frac{56 + f_1 + f_2}{2} - (14 + f_1)}{27} \times 10$$

$$\text{and } 24 = 20 + \frac{27 - f_1}{2 \times 27 - f_1 - f_2} \times 10$$

Simpling these equations, we get

$$f_1 - f_2 = 1$$

$$\text{and } 3f_1 - 2f_2 = 27.$$

**Ans.** 25, 24

26. (a) The numbers 3.2, 5.8, 7.9 and 4.5 have frequencies  $x$ ,  $(x+2)$ ,  $(x-3)$  and  $(x+6)$  respectively. If their arithmetic mean is 4.876, find the value of  $x$ .

(b) If  $M_{g,x}$  is the geometric mean of  $N$   $x$ 's and  $M_{g,y}$  is the geometric mean of  $N$   $y$ 's, then the geometric mean  $M_g$  of the  $2N$  values is given by

$$M_g^2 = M_{g,x} M_{g,y}. \quad (\text{Nagpur Univ. B.Sc., 1990})$$

(c) The weighted geometric mean of the three numbers 229, 275 and 125 is 203. The weights for the first and the second numbers are 2 and 4 respectively. Find the weight of the third. **Ans.** 3.

27. The geometric mean of 10 observations on a certain variable was calculated as 16.2. It was later discovered that one of the observations was wrongly recorded as 12.9; in fact it was 21.9. Apply appropriate correction and calculate the correct geometric mean.

**Hint.** Correct value of the geometric mean,  $G'$  is given by

$$G' = \left( \frac{(16.2)^{10} \times 21.9}{12.9} \right)^{1/10} = 17.68$$

28. A variate takes the values  $a, ar, ar^2, \dots, ar^{n-1}$  each with frequency unity. If  $A$ ,  $G$  and  $H$  are respectively the arithmetic mean, geometric mean and harmonic mean, show that

$$A = \frac{a(1-r^n)}{n(1-r)}, G = ar^{(n-1)/2}, H = \frac{an(1-r)r^{n-1}}{(1-r^n)}$$

Prove that  $G^2 = AH$ . Prove also that  $A > G > H$  unless  $r = 1$ , when all the three means coincide.

29. If  $\bar{x}_1 = \frac{1}{n} \sum_{i=1}^n x_i$ ,  $\bar{x}_2 = \frac{1}{n} \sum_{i=2}^{n+1} x_i$  and  $\bar{x}_3 = \frac{1}{n} \sum_{i=3}^{n+2} x_i$

then show that

(a)  $\bar{x}_2 = \bar{x}_1 + \frac{1}{n}(x_{n+1} - x_1)$ , and (b)  $\bar{x}_3 = \bar{x}_2 + \frac{1}{n}(x_{n+2} - x_2)$

30. A distribution  $x_1, x_2, \dots, x_n$  with frequencies  $f_1, f_2, \dots, f_n$  transformed into the distribution  $X_1, X_2, \dots, X_n$  with the same corresponding frequencies by the relation  $X_r = ax_r + b$ , where  $a$  and  $b$  are constants. Show that the mean, median and mode of the new distribution are given in terms of those of the first distribution by the same transformation. [Kanpur Univ. B.Sc., 1992]

Use the method indicated above to find the mean of the following distribution:  $x$  (duration of telephone conversation in seconds)

49.5, 149.5, 249.5, 349.5, 449.5, 549.5, 649.5, 749.5, 849.5, 949.5  
 $f$  (respective frequency)

6      28      88      180      247      260      132      48      11      5

31. If  $\bar{x}_w$  is the weighted mean of  $x_i$ 's with weights  $w_i$ , prove that

$$\left( \sum_{i=1}^n w_i \right) \left( \sum_{i=1}^n w_i (x_i - \bar{x}_w)^2 \right) = \sum_{i=1}^n \sum_{j>i}^n w_i w_j (x_i - x_j)^2, \text{ where } \sum_{i=1}^n w_i \neq 0.$$

[Allahabad Univ. B.Sc., 1992]

Hint.  $\left[ \sum_{i=1}^n \sum_{j>i}^n w_i w_j (x_i - x_j)^2 = \frac{1}{2} \sum_{i=1}^n \sum_{j=i}^n w_i w_j (x_i - x_j)^2 \right]$

$$= \frac{1}{2} \sum_{i=1}^n \cdot \sum_{j=i}^n w_i w_j \left\{ (x_i - \bar{x}_w) - (x_j - \bar{x}_w) \right\}^2$$

32. In a frequency table, the upper boundary of each class interval has a constant ratio to the lower boundary. Show that the geometric mean  $G$  may be expressed by the formula :

$$\log G = x_o + \frac{c}{N} \sum_i f_i (i-1)$$

where  $x_o$  is the logarithm of the mid-value of the first interval and  $c$  is the logarithm of the ratio between upper and lower boundaries.

[Delhi Univ. B.Sc. (Stat. Hons.), 1990, 1986]

33. Find the minimum value of :

- $f(x) = (x - 6)^2 + (x + 3)^2 + (x - 8)^2 + (x + 4)^2 + (x - 3)^2$
- $g(x) = |x - 6| + |x + 3| + |x - 8| + |x + 4| + |x - 3|$ .

[Delhi Univ. B.Sc.(Stat. Hons.), 1991]

**Hint.** The sum of squares of deviations is minimum when taken from arithmetic mean and the sum of absolute deviations is minimum when taken from median.

34. If  $A$ ,  $G$  and  $H$  be the arithmetic mean, geometric mean and harmonic mean respectively of two positive numbers  $a$  and  $b$ , then prove that :

- $A \geq G \geq H$ .

When does the equality sign hold?

- $G^2 = AH$ .

35. Calculate simple and weighted arithmetic averages from the following data and comment on them :

Designation	Monthly salary (in Rs.)	Strength of the cadre
Class I Officers	1,500	10
Class II Officers	800	20
Subordinate staff	500	70
Clerical staff	250	100
Lower staff	100	150

Ans.  $\bar{X} = \text{Rs. } 630$ ,  $\bar{X}_w = \text{Rs. } 302.86$ . Latter is more representative.

36. Treating the number of letters in each word in the following passage as the variable  $x$ , prepare the frequency distribution table and obtain its mean, median, mode.

"The reliability of data must always be examined before any attempt is made to base conclusions upon them. This is true of all data, but particularly so of numerical data, which do not carry their quality written large on them. It is a waste of time to apply the refined theoretical methods of Statistics to data which are suspect from the beginning."

Ans. Mean = 4.565, Median = 4, Mode = 3.

### OBJECTIVE TYPE QUESTIONS

I. Match the correct parts to make a valid statement :

- |                     |  |
|---------------------|--|
| (a) Arithmetic Mean | (i) $l + [f_2/(f_1 + f_2)] \times i$               |
| (b) Geometric Mean  | (ii) $(x_1 \cdot x_2 \cdot \dots \cdot x_n)^{1/n}$ |
| (c) Harmonic Mean   | (iii) $\Sigma f X / \Sigma f$                      |
| (d) Median          | (iv) $l + \frac{N/2 - c.f.}{f} \times i$           |

(e) Mode

$$(v) \left[ \frac{1}{n} \left( \frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n} \right) \right]^{-1}$$

$$(vi) l = \frac{f_1 - f_o}{2f_1 - f_o - f_2} \times i$$

**II.** Which measure of location will be suitable to compare:

- (i) heights of students in two classes;
- (ii) size of agricultural holdings;
- (iii) average sales for various years;
- (iv) intelligence of students;
- (v) per capita income in several countries;
- (vi) sale of shirts with collar size; 16",  $15\frac{1}{2}"$ , 15", 14", 13", 15";
- (vii) marks obtained 10, 8, 12, 4, 7, 11, and  $X$  ( $X < 5$ ).

**Ans.** (i) Mean, (ii) Mode, (iii) Mean, (iv) Median, (v) Mean, (vi) Mode, (vii) Median.

**III.** Which of the following are true for all sets of data?

- (i) Arithmetic Mean  $\leq$  median  $\leq$  mode,
- (ii) Arithmetic mean  $\geq$  median  $\geq$  mode,
- (iii) Arithmetic mean = median = mode
- (iv) None of these.

**IV.** Which of the following are true in respect of any distribution?

- (i) The percentile points are in the ascending order.
- (ii) The percentile points are equispaced.
- (iii) The median is the mid-point of the range and the distribution.
- (iv) A unique median value exists for each and every distribution.

**V.** Find out the missing figures :

- (a) Mean = ? (3 Median - Mode).
- (b) Mean - Mode = ? (Mean - Median).
- (c) Median = Mode + ? (Mean - Mode).
- (d) Mode = Mean - ? (Mean - Median).

**Ans.** (a) 1/2, (b) 3, (c) 2/3, (d) 3.

**VI.** Fill in blanks :

- (i) Harmonic mean of a number of observations is .....
- (ii) The geometric mean of 2, 4, 16, and 32 is .....
- (iii) The strength of 7 colleges in a city are 385; 1,748; 1,343; 1,935; 786; 2,874 and 2,108. Then the median strength is .....
- (iv) The geometric mean of a set of values lies between arithmetic mean and...
- (v) The mean and median of 100 items are 50 and 52 respectively. The value of the largest item is 100. It was later found that it is actually 110. Therefore, the true mean is ... and the true median is .....

- (vi) The algebraic sum of the deviations of 20 observations measured from 30 is 2. Therefore, mean of these observations is ....  
 (vii) The relationship between A.M., G.M. and H.M. is ....  
 (viii) The mean of 20 observations is 15. On checking it was found that two observations were wrongly copied as 3 and 6. If wrong observations are replaced by correct values 8 and 4, then the correct mean is ....

(ix) Median = .... Quartile.

(x) Median is the average suited for .... classes.

(xi) A distribution with two modes is called .... and with more than two modes is called ....

(xii) .... is not affected by extreme observations.

Ans. (ii) 8 ; (iii) 1,748 ; (iv) H.M. ; (v) 50·1, 52 ; (vi) 30·1 ;  
 (vii) A.M.  $\geq$  G.M.  $\geq$  H.M. ; (viii) 15·15 ; (ix) Second ; (x) Open end ; (xi) Bimodal, multimodal ; (xii) Median or mode.

### VII. For the questions given below, give correct answers.

(i) The algebraic sum of the deviations of a set of  $n$  values from their arithmetic mean is

(a)  $n$ , (b) 0, (c) 1, (d) none of these.

(ii) The most stable measure of central tendency is

(a) the mean, (b) the median, (c) the mode, (d) none of these.

(iii) 10 is the mean of a set of 7 observations and 5 is the mean of a set of 3 observations. The mean of a combined set is given by

(a) 15, (b) 10, (c) 8·5, (d) 7·5, (e) none of these.

(iv) The mean of the distribution, in which the value of  $x$  are 1, 2, ...,  $n$ , the frequency of each being unity is:

(a)  $n(n+1)/2$ , (b)  $n/2$ , (c)  $(n+1)/2$ , (d) none of these.

(v) The arithmetic mean of the numbers 1, 2, 3, ...,  $n$  is

(a)  $\frac{n(n+1)(2n+1)}{6}$ , (b)  $\frac{n(n+1)^2}{4}$ , (c)  $\frac{n(n+1)}{2}$ , (d) none of these.

(ii) The most stable measure of central tendency

(vi) The point of intersection of the 'less than' and the 'greater than' ogive corresponds to

(a) the mean, (b) the median, (c) the geometric mean, (d) none of these.

(vii) When  $x_i$  and  $y_i$  are two variables ( $i = 1, 2, \dots, n$ ) with G.M.'s  $G_1$  and  $G_2$  respectively then the geometric mean of  $\left(\frac{x_i}{y_i}\right)$  is

(a)  $\frac{G_1}{G_2}$ , (b) antilog  $\left(\frac{G_1}{G_2}\right)$ , (c)  $n(\log G_1 - \log G_2)$ ;

$$(d) \text{ Antilog} \left( \frac{\log G_1 - \log G_2}{2n} \right)$$

Ans. (i) (b); (ii) (a); (iii) (c); (iv) (c); (v) (d); (vi) (b); (vii) (a).

**VIII.** State which of the following statements are True and which are False. In case of false statements give the correct statement.

(i) The harmonic mean of  $n$  numbers is the reciprocal of the Arithmetic mean of the reciprocals of the numbers.

(ii) For the wholesale manufacturers interested in the type which is usually in demand, median is the most suitable average.

(iii) The algebraic sum of the deviations of a series of individual observations from their mean is always zero.

(iv) Geometric mean is the appropriate average when emphasis is on the rate of change rather than the amount of change.

(v) Harmonic mean becomes zero when one of the items is zero.

(vi) Mean lies between median and mode.

(vii) Cumulative frequency is not-decreasing.

(viii) Geometric mean is the arithmetic mean of harmonic mean and arithmetic mean.

(ix) Mean, median mode have the same unit.

(x) One quintal of wheat was purchased at 0.8 kg. per rupee and another quintal at 1.2 kg. per rupee. The average rate per rupee is 1kg.

(xi) One limitation of the median is that it cannot be calculated from a frequency distribution with open-end classes.

(xii) The arithmetic mean of a frequency distribution is always located in the class which has the greatest number of frequencies.

(xiii) In a moderately asymmetrical distribution, the mean, median and mode are the same.

(xiv) It is really immaterial in which class an item falling at the boundary between two classes is listed.

(xv) The median is not affected by extreme items.

(xvi) The median is the point about which the sum of squared deviations is minimum.

(xvii) In construction of the frequency distribution, the selection of the class interval is arbitrary.

(xviii) Usual attendance of B.Sc. class is 35 per day. So for 100 working days total attendance is 3,500.

(xix) A car travels 100 miles at a speed of 40 m.p.h. and another 400 miles at a speed of 30 m.p.h. So the average speed for the whole journey is either 35 m.p.h. or 33 m.p.h.

(xx) In calculating the mean for grouped data, the assumption is made that the mean of the items in each class is equal to the mid-value of the class.

(xxi) The geometric mean of a group of numbers is less than the arithmetic mean in all cases, except in the special case in which the numbers are all the same.

(xxii) The geometric mean equals the antilog of the arithmetic mean of the logs of the values.

(xxiii) The median may be considered more typical than the mean because the median is not affected by the size of the extremes.

(xxiv) The Harmonic Mean of a series of fractions is the same as the reciprocal of the arithmetic mean of the series.

(xxv) In a frequency distribution the true value of mode cannot be calculated exactly.

**IX.** In each of the following cases, explain whether the description applies to mean, median or both.

(i) it can be calculated from a frequency distribution with open-end classes.

(ii) the values of all items are taken into consideration in the calculation.

(iii) the values of extreme items do not influence the average.

(iv) In a distribution with a single peak and moderate skewness to the right it is closer to the concentration of the distribution.

**Ans.** (i) median, (ii) mean, (iii) median, (iv) median.

**X. Be brief in your answer :**

(a) The production in an industrial unit was 10,000 units during 1981 and in 1980 the production was 25,000 units. Hence the production has declined by 150 percent. Comment.

(b) A man travels by a car for 4 days. He travelled for 10 hours each day. He drove on the first day at the rate of 45 km per hour, second day at 40 km. per hour, third day at the rate of 38 km. per hour and the fourth day at the rate of 37 km. per hour.

Which average, harmonic mean or arithmetic mean or median will give us his average speed? Why?

(c) It is seen from records that a country does not export more than 5 % of its total production. Hence export trade is not vital to the economy of that country. Is the conclusion right?

(d) A survey revealed that the children of engineers, doctors and lawyers have high intelligence quotients. It further revealed that the grandfathers of these children were also highly intelligent. Hence the inference is that intelligence is hereditary. Do you agree?

**XI.** Do you agree with the following interpretations made on the basis of the facts given. Explain briefly your answer.

(a) The number of deaths in military in the recent war was 10 out of 1,000 while the number of deaths in Hyderabad in the same period was 18 per 1,000. Hence it is safe to join military service than to live in the city of Hyderabad.

(b) The examination result in a college  $X$  was 70% in the year 1991. In the same year and at the same examination only 500 out of 750 students were successful in college  $Y$ . Hence the teaching standard in college  $X$  was better.

(c) The average daily production in a small-scale factory in January 1991 was 4,000 candles and 3,800 candles in February 1981. So the workers were more efficient in January.

(d) The increase in the price of a commodity was 25%. Then the price decreased by 20% and again increased by 10%. So the resultant increase in the price was  $25 - 20 + 10 = 15\%$ .

(e) The rate of tomato in the first week of January was 2 kg. for a rupee and in the 2nd week was 4 kg. for a rupee. Hence the average price of tomato is  $\frac{1}{2}(2 + 4) = 3$  kg. for a rupee.

**XII.** (a) The mean mark of 100 students was given to be 40. It was found later that a mark 53 was read as 83. What is the corrected mean mark?

(b) The mean salary paid to 1,000 employees of an establishment was found to be Rs. 108.40. Later on, after disbursement of salary it was discovered that the salary of two employees was wrongly entered as Rs. 297 and Rs. 165. Their correct salaries were Rs. 197 and Rs. 185. Find the correct arithmetic mean.

(c) Twelve persons gambled on a certain night. Seven of them lost at an average rate of Rs. 10.50 while the remaining five gained at an average of Rs. 13.00. Is the information given above correct? If not, why?

## CHAPTER THREE

# Measures of Dispersion, Skewness and Kurtosis

**3-1. Dispersion.** Averages or the measures of central tendency give us an idea of the concentration of the observations about the central part of the distribution. If we know the average alone we cannot form a complete idea about the distribution as will be clear from the following example.

Consider the series (i) 7, 8, 10, 11, (ii) 3, 6, 9, 12, 15, (iii) 1, 5, 9, 13, 17. In all these cases we see that  $n$ , the number of observations is 5 and the mean is 9. If we are given that the mean of 5 observations is 9, we cannot form an idea as to whether it is the average of first series or second series or third series or of any other series of 5 observations whose sum is 45. Thus we see that the measures of central tendency are inadequate to give us a complete idea of the distribution. They must be supported and supplemented by some other measures. One such measure is *Dispersion*.

Literal meaning of dispersion is 'scatteredness'. We study dispersion to have an idea about the homogeneity or heterogeneity of the distribution. In the above case we say that series (i) is more homogeneous (less dispersed) than the series (ii) or (iii) or we say that series (iii) is more heterogeneous (more scattered) than the series (i) or (ii).

**3-2. Characteristics for an Ideal Measure of Dispersion.** The desiderata for an ideal measure of dispersion are the same as those for an ideal measure of central tendency, viz.,

- (i) It should be rigidly defined.
- (ii) It should be easy to calculate and easy to understand.
- (iii) It should be based on all the observations.
- (iv) It should be amenable to further mathematical treatment.
- (v) It should be affected as little as possible by fluctuations of sampling.

**3-3. Measures of Dispersion.** The following are the measures of dispersion:

- (i) Range,
- (ii) Quartile deviation or Semi-interquartile range,
- (iii) Mean deviation, and
- (iv) Standard deviation.

**3-4. Range.** The range is the difference between two extreme observations of the distribution. If  $A$  and  $B$  are the greatest and smallest observations respectively in a distribution, then its range is  $A - B$ .

Range is the simplest but a crude measure of dispersion. Since it is based on two extreme observations which themselves are subject to chance fluctuations, it is not at all a reliable measure of dispersion.

**3-5. Quartile Deviation.** Quartile deviation or semi-interquartile range

$Q$  is given by

$$Q = \frac{1}{2}(Q_3 - Q_1), \quad \dots(3-1)$$

where  $Q_1$  and  $Q_3$  are the first and third quartiles of the distribution respectively.

Quartile deviation is definitely a better measure than the range as it makes use of 50% of the data. But since it ignores the other 50% of the data, it cannot be regarded as a reliable measure.

**3-6. Mean Deviation.** If  $x_i | f_i, i = 1, 2, \dots, n$  is the frequency distribution, then mean deviation from the average  $A$ , (usually mean, median or mode), is given by

$$\text{Mean deviation} = \frac{1}{N} \sum_i f_i |x_i - A|, \quad \sum f_i = N \quad \dots(3-2)$$

where  $|x_i - A|$  represents the modulus or the absolute value of the deviation  $(x_i - A)$ , when the -ive sign is ignored.

Since mean deviation is based on all the observations, it is a better measure of dispersion than range or quartile deviation. But the step of ignoring the signs of the deviations  $(x_i - A)$  creates artificiality and renders it useless for further mathematical treatment.

It may be pointed out here that mean deviation is least when taken from median. (The proof is given for continuous variable in Chapter 5)

**3-7. Standard Deviation and Root Mean Square Deviation.** Standard deviation, usually denoted by the Greek letter small sigma ( $\sigma$ ), is the positive square root of the arithmetic mean of the squares of the deviations of the given values from their arithmetic mean. For the frequency distribution  $x_i | f_i, i = 1, 2, \dots, n$ ,

$$\sigma = \sqrt{\frac{1}{N} \sum_i f_i (x_i - \bar{x})^2} \quad \dots(3-3)$$

where  $\bar{x}$  is the arithmetic mean of the distribution and  $\sum f_i = N$ .

The step of squaring the deviations  $(x_i - \bar{x})$  overcomes the drawback of ignoring the signs in mean deviation. Standard deviation is also suitable for further mathematical treatment (§ 3-7-3). Moreover of all the measures, standard deviation is affected least by fluctuations of sampling.

Thus we see that standard deviation satisfies almost all the properties laid down for an ideal measure of dispersion except for the general nature of extracting the square root which is not readily comprehensible for a non-mathematical person. It may also be pointed out that standard deviation gives greater weight to extreme values and as such has not found favour with economists or businessmen who are more interested in the results of the modal class. Taking into consideration the pros and cons and also the wide applications of standard deviation in statistical theory, we may regard standard deviation as the best and the most powerful measure of dispersion!

The square of standard deviation is called the variance and is given by

$$\sigma^2 = \frac{1}{N} \sum_i f_i (x_i - \bar{x})^2 \quad \dots(3.3a)$$

**Root mean square deviation**, denoted by 's' is given by

$$s = \sqrt{\frac{1}{N} \sum_i f_i (x_i - A)^2} \quad \dots(3.4)$$

where  $A$  is any arbitrary number.  $s^2$  is called mean square deviation.

**3.7.1. Relation between  $\sigma$  and  $s$ .** By definition, we have

$$\begin{aligned} s^2 &= \frac{1}{N} \sum_i f_i (x_i - A)^2 = \frac{1}{N} \sum_i f_i (x_i - \bar{x} + \bar{x} - A)^2 \\ &= \frac{1}{N} \sum_i f_i [ (x_i - \bar{x})^2 + (\bar{x} - A)^2 + 2(\bar{x} - A)(x_i - \bar{x}) ] \\ &= \frac{1}{N} \sum_i f_i (x_i - \bar{x})^2 + (\bar{x} - A)^2 \frac{1}{N} \sum_i f_i + 2(\bar{x} - A) \sum_i f_i (x_i - \bar{x}), \end{aligned}$$

$(\bar{x} - A)$ , being constant is taken outside the summation sign. But  $\sum_i f_i (x_i - \bar{x}) = 0$ , being the algebraic sum of the deviations of the given values from their mean. Thus

$$s^2 = \sigma^2 + (\bar{x} - A)^2 = \sigma^2 + d^2, \text{ where } d = \bar{x} - A$$

Obviously  $s^2$  will be least when  $d = 0$ , i.e.,  $\bar{x} = A$ . Hence mean square deviation and consequently root mean square deviation is least when the deviations are taken from  $A = \bar{x}$ , i.e., standard deviation is the least value of root mean square deviation.

The same result could be obtained alternatively as follows:

**Mean square deviation is given by**

$$s^2 = \frac{1}{N} \sum_i f_i (x_i - A)^2$$

It has been shown in § 2.5.1 Property 2 that  $\sum_i f_i (x_i - A)^2$  is minimum when  $A = \bar{x}$ . Thus mean square deviation is minimum when  $A = \bar{x}$  and its minimum value is

$$(s^2)_{\min} = \frac{1}{N} \sum_i f_i (x_i - \bar{x})^2 = \sigma^2$$

Hence variance is the minimum value of mean square deviation or standard deviation is the minimum value of root mean square deviation.

**3.7.2. Different Formulae For Calculating Variance.** By definition, we have

$$\sigma^2 = \frac{1}{N} \sum_i f_i (x_i - \bar{x})^2$$

More precisely we write it as  $\sigma_x^2$ , i.e., variance of  $x$ . Thus

$$\sigma_x^2 = \frac{1}{N} \sum_i f_i (x_i - \bar{x})^2 \quad \dots(3.5)$$

If  $\bar{x}$  is not a whole number but comes out to be in fractions, the calculation of  $\sigma_x^2$  by using (3.5) is very cumbersome and time consuming. In order to overcome this difficulty, we shall develop different forms of the formula (3.5) which reduce the arithmetic to a great extent and are very useful for computational work. In the following sequence the summation is extended over  $i$  from 1 to  $n$ .

$$\begin{aligned}\sigma_x^2 &= \frac{1}{N} \sum_i f_i (x_i - \bar{x})^2 = \frac{1}{N} \sum_i f_i (x_i^2 + \bar{x}^2 - 2x_i \bar{x}) \\ &= \frac{1}{N} \sum_i f_i x_i^2 + \bar{x}^2 - \frac{1}{N} \sum_i f_i \cdot 2\bar{x} \cdot \frac{1}{N} \sum_i f_i x_i \\ &= \frac{1}{N} \sum_i f_i x_i^2 + \bar{x}^2 - 2\bar{x}^2 = \frac{1}{N} \sum_i f_i x_i^2 - \bar{x}^2 \quad \dots(3.6)\end{aligned}$$

$$\Rightarrow \sigma_x^2 = \frac{1}{N} \sum_i f_i x_i^2 - \left( \frac{1}{N} \sum_i f_i x_i \right)^2 \quad \dots(3.6a)$$

If the values of  $x$  and  $f$  are large the calculation of  $fx, fx^2$  is quite tedious. In that case we take the deviations from any arbitrary point 'A'. Generally the point in the middle of the distribution is much convenient though the formula is true in general. We have

$$\begin{aligned}\sigma_x^2 &= \frac{1}{N} \sum_i f_i (x_i - \bar{x})^2 = \frac{1}{N} \sum_i f_i (x_i - A + A - \bar{x})^2 \\ &= \frac{1}{N} \sum_i f_i (d_i + A - \bar{x})^2, \text{ where } d_i = x_i - A. \\ \sigma_x^2 &= \frac{1}{N} \sum_i f_i [ d_i^2 + (A - \bar{x})^2 + 2(A - \bar{x}) d_i ] \\ &= \frac{1}{N} \sum_i f_i d_i^2 + (A - \bar{x})^2 + 2(A - \bar{x}) \cdot \frac{1}{N} \sum_i f_i d_i\end{aligned}$$

We know that if  $d_i = x_i - A$  then  $\bar{x} = A + \frac{1}{N} \sum_i f_i d_i$

$$\therefore A - \bar{x} = -\frac{1}{N} \sum_i f_i d_i$$

Hence

$$\begin{aligned}\sigma_x^2 &= \frac{1}{N} \sum_i f_i d_i^2 + \left( -\frac{1}{N} \sum_i f_i d_i \right)^2 + 2 \left( -\frac{1}{N} \sum_i f_i d_i \right) \left( \frac{1}{N} \sum_i f_i d_i \right) \\ &= \frac{1}{N} \sum_i f_i d_i^2 - \left( \frac{1}{N} \sum_i f_i d_i \right)^2 \quad \dots(3.7) \\ \Rightarrow \sigma_x^2 &= \sigma_d^2 \quad [\text{On comparison with (3.6a)}]\end{aligned}$$

Hence variance and consequently standard deviation is independent of change of origin.

If we take  $d_i = (x_i - A)/h$  so that  $(x_i - A) = hd_i$ , then

$$\begin{aligned}\sigma_x^2 &= \frac{1}{N} \sum_i f_i (x_i - \bar{x})^2 = \frac{1}{N} \sum_i f_i (x_i - A + A - \bar{x})^2 \\&= \frac{1}{N} \sum_i f_i (hd_i + A - \bar{x})^2 \\&= h^2 \frac{1}{N} \sum_i f_i d_i^2 + (A - \bar{x})^2 + 2(A - \bar{x}) \cdot h \cdot \frac{1}{N} \sum_i f_i d_i\end{aligned}$$

Using  $\bar{x} = A + h \frac{\sum f_i d_i}{N}$ , we get

$$\sigma_x^2 = h^2 \left[ \frac{1}{N} \sum_i f_i d_i^2 - \left( \frac{1}{N} \sum_i f_i d_i \right)^2 \right] = h^2 \sigma_d^2, \quad \dots(3-8)$$

which shows that variance is not independent of change of scale.

Aliter. If  $d_i = \frac{x_i - A}{h}$ , then

$$x_i = A + hd_i \quad \text{and} \quad \bar{x} = A + h \cdot \frac{1}{N} \sum_i f_i d_i = A + h \bar{d}$$

Obviously

$$x_i - \bar{x} = h(d_i - \bar{d})$$

$$\therefore \sigma_x^2 = \frac{1}{N} \sum_i f_i (x_i - \bar{x})^2 = h^2 \cdot \frac{1}{N} \sum_i f_i (d_i - \bar{d})^2 = h^2 \sigma_d^2$$

Hence variance is independent of change of origin but not of scale.

**Example 3-1.** Calculate the mean and standard deviation for the following table giving the age distribution of 542 members.

Age in years :	20—30	30—40	40—50	50—60	60—70	70—80	80—90
No. of members :	3	61	132	153	140	51	2

**Solution.** Here we take  $d = \frac{x - A}{h} = \frac{x - 55}{10}$ .

Age group	Mid-value (x)	Frequency (f)	$d = \frac{x - 55}{10}$	$fd$	$fd^2$
20—30	25	3	-3	-9	27
30—40	35	61	-2	-122	244
40—50	45	132	-1	-132	132
50—60	55	153	0	0	0
60—70	65	140	1	140	140
70—80	75	51	2	102	204
80—90	85	2	3	6	18
		$N = \sum f = 542$		$\sum fd = -15$	$\sum fd^2 = 765$

$$\bar{x} = A + h \frac{\sum fd}{N} = 55 + \frac{10 \times (-15)}{542} = 55 - 0.28 = 54.72 \text{ years.}$$

$$\sigma^2 = h^2 \left[ \frac{1}{N} \sum f d^2 - \left( \frac{1}{N} \sum f d \right)^2 \right] = 100 \left[ \frac{765}{542} - (0.28)^2 \right]$$

$$= 100 \times 1.333 = 133.3$$

$$\therefore \sigma (\text{standard deviation}) = 11.55 \text{ years}$$

**Example 3-2.** Prove that for any discrete distribution standard deviation is not less than mean deviation from mean.

[Delhi Univ. B.Sc. (Stat. Hons.), 1989]

**Solution.** Let  $x_i | f_i, i = 1, 2, 3, \dots, n$  be any discrete distribution. Then we have to prove that

S.D.  $\nless$  Mean deviation from mean

$$\Rightarrow (\text{S.D.})^2 \nless (\text{Mean deviation from mean})^2$$

$$\Rightarrow (\text{S.D.})^2 \geq (\text{M. D. from mean})^2$$

$$\Rightarrow \frac{1}{N} \sum_{i=1}^n f_i (x_i - \bar{x})^2 \geq \left( \frac{1}{N} \sum_{i=1}^n f_i |x_i - \bar{x}| \right)^2$$

If we put  $|x_i - \bar{x}| = z_i$ , then we have to prove that

$$\frac{1}{N} \sum_{i=1}^n f_i z_i^2 \geq \left( \frac{1}{N} \sum_{i=1}^n f_i z_i \right)^2$$

$$\text{i.e., } \frac{1}{N} \sum_{i=1}^n f_i z_i^2 - \left( \frac{1}{N} \sum_{i=1}^n f_i z_i \right)^2 \geq 0$$

$$\text{i.e., } \frac{1}{N} \sum_{i=1}^n f_i (z_i - \bar{z})^2 \geq 0$$

$$\text{i.e., } \sigma_z^2 \geq 0,$$

which is always true. Hence the result.

**Example 3-3.** Find the mean deviation from the mean and standard deviation of A.P.  $a, a+d, a+2d, \dots, a+2nd$  and verify that the latter is greater than the former.

[Delhi Univ. B.Sc. (Stat. Hons.), 1990]

**Solution.** We know that the mean of a series in A.P. is the mean of its first and last term. Hence the mean of the given series is

$$\bar{x} = \frac{1}{2} (a + a + 2nd) = a + nd$$

$x$	$ x - \bar{x} $	$(x - \bar{x})^2$
$a$	$nd$	$n^2 d^2$
$a+d$	$(n-1)d$	$(n-1)^2 d^2$
$a+2d$	$(n-2)d$	$(n-2)^2 d^2$
$\vdots$	$\vdots$	$\vdots$
$a+(n-2)d$	$2d$	$2^2 \cdot d^2$
$a+(n-1)d$	$d$	$1^2 \cdot d^2$
$a+nd$	$0$	$0$
$a+(n+1)d$	$d$	$1^2 \cdot d^2$
$a+(n+2)d$	$2d$	$2^2 \cdot d^2$

$$\begin{array}{c} \vdots \\ a + (2n - 2)d \\ a + (2n - 2)d \\ a + 2nd \end{array}$$

$$\begin{array}{c} \vdots \\ (n - 2)d \\ (n - 1)d \\ nd \end{array}$$

$$\begin{array}{c} \vdots \\ (n - 2)^2 d^2 \\ (n - 2)^2 d^2 \\ n^2 d^2 \end{array}$$

$$\text{Mean deviation from mean} = \frac{1}{2n+1} \sum |x - \bar{x}|$$

$$\begin{aligned} &= \frac{1}{2n+1} 2d(1+2+3+\dots+n) \\ &= \frac{n(n+1)d}{(2n+1)} \end{aligned}$$

$$\sigma^2 = \frac{1}{2n+1} \sum (x - \bar{x})^2 = \frac{1}{2n+1} 2d^2 (1^2 + 2^2 + 3^2 + \dots + n^2)$$

$$= \frac{1}{2n+1} 2d^2 \cdot \frac{n(n+1)(2n+1)}{6} = \frac{n(n+1)d^2}{3}$$

Hence standard deviation

$$\sigma = \sqrt{\frac{n(n+1)}{3}} \times d$$

### Verification.

S.D. > M.D. from mean

$$\text{if } (\text{S.D.})^2 > (\text{M.D. from mean})^2$$

$$\text{i.e., if } \frac{n(n+1)d^2}{3} > \left( \frac{n(n+1)d}{2n+1} \right)^2$$

$$\text{or if } (2n+1)^2 > 3n(n+1)$$

$$\text{or if } n^2 + n + 1 > 0$$

$$\text{or if } (n + \frac{1}{2})^2 + \frac{3}{4} > 0$$

which is always true.

**Example 3-4.** Show that in a discrete series if deviations are small compared with mean  $M$  so that  $(x/M)^3$  and higher powers of  $(x/M)$  are neglected, we have

$$(i) G = M \left( 1 - \frac{1}{2} \cdot \frac{\sigma^2}{M^2} \right),$$

$$(ii) M^2 - G^2 = \sigma^2, \text{ and } (iii) H = M \left( 1 - \frac{\sigma^2}{M^2} \right),$$

where  $M$  is the arithmetic mean,  $G$ , the geometric mean,  $H$ , the harmonic mean and  $\sigma$  is the standard deviation of the distribution.

**Solution.** Let  $X_i | f_i, i = 1, 2, \dots, n$  be the given frequency distribution. Then we are given that  $x_i = X_i - M$ , i.e.,  $X_i = x_i + M$  where  $M$  is the mean of

the distribution. We have

$$\sum_i f_i x_i = \sum_i f_i (X_i - M) = 0, \quad \dots(1)$$

being the algebraic sum of the deviations of the given values from their mean. Also

$$\sum_i f_i x_i^2 = \sum_i f_i (X_i - M)^2 = \sigma^2 \quad \dots(2)$$

(i) By definition, we have

$$\begin{aligned} G &= (X_1^{f_1} \cdot X_2^{f_2} \cdots X_n^{f_n})^{1/N}, \text{ where } N = \sum f_i \\ \log G &= \frac{1}{N} \sum_i f_i \log X_i = \frac{1}{N} \sum_i f_i \log (x_i + M) \\ &= \frac{1}{N} \sum_i \left\{ f_i \log \left[ M \left( 1 + \frac{x_i}{M} \right) \right] \right\} \\ &= \frac{1}{N} \sum_i \left\{ f_i \left[ \log M + \log \left( 1 + \frac{x_i}{M} \right) \right] \right\} \\ &\approx \log M + \frac{1}{N} \sum_i f_i \log \left( 1 + \frac{x_i}{M} \right) \\ &= \log M + \frac{1}{N} \sum_i f_i \left[ \frac{x_i}{M} - \frac{1}{2} \frac{x_i^2}{M^2} + \frac{1}{3} \left( \frac{x_i}{M} \right)^3 + \dots \right], \end{aligned}$$

the expansion of  $\log \left( 1 + \frac{x_i}{M} \right)$  in ascending powers of  $(x_i/M)$  being valid since  $|x_i/M| < 1$ . Neglecting  $(x_i/M)^3$  and higher powers of  $(x_i/M)$ , we get

$$\begin{aligned} \log G &= \log M + \frac{1}{NM} \sum_i f_i x_i - \frac{1}{2M^2} \cdot \frac{1}{N} \sum_i f_i x_i^2 \\ &= \log M - \frac{\sigma^2}{2M^2}, \quad \text{(On using (1) and (2))} \\ &= \log \left( M e^{-\sigma^2/2M^2} \right) \\ \Rightarrow G &= M e^{-\sigma^2/2M^2} = M \left( 1 - \frac{\sigma^2}{2M^2} \right), \end{aligned}$$

neglecting higher powers.

$$\text{Hence } G = M \left( 1 - \frac{1}{2} \cdot \frac{\sigma^2}{M^2} \right). \quad \dots(3)$$

(ii) Squaring both sides in (3), we get

$$G^2 = M^2 \left( 1 - \frac{1}{2} \cdot \frac{\sigma^2}{M^2} \right)^2 = M^2 \left( 1 - \frac{\sigma^2}{M^2} \right) = M^2 - \sigma^2,$$

neglecting  $(\sigma/M)^4$ .

$$\therefore M^2 - G^2 = \sigma^2 \quad \dots(4)$$

(iii) By definition, harmonic mean  $H$  is given by

$$\begin{aligned} \frac{1}{H} &= \frac{1}{N} \sum_i (f_i/X_i) = \frac{1}{N} \sum_i [f_i/(x_i + M)] \\ &= \frac{1}{MN} \sum_i \frac{f_i}{[1 + (x_i/M)]} = \frac{1}{MN} \sum_i f_i \left(1 + \frac{x_i}{M}\right)^{-1} \end{aligned}$$

Since  $\left|\frac{x_i}{M}\right| < 1$ , the expansion of  $\left(1 + \frac{x_i}{M}\right)^{-1}$  in ascending powers of  $(x_i/M)$  is valid. Neglecting  $(x_i/M)^3$  and higher powers of  $(x_i/M)$ , we get

$$\begin{aligned} \frac{1}{H} &= \frac{1}{MN} \sum_i f_i \left(1 - \frac{x_i}{M} + \frac{x_i^2}{M^2}\right) \\ &= \frac{1}{M} \left( \frac{1}{N} \sum_i f_i - \frac{1}{MN} \sum_i f_i x_i + \frac{1}{M^2} \frac{1}{N} \sum_i f_i x_i^2 \right) \\ &= \frac{1}{M} \left( 1 + \frac{\sigma^2}{M^2} \right)^{-1} \quad [\text{On using (1) and (2)}] \\ \therefore H &= M \left( 1 + \frac{\sigma^2}{M^2} \right)^{-1} = M \left( 1 - \frac{\sigma^2}{M^2} \right), \end{aligned}$$

higher powers being neglected.

$$\text{Hence } H = M \left( 1 - \frac{\sigma^2}{M^2} \right) \quad \dots(5)$$

**Example 3-5.** For a group of 200 candidates, the mean and standard deviation of scores were found to be 40 and 15 respectively. Later on it was discovered that the scores 43 and 35 were misread as 34 and 53 respectively. Find the corrected mean and standard deviation corresponding to the corrected figures.

**Solution.** Let  $x$  be the given variable. We are given  $n = 200$ ,  $\bar{x} = 40$  and  $\sigma = 15$

$$\text{Now } \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \Rightarrow \sum_i x_i = n \bar{x} = 200 \times 40 = 8000$$

$$\text{Also } \sigma^2 = \frac{1}{n} \sum_i x_i^2 - \bar{x}^2$$

$$\therefore \sum_i x_i^2 = n(\sigma^2 + \bar{x}^2) = 200(225 + 1600) = 365000$$

$$\text{Corrected } \sum_i x_i = 8000 - 34 - 53 + 43 + 35 = 7991$$

$$\text{and } \text{Corrected } \sum_i x_i^2 = 365000 - (34)^2 - (53)^2 + (43)^2 + (35)^2 = 364109$$

$$\text{Hence, } \text{Corrected mean} = \frac{7991}{200} = 39.955$$

$$\text{Corrected } \sigma^2 = \frac{364109}{200} - (39.955)^2 = 1820.54 - 1596.40 = 224.14$$

$\therefore \text{Corrected standard deviation} = 14.97$

**3-7-3. Theorem.** (*Variance of the combined series*). If  $n_1, n_2$  are the sizes;  $\bar{x}_1, \bar{x}_2$  the means, and  $\sigma_1, \sigma_2$  the standard deviations of two series, then the standard deviation  $\sigma$  of the combined series of size  $n_1 + n_2$  is given by

$$\sigma^2 = \frac{1}{n_1 + n_2} [n_1(\sigma_1^2 + d_1^2) + n_2(\sigma_2^2 + d_2^2)] \quad \dots(3.9)$$

where  $d_1 = \bar{x}_1 - \bar{x}$ ,  $d_2 = \bar{x}_2 - \bar{x}$

and  $\bar{x} = \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2}{n_1 + n_2}$ , is the mean of the combined series.

**Proof.** Let  $x_{1i}; i = 1, 2, \dots, n_1$  and  $x_{2j}; j = 1, 2, \dots, n_2$ , be the two series then

$$\left. \begin{array}{l} \bar{x}_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} x_{1i} \\ \bar{x}_2 = \frac{1}{n_2} \sum_{j=1}^{n_2} x_{2j} \end{array} \right\} \dots(*) \quad \text{and} \quad \left. \begin{array}{l} \sigma_1^2 = \frac{1}{n_1} \sum_{i=1}^{n_1} (x_{1i} - \bar{x}_1)^2 \\ \sigma_2^2 = \frac{1}{n_2} \sum_{j=1}^{n_2} (x_{2j} - \bar{x}_2)^2 \end{array} \right\} \dots(**)$$

The mean  $\bar{x}$  of the combined series is given by

$$\bar{x} = \frac{1}{n_1 + n_2} \left[ \sum_{i=1}^{n_1} x_{1i} + \sum_{j=1}^{n_2} x_{2j} \right] = \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2}{n_1 + n_2} \quad [\text{From } (*)]$$

The variance  $\sigma^2$  of the combined series is given by

$$\sigma^2 = \frac{1}{n_1 + n_2} \left[ \sum_{i=1}^{n_1} (x_{1i} - \bar{x})^2 + \sum_{j=1}^{n_2} (x_{2j} - \bar{x})^2 \right] \quad \dots(3.10)$$

Now

$$\begin{aligned} \sum_{i=1}^{n_1} (x_{1i} - \bar{x})^2 &= \sum_{i=1}^{n_1} (x_{1i} - \bar{x}_1 + \bar{x}_1 - \bar{x})^2 \\ &= \sum_{i=1}^{n_1} (x_{1i} - \bar{x}_1)^2 + n_1 (\bar{x}_1 - \bar{x})^2 + 2 (\bar{x}_1 - \bar{x}) \sum_{i=1}^{n_1} (x_{1i} - \bar{x}_1). \end{aligned} \quad \dots(3.10a)$$

But  $\sum_{i=1}^{n_1} (x_{1i} - \bar{x}_1) = 0$ , being the algebraic sum of the deviations the values of first series from their mean. Hence from (3.10a), on using (\*\*), we get

$$\sum_{i=1}^{n_1} (x_{1i} - \bar{x})^2 = n_1 \sigma_1^2 + n_1 (\bar{x}_1 - \bar{x})^2 = n_1 \sigma_1^2 + n_1 d_1^2 \quad \dots(3.10b)$$

where  $d_1 = \bar{x}_1 - \bar{x}$ .

Similarly, we get

$$\begin{aligned} \sum_{j=1}^{n_2} (x_{2j} - \bar{x})^2 &= \sum_{j=1}^{n_2} (x_{2j} - \bar{x}_2 + \bar{x}_2 - \bar{x})^2 \\ &= \sum_{j=1}^{n_2} (x_{2j} - \bar{x}_2)^2 + n_2 (\bar{x}_2 - \bar{x})^2 = n_2 \sigma_2^2 + n_2 d_2^2 \end{aligned} \quad \dots(3.10c)$$

where  $d_2 = \bar{x}_2 - \bar{x}$ .

Substituting from (3.10b) and (3.10c) in (3.10), we get the required formula

$$\sigma^2 = \frac{1}{n_1 + n_2} [n_1(\sigma_1^2 + d_1^2) + n_2(\sigma_2^2 + d_2^2)]$$

This formula can be simplified still further. We have

$$\begin{aligned} d_1 &= \bar{x}_1 - \bar{x} = \bar{x}_1 - \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2}{n_1 + n_2} = \frac{n_2 (\bar{x}_1 - \bar{x}_2)}{n_1 + n_2} \\ d_2 &= \bar{x}_2 - \bar{x} = \bar{x}_2 - \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2}{n_1 + n_2} = \frac{n_1 (\bar{x}_2 - \bar{x}_1)}{n_1 + n_2} \end{aligned}$$

Hence

$$\begin{aligned} \sigma^2 &= \frac{1}{n_1 + n_2} \left[ n_1 \sigma_1^2 + n_2 \sigma_2^2 + \left\{ \frac{n_1 n_2^2 (\bar{x}_1 - \bar{x}_2)^2}{(n_1 + n_2)^2} + \frac{n_2 n_1^2 (\bar{x}_2 - \bar{x}_1)^2}{(n_1 + n_2)^2} \right\} \right] \\ &= \frac{1}{n_1 + n_2} \left[ n_1 \sigma_1^2 + n_2 \sigma_2^2 + \frac{n_1 n_2}{n_1 + n_2} (\bar{x}_1 - \bar{x}_2)^2 \right] \end{aligned} \quad \dots(3.11)$$

**Remark.** The formula (3.9) can be easily generalised to the case of more than two series. If  $n_i, \bar{x}_i$  and  $\sigma_i, i = 1, 2, \dots, k$  are the sizes, means and standard deviations respectively of  $k$ -component series then the standard deviation  $\sigma$  of

the combined series of size  $\sum_{i=1}^k n_i$  is given by

$$\sigma^2 = \frac{1}{n_1 + n_2 + \dots + n_k} [n_1(\sigma_1^2 + d_1^2) + n_2(\sigma_2^2 + d_2^2) + \dots + n_k(\sigma_k^2 + d_k^2)] \quad \dots(3.12)$$

where

$$d_i = \bar{x}_i - \bar{x}; \quad i = 1, 2, \dots, k$$

and

$$\bar{x} = \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2 + \dots + n_k \bar{x}_k}{n_1 + n_2 + \dots + n_k}$$

**Example 3.6.** The first of the two samples has 100 items with mean 15 and standard deviation 3. If the whole group has 250 items with mean 15.6 and standard deviation  $\sqrt{13.44}$ , find the standard deviation of the second group.

**Solution.** Here we are given

$$n_1 = 100, \bar{x}_1 = 15 \text{ and } \sigma_1 = 3$$

$$n = n_1 + n_2 = 250, \bar{x} = 15.6, \text{ and } \sigma = \sqrt{13.44}$$

We want  $\sigma_2$ .

Obviously  $n_2 = 250 - 100 = 150$ . We have

$$\bar{x} = \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2}{n_1 + n_2} \Rightarrow 15.6 = \frac{100 \times 15 + 150 \times \bar{x}_2}{250}$$

$$\Rightarrow 150 \bar{x}_2 = 250 \times 15.6 - 1500 = 2400$$

$$\therefore \bar{x}_2 = \frac{2400}{150} = 16$$

$$\text{Hence } d_1 = \bar{x}_1 - \bar{x} = 15 - 15.6 = -0.6$$

$$\text{and } d_2 = \bar{x}_2 - \bar{x} = 16 - 15.6 = 0.4$$

The variance  $\sigma^2$  of the combined group is given by the formula :

$$(n_1 + n_2)\sigma^2 = n_1(\sigma_1^2 + d_1^2) + n_2(\sigma_2^2 + d_2^2)$$

$$\Rightarrow 250 \times 13.44 = 100(9 + 0.36) + 150(\sigma_2^2 + 0.16)$$

$$\therefore 150 \sigma_2^2 = 250 \times 13.44 - 100 \times 9.36 - 150 \times 0.16 \\ = 3360 - 936 - 24 = 2400$$

$$\therefore \sigma_2^2 = \frac{2400}{150} = 16$$

$$\text{Hence } \sigma_2 = \sqrt{16} = 4$$

**3-8. Co-efficient of Dispersion.** Whenever we want to compare the variability of the two series which differ widely in their averages or which are measured in different units, we do not merely calculate the measures of dispersion but we calculate the co-efficients of dispersion which are pure numbers independent of the units of measurement. The co-efficients of dispersion (C.D.) based on different measures of dispersion are as follows :

1. C.D. based upon range  $= \frac{A - B}{A + B}$ , where  $A$  and  $B$  are the greatest and the smallest items in the series.

2. Based upon quartile deviation :

$$\text{C.D.} = \frac{(Q_3 - Q_1)/2}{(Q_3 + Q_1)/2} = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

3. Based upon mean deviation :

$$\text{C.D.} = \frac{\text{Mean deviation}}{\text{Average from which it is calculated}}$$

4. Based upon standard deviation :

$$\text{C.D.} = \frac{\text{S.D.}}{\text{Mean}} = \frac{\sigma}{\bar{x}}$$

**3-8-1. Co-efficient of Variation.** 100 times the co-efficient of dispersion based upon standard deviation is called co-efficient of variation (C.V.).

$$\text{C.V.} = 100 \times \frac{\sigma}{\bar{x}} \quad . \quad (3-13)$$

According to Professor Karl Pearson who suggested this measure, C.V. is the percentage variation in the mean, standard deviation being considered as the total variation in the mean.

For comparing the variability of two series, we calculate the co-efficient of variations for each series. The series having greater C.V. is said to be more variable than the other and the series having lesser C.V. is said to be

more consistent (or homogenous) than the other.

**Example 3.7.** An analysis of monthly wages paid to the workers of two firms A and B belonging to the same industry gives the following results :

	Firm A	Firm B
Number of workers	500	600
Average monthly wage	Rs. 186.00	Rs. 175.00
Variance of distribution of wages	81	100

(i) Which firm, A or B, has a larger wage bill?

(ii) In which firm, A or B, is there greater variability in individual wages?

(iii) Calculate (a) the average monthly wage, and (b) the variance of the distribution of wages, of all the workers in the firms A and B taken together.

**Solution.**

(i) Firm A :

No. of wage-earners (say)  $n_1 = 500$

Average monthly wages (say)  $\bar{x}_1 = \text{Rs. } 186$

$$\text{Average monthly wage} = \frac{\text{Total wages paid}}{\text{No. of workers}}$$

Hence total wages paid to the workers =  $n_1 \bar{x}_1 = 500 \times 186 = \text{Rs. } 93,000$

**Firm B**

No. of wage-earners (say)  $n_2 = 600$

Average monthly wages (say)  $\bar{x}_2 = \text{Rs. } 175$

$\therefore$  Total wages paid to the workers =  $n_2 \bar{x}_2 = 600 \times 175 = \text{Rs. } 1,05,000$

Thus we see that the firm B has larger wage bill.

(ii) Variance of distribution of wages in firm A (say)  $\sigma_1^2 = 81$

Variance of distribution of wages in firm B (say)  $\sigma_2^2 = 100$

$$\text{C.V. of distribution of wages for firm A} = 100 \times \frac{\sigma_1}{\bar{x}_1} = \frac{100 \times 9}{186} = 4.84$$

$$\text{C.V. of distribution of wages for firm B} = 100 \times \frac{\sigma_2}{\bar{x}_2} = \frac{100 \times 10}{175} = 5.71$$

Since C.V. for firm B is greater than C.V. for firm A, firm B has greater variability in individual wages.

(iii) (a) The average monthly wages (say)  $\bar{x}$ , of all the workers in the two firms A and B taken together is given by

$$\bar{x} = \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2}{n_1 + n_2} = \frac{500 \times 186 + 600 \times 175}{500 + 600} = \frac{198000}{1100} = \text{Rs. } 180$$

(b) The combined variance  $\sigma^2$  is given by the formula

$$\sigma^2 = \frac{1}{n_1 + n_2} [n_1 (\sigma_1^2 + d_1^2) + n_2 (\sigma_2^2 + d_2^2)]$$

where  $d_1 = \bar{x}_1 - \bar{x}$  and  $d_2 = \bar{x}_2 - \bar{x}$

Here  $d_1 = 186 - 180 = 6$  and  $d_2 = 175 - 180 = -5$

$$\text{Hence } \sigma^2 = \frac{500(81+36) + 600(100+25)}{500+600} = \frac{133500}{1100} = 121.36$$

### EXERCISE 3 (a)

1. (a) Explain with suitable examples the term 'dispersion'. State the relative and absolute measures of dispersion and describe the merits and demerits of standard deviation.

(b) Explain the main difference between mean deviation and standard deviation. Show that standard deviation is independent of change of origin and scale.

(c) Distinguish between absolute and relative measures of dispersion.

2. (a) Explain the graphical method of obtaining median and quartile deviation. (Calicut Univ.B.Sc., April 1989)

(b) Compute quartile deviation graphically for the following data :

Marks	: 20 - 30	30 - 40	40 - 50	50 - 60	60 - 70	70 & over
Number of students	: 5	20	14	10	8	5

3. (a) Show that for raw data mean deviation is minimum when measured from the median.

(b) Compute a suitable measure of dispersion for the following grouped frequency distribution giving reasons :

Classes	Frequency
Less than 20	30
20 - 30	20
30 - 40	15
40 - 50	10
50 - 60	5

(c) Age distribution of hundred life insurance policyholders is as follows:

Age as on nearest birthday	Number
17 - 19.5	9
20 - 25.5	16
26 - 35.5	12
36 - 40.5	26
41 - 50.5	14
51 - 55.5	12
56 - 60.5	6
61 - 70.5	5

Calculate mean deviation from median age.

Ans. Median = 38.25, M.D. = 10.605

4. Prove that the mean deviation about the mean  $\bar{x}$  of the variate  $x$ , the frequency of whose  $i$ th size  $x_i$  is  $f_i$  is given by

$$\frac{2}{N} \left[ \bar{x} \sum_{x_i < \bar{x}} f_i - \sum_{x_i > \bar{x}} f_i x_i \right]$$

**Hint.** Mean deviation about mean

$$\begin{aligned} &= \frac{1}{N} \left[ \sum_{x_i < \bar{x}} f_i (\bar{x} - x_i) + \sum_{x_i > \bar{x}} f_i (x_i - \bar{x}) \right] \\ &= \frac{1}{N} \left[ -\sum_{x_i < \bar{x}} f_i (x_i - \bar{x}) + \sum_{x_i > \bar{x}} f_i (x_i - \bar{x}) \right] \end{aligned}$$

Since  $(\sum f_i (x_i - \bar{x})) = 0$ ,

$$\sum_{x_i > \bar{x}} f_i (x_i - \bar{x}) + \sum_{x_i < \bar{x}} f_i (x_i - \bar{x}) = 0$$

$$\therefore M.D. = \frac{1}{N} \left( -\sum_{x_i < \bar{x}} f_i (x_i - \bar{x}) - \sum_{x_i > \bar{x}} f_i (x_i - \bar{x}) \right) = -\frac{2}{N} \left( \sum_{x_i < \bar{x}} f_i (x_i - \bar{x}) \right)$$

**5.** What is standard deviation? Explain its superiority over other measures of dispersion.

**6.** Calculate the mean and standard deviation of the following distribution:

x : 2.5 — 7.5    7.5 — 12.5    12.5 — 17.5    17.5 — 22.5

f :        12                      28                      65                      121

x : 22.5 — 27.5    27.5 — 32.5    32.5 — 37.5    37.5 — 42.5    42.5 — 47.5

f :        175                      198                      176                      120                      66

x : 47.5 — 52.5    52.5 — 57.5    57.5 — 62.5

f :        27                      9                              3

**Ans.** Mean = 30.005, Standard Deviation = 0.01

**7.** Explain clearly the ideas implied in using arbitrary working origin, and scale for the calculation of the arithmetic mean and standard deviation of a frequency distribution. The values of the arithmetic mean and standard deviation of the following frequency distribution of a continuous variable derived from the analysis in the above manner are 40.604 lb. and 7.92 lb. respectively.

x :	-3	-2	-1	0	1	2	3	4		Total
f :	3	15	45	57	50	36	25	9		240

Determine the actual class intervals.

**8. (a).** The arithmetic mean and variance of a set of 10 figures are known to be 17 and 33 respectively. Of the 10 figures, one figure (i.e., 26) was subsequently found inaccurate, and was weeded out. What is the resulting (a) arithmetic mean and (b) standard deviation. (M.S. Baroda U. B.Sc. 1993)

**(b)** The mean and standard deviation of 20 items is found to be 10 and 2 respectively. At the time of checking it was found that one item 8 was incorrect. Calculate the mean and standard deviation if

- (i) the wrong item is omitted, and
- (ii) it is replaced by 12.

(c) For a frequency distribution of marks in Statistics of 200 candidates (grouped in intervals 0-5, 5-10,..., etc.), the mean and standard deviation were found to be 40 and 15 respectively. Later it was discovered that the score 43 was misread as 53 in obtaining the frequency distribution. Find the corrected mean and standard deviation corresponding to the corrected frequency distribution.

**Ans.** Mean = 39.95, S.D. = 14.974.

**9. (a)** Complete a table showing the frequencies with which words of different numbers of letters occur in the extract reproduced below (omitting punctuation marks) treating as the variable the number of letters in each word, and obtain the mean, median and coefficient of variation of the distribution :

"Her eyes were blue : blue as autumn distance-blue as the blue we see, between the retreating mouldings of hills and woody slopes on a sunny September morning : a misty and shady blue, that had no beginning or surface, and was looked into rather than at."

**Ans.** Mean = 4.35, Median = 4,  $\sigma$  = 2.23 and C.V. = 51.26

**(b)** Treating the number of letters in each word in the following passage as the variable  $x$ , prepare the frequency distribution table and obtain its mean, median, mode and variance.

"The reliability of data must always be examined before any attempt is made to base conclusions upon them. This is true of all data, but particularly so of numerical data, which do not carry their quality written large on them. It is a waste of time to apply the refined theoretical methods of Statistics to data which are suspect from the beginning."

**Ans.** Mean = 4.565, Median = 4, Mode = 3, S.D. = 2.673.

**10.** The mean of 5 observations is 4.4 and variance is 8.24. If three of the five observation are 1, 2 and 6, find the other two.

**11. (a)** Scores of two golfers for 24 rounds were as follows :

Golfer A : 74, 75, 78, 72, 77, 79, 78, 81, 76, 72, 77, 74, 70, 78, 79, 80, 81, 74, 80, 75, 71, 73.

Golfer B : 86, 84, 80, 88, 89, 85, 86, 82, 82, 79, 86, 80, 82, 76, 86, 89, 87, 83, 80, 88, 86, 81, 81, 87

Find which golfer may be considered to be a more consistent player ?

**Ans.** Golfer B is more consistent player.

**(b).** The sum and sum of squares corresponding to length  $X$  (in cms.) and weight  $Y$  (in gms.) of 50 tapioca tubers are given below :

$$\Sigma X = 212, \quad \Sigma X^2 = 902.8$$

$$\Sigma Y = 261, \quad \Sigma Y^2 = 1457.6$$

Which is more varying, the length or weight.

**12. (a)** Lives of two models of refrigerators turned in for new models in a recent survey are

Life (No. of years)	Model A	Model B
0 - 2	5	2
2 - 4	16	7

4 - 6	13	12
6 - 8	7	19
8 - 10	5	9
10 - 12	4	1

What is the average life of each model of these refrigerators? Which model shows more uniformity?

Ans: C.V. (Model A)=54.9%, C.V. (Model B)=3.62%

(b) Goals scored by two teams A and B in a football season were as follows :

No. of goals scored in a match	No. of matches	
	A	B
0	27	17
1	9	9
2	8	6
3	5	5
4	4	3

(Sri Venkateswara, U. B.Sc. Sept. 1992)

Find out which team is more consistent.

Ans. Team A : C.V.=122.0, Team B : C.V. = 108.3.

(c) An analysis of monthly wages paid to the workers in two firms, A and B belonging to the same industry, gave the following results :

	Firm A	Firm B
No. of wage-earners	986	548
Average monthly wages	Rs. 52.5	Rs. 47.5
Variance of distribution of wages	100	121

- (i) Which firm, A or B, pays out larger amount as monthly wages?
- (ii) In which firm A or B, is there greater variability in individual wages?
- (iii) What are the measures of average monthly wages and the variability in individual wages, of all the workers in the two firms, A and B taken together.

Ans. (i) Firm B pays a larger amount as monthly wages.

(ii) There is greater variability in individual wages in firm B.

(iii) Combined arithmetic mean = Rs. 49.87.

Combined standard deviation = Rs. 10.82.

14. (a) The following data give the arithmetic averages and standard deviations of three sub-groups. Calculate the arithmetic average and standard deviation of the whole group.

Sub-group	No. of men	Average wages (Rs.)	Standard deviation (Rs.)
A	50	61.0	8.0
B	100	70.0	9.0
C	120	80.5	10.0

Ans. Combined Mean = 73, Combined S.D.=11.9.

(b) Find the missing information from the following data :

	<i>Group I</i>	<i>Group II</i>	<i>Group III</i>	<i>Combined</i>
Number	50	?	90	200
Standard Deviation	6	7	?	7.746
Mean	113	?	115	116

Ans.  $n_2 = 60$ ,  $\bar{x}_2 = 120$  and  $\sigma_3 = 8$

15. A collar manufacturer is considering the production of a new style collar to attract young men. The following statistics of neck circumference are avail based on the measurement of a typical group of students :

Mid-value : 12.5 13.0 13.5 14.0 14.5 15.0 15.5 16.0

in inches

No. of students : 4 19 30 63 66 29 18 1

Compute the mean and standard deviation and use the criterion  $\bar{x} \pm$  obtain the largest and smallest size of collar he should make in order to meet needs of practically all his customers bearing in mind that the collars are worn on average 3/4 inch larger than neck size. (Nagpur Univ. B.Sc., 1992)

Ans. Mean = 14.232, S.D. = 0.72, largest size = 17.14", smallest size = 12.83"

16. (a) A frequency distribution is divided into two parts. The mean and standard deviation of the first part are  $m_1$  and  $s_1$  and those of the second part are  $m_2$  and  $s_2$  respectively. Obtain the mean and standard deviation for the combined distribution. [Delhi Univ. B.Sc.(Stat.Hons.), 1986]

(b) The means of two samples of size 50 and 100 respectively are 54.1 and 50.3 and the standard deviations are 8 and 7. Obtain the mean and standard deviation of the sample of size 150 obtained by combining the two samples.

Ans. Combined mean = 51.57. Combined S.D. = 7.5 approx.

(c) A distribution consists of three components with frequencies 200, 250 and 300 having means 25, 10 and 15 and standard deviations 3, 4 and 5 respectively.

Show that the mean of the combined group is 16 and its standard deviation 7.2 approximately. (Bangalore Univ. B.Sc. 1992)

17. In a certain test for which the pass marks is 30, the distribution of marks of passing candidates classified by sex (boys and girls) were as given below :

<i>Marks</i>	<i>Frequency</i>	
	<i>Boys</i>	<i>Girls</i>
30-34	5	15
35-39	10	20
40-44	15	30
45-49	30	20
50-54	5	5
55-59	5	-
<b>Total</b>	<b>70</b>	<b>90</b>

The overall means and standard deviation of marks for boys including the 30 failed were 38 and 10. The corresponding figures for girls including the 10 failed were 35 and 9.

(i) Find the mean and standard deviation of marks obtained by the 30 boys who failed in the test.

(ii) The moderation committee argued that percentage of passes among girls is higher because the girls are very studious and if the intention is to pass those who are really intelligent, a higher pass marks should be used for girls. Without questioning the propriety of this argument, suggest what the pass mark should be which would allow only 70% of the girls to pass.

(iii) The prize committee decided to award prizes to the best 40 candidates (irrespective of sex) judged on the basis of marks obtained in the test. Estimate the number of girls who would receive prizes.

**Ans.** (i)  $\bar{x} = 22.83$ ,  $\sigma_2 = 8.27$  (ii) 39 (iii) 15

**18.** Find the mean and variance of first  $n$ -natural numbers.

(Agra Univ. B.Sc., 1993)

$$\text{Ans. } \bar{x} = \frac{n+1}{2}, \sigma_2 = \frac{n^2 - 1}{12}$$

**19.** In a frequency distribution, the  $n$  intervals are 0 to 1, 1 to 2, ...,  $(n-1)$  to  $n$  with equal frequencies. Find the mean deviation and variance.

**20.** If the mean and standard deviation of a variable  $x$  are  $m$  and  $\sigma$  respectively, obtain the mean and standard deviation of  $(ax + b)/c$ , where  $a$ ,  $b$  and  $c$  are constants.

$$\text{Ans. } \bar{u} = \frac{1}{c} (a\bar{x} + b), \sigma_u = \left| \frac{a}{c} \right| \sigma$$

**21.** In a series of measurements we obtain  $m_1$  values of magnitude  $x_1$ ,  $m_2$  values of magnitude  $x_2$ , and so on. If  $\bar{x}$  is the mean value of all the measurements, prove that the standard deviation is

$$\sqrt{\frac{\sum m_r (k - x_r)^2}{\sum m_r}} = \delta^2$$

where  $\bar{x} = k + \delta$  and  $k$  is any constant.

(Delhi Univ. B.Sc. (Stat. Hons.), 1992)

**22. (a)** Show that in a discrete series if deviations are small compared with mean  $M$  so that  $(x/M)^2$  and higher powers of  $(x/M)$  are neglected, prove that

$$(i) MH = G^2 \quad (II) M - 2G + H = 0,$$

where  $G$  is geometric mean and  $H$  is harmonic mean.

**(b)** The mean and standard deviation of a variable  $x$  are  $m$  and  $\sigma$  respectively. If the deviations are small compared with the value of the mean, show that

$$(i) \text{ Mean } (\sqrt{x}) = \sqrt{m} \left( 1 - \frac{\sigma^2}{8m^2} \right)$$

$$(ii) \text{ Mean } \left( \frac{1}{\sqrt{x}} \right) = \frac{1}{\sqrt{m}} \left( 1 + \frac{3\sigma^2}{8m^2} \right) \text{ approximately.}$$

(M.S. Baroda U. B.Sc. 1993)

(c) If the deviation  $X_i = x_i - M$  is very small in comparison with mean  $M$  and  $(X_i/M)^2$  and higher powers of  $(X_i/M)$  are neglected, prove that

$$V = \sqrt{\frac{2(M - G)}{M}}$$

where  $G$  is the geometric mean of the values  $x_1, x_2, \dots, x_n$  and  $V$  is the coefficient of dispersion ( $\sigma/M$ ). (Lucknow Univ. B.Sc., 1993)

23. From a sample of observations the arithmetic mean and variance are calculated. It is then found that one of the values,  $x_1$ , is in error and should be replaced by  $x_1'$ . Show that the adjustment to the variance to correct this error is

$$\frac{1}{n} (x_1' - x_1) \cdot \left( x_1' + x_1 + \frac{x_1' - x_1 + 2T}{n} \right)$$

where  $T$  is the total of the original results.

(Meerut Univ. B.Sc., 1992; Delhi Univ. B.Sc. (Stat. Hons.), 1989, 1985)

$$\begin{aligned} \text{Hint. } \sigma^2 &= \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2 \\ &= \frac{1}{n} \left( x_1^2 + x_2^2 + \dots + x_n^2 \right) - \frac{T^2}{n^2} \end{aligned}$$

where  $T = x_1 + x_2 + \dots + x_n$ .

Let  $\sigma_1^2$  be the corrected variance. Then

$$\sigma_1^2 = \frac{1}{n} \left\{ x_1'^2 + x_2^2 + \dots + x_n^2 \right\} - \left\{ \frac{T - x_1 + x_1'}{n} \right\}^2$$

Adjustment to the variance to correct the error is :

$$\begin{aligned} \sigma_1^2 - \sigma^2 &= \frac{1}{n} \left\{ x_1'^2 - x_1^2 \right\} - \frac{1}{n^2} \left\{ (T - x_1 + x_1')^2 - T^2 \right\} \\ &= \frac{1}{n} \left\{ x_1' + x_1 \right\} \left\{ x_1' - x_1 \right\} - \frac{1}{n^2} \left\{ (x_1' - x_1) \times (2T - x_1 + x_1') \right\} \end{aligned}$$

24. Show that, if the variable takes the values  $0, 1, 2, \dots, n$  with frequencies proportional to the binomial coefficients " $C_0, C_1, C_2, \dots, C_n$ " respectively then the mean of the distribution is  $(n/2)$ , the mean square deviation about  $x=0$  is  $(n+1)/4$  and the variance is  $n/4$ .

(Delhi Univ. B.Sc. (Stat. Hons.), 1991)

$$\begin{aligned} \text{Hint. } N &= \sum f = "C_0 + "C_1 + "C_2 + \dots + "C_n = (1+1)^n = 2^n \\ \sum fx &= 0."C_0 + 1."C_1 + 2."C_2 + 3."C_3 + \dots + n."C_n \\ &= n \left\{ 1 + (n-1) + \frac{(n-1)(n-2)}{2!} + \dots + 1 \right\} \\ &= n (1+1)^{n-1} = n \cdot 2^{(n-1)} \end{aligned}$$

$$\text{Hence mean } (\bar{x}) = \frac{1}{N} \sum fx = \frac{n \cdot 2^{(n-1)}}{2^n} = \frac{n}{2}$$

The mean square deviation  $s^2$ , (say), about the point  $x=0$  is given by,

$$\begin{aligned}
 s^2 &= \frac{1}{N} \sum f_i x_i^2 = \frac{1}{2^n} [1^2 \cdot n C_1 + 2^2 \cdot n C_2 + 3^2 \cdot n C_3 + \dots + n^2 \cdot n C_n] \\
 &= \frac{n}{2^n} [1 + 2(n-1) + \frac{3}{2}(n-1)(n-2) + \dots + n] \\
 &= \frac{n}{2^n} \left[ \left\{ 1 + (n-1) + \frac{(n-1)(n-2)}{2!} + \dots + 1 \right\} \right. \\
 &\quad \left. + \{(n-1) + (n-1)(n-2) + \dots + (n-1)\} \right] \\
 &= \frac{n}{2^n} \left[ (n-1)C_0 + (n-1)C_1 + (n-1)C_2 + \dots + (n-1)C_{n-1} \right] \\
 &\quad + \{(n-1)(n-2)C_0 + (n-2)C_1 + \dots + (n-2)C_{n-2}\}] \\
 &= \frac{n}{2^n} [(1+1)^{n-1} + (n-1)(1+1)^{n-2}] = \frac{n(n+1)}{4} \\
 \therefore \quad \sigma^2 &= \frac{n(n+1)}{4} - \frac{n^2}{4} = \frac{n}{4}.
 \end{aligned}$$

25. (a) Let  $r$  be the range and  $s$  be the standard deviation of a set of observations  $x_1, x_2, \dots, x_n$ ; then prove by general reasoning or otherwise that  $s \leq r$ .

**Hint.** Since  $x_i - \bar{x} \leq r$ ,  $i = 1, 2, \dots, n$ , we have

$$\begin{aligned}
 s^2 &= \frac{1}{N} \sum_{i=1}^n f_i (x_i - \bar{x})^2 \leq \frac{1}{N} \sum_{i=1}^n f_i (r^2) \\
 \Rightarrow \quad s^2 &\leq r^2 \frac{1}{N} \sum_{i=1}^n f_i = r^2 \quad \Rightarrow \quad s \leq r
 \end{aligned}$$

(b) Let  $r$  be the range and

$$S = \left( \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^{\frac{1}{2}}$$

be the standard deviation of a set of observations  $x_1, x_2, \dots, x_n$ , then prove that

$$S \leq r \left( \frac{n}{n-1} \right)^{\frac{1}{2}} \quad [\text{Punjab Univ. B.Sc (Stat. Hons.), 1993}]$$

**3.9. Moments.** The  $r$ th moment of a variable  $x$  about any point  $x = A$ , usually denoted by  $\mu'_r$ , is given by

$$\mu'_r = \frac{1}{N} \sum_i f_i (x_i - A)^r, \quad \sum_i f_i = N \quad \dots (3.14)$$

$$= \frac{1}{N} \sum_i f_i d_i^r, \quad \dots (3.14a)$$

where  $d_i = x_i - A$ .

The  $r$ th moment of a variable about the mean  $\bar{x}$ , usually denoted by  $\mu_r$ , is given by

$$\mu_r = \frac{1}{N} \sum_i f_i (x_i - \bar{x})^r = \frac{1}{N} \sum_i f_i z_i^r \quad \dots (3.15)$$

where  $z_i = x_i - \bar{x}$ .

In particular

$$\mu_0 = \frac{1}{N} \sum_i f_i (x_i - \bar{x})^0 = \frac{1}{N} \sum_i f_i = 1$$

and  $\mu_1 = \frac{1}{N} \sum_i f_i (x_i - \bar{x}) = 0$ , being the algebraic sum of deviations from the mean. Also

$$\mu_2 = \frac{1}{N} \sum_i f_i (x_i - \bar{x})^2 = \sigma^2 \quad \dots(3.16)$$

These results, viz.,  $\mu_0 = 1$ ,  $\mu_1 = 0$ , and  $\mu_2 = \sigma^2$ , are of fundamental importance and should be committed to memory.

We know that if  $d_i = x_i - A$ , then

$$\bar{x} = A + \frac{1}{N} \sum_i f_i d_i = A + \mu_1' \quad \dots(3.17)$$

**3.9.1. Relation between moments about mean in terms of moments about any point and vice versa.**

We have

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

Using (3.17), we get

$$\begin{aligned} \mu_r &= \frac{1}{N} \sum_i f_i (d_i - \mu_1')^r \\ &= \frac{1}{N} \sum_i f_i [d_i^r - C_1 d_i^{r-1} \mu_1' + C_2 d_i^{r-2} \mu_1'^2 - C_3 d_i^{r-3} \mu_1'^3 + \dots + (-1)^r \mu_1'^r] \\ &= \mu_r' - C_1 \mu_{r-1}' \mu_1' + C_2 \mu_{r-2}' \mu_1'^2 - \dots + (-1)^r \mu_1'^r \quad [\text{On using (3.14a)}] \end{aligned} \quad \dots(3.18)$$

In particular, on putting  $r = 2, 3$  and  $4$  in (3.18), we get

$$\begin{aligned} \mu_2 &= \mu_2' - \mu_1'^2 \\ \mu_3 &= \mu_3' - 3\mu_2' \mu_1' + 2\mu_1'^3 \\ \mu_4 &= \mu_4' - 4\mu_3' \mu_1' + 6\mu_2' \mu_1'^2 - 3\mu_1'^4 \end{aligned} \quad \dots(3.19)$$

Conversely,

$$\begin{aligned} \mu_r' &= \frac{1}{N} \sum_i f_i (x_i - A)^r = \frac{1}{N} \sum_i f_i (x_i - \bar{x} + \bar{x} - A)^r \\ &= \frac{1}{N} \sum_i f_i (z_i + \mu_1')^r \end{aligned}$$

where  $x_i - \bar{x} = z_i$  and  $\bar{x} = A + \mu_1'$

$$\text{Thus } \mu_r' = \frac{1}{N} \sum_i f_i (z_i' + 'C_1 z_i^{r-1} \mu_1' + 'C_2 z_i^{r-2} \mu_1'^2 + \dots + \mu_1'^r)$$

$$= \mu_r + 'C_1 \mu_{r-1} \mu_1' + 'C_2 \mu_{r-2} \mu_1'^2 + \dots + \mu_1'^r. \text{ [From (3.15)]}$$

In particular, putting  $r = 2, 3$  and  $4$  and noting that  $\mu_1 = 0$ , we get

$$\mu_2' = \mu_2 + \mu_1'^2$$

$$\mu_3' = \mu_3 + 3\mu_2 \mu_1' + \mu_1'^3 \quad \dots (3.20)$$

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

These formulae enable us to find the moments about any point, once the mean and the moments about mean are known.

### 3.9.2 Effect of Change of Origin and Scale on Moments.

Let  $u = \frac{x - A}{h}$ , so that  $x = A + hu$ ,  $\bar{x} = A + h\bar{u}$  and  $x - \bar{x} = h(u - \bar{u})$

Thus,  $r$ th moment of  $x$  about any point  $x = A$  is given by

$$\mu_r' = \frac{1}{N} \sum_i f_i (x_i - A)^r = \frac{1}{N} \sum_i f_i (hu_i)^r = h^r \cdot \frac{1}{N} \sum_i f_i u_i^r$$

And the  $r$ th moment of  $x$  about mean is

$$\begin{aligned} \mu_r &= \frac{1}{N} \sum_i f_i (x_i - \bar{x})^r = \frac{1}{N} \sum_i f_i [h(u_i - \bar{u})]^r \\ &= h^r \frac{1}{N} \sum_i f_i (u_i - \bar{u})^r \end{aligned}$$

Thus the  $r$ th moment of the variable  $x$  about mean is  $h^r$  times the  $r$ th moment of the variable  $u$  about its mean.

**3.9.3. Sheppard's Corrections for Moments.** In case of grouped frequency distribution, while calculating moments we assume that the frequencies are concentrated at the middle point of the class intervals. If the distribution is symmetrical or slightly symmetrical and the class intervals are not greater than one-twentieth of the range, this assumption is very nearly true. But since the assumption is not in general true, some error, called the 'grouping error', creeps into the calculation of the moments. W.F. Sheppard proved that if

(i) the frequency distribution is continuous, and

(ii) the frequency tapers off to zero in both directions,

the effect due to grouping at the mid-point of the intervals can be corrected by the following formulae, known as Sheppard's corrections :

$$\mu_2 (\text{corrected}) = \mu_2 - \frac{h^2}{12} \quad \dots (3.21)$$

$$\mu_3 (\text{corrected}) = \mu_3$$

$$\mu_4 (\text{corrected}) = \mu_4 - \frac{1}{2} h^2 \mu_2 + \frac{7}{240} h^4$$

where  $h$  is the width of the class interval.

### 3-9-4. Charlier's Checks.

The following identities

$$\sum f(x+1) = \sum fx + N; \quad \sum f(x+1)^2 = \sum fx^2 + 2\sum fx + N$$

$$\sum f(x+1)^3 = \sum fx^3 + 3\sum fx^2 + 3\sum fx + N$$

$$\sum f(x+1)^4 = \sum fx^4 + 4\sum fx^3 + 6\sum fx^2 + 4\sum fx + N,$$

are often used in checking the accuracy in the calculation of first four moments and are known as Charlier's Checks.

**3-10. Pearson's  $\beta$  and  $\gamma$  Coefficients.** Karl Pearson defined the following four coefficients, based upon the first four moments about mean :

$$\beta_1 = \frac{\mu_3^2}{\mu_2^2}, \quad \gamma_1 = +\sqrt{\beta_1} \text{ and } \beta_2 = \frac{\mu_4}{\mu_2^2}, \quad \gamma_2 = \beta_2 - 3 \quad \dots (3-22)$$

It may be pointed out that these coefficients are pure numbers independent of units of measurement. The practical utility of these coefficients is discussed in § 3-13 and § 3-14.

**Remark.** Sometimes, another coefficient based on moments, viz., Alpha ( $\alpha$ ) coefficient is used. Alpha coefficients are defined as :

$$\alpha_1 = \frac{\mu_1}{\sigma} = 0, \quad \alpha_2 = \frac{\mu_2}{\sigma^2} = 1, \quad \alpha_3 = \frac{\mu_3}{\sigma^3} = \sqrt{\beta_1} = \gamma_1, \quad \alpha_4 = \frac{\mu_4}{\sigma^4} = \beta_2$$

**3-11. Factorial Moments.** Factorial moment of order  $r$  about the origin of the frequency distribution  $x_i | f_i$ , ( $i = 1, 2, \dots, n$ ), is defined as

$$\mu_{(r)}' = \frac{1}{N} \sum_{i=1}^n f_i x_i^{(r)} \quad \dots (3-23)$$

where  $x^{(r)} = x(x-1)(x-2)\dots(x-r+1)$  and  $N = \sum_{i=1}^n f_i$

Thus the factorial moment of order  $r$  about any point  $x = a$  is given by

$$\mu_{(r)}' = \frac{1}{N} \sum_i f_i (x_i - a)^{(r)} \quad \dots (3-24)$$

where  $(x-a)^{(r)} = (x-a)(x-a-1)\dots(x-a-r+1)$

In particular from (3-23), we have

$$\mu_{(1)}' = \frac{1}{N} \sum_i f_i x_i = \mu_1' \text{ (about origin) } = \text{Mean } (\bar{x})$$

$$\mu_{(2)}' = \frac{1}{N} \sum_i f_i x_i^{(2)} = \frac{1}{N} \sum_i f_i x_i (x_i - 1)$$

$$= \frac{1}{N} \sum_i f_i x_i^2 - \frac{1}{N} \sum_i f_i x_i = \mu_2' - \mu_1'$$

$$\begin{aligned}\mu_{(3)'} &= \frac{1}{N} \sum_i f_i x_i^{(3)} = \frac{1}{N} \sum_i f_i x_i (x_i - 1)(x_i - 2) \\&= \frac{1}{N} \sum_i f_i x_i^3 - 3 \cdot \frac{1}{N} \sum_i f_i x_i^2 + 2 \cdot \frac{1}{N} \cdot \sum_i f_i x_i \\&= \mu_3' - 3\mu_2' + 2\mu_1' \\ \mu_{(4)'} &= \frac{1}{N} \sum_i f_i x_i^{(4)} = \frac{1}{N} \sum_i f_i x_i (x_i - 1)(x_i - 2)(x_i - 3) \\&= \frac{1}{N} \sum_i f_i x_i (x_i^3 - 6x_i^2 + 11x_i - 6) \\&= \frac{1}{N} \sum_i f_i x_i^4 - 6 \cdot \frac{1}{N} \sum_i f_i x_i^3 + 11 \cdot \frac{1}{N} \sum_i f_i x_i^2 - 6 \cdot \frac{1}{N} \sum_i f_i x_i \\&= \mu_4' - 6\mu_3' + 11\mu_2' - 6\mu_1'\end{aligned}$$

Conversely, we will get

$$\begin{aligned}\mu_1' &= \mu_{(1)'} \\ \mu_2' &= \mu_{(2)'} + \mu_{(1)'} \\ \mu_3' &= \mu_{(3)'} + 3\mu_{(2)'} + \mu_{(1)'} \\ \mu_4' &= \mu_{(4)'} + 6\mu_{(3)'} + 7\mu_{(2)'} + \mu_{(1)'}\end{aligned}\dots (3.25)$$

**3.12. Absolute Moments.** For the frequency distribution  $x_i / f_i$ ,  $i = 1, 2, \dots, n$ , the  $r$ th absolute moment of the variable about the origin is given by

$$\frac{1}{N} \sum_{i=1}^n f_i |x_i'|, N = \sum f_i \dots (3.26)$$

where  $|x_i'|$  represents the absolute or modulus value of  $x_i'$ .

The  $r$ th absolute moment of the variable about the mean  $\bar{x}$  is given by

$$\frac{1}{N} \sum_{i=1}^n f_i |x_i - \bar{x}|^r \dots (3.26a)$$

**Example 3.8.** The first four moments of a distribution about the value 4 of the variable are  $-1.5, 17, -30$  and  $108$ . Find the moments about mean,  $\beta_1$  and  $\beta_2$ .

Find also the moments about (i) the origin, and (ii) the point  $x = 2$ .

**Solution.** In the usual notations, we are given  $A = 4$  and

$$\mu_1' = -1.5, \mu_2' = 17, \mu_3' = -30 \text{ and } \mu_4' = 108.$$

Moments about mean :  $\mu_1 = 0$

$$\mu_2 = \mu_2' - \mu_1'^2 = 17 - (-1.5)^2 = 17 - 2.25 = 14.75$$

$$\begin{aligned}\mu_3 &= \mu_3' - 3\mu_2' \mu_1' + 2\mu_1'^3 \\&= -30 - 3 \times (17) \times (-1.5) + 2 \times (-1.5)^3 \\&= -30 + 76.5 - 6.75 = 39.75\end{aligned}$$

$$\begin{aligned}\mu_4' &= \mu_4 - 4\mu_3'\mu_1' + 6\mu_2'\mu_1'^2 - 3\mu_1'^4 \\&= 108 - 4(-30)(-1.5) + 6(17)(-1.5)^2 - 3(-1.5)^4 \\&= 108 - 180 + 229.5 - 15.1875 = 142.3125\end{aligned}$$

Hence  $\beta_1 = \frac{\mu_3^2}{\mu_2^2} = \frac{(39.75)^2}{(14.75)^3} = 0.4924$

$$\beta_2 = \frac{\mu_4'}{\mu_2^2} = \frac{(142.3125)}{(14.75)^2} = 0.6541$$

Also  $\bar{x} = A + \mu_1' = 4 + (-1.5) = 2.5$

*Moments about origin.* We have

$$\bar{x} = 2.5, \mu_2 = 14.75, \mu_3 = 39.75 \text{ and } \mu_1 = 142.31 \text{ (approx).}$$

We know  $\bar{x} = A + \mu_1'$ , where  $\mu_1'$  is the first moment about the point  $x = A$ . Taking  $A = 0$ , we get the first moment about origin as  $\mu_1' = \text{mean} = 2.5$ .

Using (3-20), we get

$$\begin{aligned}\mu_2' &= \mu_2 + \mu_1'^2 = 14.75 + (2.5)^2 = 14.75 + 6.25 = 21 \\ \mu_3' &= \mu_3 + 3\mu_2\mu_1' + \mu_1'^3 = 39.75 + 3(14.75)(2.5) + (2.5)^3 \\&= 39.75 + 110.625 + 15.625 = 166 \\ \mu_4' &= \mu_4 + 4\mu_3\mu_1' + 6\mu_2\mu_1'^2 + \mu_1'^4 \\&= 142.3125 + 4(39.75)(2.5) + 6(14.75)(2.5)^2 + (2.5)^4 \\&= 142.3125 + 397.5 + 553.125 + 39.0625 \\&= 1132.\end{aligned}$$

*Moments about the point  $x = 2$ .* We have  $\bar{x} = A + \mu_1'$ . Taking  $A = 2$ , the first moment about the point  $x = 2$  is

$$\mu_1' = \bar{x} - 2 = 2.5 - 2 = 0.5$$

Hence

$$\begin{aligned}\mu_2' &= \mu_2 + \mu_1'^2 = 14.75 + 0.25 = 15 \\ \mu_3' &= \mu_3 + 3\mu_2\mu_1' + \mu_1'^3 = 39.75 + 3(14.75)(0.5) + (0.5)^3 \\&= 39.75 + 22.125 + 0.125 = 62 \\ \mu_4' &= \mu_4 + 4\mu_3\mu_1' + 6\mu_2\mu_1'^2 + \mu_1'^4 \\&= 142.3125 + 4(39.75)(0.5) + 6(14.75)(0.5)^2 + (0.5)^4 \\&= 142.3125 + 79.5 + 22.125 + 0.0625 \\&= 244\end{aligned}$$

**Example 3-9.** Calculate the first four moments of the following distribution about the mean and hence find  $\beta_1$  and  $\beta_2$ .

$x:$	0	1	2	3	4	5	6	7	8
$f:$	1	8	28	56	70	56	28	8	1

**Solution.****CALCULATION OF MOMENTS**

<i>x</i>	<i>f</i>	<i>d = x - 4</i>	<i>fd</i>	<i>fd<sup>2</sup></i>	<i>fd<sup>3</sup></i>	<i>fd<sup>4</sup></i>
0	1	-4	-4	16	-64	256
1	8	-3	-24	72	-216	648
2	28	-2	-56	112	-224	448
3	56	-1	-56	56	-56	56
4	70	0	0	0	0	0
5	56	1	56	56	56	56
6	28	2	56	112	224	448
7	8	3	24	72	216	648
8	1	4	4	16	64	256
Total	256	0	0	512	0	2,816

Moments about the points  $x = 4$  are

$$\mu_1' = \frac{1}{N} \sum fd = 0, \quad \mu_2' = \frac{1}{N} \sum fd^2 = \frac{512}{256} = 2,$$

$$\mu_3' = \frac{1}{N} \sum fd^3 = 0 \text{ and } \mu_4' = \frac{1}{N} \sum fd^4 = \frac{2816}{256} = 11$$

Moments about mean are :

$$\mu_1 = 0, \quad \mu_2 = \mu_2' - \mu_1'^2 = 2$$

$$\mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2\mu_1'^3 = 0$$

$$\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'\mu_1'^2 - 3\mu_1'^4 = 11$$

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = 0, \quad \beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{11}{4} = 2.75$$

**Example 3-10** For a distribution the mean is 10, variance is 16,  $\gamma_1$  is +1 and  $\beta_2$  is 4. Obtain the first four moments about the origin, i.e., zero. Comment upon the nature of distribution.

**Solution.** We are given

$$\text{Mean} = 10, \quad \mu_2 = 16, \quad \gamma_1 = +1, \quad \beta_2 = 4$$

First four moments about origin ( $\mu_1'$ ,  $\mu_2'$ ,  $\mu_3'$ ,  $\mu_4'$ )

$$\mu_1' = \text{First moment about origin} = \text{Mean} = 10$$

$$\mu_2 = \mu_2' - \mu_1'^2 \Rightarrow \mu_2' = \mu_2 + \mu_1'^2 \Rightarrow \mu_2' = 16 + 10^2 = 116$$

$$\text{we have } \gamma_1 = +1 \Rightarrow \frac{\mu_3}{\mu_2^{3/2}} = 1$$

$$\Rightarrow \mu_3 = \mu_2^{3/2} = (16)^{3/2} = 4^3 = 64$$

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

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

$$= 64 + 3 \times 116 \times 10 - 2 \times 1000 = 3544 - 2000 = 1544$$

$$\text{Now } \beta_2 = \frac{\mu_4}{\mu_2^2} = 4 \Rightarrow \mu_4 = 4 \times 16^2 = 1024$$

$$\text{and } \mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'\mu_1'^2 - \mu_1'^4$$

$$\Rightarrow \mu_4' = 1024 + 4 \times 1544 \times 10 - 6 \times 116 \times 100 + 3 \times 10000 \\ = 92784 - 69600 = 23184.$$

*Comments of Nature of distribution:* [c.f. § 3.13 and § 3.14] Since  $\gamma_1 = +1$ , the distribution is moderately positively skewed, i.e., if we draw the curve for the given distribution, it will have longer tail towards the right. Further since  $\beta_2 = 4 > 3$ , the distribution is leptokurtic, i.e., it will be more peaked than the normal curve.

**Example 3.1.1.** If for a random variable  $x$ , the absolute moment of order  $k$  exists for ordinary  $k = 1, 2, \dots, n-1$ , then the following inequalities

$$(i). \beta_k^{2k} \leq \beta_{k-1}^k \cdot \beta_{k+1}^k, \quad (ii). \beta_k^k \leq \beta_{k+1}^{k+1}$$

holds for  $k=1, 2, \dots, n-1$ , where  $\beta_k$  is the  $k$ th absolute moment about the origin.

[Delhi Univ. B.Sc. (Stat.Hons.) 1989]

**Solution.** If  $x_i | f_i, i = 1, 2, \dots, n$  is the given frequency distribution, then

$$\beta_k = \frac{1}{N} \sum f_i | x_i^k | \quad \dots(1)$$

Let  $u$  and  $v$  be arbitrary real numbers, then the expression

$$\sum_{i=1}^n f_i \left[ u |x_i^{(k-1)/2}| + v |x_i^{(k+1)/2}| \right]^2 \text{ is non-negative.}$$

$$\Rightarrow \sum_{i=1}^n f_i \left[ u |x_i|^{(k-1)/2} + v |x_i|^{(k+1)/2} \right]^2 \geq 0$$

$$\Rightarrow u^2 \sum f_i |x_i|^{k-1} + v^2 \sum f_i |x_i|^{k+1} + 2uv \sum f_i |x_i|^k \geq 0$$

Dividing throughout by  $N$  and using relation (1), we get

$$u^2 \beta_{k-1} + v^2 \beta_{k+1} + 2uv \beta_k \geq 0, \text{ i.e., } u^2 \beta_{k-1} + 2uv \beta_k + v^2 \beta_{k+1} \geq 0 \quad \dots(2)$$

We know that the condition for the expression  $a.x^2 + 2axy + b.y^2$  to be non-negative for all values of  $x$  and  $y$  is that

$$\begin{vmatrix} a & h \\ h & b \end{vmatrix} \geq 0$$

Using this result, we get from (2)

$$\begin{vmatrix} \beta_{k-1} & \beta_k \\ \beta_k & \beta_{k+1} \end{vmatrix} \geq 0$$

$$\Rightarrow \beta_{k-1} \cdot \beta_{k+1} - \beta_k^2 \geq 0 \quad \dots(3)$$

Raising both sides of (3) to power  $k$ , we get

$$\beta_k^{2k} \geq \beta_{k-1}^k \cdot \beta_{k+1}^k \quad \dots(4)$$

Putting  $k=1, 2, \dots, k-1, k$  successively in (4), we get

$$\beta_1^2 \leq \beta_0 \beta_2$$

$$\beta_2^4 \leq \beta_1^2 \beta_3^2$$

$$\beta_3^6 \leq \beta_2^3 \beta_4^3$$

.....

$$\beta_{k-1}^{2(k-1)} \leq \beta_{k-2}^{k-1} \cdot \beta_k^{k-1}$$

$$\beta_k^{2k} \leq \beta_{k-1}^k \beta_{k+1}^k$$

Multiplying these inequalities and noting that  $\beta_0 = 1$ , we get

$$\beta_k^{k+1} \leq \beta_{k+1}^k \text{ for } k = 1, 2, \dots, n-1.$$

Raising both sides of the inequality to the power  $\frac{1}{k(k+1)}$ , we get

$$\beta_k^{1/k} \leq \beta_{k+1}^{1/(k+1)} \quad \dots (5)$$

**Remark.** Result (5) shows that  $\beta_k^{1/k}$  is an increasing function of  $k$ .

### EXERCISE 3 (b)

1. (a) Define the raw and central moments of a frequency distribution. Obtain the relation between the central moments of order  $r$  in terms of the raw moments. What are Sheppard's corrections to the central moments?

(b) Define moments. Establish the relationship between the moments about mean, i.e., Central Moments in terms of moments about any arbitrary point and vice versa.

The first three moments of a distribution about the value 2 of the variable are 1, 16 and -40. Show that the mean is 3, the variance is 15 and  $\mu_3 = -86$ . Also show that the first three moments about  $x = 0$  are 3, 24 and 76.

(c) For a distribution the mean is 10, variance is 16,  $\gamma_1$  is +1 and  $\beta_2$  is 4. Find the first four moments about the origin.

**Ans.**  $\mu_1' = 10$ ,  $\mu_2' = 116$ ,  $\mu_3' = 1544$  and  $\mu_4' = 23184$ .

(d) (i) Define 'moment'. What is its use? Express first four central moments in terms of moments about the origin. What is the effect of change of origin and scale on  $\mu_3$ ?

(ii) The first three moments of a distribution about the point  $X = 7$  are 3, 11 and 15 respectively. Obtain mean, variance and  $\beta_1$ .

2. The first four moments of distribution about the value 5 of the variable are 2, 20, 40 and 50. Obtain as far as possible, the various characteristics of the distribution on the basis of the information given.

**Ans.** Mean = 7,  $\mu_2 = 16$ ,  $\mu_3 = -64$ ,  $\mu_4 = 162$ ,  $\beta_1 = 1$  and  $\beta_2 = 0.63$ .

3. (a) If the first four moments of a distribution about the value 5 are equal to -4, 22, -117 and 560, determine the corresponding moments (i) about the mean, (ii) about zero.

(b) What is Sheppard's correction? What will be the corrections for the first four moments?

The first four moments of a distribution about  $x = 4$  are 1, 4, 10, 45. Show that the mean is 5 and the variance is 3 and  $\mu_3$  and  $\mu_4$  are 0 and 26 respectively,

(c) In certain distribution, the first four moments about the point 4 are -1.5, 17, -13 and 308. Calculate  $\beta_1$  and  $\beta_2$ .

(d) The first four moments of a frequency distribution about the point 5 are -0.55, 4.46, -0.43 and 68.52. Find  $\beta_1$  and  $\beta_2$ .

**Ans.**  $\mu_2 = 4.1575$ ,  $\mu_3 = 6.5962$ ,  $\mu_4 = 75.3944$ ,  $\beta_1 = 0.6055$ ,  $\beta_2 = 4.3619$ .

4. (a) For the following data, calculate (i) Mean, (ii) Median, (iii) Semi-inter-quartile range, (iv) Coefficient of variation, and (v)  $\beta_1$  and  $\beta_2$  coefficients.

Wages in Rupees :	170—	180—	190—	200—	210—	220—	230—	240—
No. of Persons :	52	68	85	92	100	95	70	28

Ans. Mean = 209 (approx.); Median = 209.8; Q.D. = 15.8;  $\sigma = 19.7$ ; C.V. = 9.4;  $\beta_1 = 0.003$ ;  $\beta_2 = 26.105$ .

(b) Find the second, third and fourth central moments of the frequency distribution given below. Hence find (i) a measure of skewness ( $\gamma_1$ ) and (ii) measure of kurtosis ( $\gamma_2$ ).

Class Limits	Frequency
100.0 – 114.9	5
115.0 – 119.9	15
120.0 – 124.9	20
125.0 – 129.9	35
130.0 – 134.9	10
135.0 – 139.9	10
140.0 – 144.9	5

Also apply Sheppard's corrections for moments.

Ans.  $\mu_2 = 2.16$ ,  $\mu_3 = 0.804$ ,  $\mu_4 = 12.5232$

$$\gamma_1 = \sqrt{\beta_1} = 0.25298; \gamma_2 = \beta_2 - 3 = -0.317.$$

(c) The standard deviation of a symmetrical distribution is 5. What must be the value of the fourth moment about the mean in order that the distribution be (i) leptokurtic, (ii) mesokurtic, and (iii) platykurtic?

Hint :  $\mu_1 = \mu_3 = 0$  (because distribution is symmetrical).

$$\sigma = 5 \Rightarrow \sigma^2 = \mu_2 = 25$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{\mu_4}{625}$$

(i) Distribution is leptokurtic if  $\beta_2 > 3$ , i.e., if  $\frac{\mu_4}{625} > 3 \Rightarrow \mu_4 > 1875$

(ii) Distribution is mesokurtic if  $\beta_2 = 3 \Rightarrow \text{if } \mu_4 = 1875$

(iii) Distribution is platykurtic if  $\beta_2 < 3 \Rightarrow \text{if } \mu_4 < 1875$

5. Show that for discrete distribution  $\beta_2 > 1$ .

[Allahabad Univ. M.A., 1993; Delhi Univ. B.Sc. (Stat. Hons), 1992]

Hint. We have to show that  $\mu_4/\mu_2^2 > 1$ , i.e.,  $\mu_4 > \mu_2^2$ . If  $x_i$  /  $f_i$ ,  $i = 1, 2, \dots$ ,

$n$ , be the given discrete distribution, then we have to prove that

$$\frac{1}{N} \sum_i f_i (x_i - \bar{x})^4 > \left( \frac{1}{N} \sum_i f_i (x_i - \bar{x})^2 \right)^2$$

Putting  $(x_i - \bar{x})^2 = z_i$ , we have to show that

$$\frac{1}{N} \sum_i f_i z_i^2 > \left( \frac{1}{N} \sum_i f_i z_i \right)^2$$

$$\text{i.e., } \frac{1}{N} \sum_i f_i z_i^2 - \left( \frac{1}{N} \sum_i f_i z_i \right)^2 > 0$$

$$\text{i.e., } \sigma_z^2 > 0,$$

which is always true, since variance is always positive.

Hence  $\beta_2 > 1$ .

6. (a) The scores in Economics of 250 candidates appearing at an examination have

$$\text{Mean} = \bar{x} = 39.72$$

$$\text{Variance} = \sigma^2 = 97.80$$

$$\text{Third Central moment} = \mu_3 = +114.18$$

$$\text{Fourth central moment} = \mu_4 = 28,396.14$$

It was later found on scrutiny that the score 61 of a candidate has been wrongly recorded as 51. Make necessary corrections in the given values of the mean and the central moment. (Gujarat Univ. M.A., 1993)

(b) For a distribution of 250 heights, calculations showed that the mean, standard deviation,  $\beta_1$  and  $\beta_2$  were 54 inches, 3 inches 0 and 3 inches respectively. It was, however, discovered on checking that the two items 64 and 52 in the original data were wrongly written in place of correct values 62 and 52 inches respectively. Calculate the correct frequency constants.

Ans. Correct Mean = 54, S.D. = 2.97,  $\mu_3 = -2.18$ ,  $\mu_4 = 218.42$ ,  $\beta_1 = 0.0070$  and  $\beta_2 = 2.81$

7. In calculating the moments of a frequency distribution based on 100 observations, the following results are obtained :

$$\text{Mean} = 9, \text{Variance} = 19, \beta_1 = 0.7 (\mu_3 + \text{ive}), \beta_2 = 4$$

But later on it was found that one observation 12 was read as 21. Obtain the correct value of the first four central moments.

Ans. Corrected mean = 8.91,  $\mu_2 = 17.64$ ,  $\mu_3 = 57.05$ ,  $\mu_4 = 1257.15$ ,  $\beta_1 = 0.59$  and  $\beta_2 = 4.04$ .

8. (a) Show that if a range of six times the standard deviation covers at least 18 class intervals, Sheppard's correction will make a difference of less than 0.5 percent in the corrected value of the standard deviation.

**Hint.** If  $h$  is the magnitude of the class interval, then we want :

$$6\sigma > 18h \Rightarrow \sigma > 3h \Rightarrow h^2 < \frac{1}{9}\sigma^2 \Rightarrow -h^2 > -\frac{1}{9}\sigma^2$$

$$\therefore \mu_2(\text{corrected}) = \mu_2 - \frac{h^2}{12} \geq \sigma^2 - \frac{1}{9 \times 12} \sigma^2 = \sigma^2 \left( 1 - \frac{1}{108} \right)$$

$$\Rightarrow \text{s.d. (corrected)} \geq \sigma \left( 1 - \frac{1}{108} \right)^{1/2} = \sigma \left( 1 - \frac{1}{2} \times \frac{1}{108} \right)$$

$$\therefore \text{Required adjustment} = \sigma - \sigma(\text{corrected}) < \frac{\sigma}{216} < \frac{\sigma}{200} = \frac{1}{2} \% \text{ of s.d.}$$

(b) Show that, if the class intervals of a grouped distribution is less than one-third of the calculated standard deviation, Sheppard's adjustment makes a difference of less than  $\frac{1}{2}\%$  in the estimate of the standard deviation

9. (a) If  $\delta_r$  is the  $r$ th absolute moment about zero, use the mean value of

$$[u|x|^{(r-1)/2} + v|x|^{(r+1)/2}]^2$$

to show that

$$(\delta_r)^{2r} \leq (\delta_{r-1})^r (\delta_{r+1})^r$$

From this derive the following inequalities :

$$(i) (\delta_r)^{r+1} \leq (\delta_{r+1})^r, (ii) (\delta_r)^{1/r} \leq (\delta_{r+1})^{1/(r+1)}$$

(b) For a random variable  $X$  moments of all order exist. Denoting by  $\mu_j$  and  $\delta_j$ , the  $j$ th central moment and  $j$ th absolute moment respectively, show that

$$(i) (\mu_{2j+1})^2 \leq \mu_{2j} \mu_{2j+2},$$

$$(ii) (\delta_j)^{1/j} \leq (\delta_{j+1})^{1/(j+1)} \quad (\text{Karnataka Univ. B.Sc., 1993})$$

10. If  $\beta_1$  and  $\beta_2$  are the Pearson's coefficients of skewness and kurtosis respectively, show that  $\beta_2 > \beta_1 + 1$ . (Bangalore Univ. B.Sc., 1993)

3.13. Skewness. Literally, skewness means '*lack of symmetry*'. We study skewness to have an idea about the shape of the curve which we can draw with the help of the given data. A distribution is said to be skewed if

(i) Mean, median and mode fall at different points,

i.e., Mean  $\neq$  Median  $\neq$  Mode,

(ii) Quartiles are not equidistant from median, and

(iii) The curve drawn with the help of the given data is not symmetrical but stretched more to one side than to the other.

**Measures of Skewness.** Various measures of skewness are

$$(1) S_k = M - M_d \quad (2) S_k = M - M_0,$$

where  $M$  is the mean,  $M_d$ , the median and  $M_0$ , the mode of the distribution.

$$(3) S_k = (Q_3 - M_d) - (M_d - Q_1).$$

These are the absolute measures of skewness. As in dispersion, for comparing two series we do not calculate these absolute measures but we calculate the relative measures called the *co-efficients of skewness* which are pure numbers independent of units of measurement. The following are the *coefficients of Skewness*.

1. Prof. Karl Pearson's Coefficient of Skewness.

$$S_k = \frac{(M - M_0)}{\sigma} \quad \dots (3.27)$$

where  $\sigma$  is the standard deviation of the distribution.

If mode is ill-defined, then using the relation,  $M_0 = 3M_d - 2M$ , for a moderately asymmetrical distribution, we get

$$S_k = \frac{3(M - M_d)}{\sigma} \quad \dots (3.27a)$$

The limits for Karl Pearson's coefficient of skewness are  $\pm 3$ . In practice, these limits are rarely attained.

Skewness is positive if  $M > M_d$  or  $M < M_d$  and negative if  $M < M_d$  or  $M > M_d$ .

### II. Prof. Bowley's Coefficient of Skewness. Based on quartiles,

$$S_k = \frac{(Q_3 - M_d) - (M_d - Q_1)}{(Q_3 - M_d) + (M_d - Q_1)} = \frac{Q_3 + Q_1 - 2M_d}{Q_3 - Q_1} \quad \dots (3-28)$$

**Remarks 1.** Bowley's coefficient of skewness is also known as *Quartile coefficient of skewness* and is especially useful in situations where quartiles and median are used, viz.,

(i) When the mode is ill-defined and extreme observations are present in the data.

(ii) When the distribution has open end classes or unequal class intervals.

In these situations Pearson's coefficient of skewness cannot be used.

2. From (3-28), we observe that

$$S_k = 0, \text{ if } Q_3 - M_d = M_d - Q_1$$

This implies that for a symmetrical distribution ( $S_k = 0$ ), median is equidistant from the upper and lower quartiles. Moreover skewness is positive if :

$$Q_3 - M_d > M_d - Q_1 \Rightarrow Q_3 + Q_1 > 2M_d$$

and skewness is negative if

$$Q_3 - M_d < M_d - Q_1 \Rightarrow Q_3 + Q_1 < 2M_d$$

3. *Limits for Bowley's Coefficient of Skewness.* We know that for two real positive numbers  $a$  and  $b$  (i.e.,  $a > 0$  and  $b > 0$ ), the modulus value of the difference  $(a - b)$  is always less than or equal to the modulus value of the sum  $(a + b)$ , i.e.,

$$|a - b| \leq |a + b| \Rightarrow \left| \frac{a - b}{a + b} \right| \leq 1 \quad \dots (*)$$

We also know that  $(Q_3 - M_d)$  and  $(M_d - Q_1)$  are both non-negative. Thus, taking  $a = Q_3 - M_d$  and  $b = M_d - Q_1$  in (\*), we get

$$\begin{aligned} & \left| \frac{(Q_3 - M_d) - (M_d - Q_1)}{(Q_3 - M_d) + (M_d - Q_1)} \right| \leq 1 \\ \Rightarrow & |S_k (\text{Bowley})| \leq 1 \\ \Rightarrow & -1 \leq S_k (\text{Bowley}) \leq 1. \end{aligned}$$

Thus, Bowley's coefficient of skewness ranges from -1 to 1.

Further, we note from (3-28) that :

$$S_k = +1, \text{ if } M_d - Q_1 = 0, \text{ i.e., if } M_d = Q_1$$

$$S_k = -1, \text{ if } Q_3 - M_d = 0, \text{ i.e., if } Q_3 = M_d.$$

4. It should be clearly understood that the values of the coefficients of skewness obtained by Bowley's formula and Pearson's formula are not comparable, although in each case,  $S_k = 0$  implies the absence of skewness, i.e., the distribution is symmetrical. It may even happen that one of them gives positive skewness while the other gives negative skewness.

5. In Bowley's coefficient of skewness the disturbing factor of variation is eliminated by dividing the absolute measure of skewness, viz.,  $(Q_3 - M_d) - (M_d - Q_1)$  by the measure of dispersion  $(Q_3 - Q_1)$ , i.e., quartile range.

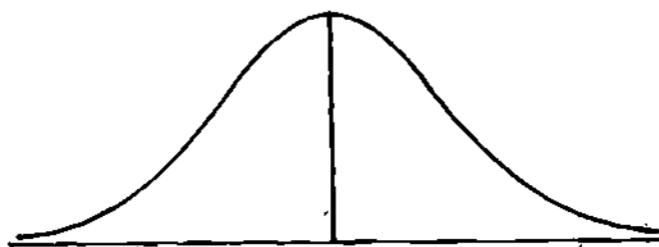
6. The only and perhaps quite serious limitations of this coefficient is that it is based only on the central 50% of the data and ignores the remaining 50% of the data towards the extremes.

### III. Based upon moments, co-efficient of skewness is

$$S_k = \frac{\sqrt{\beta_1} (\beta_2 + 3)}{2 (5\beta_2 - 6\beta_1 - 9)} \quad \dots(3-29)$$

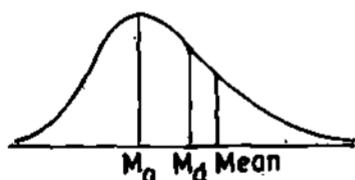
where symbols have their usual meaning. Thus  $S_k = 0$  if either  $\beta_1 = 0$  or  $\beta_2 = -3$ . But since  $\beta_2 = \mu_4/\mu_2^2$ , cannot be negative,  $S_k = 0$  if and only if  $\beta_1 = 0$ . Thus for a symmetrical distribution  $\beta_1 = 0$ . In this respect  $\beta_1$  is taken to be a measure of skewness. The co-efficient, in (3-29) is to be regarded as without sign.

We observe in (3-27) and (3-28) that skewness can be positive as well as negative. The skewness is positive if the larger tail of the distribution lies towards

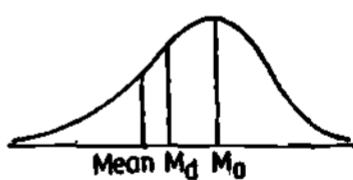


$$\bar{x} \text{ (Mean)} = M_0 = M_d \\ (\text{Symmetrical Distribution})$$

the higher values of the variate (the right), i.e., if the curve drawn with the help of the given data is stretched more to the right than to the left and is negative



(Positively Skewed Distribution)



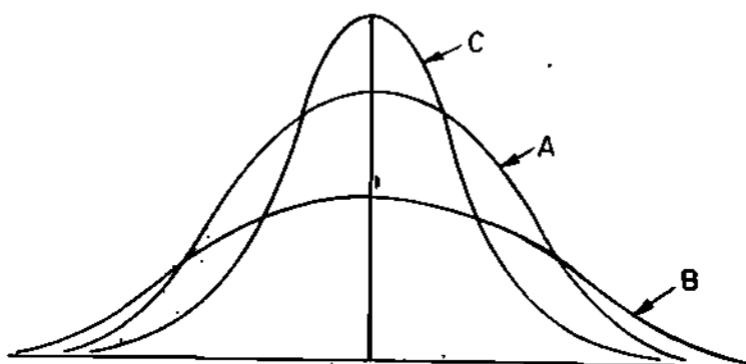
(Negatively Skewed Distribution)

in the contrary case.

**3-14. Kurtosis.** If we know the measures of central tendency, dispersion and skewness, we still cannot form a complete idea about the distribution as will be clear from the following figure in which all the three curves A, B and C are symmetrical about the mean ' $m$ ' and have the same range.

In addition to these measures we should know one more measure which Prof. Karl Pearson calls as the 'Convexity of curve' or *Kurtosis*. Kurtosis enables us to have an idea about the flatness or peakedness of the curve. It is measured by the co-efficient  $\beta_2$  or its derivation  $\gamma_2$  given by

$$\beta_2 = \mu_4/\mu_2^2, \gamma_2 = \beta_2 - 3$$



Curve of the type 'A' which is neither flat nor peaked is called the *normal curve* or *mesokurtic curve* and for such a curve  $\beta_2 = 3$ , i.e.,  $\gamma_2 = 0$ . Curve of the type 'B' which is flatter than the normal curve is known as *platykurtic* and for such a curve  $\beta_2 < 3$ , i.e.,  $\gamma_2 < 0$ . Curve of the type 'C' which is more peaked than the normal curve is called *leptokurtic* and for such a curve  $\beta_2 > 3$ , i.e.,  $\gamma_2 > 0$ .

### EXERCISE 3 (c)

1. What do you understand by skewness? How is it measured? Distinguish clearly, by giving figures, between positive and negative skewness.
2. Explain the methods of measuring skewness and kurtosis of a frequency distribution.
3. Show that for any frequency distribution :
  - (i) Kurtosis is greater than unity.
  - (ii) Co-efficient of skewness is less than 1 numerically.
4. Why do we calculate in general, only the first four moments about mean of a distribution and not the higher moments?
5. (a) Obtain Karl Pearson's measure of skewness for the following data:

Values	Frequency	Values	Frequency
5 - 10	6	25 - 30	15
10 - 15	8	30 - 35	11
15 - 20	17	35 - 40	2
20 - 25	21		

(b) Assume that a firm has selected a random sample of 100 from its production line and has obtain the data shown in the table below :

Class interval	Frequency	Class interval	Frequency
130 - 134	3	150 - 154	19
135 - 139	12	155 - 159	12
140 - 144	21	160 - 164	5
145 - 149	28		

Compute the following :

- (a) The arithmetic mean, (b) the standard deviation,
- (c) Karl Pearson's coefficient of skewness.

Ans. (a) 147.2, (b) 7.2083, (c) 0.0711

6. (a) For the frequency distribution given below, calculate the coefficient of skewness based on quartiles.

Annual Sales (Rs. '000)	No. of Firms	Annual Sales (Rs. '000)	No. of firms
Less than 20	30	Less than 70	644
Less than 30	225	Less than 80	650
Less than 40	465	Less than 90	665
Less than 50	580	Less than 100	680
Less than 60	634		

(b) (i) Karl Pearson's coefficient of skewness of a distribution is 0.32, its s.d. is 6.5 and mean is 29.6. Find the mode of the distribution.

(ii) If the mode of the above distribution is 24.8, what will be the s.d. ?

7. (a) In a frequency distribution, the co-efficient of skewness based upon the quartiles is 0.6. If the sum of the upper and lower quartiles is 100 and median is 38, find the value of the upper and lower quartiles.

Hint. We are given

$$S_k = \frac{Q_3 + Q_1 - 2 \text{Md}}{Q_3 - Q_1} = 0.6 \quad \dots (*)$$

Also  $Q_3 + Q_1 = 100$  and Median = 38

Substituting (\*), we get

$$\frac{100 - 2 \times 38}{Q_3 - Q_1} = 0.6$$

$$\Rightarrow Q_3 - Q_1 = 40$$

Simplifying we get  $Q_1 = 30, Q_3 = 70$

- (b) A frequency distribution gives the following results :  
 (i) C.V. = 5              (ii) Karl Pearson's co-efficient of skewness = 0.5  
 (iii)  $\sigma = 2$ .

Find the mean and mode of the distribution.

- (c) find the C.V. of a frequency distribution given that its mean is 120, mode is 123 and Karl Pearson's co-efficient of skewness is -0.3.

Ans. C.V. = 8.33

- (d) The first three moments of distribution about the value 2 are 1, 16 and 40 respectively. Examine the skewness of the distribution.

8. The first three moments about the origin 51 Kg, calculated from the data on the weights of 25 college students are

$$\mu_1' = +0.4 \text{ kg.}, \sqrt{\mu_2'} = 1.2 \text{ kg. and } (\mu_3')^{1/2} = -0.25 \text{ kg.}$$

Determine the mean, the standard deviation and coefficient of skewness.

9. The first three moments about the origin are given by

$$\mu_1' = \frac{n+1}{2}, \mu_2' = \frac{(n+1)(2n+1)}{6} \text{ and } \mu_3' = \frac{n(9n+1)^2}{4}$$

Examine the skewness of the data.

10. Find out the kurtosis of the data given below :

<i>Class interval</i>	0 – 10	10 – 20	20 – 30	30 – 40
<i>Frequency</i>	1	3	4	2

11. Data were obtained for distribution of passengers, entering Bombay local trains over time at intervals of 15 minutes for morning and evening rush hours separately, and the following results were obtained.

	<i>Morning hours</i>	<i>Evening hours</i>
Arithmetic mean (Peak Hours)	8 hrs. 38 min.	5 hrs. 40 min.
Standard deviation	38.5 min.	34.9 min.
Coefficient of skewness (in 15 min. unit)	- 0.32	+ 0.17
Kurtosis measure	2.0	2.2

Interpret the result and discuss giving reasons, whether you approve of the measure of 'peak hour'.

12. (a) The standard deviation of a symmetrical distribution is 5. What must be the value of the fourth moment about the mean in order that the distribution be (i) Leptokurtic, (ii) mesokurtic, and (iii) platykurtic.

Hint.  $\mu_1 = \mu_3 = 0$  (Because distribution is symmetrical),  $\sigma = 5 \Rightarrow \sigma^2 = \mu_2 = 25$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{\mu_4}{625}$$

- (i) Distt. is leptokurtic if  $\beta_2 > 3$  i.e., if  $\frac{\mu_4}{625} > 3 \Rightarrow \mu_4 > 1875$   
 (ii) Distt. is mesokurtic if  $\beta_2 = 3 \Rightarrow \text{if } \frac{\mu_4}{625} < 1875$

(iii) Distt. is platykurtic if  $\beta_2 < 3 \Rightarrow \mu_4 < 1875$ .

(b) Find the second, third and fourth central moments of the frequency distribution given below. Hence, find (i) a measure of skewness, and (ii) a measure of kurtosis ( $\gamma_2$ ).

<i>Class limits</i>	<i>Frequency</i>
110.0 — 114.9	5
115.0— 119.9	15
120.0— 124.9	20
125.0— 129.9	35
130.0— 134.9	10
135.0— 139.9	10
140.0— 144.9	5

Ans.  $\mu_2 = 2.16$ ,  $\mu_3 = 0.804$ ,  $\mu_4 = 12.5232$ .

$$\gamma_1 = \sqrt{\beta_1} = 0.25298 ; \gamma_2 = \beta_2 - 3 = -0.317.$$

13. (a) Define Pearsonian coefficients  $\beta_1$  and  $\beta_2$  and discuss their utility in statistics. [Delhi Univ. B.Sc. (Hons.), 1993]

(b) What do you mean by skewness and kurtosis of a distribution? Show that the Pearson's Beta coefficients satisfy the inequality  $\beta_2 - \beta_1 - 1 \geq 0$ . Also deduce that  $\beta_2 \geq 1$ . [Delhi Univ. B.Sc. (Stat. Hons.), 1991]

(c) Define the Pearson's coefficients  $\gamma_1$  and  $\gamma_2$  and discuss their utility in Statistics.

### OBJECTIVE TYPE QUESTIONS

I. Match the correct parts to make a valid statement.

(a) Range (i)  $(Q_3 - Q_1)/2$

(b) Quartile Deviation (ii)  $\sqrt{\frac{1}{N} \sum f_i (x_i - \bar{x})^2}$

(c) Mean Deviation (iii)  $\frac{S.D.}{Mean} \times 100$

(d) Standard Deviation (iv)  $\frac{1}{N} \sum f_i |(x_i - \bar{x})|$

(e) Coefficient of Variation (v)  $X_{max} - X_{min}$

II. Which value of 'a' gives the minimum?

(i) Mean square deviation from 'a'

(ii) Mean deviation from 'a'

III. Mean of 100 observations is 50 and S.D. is 10. What will be the new mean and S.D., if

(i) 5 is added to each observation,

(ii) each observation is multiplied by 3,

(iii) 5 is subtracted from each observation and then it is divided by 4?

IV. Fill in the blanks :

(i) (a) Absolute sum of deviation is minimum from.....

- (b) Least value of root mean square deviation is .....
- (ii) The sum of squares of deviations is least when measured from .....
- (iii) The sum of 10 items is 12 and the sum of their squares is 16.9.
- (iv) In any distribution, the standard deviation is always ..... the mean deviation.
- (v) The relationship between root mean square deviation and standard deviation  $\sigma$  is .....
- (vi) If 25% of the items are less than 10 and 25% are more than 40, the coefficient of quartile deviation is .....
- (vii) The median and standard deviation of a distribution are 20 and 4 respectively. If each item is increased by 2, the median will be ..... and the new standard deviation will be .....
- (viii) In a symmetric distribution, the mean and the mode are .....
- (ix) In symmetric distribution, the upper and the lower quartiles are equidistant from .....
- (x) If the mean, mode and standard deviation of a frequency distribution are 41, 45 and 8 respectively, then its Pearson's coefficient of skewness is .....
- (xi) For a symmetrical distribution  $\beta_1 = \dots$
- (xii) If  $\beta_2 > 3$  the distribution is said to be .....
- (xiii) For a symmetric distribution  $\mu_2 = \dots$   
 $\mu_{2n+1} = \dots$
- (xiv) If the mean and the mode of a given distribution are equal then its coefficient of skewness is .....
- (xv) If the kurtosis of a distribution is 3, it is called ..... distribution.
- (xvi) In a perfectly symmetrical distribution 50% of items are above 60 and 75% items are below 75. Therefore, the coefficient of quartile deviation is ..... and coefficient of skewness is .....
- (xvii) Relation between  $\beta_1$  and  $\beta_2$  is given by .....
- V. For the following questions give correct answers :
- (i) Sum of absolute deviations about median is  
 (a) Least, (b) greatest, (c) zero, (d) equal.
- (ii) The sum of squares of deviations is least when measured from  
 (a) Median, (b), (c) Mean, (d) Mode, (e) none of them.
- (iii) In any discrete series (when all the values are not same) the relationship between M.D. about mean and S.D. is  
 (a)  $M.D. = S.D.$ , (b)  $M.D. \geq S.D.$ , (c)  $M.D. < S.D.$ ,  
 (d)  $M.D. \leq S.D.$ , (e) None of these.
- (iv) If each of a set of observations of a variable is multiplied by a constant (non-zero) value, the variance of the resultant variable.  
 (a) is unaltered, (b) increases (c) decreases, (d) is unknown.

- (v) The appropriate measure whenever the extreme items are to be disregarded and when the distribution contains indefinite classes at the end is  
 (a) Median, (b) Mode, (c) Quartile deviation,  
 (d) Standard Deviation
- (vi) A.M., G.M. and H.M. in any series are equal when  
 (a) the distribution is symmetric, (b) all the values are same,  
 (c) the distribution is positively skewed,  
 (d) the distribution is unimodal.
- (vii) The limits for quartile coefficient of skewness are  
 (a)  $\pm 3$ , (b) 0 and 3, (c)  $\pm 1$ , (d)  $\pm \infty$
- (viii) The statement that the variance is equal to the second central moment  
 (a) always true, (b) sometimes true, (c) never true,  
 (d) ambiguous.
- (ix) The standard deviation of a distribution is 5. The value of the fourth central moment ( $\mu_4$ ), in order that the distribution be mesokurtic should be  
 (a) equal to 3, (b) greater than 1,875, (c) equal to 1,875,  
 (d) less than 1,875.
- (x) In a frequency curve of scores the mode was found to be higher than the mean. This shows that the distribution is  
 (a) Symmetric, (b) negatively skewed, (c) positively skewed,  
 (d) normal.
- (xi) For any frequency distribution, the kurtosis is  
 (a) greater than 1, (b) less than 1, (c) equal to 1.
- (xii) The measure of kurtosis is  
 (a)  $\beta_2 = 0$ , (b)  $\beta_2 = 3$ , (c)  $\beta_2 = 4$ .
- (xiii) For the distribution  
 (a)  $\mu_4 = 0$ , (b) Median = 0,  
 (c) The distribution of  $x$  is symmetrical.

$X:$	-4	-3	-2	-1	0	1	2	3	4	Total
$f:$	2	4	5	7	10	7	5	4	2	46

- (xiv) For a symmetric distribution

(a)  $\mu_2 = 0$ , (b)  $\mu_2 > 0$ ; (c)  $\mu_3 > 0$

VI. State which of the following statements are True and which False. In each of false statements give the correct statement.

- (i) Mean, standard deviation and variance have the same unit.  
 (ii) Standard deviation of every distribution is unique and always exists.  
 (iii) Median is the value of the variance which divides the total frequency it two equal parts.  
 (iv) Mean - Mode = 3 (mean - median) is often approximately satisfied.  
 (v) Mean deviation =  $\frac{4}{\zeta}$  (standard deviation) is always satisfied.  
 (vi)  $\beta_2 \geq 1$  is always satisfied  
 (vii)  $\beta_1 = 0$  is a conclusive test for a distribution to be symmetrical.

*Theory of Probability*

**4.1. Introduction.** If an experiment is repeated under essentially homogeneous and similar conditions we generally come across two types of situations:

- (i) The result or what is usually known as the '*outcome*' is unique or certain.
- (ii) The result is not unique but may be one of the several possible outcomes.

The phenomena covered by (i) are known as '*deterministic*' or '*predictable*' phenomena. By a deterministic phenomenon we mean one in which the result can be predicted with certainty. For example :

- (a) For a perfect gas,

$$V \propto \frac{1}{P} \quad i.e., PV = \text{constant},$$

provided the temperature remains the same.

- (b) The velocity '*v*' of a particle after time '*t*' is given by

$$v = u + at$$

where *u* is the initial velocity and *a* is the acceleration. This equation uniquely determines *v* if the right-hand quantities are known.

- (c) Ohm's Law, viz.,  $C = \frac{E}{R}$

where *C* is the flow of current, *E* the potential difference between the two ends of the conductor and *R* the resistance, uniquely determines the value *C* as soon as *E* and *R* are given.

A deterministic model is defined as a model which stipulates that the conditions under which an experiment is performed determine the outcome of the experiment. For a number of situations the deterministic model suffices. However, there are phenomena [as covered by (ii) above] which do not lend themselves to deterministic approach and are known as '*unpredictable*' or '*probabilistic*' phenomena. For example :

- (i) In tossing of a coin one is not sure if a head or tail will be obtained.
- (ii) If a light tube has lasted for *t* hours, nothing can be said about its further life. It may fail to function any moment.

In such cases we talk of chance or probability which is taken to be a quantitative measure of certainty.

**4.2. Short History.** Galileo (1564-1642), an Italian mathematician, was the first to attempt at a quantitative measure of probability while dealing with some problems related to the theory of dice in gambling. But the first foundation of the mathematical theory of probability was laid in the mid-seventeenth century by two French mathematicians, B. Pascal (1623-1662) and P. Fermat (1601-1665), while

solving a number of problems posed by French gambler and noble man Chevalier-De-Mere to Pascal. The famous '*problem of points*' posed by De-Mere to Pascal is : "Two persons play a game of chance. The person who first gains a certain number of points wins the stake. They stop playing before the game is completed. How is the stake to be decided on the basis of the number of points each has won?" The two mathematicians after a lengthy correspondence between themselves ultimately solved this problem and this correspondence laid the first foundation of the science of probability. Next stalwart in this field was J. Bernoulli (1654-1705) whose '*Treatise on Probability*' was published posthumously by his nephew N. Bernoulli in 1713. De-Moivre (1667-1754) also did considerable work in this field and published his famous '*Doctrine of Chances*' in 1718. Other main contributors are : T. Bayes (Inverse probability), P.S. Laplace (1749-1827) who after extensive research over a number of years finally published '*Theorie analytique des probabilités*' in 1812. In addition to these, other outstanding contributors are Levy, Mises and R.A. Fisher.

Russian mathematicians also have made very valuable contributions to the modern theory of probability. Chief contributors, to mention only a few of them are : Chebyshev (1821-94) who founded the Russian School of Statisticians; A. Markoff (1856-1922); Liapounoff (Central Limit Theorem); A. Khintchine (Law of Large Numbers) and A. Kolmogórov, who axiomised the calculus of probability.

**4.3. Definitions of Various Terms.** In this section we will define and explain the various terms which are used in the definition of probability.

**Trial and Event.** Consider an experiment which, though repeated under essentially identical conditions, does not give unique results but may result in any one of the several possible outcomes. The experiment is known as a *trial* and the outcomes are known as *events* or *cases*. For example :

- (i) Throwing of a die is a trial and getting 1 (or 2 or 3, ... or 6) is an event.
- (ii) Tossing of a coin is a trial and getting head (*H*) or tail (*T*) is an event.
- (iii) Drawing two cards from a pack of well-shuffled cards is a trial and getting a king and a queen are events.

**Exhaustive Events.** The total number of possible outcomes in any trial is known as exhaustive events or exhaustive cases. For example :

- (i) In tossing of a coin there are two exhaustive cases, viz., head and tail (the possibility of the coin standing on an edge being ignored).
- (ii) In throwing of a die, there are six exhaustive cases since any one of the 6 faces 1, 2, ..., 6 may come uppermost.
- (iii) In drawing two cards from a pack of cards the exhaustive number of cases is  ${}^{52}C_2$ , since 2 cards can be drawn out of 52 cards in  ${}^{52}C_2$  ways.
- (iv) In throwing of two dice, the exhaustive number of cases is  $6^2 = 36$ , since any of the 6 numbers 1 to 6 on the first die can be associated with any of the six numbers on the other die.

In general in throwing of  $n$  dice the exhaustive number of cases is  $6^n$ .

**Favourable Events or Cases.** The number of cases favourable to an event in a trial is the number of outcomes which entail the happening of the event. For example,

(i) In drawing a card from a pack of cards the number of cases favourable to drawing of an ace is 4, for drawing a spade is 13 and for drawing a red card is 26.

(ii) In throwing of two dice, the number of cases favourable to getting the sum 5 is : (1,4) (4,1) (2,3) (3,2), i.e., 4.

**Mutually exclusive events.** Events are said to be *mutually exclusive* or *incompatible* if the happening of any one of them precludes the happening of all the others, i.e., if no two or more of them can happen simultaneously in the same trial. For example :

(i) In throwing a die all the 6 faces numbered 1 to 6 are mutually exclusive since if any one of these faces comes, the possibility of others, in the same trial, is ruled out.

(ii) Similarly in tossing a coin the events head and tail are mutually exclusive.

**Equally likely events.** Outcomes of a trial are set to be equally likely if taking into consideration all the relevant evidences, there is no reason to expect one in preference to the others. For example

(i) In tossing an unbiased or uniform coin, head or tail are equally likely events.

(ii) In throwing an unbiased die, all the six faces are equally likely to come.

**Independent events.** Several events are said to be independent if the happening (or non-happening) of an event is not affected by the supplementary knowledge concerning the occurrence of any number of the remaining events. For example

(i) In tossing an unbiased coin the event of getting a head in the first toss is independent of getting a head in the second, third and subsequent throws.

(ii) If we draw a card from a pack of well-shuffled cards and replace it before drawing the second card, the result of the second draw is independent of the first draw. But, however, if the first card drawn is not replaced then the second draw is dependent on the first draw.

### •3.1. Mathematical or Classical or 'a priori' Probability

**Definition.** If a trial results in  $n$  exhaustive, mutually exclusive and equally likely cases and  $m$  of them are favourable to the happening of an event  $E$ , then the probability ' $p$ ' of happening of  $E$  is given by

$$p = P(E) = \frac{\text{Favourable number of cases}}{\text{Exhaustive number of cases}} = \frac{m}{n} \quad \dots(4.1)$$

Sometimes we express (4.1) by saying that 'the odds in favour of  $E$  are  $m : (n - m)$ ' or the odds against  $E$  are  $(n - m) : n$ '.

Since the number of cases favourable to the 'non-happening' of the event  $E$  are. ( $n - m$ ), the probability ' $q$ ' that  $E$  will not happen is given by

$$q = \frac{n-m}{n} = 1 - \frac{m}{n} = 1 - p \Rightarrow p + q = 1 \quad \dots(4.1a)$$

Obviously  $p$  as well as  $q$  are non-negative and cannot exceed unity, i.e.,  $0 \leq p \leq 1$ ,  $0 \leq q \leq 1$ .

**Remarks.** 1. Probability ' $p$ ' of the happening of an event is also known as the probability of success and the probability ' $q$ ' of the non-happening of the event as the probability of failure.

2. If  $P(E) = 1$ ,  $E$  is called a *certain event* and if  $P(E) = 0$ ,  $E$  is called an *impossible event*.

3. **Limitations of Classical Definition.** This definition of Classical Probability breaks down in the following cases :

(i) If the various outcomes of the trial are not equally likely or equally probable. For example, the probability that a candidate will pass in a certain test is not 50% since the two possible outcomes, viz., success and failure (excluding the possibility of a compartment) are not equally likely.

(ii) If the exhaustive number of cases in a trial is infinite.

#### 4.3-2. Statistical or Empirical Probability

**Definition (Von Mises).** If a trial is repeated a number of times under essentially homogeneous and identical conditions, then the limiting value of the ratio of the number of times the event happens to the number of trials, as the number of trials become indefinitely large, is called the probability of happening of the event. (It is assumed that the limit is finite and unique).

Symbolically, if in  $n$  trials an event  $E$  happens  $m$  times, then the probability ' $p$ ' of the happening of  $E$  is given by

$$p = P(E) = \lim_{n \rightarrow \infty} \frac{m}{n} \quad \dots(4.2)$$

**Example 4.1.** What is the chance that a leap year selected at random will contain 53 Sundays?

**Solution.** In a leap year (which consists of 366 days) there are 52 complete weeks and 2 days over. The following are the possible combinations for these two 'over' days:

(i) Sunday and Monday, (ii) Monday and Tuesday, (iii) Tuesday and Wednesday, (iv) Wednesday and Thursday, (v) Thursday and Friday, (vi) Friday and Saturday, and (vii) Saturday and Sunday.

In order that a leap year selected at random should contain 53 Sundays, one of the two 'over' days must be Sunday. Since out of the above 7 possibilities, 2 viz., (i) and (vii), are favourable to this event,

$$\therefore \text{Required probability} = \frac{2}{7}$$

**Example 4.2.** A bag contains 3 red, 6 white and 7 blue balls. What is the probability that two balls drawn are white and blue?

**Solution.** Total number of balls =  $3 + 6 + 7 = 16$ .

Now, out of 16 balls, 2 can be drawn in  ${}^{16}C_2$  ways.

$$\therefore \text{Exhaustive number of cases} = {}^{16}C_2 = \frac{16 \times 15}{2} = 120.$$

Out of 6 white balls 1 ball can be drawn in  ${}^6C_1$  ways and out of 7 blue balls 1 ball can be drawn in  ${}^7C_1$  ways. Since each of the former cases can be associated with each of the latter cases, total number of favourable cases is :  ${}^6C_1 \times {}^7C_1 = 6 \times 7 = 42$ .

$$\therefore \text{Required probability} = \frac{42}{120} = \frac{7}{20}.$$

**Example 4.3. (a)** Two cards are drawn at random from a well-shuffled pack of 52 cards. Show that the chance of drawing two aces is  $1/221$ .

(b) From a pack of 52 cards, three are drawn at random. Find the chance that they are a king, a queen and a knave.

(c) Four cards are drawn from a pack of cards. Find the probability that

(i) all are diamond, (ii) there is one card of each suit, and (iii) there are two spades and two hearts.

**Solution.** (a) From a pack of 52 cards 2 cards can be drawn in  ${}^{52}C_2$  ways, all being equally likely.

$$\therefore \text{Exhaustive number of cases} = {}^{52}C_2$$

In a pack there are 4 aces and therefore 2 aces can be drawn in  ${}^4C_2$  ways.

$$\therefore \text{Required probability} = \frac{{}^4C_2}{{}^{52}C_2} = \frac{4 \times 3}{2} \times \frac{2}{52 \times 51} = \frac{1}{221}$$

$$(b) \text{ Exhaustive number of cases} = {}^{52}C_3$$

A pack of cards contains 4 kings, 4 queens and 4 knaves. A king, a queen and a knave can each be drawn in  ${}^4C_1$  ways and since each way of drawing a king can be associated with each of the ways of drawing a queen and a knave, the total number of favourable cases =  ${}^4C_1 \times {}^4C_1 \times {}^4C_1$

$$\therefore \text{Required probability} = \frac{{}^4C_1 \times {}^4C_1 \times {}^4C_1}{{}^{52}C_3} = \frac{4 \times 4 \times 4 \times 6}{52 \times 51 \times 50} = \frac{16}{5525}$$

$$(c) \text{ Exhaustive number of cases} = {}^{52}C_4$$

$$(i) \text{ Required probability} = \frac{{}^{13}C_4}{{}^{52}C_4}$$

$$(ii) \text{ Required probability} = \frac{{}^{13}C_1 \times {}^{13}C_1 \times {}^{13}C_1 \times {}^{13}C_1}{{}^{52}C_4}$$

$$(iii) \text{ Required probability} = \frac{{}^{13}C_2 \times {}^{13}C_2}{{}^{52}C_4}$$

**Example 4-4.** What is the probability of getting 9 cards of the same suit in one hand at a game of bridge?

**Solution.** One hand in a game of bridge consists of 13 cards.

$$\therefore \text{Exhaustive number of cases} = {}^{52}C_{13}$$

Number of ways in which, in one hand, a particular player gets 9 cards of one suit are  ${}^{13}C_9$ , and the number of ways in which the remaining 4 cards are of some other suit are  ${}^{39}C_4$ . Since there are 4 suits in a pack of cards, total number of favourable cases =  $4 \times {}^{13}C_9 \times {}^{39}C_4$ .

$$\therefore \text{Required probability} = \frac{4 \times {}^{13}C_9 \times {}^{39}C_4}{{}^{52}C_{13}}$$

**Example 4-5.** (a) Among the digits 1, 2, 3, 4, 5, at first one is chosen and then a second selection is made among the remaining four digits. Assuming that all twenty possible outcomes have equal probabilities, find the probability that an odd digit will be selected (i) the first time, (ii) the second time, and (iii) both times.

(b) From 25 tickets, marked with the first 25 numerals, one is drawn at random. Find the chance that

- (i) it is a multiple of 5 or 7,
- (ii) it is a multiple of 3 or 7.

**Solution.** (a) Total number of cases =  $5 \times 4 = 20$

(i) Now there are 12 cases in which the first digit drawn is odd, viz., (1, 2), (1, 3), (1, 4), (1, 5), (3, 1), (3, 2), (3, 4), (3, 5), (5, 1), (5, 2), (5, 3) and (5, 4).

$\therefore$  The probability that the first digit drawn is odd

$$= \frac{12}{20} = \frac{3}{5}$$

(ii) Also there are 12 cases in which the second digit drawn is odd, viz., (2, 1), (2, 3), (4, 1), (4, 3), (5, 1), (5, 3), (1, 5), (2, 5), (3, 5) and (4, 5).

$\therefore$  The probability that the second digit drawn is odd

$$= \frac{12}{20} = \frac{3}{5}$$

(iii) There are six cases in which both the digits drawn are odd, viz., (1, 3), (1, 5), (3, 1), (3, 5), (5, 1) and (5, 3).

$\therefore$  The probability that both the digits drawn are odd

$$= \frac{6}{20} = \frac{3}{10}$$

(b) (i) Numbers (out of the first 25 numerals) which are multiples of 5 are 5, 10, 15, 20 and 25, i.e., 5 in all and the numbers which are multiples of 7 are 7, 14 and 21, i.e., 3 in all. Hence required number of favourable cases are  $5+3=8$ .

$$\therefore \text{Required probability} = \frac{8}{25}$$

(ii) Numbers (among the first 25 numerals) which are multiples of 3 are 3, 6, 9, 12, 15, 18, 21, 24, i.e., 8 in all, and the numbers which are multiples of 7 are 7,

14, 21, i.e., 3 in all. Since the number 21 is common in both the cases, the required number of distinct favourable cases is  $8 + 3 - 1 = 10$ .

$$\therefore \text{Required probability} = \frac{10}{25} = \frac{2}{5}$$

**Example 4.6.** A committee of 4 people is to be appointed from 3 officers of the production department, 4 officers of the purchase department, two officers of the sales department and 1 chartered accountant. Find the probability of forming the committee in the following manner:

- (i) There must be one from each category.
- (ii) It should have at least one from the purchase department.
- (iii) The chartered accountant must be in the committee.

**Solution.** There are  $3+4+2+1=10$  persons in all and a committee of 4 people can be formed out of them in  ${}^{10}C_4$  ways. Hence exhaustive number of cases is

$${}^{10}C_4 = \frac{10 \times 9 \times 8 \times 7}{4!} = 210$$

(i) Favourable number of cases for the committee to consist of 4 members, one from each category is :

$${}^3C_1 \times {}^3C_1 \times {}^2C_1 \times 1 = 4 \times 3 \times 2 = 24$$

$$\therefore \text{Required probability} = \frac{24}{210} = \frac{8}{70}$$

- (ii)  $P$  [Committee has at least one purchase officer]  
 $= 1 - P$  [Committee has no purchase officer]

In order that the committee has no purchase officer, all the 4 members are to be selected from amongst officers of production department, sales department and chartered accountant, i.e., out of  $3+2+1=6$  members and this can be done in

$${}^6C_4 = \frac{6 \times 5}{1 \times 2} = 15 \text{ ways. Hence}$$

$$P [\text{Committee has no purchase officer}] = \frac{15}{210} = \frac{1}{14}$$

$$\therefore P [\text{Committee has at least one purchase officer}] = 1 - \frac{1}{14} = \frac{13}{14}$$

(iii) Favourable number of cases that the committee consists of a chartered accountant as a member and three others are :

$$1 \times {}^9C_3 = \frac{9 \times 8 \times 7}{1 \times 2 \times 3} = 84 \text{ ways,}$$

since a chartered accountant can be selected out of one chartered accountant in only 1 way and the remaining 3 members can be selected out of the remaining  $10 - 1 = 9$  persons in  ${}^9C_3$  ways. Hence the required probability  $= \frac{84}{210} = \frac{2}{5}$ .

**Example 4.7.** (a) If the letters of the word 'REGULATIONS' be arranged at random, what is the chance that there will be exactly 4 letters between R and E?

(b) What is the probability that four S's come consecutively in the word 'MISSISSIPPI'?

**Solution.** (a) The word 'REGULATIONS' consists of 11 letters. The two letters R and E can occupy  ${}^{11}P_2$ , i.e.,  $11 \times 10 = 110$  positions.

The number of ways in which there will be exactly 4 letters between R and E are enumerated below:

- (i) R is in the 1st place and E is in the 6th place.
- (ii) R is in the 2nd place and E is in the 7th place.

...	...	...
...	...	...
...	...	...

- (vi) R is in the 6th place and E is in the 11th place.

Since R and E can interchange their positions, the required number of favourable cases is  $2 \times 6 = 12$

$$\therefore \text{The required probability} = \frac{12}{110} = \frac{6}{55}.$$

(b) Total number of permutations of the 11 letters of the word 'MISSISSIPPI', in which 4 are of one kind (viz., S), 4 of other kind (viz., I), 2 of third kind (viz., P) and 1 of fourth kind (viz., M) are

$$\frac{11!}{4! 4! 2! 1!}$$

Following are the 8 possible combinations of 4 S's coming consecutively:

- (i) S S S S
- (ii) — S S S S
- (iii) — — S S S S
- ⋮ ⋮ ⋮ ⋮
- (viii) — — — — — — S S S S

Since in each of the above cases, the total number of arrangements of the remaining 7 letters, viz., MIIIPPI of which 4 are of one kind, 2 of other kind and one of third kind are  $\frac{7!}{4! 2! 1!}$ , the required number of favourable cases

$$= \frac{8 \times 7!}{4! 2! 1!}$$

$$\therefore \text{Required probability} = \frac{8 \times 7!}{4! 2! 1!} + \frac{11!}{4! 4! 2! 1!} \\ = \frac{8 \times 7! \times 4!}{11!} = \frac{4}{165}$$

**Example 4.6.** Each coefficient in the equation  $ax^2 + bx + c = 0$  is determined by throwing an ordinary die. Find the probability that the equation will have real roots.  
[Madras Univ. B. Sc. (Stat. Main), 1992]

**Solution.** The roots of the equation  $ax^2 + bx + c = 0$  ...(\*) will be real if its discriminant is non-negative, i.e., if

$$b^2 - 4ac \geq 0 \Rightarrow b^2 \geq 4ac$$

Since each co-efficient in equation (\*) is determined by throwing an ordinary die, each of the co-efficients  $a$ ,  $b$  and  $c$  can take the values from 1 to 6.

∴ Total number of possible outcomes (all being equally likely)

$$= 6 \times 6 \times 6 = 216$$

The number of favourable cases can be enumerated as follows:

$ac$	$a$	$c$	$4ac$	$b$	No. of cases
( so that $b^2 \geq 4ac$ )					
1	1	1	4	2, 3, 4, 5	$1 \times 5 = 5$
2	(i) 1	2	8	3, 4, 5, 6	$2 \times 4 = 8$
	(ii) 2	1			
3	(i) 1	3	12	4, 5, 6	$2 \times 3 = 6$
	(ii) 3	1			
4	(i) 1	4	16	4, 5, 6	$3 \times 3 = 9$
	(ii) 4	1			
	(iii) 2	2			
5	(i) 1	5	20	5, 6	$2 \times 2 = 4$
	(ii) 5	1			
6	(i) 1	6	24	5, 6	$4 \times 2 = 8$
	(ii) 6	1			
	(iii) 3	2			
	(iv) 2	3			
7	( $ac = 7$ is not possible )				
8	(i) 2	4	32	6	$2 \times 1 = 2$
	(ii) 4	2			
9	3	3	36	6	$\frac{1}{Total = 43}$

Since  $b^2 \geq 4ac$  and since the maximum value of  $b^2$  is 36,  $ac = 10, 11, 12, \dots$  etc. is not possible.

Hence total number of favourable cases = 43.

$$\therefore \text{Required probability} = \frac{43}{216}$$

**Example 4.9.** The sum of two non-negative quantities is equal to  $2n$ . Find the chance that their product is not less than  $\frac{3}{4}$  times their greatest product.

**Solution.** Let  $x > 0$  and  $y > 0$  be the given quantities so that  $x + y = 2n$ .

We know that the product of two positive quantities whose sum is constant is greatest when the quantities are equal. Thus the product of  $x$  and  $y$  is maximum when  $x = y = n$ .

∴ Maximum product =  $n \cdot n = n^2$

Now  $P \left[ xy < \frac{3}{4} n^2 \right] = P \left[ xy \geq \frac{3}{4} n^2 \right] = P \left[ x(2n-x) \geq \frac{3}{4} n^2 \right]$   
 $= P \left[ (4x^2 - 8nx + 3n^2) \leq 0 \right]$   
 $= P \left[ (2x-3n)(2x-n) \leq 0 \right]$   
 $= P \left[ x \text{ lies between } \frac{n}{2} \text{ and } \frac{3n}{2} \right]$

∴ Favourable range =  $\frac{3n}{2} - \frac{n}{2} = n$   
 Total range =  $2n$

∴ Required probability =  $\frac{n}{2n} = \frac{1}{2}$

**Example 4-10.** Out of  $(2n+1)$  tickets consecutively numbered three are drawn at random. Find the chance that the numbers on them are in A.P.

[Calicut Univ. B.Sc., 1991; Delhi Univ. B.Sc.(Stat. Hons.), 1992]

**Solution.** Since out of  $(2n+1)$  tickets, 3 tickets can be drawn in  $^{2n+1}C_3$  ways,

$$\begin{aligned}\text{Exhaustive number of cases} &= {}^{2n+1}C_3 = \frac{(2n+1) \cdot 2n \cdot (2n-1)}{3!} \\ &= \frac{n(4n^2-1)}{3}\end{aligned}$$

To find the favourable number of cases we are to enumerate all the cases in which the numbers on the drawn tickets are in A.P with common difference, (say  $d = 1, 2, 3, \dots, n-1, n$ ).

If  $d = 1$ , the possible cases are as follows:

$$\left. \begin{array}{c} 1, 2, 3 \\ 2, 3, 4 \\ \vdots \quad \vdots \quad \vdots \\ 2n-1, n, 2n+1 \end{array} \right\}, \text{i.e., } (2n-1) \text{ cases in all}$$

If  $d = 2$ , the possible cases are as follows:

$$\left. \begin{array}{c} 1, 3, 5 \\ 2, 4, 6 \\ \vdots \quad \vdots \quad \vdots \\ 2n-3, 2n-1, 2n+1 \end{array} \right\}, \text{i.e., } (2n-3) \text{ cases in all}$$

and so on.

If  $d = n - 1$ , the possible cases are as follows:

$$\left. \begin{array}{ll} 1, n, & 2n-1 \\ 2, n+1, & 2n \\ 3, n+2, & 2n+1 \end{array} \right\}, \text{ i.e., 3 cases in all}$$

If  $d = n$ , there is only one case, viz.,  $(1, n+1, 2n+1)$ .

Hence total number of favourable cases

$$\begin{aligned} &= (2n-1) + (2n-3) + \dots + 5 + 3 + 1 \\ &= 1 + 3 + 5 + \dots + (2n-1), \end{aligned}$$

which is a series in A.P. with  $d = 2$  and  $n$  terms.

$$\therefore \text{Number of favourable cases} = \frac{n}{2} [1 + (2n-1)] = n^2$$

$$\therefore \text{Required probability} = \frac{n^2}{n(4n^2-1)/3} = \frac{3n}{(4n^2-1)}$$

#### EXERCISE 4 (a)

1. (a) Give the classical and statistical definitions of probability. What are the objections raised in these definitions?

[Delhi Univ. B.Sc. (Stat. Hons.), 1988, 1985]

(b) When are a number of cases said to be equally likely? Give an example each of the following :

- (i) the equally likely cases,
- (ii) four cases which are not equally likely, and
- (iii) five cases in which one case is more likely than the other four.

(c) What is meant by mutually exclusive events? Give an example of

- (i) three mutually exclusive events,
- (ii) three events which are not mutually exclusive.

[Meerut Univ. B.Sc. (Stat.), 1987]

(d) Can

- (i) events be mutually exclusive and exhaustive?
- (ii) events be exhaustive and independent?
- (iii) events be mutually exclusive and independent?
- (iv) events be mutually exhaustive, exclusive and independent?

2. (a) Prove that the probability of obtaining a total of 9 in a single throw with two dice is one by nine.

(b) Prove that in a single throw with a pair of dice the probability of getting the sum of 7 is equal to  $1/6$  and the probability of getting the sum of 10 is equal to  $1/12$ .

(c) Show that in a single throw with two dice, the chance of throwing more than seven is equal to that of throwing less than seven.

Ans. 5/12

[Delhi Univ. B.Sc., 1987, 1985]

(d) In a single throw with two dice, what is the number whose probability is minimum?

(e) Two persons A and B throw three dice (six faced). If A throws 14, find B's chance of throwing a higher number. [Meerut Univ. B.Sc.(Stat.), 1987]

3. (a) A bag contains 7 white, 6 red and 5 black balls. Two balls are drawn at random. Find the probability that they will both be white.

Ans. 21/153

(b) A bag contains 10 white, 6 red, 4 black and 7 blue balls. 5 balls are drawn at random. What is the probability that 2 of them are red and one black?

Ans.  ${}^6C_2 \times {}^4C_1 / {}^{21}C_5$

4. (a) From a set of raffle tickets numbered 1 to 100, three are drawn at random. What is the probability that all the tickets are odd-numbered?

Ans.  ${}^{50}C_3 / {}^{100}C_3$

(b) A number is chosen from each of the two sets :

(1, 2, 3, 4, 5, 6, 7, 8, 9); (4, 5, 6, 7, 8, 9)

If  $p_1$  is the probability that the sum of the two numbers be 10 and  $p_2$  the probability that their sum be 8, find  $p_1 + p_2$ .

(c) Two different digits are chosen at random from the set 1, 2, 3, ..., 8. Show that the probability that the sum of the digits will be equal to 5 is the same as the probability that their sum will exceed 13, each being 1/14. Also show that the chance of both digits exceeding 5 is 3/28. [Nagpur Univ. B.Sc., 1992]

5. What is the chance that (i) a leap year selected at random will contain 53 Sundays? (ii) a non-leap year selected at random would contain 53 Sundays.

Ans. (i) 2/7, (ii) 1/7

6. (a) What is the probability of having a knave and a queen when two cards are drawn from a pack of 52 cards?

Ans. 8/663

(b) Seven cards are drawn at random from a pack of 52 cards. What is the probability that 4 will be red and 3 black?

Ans.  ${}^{26}C_4 \times {}^{26}C_3 / {}^{52}C_7$

(c) A card is drawn from an ordinary pack and a gambler bets that it is a spade or an ace. What are the odds against his winning the bet?

Ans. 9:4

(d) Two cards are drawn from a pack of 52 cards. What is the chance that

(i) they belong to the same suit?

(ii) they belong to different suits and different denominations.

[Bombay Univ. B.Sc., 1986]

7. (a) If the letters of the word RANDOM be arranged at random, what is the chance that there are exactly two letters between A and O.

(b) Find the probability that in a random arrangement of the letters of the word 'UNIVERSITY', the two I's do not come together.

(c) In random arrangements of the letters of the word 'ENGINEERING', what is the probability that vowels always occur together?

[Kurushetra Univ. B.E., 1991]

(d) Letters are drawn one at a time from a box containing the letters A, H, M, O, S, T. What is the probability that the letters in the order drawn spell the word 'Thomas'?

8. A letter is taken out at random out of 'ASSISTANT' and a letter out of 'STATISTIC'. What is the chance that they are the same letters?

9. (a) Twelve persons amongst whom are  $x$  and  $y$  sit down at random at a round table. What is the probability that there are two persons between  $x$  and  $y$ ?

(b)  $A$  and  $B$  stand in a line at random with 10 other people. What is the probability that there will be 3 persons between  $A$  and  $B$ ?

10. (a) If from a lot of 30 tickets marked 1, 2, 3, ..., 30 four tickets are drawn, what is the chance that those marked 1 and 2 are among them?

Ans. 2/145

(b) A bag contains 50 tickets numbered 1, 2, 3, ..., 50 of which five are drawn at random and arranged in ascending order of the magnitude ( $x_1 < x_2 < x_3 < x_4 < x_5$ ). What is the probability that  $x_3 = 30$ ?

Hint. (a) Exhaustive number of cases =  ${}^{30}C_4$

If, of the four tickets drawn, two tickets bear the numbers 1 and 2, the remaining 2 must have come out of 28 tickets numbered from 3 to 30 and this can be done in  ${}^{28}C_2$  ways.

∴ Favourable number of cases =  ${}^{28}C_2$

(b) Exhaustive number of cases =  ${}^{50}C_5$

If  $x_3 = 30$ , then the two tickets with numbers  $x_1$  and  $x_2$  must have come out of 29 tickets numbered from 1 to 29 and this can be done in  ${}^{29}C_2$  ways, and the other two tickets with numbers  $x_4$  and  $x_5$  must have been drawn out of 20 tickets numbered from 31 to 50 and this can be done in  ${}^{20}C_2$  ways.

∴ No. of favourable cases =  ${}^{29}C_2 \times {}^{20}C_2$ .

11. Four persons are chosen at random from a group containing 3 men, 2 women and 4 children. Show that the chance that exactly two of them will be children is  $10/21$ .

[Delhi Univ. B.A.1988]

$$\text{Ans. } \frac{{}^4C_2 \times {}^3C_2}{{}^9C_4} = \frac{10}{21}$$

12. From a group of 3 Indians, 4 Pakistanis and 5 Americans a sub-committee of four people is selected by lots. Find the probability that the sub-committee will consist of

(i) 2 Indians and 2 Pakistanis

(ii) 1 Indian, 1 Pakistani and 2 Americans

$$\text{Ans. (i) } \frac{^3C_2 \times ^4C_2}{^{12}C_4}, \quad \text{(ii) } \frac{^3C_1 \times ^4C_1 \times ^5C_2}{^{12}C_4}, \quad \text{(iii) } \frac{^5C_4}{^{12}C_4}$$

[Madras Univ. B.Sc.(Main Stat.), 1987]

13. In a box there are 4 granite stones, 5 sand stones and 6 bricks of identical size and shape. Out of them 3 are chosen at random. Find the chance that :

(i) They all belong to different varieties.

(ii) They all belong to the same variety.

(iii) They are all granite stones. (Madras Univ. B.Sc., Oct. 1992)

14. If  $n$  people are seated at a round table, what is the chance that two named individuals will be next to each other?

Ans.  $\frac{2}{(n-1)}$

15. Four tickets marked 00, 01, 10 and 11 respectively are placed in a bag. A ticket is drawn at random five times, being replaced each time. Find the probability that the sum of the numbers on tickets thus drawn is 23.

[Delhi Univ. B.Sc.(Subs.), 1988]

16. From a group of 25 persons, what is the probability that all 25 will have different birthdays? Assume a 365 day year and that all days are equally likely.

[Delhi Univ. B.Sc.(Maths Hons.), 1987]

Hint.  $(365 \times 364 \times \dots \times 341) + (365)^{25}$

**4-4. Mathematical Tools : Preliminary Notions of Sets.** The set theory was developed by the German mathematician, G. Cantor (1845-1918).

**4-4.1. Sets and Elements of Sets.** A set is a well defined collection or aggregate of all possible objects having given properties and specified according to a well defined rule. The objects comprising a set are called elements, members or points of the set. Sets are often denoted by capital letters, viz.,  $A, B, C$ , etc. If  $x$  is an element of the set  $A$ , we write symbolically  $x \in A$  ( $x$  belongs to  $A$ ). If  $x$  is not a member of the set  $A$ , we write  $x \notin A$  ( $x$  does not belong to  $A$ ). Sets are often described by describing the properties possessed by their members. Thus the set  $A$  of all non-negative rational numbers with square less than 2 will be written as  $A = \{x : x \text{ rational, } x \geq 0, x^2 < 2\}$ .

If every element of the set  $A$  belongs to the set  $B$ , i.e., if  $x \in A \Rightarrow x \in B$ , then we say that  $A$  is a subset of  $B$  and write symbolically  $A \subseteq B$  ( $A$  is contained in  $B$ ) or  $B \supseteq A$  ( $B$  contains  $A$ ). Two sets  $A$  and  $B$  are said to be equal or identical if  $A \subseteq B$  and  $B \subseteq A$  and we write  $A = B$  or  $B = A$ .

A null or an empty set is one which does not contain any element at all and is denoted by  $\emptyset$ .

**Remarks.** 1. Every set is a subset of itself.

2. An empty set is subset of every set.

3. A set containing only one element is conceptually distinct from the element itself, but will be represented by the same symbol for the sake of convenience.

4. As will be the case in all our applications of set theory, especially to probability theory, we shall have a fixed set  $S$  (say) given in advance, and we shall

be concerned only with subsets of this given set. The underlying set  $S$  may vary from one application to another, and it will be referred to as *universal set* of each particular discourse.

#### 4-4-2. Operation on Sets

The union of two given sets  $A$  and  $B$ , denoted by  $A \cup B$ , is defined as a set consisting of all those points which belong to either  $A$  or  $B$  or both. Thus symbolically,

$$A \cup B = \{x : x \in A \text{ or } x \in B\}.$$

Similarly

$$\bigcup_{i=1}^n A_i = \{x : x \in A_i \text{ for at least one } i = 1, 2, \dots, n\}$$

The intersection of two sets  $A$  and  $B$ , denoted by  $A \cap B$ , is defined as a set consisting of all those elements which belong to both  $A$  and  $B$ . Thus

$$A \cap B = \{x : x \in A \text{ and } x \in B\}.$$

Similarly

$$\bigcap_{i=1}^n A_i = \{x : x \in A_i \text{ for all } i = 1, 2, \dots, n\}$$

For example, if  $A = \{1, 2, 5, 8, 10\}$  and  $B = \{2, 4, 8, 12\}$ , then

$$A \cup B = \{1, 2, 4, 5, 8, 10, 12\} \text{ and } A \cap B = \{2, 8\}.$$

If  $A$  and  $B$  have no common point, i.e.,  $A \cap B = \emptyset$ , then the sets  $A$  and  $B$  are said to be *disjoint, mutually exclusive* or *non-overlapping*.

The relative difference of a set  $A$  from another set  $B$ , denoted by  $A - B$  is defined as a set consisting of those elements of  $A$  which do not belong to  $B$ . Symbolically,

$$A - B = \{x : x \in A \text{ and } x \notin B\}.$$

The complement or negative of any set  $A$ , denoted by  $\bar{A}$  is a set containing all elements of the universal set  $S$ , (say), that are not elements of  $A$ , i.e.,  $\bar{A} = S - A$ .

#### 4-4-3. Algebra of Sets

Now we state certain important properties concerning operations on sets. If  $A$ ,  $B$  and  $C$  are the subsets of a universal set  $S$ , then the following laws hold:

$$\text{Commutative Law} : A \cup B = B \cup A, A \cap B = B \cap A$$

$$\text{Associative Law} : (A \cup B) \cup C = A \cup (B \cup C)$$

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

$$\text{Distributive Law} : A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

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

$$\text{Complementary Law} : A \cup \bar{A} = S, A \cap \bar{A} = \emptyset$$

$$A \cup S = S, (\because A \subseteq S), A \cap S = A$$

$$A \cup \emptyset = A, A \cap \emptyset = \emptyset$$

$$\text{Difference Law} : A - B = A \cap \bar{B}$$

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

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

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

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

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

*De-Morgan's Law—Dualization Law.*

$$\overline{(A \cup B)} = \overline{A} \cap \overline{B} \text{ and } \overline{(A \cap B)} = \overline{A} \cup \overline{B}$$

More generally

$$\overline{\left( \bigcup_{i=1}^n A_i \right)} = \bigcap_{i=1}^n \overline{A_i} \quad \text{and} \quad \overline{\left( \bigcap_{i=1}^n A_i \right)} = \bigcup_{i=1}^n \overline{A_i}$$

*Involution Law* :  $\overline{\overline{A}} = A$

*Idempotency Law* :  $A \cup A = A, A \cap A = A$

#### 4.4.4. Limit of Sequence of Sets

Let  $\{A_n\}$  be a sequence of sets in  $S$ . The *limit supremum* or *limit superior* of the sequence, usually written as  $\limsup A_n$ , is the set of all those elements which belong to  $A_n$  for infinitely many  $n$ . Thus

$$\limsup_{n \rightarrow \infty} A_n = \{x : x \in A_n \text{ for infinitely many } n\} \quad \dots(4.3)$$

The set of all those elements which belong to  $A_n$  for all but a finite number of  $n$  is called *limit infimum* or *limit inferior* of the sequence and is denoted by  $\liminf A_n$ . Thus

$$\liminf_{n \rightarrow \infty} A_n = \{x : x \in A_n \text{ for all but a finite number of } n\} \quad \dots(4.3a)$$

The sequence  $\{A_n\}$  is said to have a limit if and only if  $\limsup A_n = \liminf A_n$  and this common value gives the limit of the sequence.

$$\text{Theorem 4.1. } \limsup A_n = \bigcap_{m=1}^{\infty} \left( \bigcup_{n=m}^{\infty} A_n \right)$$

$$\text{and } \liminf A_n = \bigcup_{m=1}^{\infty} \left( \bigcap_{n=m}^{\infty} A_n \right)$$

**Def.**  $\{A_n\}$  is a monotone (infinite) sequence of sets if either

(i)  $A_n \subset A_{n+1} \forall n$  or (ii)  $A_n \supset A_{n+1} \forall n$ .

In the former case the sequence  $\{A_n\}$  is said to be *non-decreasing sequence* and is usually expressed as  $A_n \uparrow$  and in the latter case it is said to be *non-increasing sequence* and is expressed as  $A_n \downarrow$ .

For a monotone sequence (non-increasing or non-decreasing), the limit always exists and we have,

$$\lim_{n \rightarrow \infty} A_n = \begin{cases} \bigcup_{n=1}^{\infty} A_n & \text{in case (i), i.e., } A_n \uparrow \\ \bigcap_{n=1}^{\infty} A_n & \text{in case (ii), i.e., } A_n \downarrow \end{cases}$$

**4.4.5. Classes of Sets.** A group of sets will be termed as a *class* (of sets). Below we shall define some useful types of classes.

A ring  $R$  is a non-empty class of sets which is closed under the formation of 'finite unions' and 'difference'.

i.e., if  $A \in R$ ,  $B \in R$ , then  $A \cup B \in R$  and  $A - B \in R$ .

Obviously  $\emptyset$  is a member of every ring.

A field  $F$  (or an algebra) is a non-empty class of sets which is closed under the formation of finite unions and under complementation. Thus

(i)  $A \in F$ ,  $B \in F \Rightarrow A \cup B \in F$  and

(ii)  $A \in F \Rightarrow \bar{A} \in F$ .

A  $\sigma$ -ring  $C$  is a non-empty class of sets which is closed under the formation of 'countable unions' and 'difference'. Thus

(i)  $A_i \in C$ ,  $i = 1, 2, \dots \Rightarrow \bigcup_{i=1}^{\infty} A_i \in C$

(ii)  $A \in C$ ,  $B \in C \Rightarrow A - B \in C$ .

More precisely  $\sigma$ -ring is a ring which is closed under the formation of countable unions.

A  $\sigma$  field (or  $\sigma$ -algebra)  $B$  is a non-empty class of sets that is closed under the formation of 'countable unions' and complementations,

i.e.,

(i)  $A_i \in B$ ,  $i = 1, 2, \dots \Rightarrow \bigcup_{i=1}^{\infty} A_i \in B$ .

(ii)  $A \in B \Rightarrow \bar{A} \in B$ .

$\sigma$ -field may also be defined as a field which is closed under the formation of countable unions.

**4.5. Axiomatic Approach to Probability.** The axiomatic approach to probability, which closely relates the theory of probability with the modern metric theory of functions and also set theory, was proposed by A.N. Kolmogorov, a Russian mathematician, in 1933. The axiomatic definition of probability includes 'both' the classical and the statistical definitions as particular cases and overcomes the deficiencies of each of them. On this basis, it is possible to construct a logically perfect structure of the modern theory of probability and at the same time to satisfy the enhanced requirements of modern natural science. The axiomatic development of mathematical theory of probability relies entirely upon the logic of deduction.

The diverse theorems of probability, as were available prior to 1933, were finally brought together into a unified axiomised system in 1933. It is important to remark that probability theory, in common with all axiomatic mathematical systems, is concerned solely with relations among undefined things.

The axioms thus provide a set of rules which define relationships between abstract entities. These rules can be used to deduce theorems, and the theorems can

be brought together to deduce more complex theorems. These theorems have no empirical meaning although they can be given an interpretation in terms of empirical phenomenon. The important point, however, is that the mathematical development of probability theory is, in no way, conditional upon the interpretation given to the theory.

More precisely, under axiomatic approach, the probability can be deduced from mathematical concepts. To start with some concepts are laid down. Then some statements are made in respect of the properties possessed by these concepts. These properties, often termed as "*axioms*" of the theory, are used to frame some theorems. These theorems are framed without any reference to the real world and are deductions from the axioms of the theory.

**4-5-1. Random Experiment, Sample Space.** The theory of probability provides *mathematical models* for "real-world phenomenon" involving games of chance such as the tossing of coins and dice. We feel intuitively that statements such as

- (i) "The probability of getting a "head" in one toss of an unbiased coin is  $1/2$ "
- (ii) "The probability of getting a "four" in a single toss of an unbiased die is  $1/6$ ".

should hold. We also feel that the probability of obtaining either a "5" or a "6" in a single throw of a die, should be the sum of the probabilities of a "5" and a "6", viz.,  $1/6 + 1/6 = 1/3$ . That is, probabilities should have some kind of *additive* property. Finally, we feel that in a large number of repetitions of, for example, a coin tossing experiment, the proportion of heads should be approximately  $1/2$ . That is, the probability should have a *frequency interpretation*.

To deal with these properties sensibly, we need a *mathematical description* or *model* for the probabilistic situation we have. Any such probabilistic situation is referred to as a *random experiment*, denoted by  $E$ .  $E$  may be a coin or die throwing experiment, drawing of cards from a well-shuffled pack of cards, an agricultural experiment to determine the effects of fertilizers on yield of a commodity, and so on.

Each performance in a random experiment is called a *trial*. That is, all the trials conducted under the same set of conditions form a random experiment. The result of a trial in a random experiment is called an *outcome*, an *elementary event* or a *sample point*. The totality of all possible outcomes (*i.e.*, sample points) of a random experiment constitutes the *sample space*.

Suppose  $e_1, e_2, \dots, e_n$  are the possible outcomes of a random experiment  $E$  such that no two or more of them can occur simultaneously and exactly one of the outcomes  $e_1, e_2, \dots, e_n$  must occur. More specifically, with an experiment  $E$ , we associate a set  $S = (e_1, e_2, \dots, e_n)$  of possible outcomes with the following properties:

- (i) each element of  $S$  denotes a possible outcome of the experiment,

(ii) any trial results in an outcome that corresponds to one and only one element of the set  $S$ .

The set  $S$  associated with an experiment E, real or conceptual, satisfying the above two properties is called the *sample space* of the experiment.

**Remarks.** 1. The sample space serves as universal set for all questions concerned with the experiment.

2. A sample space  $S$  is said to be finite (infinite) sample sapce if the number of elements in  $S$  is finite (infinite). For example, the sample space associated with the experiment of throwing the coin until a head appears, is infinite, with possible sample points

$$\{\omega_1, \omega_2, \omega_3, \omega_4, \dots\}$$

where  $\omega_1 = H$ ,  $\omega_2 = TH$ ,  $\omega_3 = TTH$ ,  $\omega_4 = TTHH$ , and so on,  $H$  denoting a head and  $T$  a tail.

3. A sample space is called discrete if it contains only finitely or infinitely many points which can be arranged into a simple sequence  $\omega_1, \omega_2, \dots$ , while a sample space containing non-denumerable number of points is called a continuous sample space. In this book, we shall restrict ourselves to discrete sample spaces only.

**4.5.2. Event.** Every non-empty subset  $A$  of  $S$ , which is a disjoint union of single element subsets of the sample space  $S$  of a random experiment E is called an event. The notion of an event may also be defined as follows:

"Of all the possible outcomes in the sample space of an experiment, some outcomes satisfy a specified description, which we call an event."

**Remarks.** 1. As the empty set  $\phi$  is a subset of  $S$ ,  $\phi$  is also an event, known as *impossible event*.

2. An event  $A$ , in particular, can be a single element subset of  $S$ , in which case it is known as *elementary event*.

**4.5.3. Some Illustrations — Examples.** We discuss below some examples to illustrate the concepts of sample space and event.

1. Consider tossing of a coin singly. The possible outcomes for this experiment are (writing  $H$  for a "head" and  $T$  for a "tail") :  $H$  and  $T$ . Thus the sample space  $S$  consists of two points  $\{\omega_1, \omega_2\}$ , corresponding to each possible outcome or elementary event listed.

i.e.,  $S = \{\omega_1, \omega_2\} = \{H, T\}$  and  $n(S) = 2$ ,

where  $n(S)$  is the total number of sample points in  $S$ .

If we consider two tosses of a coin, the possible outcomes are  $HH$ ,  $HT$ ,  $TH$ ,  $TT$ . Thus, in this case the sample space  $S$  consists of four points  $\{\omega_1, \omega_2, \omega_3, \omega_4\}$ , corresponding to each possible outcome listed and  $n(S)=4$ . Combinations of these outcomes form what we call events. For example, the event of getting at least one head is the set of the outcomes  $\{HH, HT, TH\} = \{\omega_1, \omega_2, \omega_3\}$ . Thus, mathematically, the events are subsets of  $S$ .

2. Let us consider a single toss of a die. Since there are six possible outcomes, our sample space  $S$  is now a space of six points  $\{\omega_1, \omega_2, \dots, \omega_6\}$  where  $\omega_i$  corresponds to the appearance of number  $i$ . Thus  $S = \{\omega_1, \omega_2, \dots, \omega_6\} = \{1, 2, \dots, 6\}$  and  $n(S) = 6$ . The event that the outcome is even is represented by the set of points  $\{\omega_2, \omega_4, \omega_6\}$ .

3. A coin and a die are tossed together. For this experiment, our sample space consists of twelve points  $\{\omega_1, \omega_2, \dots, \omega_{12}\}$  where  $\omega_i$  ( $i = 1, 2, \dots, 6$ ) represents a head on coin together with appearance of  $i$ th number on the die and  $\omega_i$  ( $i = 7, 8, \dots, 12$ ) represents a tail on coin together with the appearance of  $i$ th number on die. Thus

$$S = \{\omega_1, \omega_2, \dots, \omega_{12}\} = \{(H, T) \times (1, 2, \dots, 6)\} \text{ and } n(S) = 12$$

**Remark.** If the coin and die are unbiased, we can see intuitively that in each of the above examples, the outcomes (sample points) are equally likely to occur.

4. Consider an experiment in which two balls are drawn one by one from an urn containing 2 white and 4 blue balls such that when the second ball is drawn, the first is *not* replaced.

Let us number the six balls as 1, 2, 3, 4, 5 and 6, numbers 1 and 2 representing a white ball and numbers 3, 4, 5, and 6 representing a blue ball. Suppose in a draw we pick up balls numbered 2 and 6. Then (2,6) is called an outcome of the experiment. It should be noted that the outcome (2,6) is different from the outcome (6,2) because in the former case ball No. 2 is drawn first and ball No. 6 is drawn next while in the latter case, 6th ball is drawn first and the second ball is drawn next.

The sample space consists of thirty points as listed below:

$$\begin{array}{lllll} \omega_1 = (1, 2) & \omega_2 = (1, 3) & \omega_3 = (1, 4) & \omega_4 = (1, 5) & \omega_5 = (1, 6) \\ \omega_6 = (2, 1) & \omega_7 = (2, 3) & \omega_8 = (2, 4) & \omega_9 = (2, 5) & \omega_{10} = (2, 6) \\ \omega_{11} = (3, 1) & \omega_{12} = (3, 2) & \omega_{13} = (3, 4) & \omega_{14} = (3, 5) & \omega_{15} = (3, 6) \\ \omega_{16} = (4, 1) & \omega_{17} = (4, 2) & \omega_{18} = (4, 3) & \omega_{19} = (4, 5) & \omega_{20} = (4, 6) \\ \omega_{21} = (5, 1) & \omega_{22} = (5, 2) & \omega_{23} = (5, 3) & \omega_{24} = (5, 4) & \omega_{25} = (5, 6) \\ \omega_{26} = (6, 1) & \omega_{27} = (6, 2) & \omega_{28} = (6, 3) & \omega_{29} = (6, 4) & \omega_{30} = (6, 5) \end{array}$$

Thus

$$\begin{aligned} S &= \{\omega_1, \omega_2, \omega_3, \dots, \omega_{30}\} \text{ and } n(S) = 30 \\ \Rightarrow S &= \{1, 2, 3, 4, 5, 6\} \times \{1, 2, 3, 4, 5, 6\} \\ &\quad - \{(1, 1), (2, 2), (3, 3), (4, 4), (5, 5), (6, 6)\} \end{aligned}$$

The event

- (i) the first ball drawn is white
- (ii) the second ball drawn is white
- (iii) both the balls drawn are white
- (iv) both the balls drawn are black

are represented respectively by the following sets of points:

$$\begin{aligned} &\{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6, \omega_7, \omega_8, \omega_9, \omega_{10}\}, \\ &\{\omega_1, \omega_6, \omega_{11}, \omega_{12}, \omega_{16}, \omega_{17}, \omega_{21}, \omega_{22}, \omega_{26}, \omega_{27}\}, \end{aligned}$$

$\{\omega_1, \omega_6\}$ , and

$\{\omega_{13}, \omega_{14}, \omega_{15}, \omega_{18}, \omega_{19}, \omega_{20}, \omega_{23}, \omega_{24}, \omega_{25}, \omega_{28}, \omega_{29}, \omega_{30}\}$ .

5. Consider an experiment in which two dice are tossed. The sample space  $S$  for this experiment is given by

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

and  $n(S) = 6 \times 6 = 36$ .

Let  $E_1$  be the event that 'the sum of the spots on the dice is greater than 12',  $E_2$  be the event that 'the sum of spots on the dice is divisible by 3', and  $E_3$  be the event that 'the sum is greater than or equal to two and is less than or equal to 12'. Then these events are represented by the following subsets of  $S$ :

$$E_1 = \{\emptyset\}, E_3 = S \text{ and}$$

$$\begin{aligned} E_2 &= \{(1, 2), (1, 5), (2, 1), (2, 4), (3, 3), (3, 6), (4, 2), \\ &\quad (4, 5), (5, 1), (5, 4), (6, 3), (6, 6)\} \end{aligned}$$

Thus  $n(E_1) = 0$ ,  $n(E_2) = 12$ , and  $n(E_3) = 36$

Here  $E_1$  is an 'impossible event' and  $E_3$  a 'certain event'.

6. Let  $E$  denote the experiment of tossing a coin three times in succession or tossing three coins at a time. Then the sample space  $S$  is given by

$$\begin{aligned} S &= \{H, T\} \times \{H, T\} \times \{H, T\} \\ &= \{H, T\} \times \{HH, HT, TH, TT\} \\ &= \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\} \\ &= \{\omega_1, \omega_2, \omega_3, \dots, \omega_8\}, \text{ say.} \end{aligned}$$

If  $E_1$  is the event that 'the number of heads exceeds the number of tails',  $E_2$ , the event of 'getting two heads' and  $E_3$ , the event of getting 'head in the first trial' then these are represented by the following sets of points :

$$E_1 = \{\omega_1, \omega_2, \omega_3, \omega_5\},$$

$$E_2 = \{\omega_2, \omega_3, \omega_5\}$$

and  $E_3 = \{\omega_1, \omega_2, \omega_3, \omega_4\}$ .

7. In the foregoing examples the sample space is finite. To construct an experiment in which the sample space is countably infinite, we toss a coin repeatedly until head or tail appears twice in succession. The sample space of all the possible outcomes may be represented as :

$$S = \{HH, TT, THH, HTT, HTHH, THTT, THTHH, HTHTT, \dots\}.$$

4-5-4. Algebra of Events. For events  $A, B, C$

$$(i) A \cup B = \{\omega \in S : \omega \in A \text{ or } \omega \in B\}$$

$$(ii) A \cap B = \{\omega \in S : \omega \in A \text{ and } \omega \in B\}$$

$$(iii) \bar{A} (\text{A complement}) = \{\omega \in S : \omega \notin A\}$$

$$(iv) A - B = \{\omega \in S : \omega \in A \text{ but } \omega \notin B\}$$

$$(v) \text{Similar generalisations for } \bigcup_{i=1}^n A_i, \bigcap_{i=1}^n A_i, \cup_i A_i \text{ etc.}$$

$$(vi) A \subset B \Rightarrow \text{for every } \omega \in A, \omega \in B.$$

- (vii)  $B \supset A \Rightarrow A \subset B$ .  
(viii)  $A = B$  if and only if  $A$  and  $B$  have the same elements, i.e., if  $A \subset B$  and  $B \subset A$ .  
(ix)  $A$  and  $B$  disjoint (mutually exclusive)  $\Rightarrow A \cap B = \emptyset$  (null set).  
(x)  $A \cup B$  can be denoted by  $A + B$  if  $A$  and  $B$  are disjoint.  
(xi)  $A \Delta B$  denotes those  $\omega$  belonging to exactly one of  $A$  and  $B$ , i.e.,  

$$A \Delta B = A \bar{B} \cup \bar{A}B$$

**Remark.** Since the events are subsets of  $S$ , all the laws of set theory viz., commutative laws, associative laws, distributive laws, De-Morgan's law, etc., hold for algebra of events.

Table – Glossary of Probability Terms

Statement	Meaning in terms of set theory
1. At least one of the events $A$ or $B$ occurs.	$\omega \in A \cup B$
2. Both the events $A$ and $B$ occur.	$\omega \in A \cap B$
3. Neither $A$ nor $B$ occurs	$\omega \in \bar{A} \cap \bar{B}$
4. Event $A$ occurs and $B$ does not occur	$\omega \in A \cap \bar{B}$
5. Exactly one of the events $A$ or $B$ occurs.	$\omega \in A \Delta B$
6. Not more than one of the events $A$ or $B$ occurs.	$\omega \in (A \cap \bar{B}) \cup (\bar{A} \cap B) \cup (\bar{A} \cap \bar{B})$
7. If event $A$ occurs, so does $B$	$A \subset B$
8. Events $A$ and $B$ are mutually exclusive.	$A \cap B = \emptyset$
9. Complementary event of $A$ .	$\bar{A}$
10. Sample space	universal set $S$

**Example 4-11.**  $A$ ,  $B$  and  $C$  are three arbitrary events. Find expressions for the events noted below, in the context of  $A$ ,  $B$  and  $C$ .

- (i) only  $A$  occurs,
- (ii) Both  $A$  and  $B$ , but not  $C$ , occur,
- (iii) All three events occur,
- (iv) At least one occurs,
- (v) At least two occur,
- (vi) One and no more occurs,
- (vii) Two and no more occur,
- (viii) None occurs.

**Solution.**

- (i)  $A \cap \bar{B} \cap \bar{C}$ ,
- (ii)  $A \cap B \cap \bar{C}$ ,
- (iii)  $A \cap B \cap C$ ,
- (iv)  $A \cup B \cup C$ .

- (v)  $(A \cap B \cap \bar{C}) \cup (A \cap \bar{B} \cap C) \cup (\bar{A} \cap B \cap C) \cup (A \cap B \cap C)$   
 (vi)  $(A \cap \bar{B} \cap \bar{C}) \cup (\bar{A} \cap B \cap \bar{C}) \cup (\bar{A} \cap \bar{B} \cap C)$   
 (vii)  $(A \cap B \cap \bar{C}) \cup (\bar{A} \cap B \cap C) \cup (A \cap \bar{B} \cap C)$   
 (viii)  $\bar{A} \cap \bar{B} \cap \bar{C}$  or  $\overline{A \cup B \cup C}$

### EXERCISE 4(b)

1. (i) If  $A, B$  and  $C$  are any three events, write down the theoretical expressions for the following events:

- (a) Only  $A$  occurs, (b)  $A$  and  $B$  occur but  $C$  does not,
- (c)  $A, B$ , and  $C$  all the three occur, (d) at least one occurs
- (e) At least two occur, (f) one does not occur,
- (g) Two do not occurs, and (h) None occurs.

(ii)  $A, B$  and  $C$  are three events. Express the following events in appropriate symbols:

- (a) Simultaneous occurrence of  $A, B$  and  $C$ .
- (b) Occurrence of at least one of them.
- (c)  $A, B$  and  $C$  are mutually exclusive events.
- (d) Every point of  $A$  is contained in  $B$ .
- (e) The event  $B$  but not  $A$  occurs. [Gauhati Univ. B.Sc., Oct. 1990]

2. A sample space  $S$  contains four points  $x_1, x_2, x_3$  and  $x_4$  and the values of a set function  $P(A)$  are known for the following sets :

$$A_1 = (x_1, x_2) \text{ and } P(A_1) = \frac{4}{10}; \quad A_2 = (x_3, x_4) \text{ and } P(A_2) = \frac{6}{10};$$

$$A_3 = (x_1, x_2, x_3) \text{ and } P(A_3) = \frac{4}{10}; \quad A_4 = (x_2, x_3, x_4) \text{ and } P(A_4) = \frac{7}{10}$$

Show that :

(i) the total number of sets (including the "null" set of number points) of points of  $x$  is 16.

(ii) Although the set containing no sample point has zero probability, the converse is not always true, i.e., a set may have zero probability and yet it may be the set of a number of points.

3. Describe explicitly the sample spaces for each of the following experiments:

- (i) The tossing of four coins.
- (ii) The throwing of three dice.
- (iii) The tossing of ten coins with the aim of observing the numbers of tails coming up.
- (iv) Two cards are selected from a standard deck of cards.
- (v) Four successive draws (a) with replacement, and (b) without replacement, from a bag containing fifty coloured balls out of which ten are white, twenty blue and twenty red.
- (vi) A survey of families with two children is conducted and the sex of the children (the older child first) is recorded.
- (vii) A survey of families with three children is made and the sex of the children (in order of age, oldest child first) are recorded.

(viii) Three distinguishable objects are distributed in three numbered cells.

(ix) A poker hand (five cards) is dealt from an ordinary deck of cards.

(x) Selecting  $r$  screws from the lot produced by a machine, a screw can be defective or non-defective.

4. In an experiment a coin is thrown five times. Write down the sample space. How many points are there in the sample space?

5. Describe sample space appropriate in each of the following cases :

(i)  $n$ -tosses of a coin with head or tails as outcome in each toss.

(ii) Successive tosses of a coin until a head turns up.

(iii) A survey of families with two children is conducted and the sex of the children (the older child first) is recorded.

(iv) Two successive draws, (a) with replacement (b) without replacement, from a bag containing 4 coloured toys out of which one is white, one black and 2 red toys.

[M.S.Baroda Univ. B.Sc., 1991]

6. (a) An experiment consists of tossing an unbiased coin until the same result appears twice on succession for the first time. To every possible outcome requiring  $n$  tosses attribute probability  $1/2^n$ . Describe the sample space.

(b) A coin is tossed until there are either two consecutive heads or two consecutive tails or the number of tosses becomes five. Describe the sample space along with the probability associated with each sample point, if every sequence of  $n$  tosses has probability  $2^{-n}$ .

[Civil Services (main), 1983]

7. Urn 1 contains two white, one red and 3 black balls. Urn 2 contains one white, 3 red and 2 black balls. An experiment consists of first selecting an urn and then drawing a ball from this urn. Define a suitable sample space for this experiment.

8. Suppose an experiment has  $n$  outcomes  $A_1, A_2, \dots, A_n$  and that it is repeated  $r$  times. Let  $x_1, x_2, \dots, x_r$  record the number of occurrences of  $A_1, A_2, \dots, A_n$ . Describe the sample space. Show that the number of sample points is

$$\binom{n+r-1}{r-1}$$

9. A manufacturer buys parts from four different vendors numbered 1, 2, 3 and 4. Referring to orders placed on two successive days, (1,4) denotes the event that on the first day, the order was given to vendor 1 and on the second day it was given to vendor 4. Letting  $A$  represent the event that vendor 1 gets at least one of these two orders,  $B$  the event that the same vendor gets both orders and  $C$  the event that vendors 1 and 3 do not get either order. List the elements of :

(a) entire sample space, (b)  $A$ , (c)  $B$ , (d)  $C$ , (e)  $\bar{A}$ , (f)  $\bar{B}$ ,

(g)  $B \cup C$ , (h)  $A \cap B$ , (i)  $A \cap C$ , (j)  $\bar{A} \cup \bar{B}$ , and (k)  $A - B$

[Hint. (a) The elements of entire sample space are

(1,1); (1, 2); (1, 3); (1, 4); (2, 1); (2, 2); (2, 3); (2, 4);

(3, 1); (3, 2); (3, 3); (3, 4); (4, 1); (4, 2); (4, 3); (4, 4).

- (b) The elements of  $A$  are  
 $(1, 1); (1, 2); (1, 3); (1, 4); (2, 1); (3, 1); (4, 1);$
- (c) The elements of  $B$  are  $(1, 1); (2, 2); (3, 3)$  and  $(4, 4)$ .
- (d) The elements of  $C$  are  $(2, 2); (2, 4); (4, 2); (4, 4)$ .
- (e) The elements of  $\bar{A}$  are :  
 $(2, 2); (2, 3); (2, 4); (3, 2); (3, 3); (3, 4); (4, 2); (4, 3); (4, 4).$
- (f) The elements of  $\bar{B}$  are :  
 $(1, 2); (1, 3); (1, 4); (2, 1); (2, 3); (2, 4);$   
 $(3, 1); (3, 2); (3, 4); (4, 1); (4, 2); (4, 3).$
- (g) The elements of  $B \cup C$  are  $(1, 1); (2, 2); (3, 3); (4, 4); (2, 4); (4, 2).$
- (h) The elements of  $A \cap B$  are  $(1, 1).$
- (i)  $A \cap C = \emptyset$
- (j) Since  $\overline{A \cup B} = \bar{A} \cap \bar{B}$ . The elements of  $\overline{A \cup B}$  are  $(2, 3); (2, 4); (3, 2);$   
 $(3, 4); (4, 2); (4, 3).$
- (k) The elements of  $A - B$  are  $(1, 2); (1, 3); (1, 4); (2, 1); (3, 1); (4, 1).$

**4.6. Probability — Mathematical Notion.** We are now set to give the mathematical notion of the occurrence of a random phenomenon and the mathematical notion of probability. Suppose in a large number of trials the sample space  $S$  contains  $N$  sample points. The event  $A$  is defined by a description which is satisfied by  $N_A$  of the occurrences. The frequency interpretation of the probability  $P(A)$  of the event  $A$ , tells us that  $P(A) = N_A/N$ .

A purely mathematical definition of probability cannot give us the actual value of  $P(A)$  and this must be considered as a function defined on all events. With this in view, a mathematical definition of probability is enunciated as follows:

*"Given a sample description space, probability is a function which assigns a non-negative real number to every event  $A$ , denoted by  $P(A)$  and is called the probability of the event  $A$ ."*

**4.6.1. Probability Function.**  $P(A)$  is the probability function defined on a  $\sigma$ -field  $B$  of events if the following properties or axioms hold :

1. For each  $A \in B$ ,  $P(A)$  is defined, is real and  $P(A) \geq 0$

2.  $P(S) = 1$

3. If  $\{A_n\}$  is any finite or infinite sequence of disjoint events in  $B$ , then

$$P(\bigcup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i) \quad \dots(4.4)$$

The above three axioms are termed as the axiom of positiveness, certainty and union (additivity), respectively.

**Remarks.** 1. The set function  $P$  defined on  $\sigma$ -field  $B$ , taking its values in the real line and satisfying the above three axioms is called the probability measure.

2. The same definition of probability applies to *uncountable sample space* except that special restrictions must be placed on  $S$  and its subsets. It is important to realise that for a complete description of a probability measure, three things must

be specified, viz., the sample space  $S$ , the  $\sigma$ -field ( $\sigma$ -algebra)  $B$  formed from certain subset of  $S$  and set function  $P$ . The triplet  $(S, B, P)$  is often called the *probability space*. In most elementary applications,  $S$  is finite and the  $\sigma$ -algebra  $B$  is taken to be the collection of all subsets of  $S$ .

3. It is interesting to see that there are some formal statements of the properties of events derived from the frequency approach. Since  $P(A) = N_A/N$ , it is easy to see that  $P(A) \geq 0$ , as in Axiom 1. Next since  $N_S = N$ ,  $P(S) = 1$ , as in Axiom 2. In case of two mutually exclusive (or disjoint) events  $A$  and  $B$  defined by sample points  $N_A$  and  $N_B$ , the sample points belonging to  $A \cup B$  are  $N_A + N_B$ . Therefore,

$$P(A \cup B) = \frac{N_A + N_B}{N} = \frac{N_A}{N} + \frac{N_B}{N} = P(A) + P(B), \text{ as in axiom 3.}$$

**Extended Axiom of Addition.** If an event  $A$  can materialise in the occurrence of any one of the pairwise disjoint events  $A_1, A_2, \dots$  so that

$$A = \bigcup_{i=1}^{\infty} A_i; A_i \cap A_j = \emptyset \quad (i \neq j)$$

then

$$P(A) = P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i) \quad \dots(1)$$

**Axiom of Continuity.** If  $B_1, B_2, \dots, B_n, \dots$  be a countable sequences of events such that

$$(i) B_i \supset B_{i+1}, \quad (i = 1, 2, 3, \dots)$$

and

$$(ii) \bigcap_{n=1}^{\infty} B_n = \emptyset$$

i.e., if each succeeding event implies the preceding event and if their simultaneous occurrence is an impossible event then

$$\lim_{n \rightarrow \infty} P(B_n) = 0 \quad \dots(2)$$

We shall now prove that these two axioms, viz., the extended axiom of addition and axiom of continuity are equivalent, i.e., each implies the other, i.e., (1)  $\Leftrightarrow$  (2).

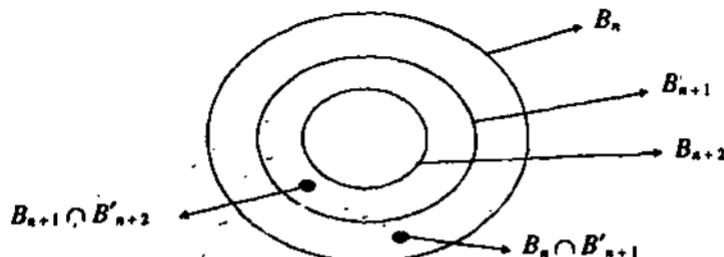
**Theorem 4.1.** *Axiom of continuity follows from the extended axiom of addition and vice versa.*

**Proof.** (a) (1)  $\Rightarrow$  (2). Let  $\{B_n\}$  be a countable sequence of events such that

$$B_1 \supset B_2 \supset B_3 \supset \dots \supset B_n \supset B_{n+1} \supset \dots$$

and let for any  $n \geq 1$ ,

$$\bigcap_{k \geq n} B_k = \emptyset \quad (*)$$



Then it is obvious from the diagram that

$$B_n = B_n B'_{n+1} \cup B_{n+1} B'_{n+2} \cup \dots \cup (\bigcap_{k \geq n} B_k)$$

$$\Rightarrow B_n = (\bigcup_{k=n}^{\infty} B_k B'_{k+1} \cup (\bigcap_{k \geq n} B_k),$$

where the events  $B_k B'_{k+1}$ ; ( $k=n, n+1, \dots$ ) are pairwise disjoint and each is disjoint with  $\bigcap_{k \geq n} B_k$ .

Thus  $B_n$  has been expressed as the countable union of pairwise disjoint events and hence by the extended axiom of addition, we get

$$P(B_n) = \sum_{k=n}^{\infty} P(B_k B'_{k+1}) + P(\bigcap_{k \geq n} B_k)$$

$$= \sum_{k=n}^{\infty} P(B_k B'_{k+1}), \quad (**)$$

since, from (\*)

$$P(\bigcap_{k \geq n} B_k) = P(\emptyset) = 0$$

Further, from (\*\*), since

$$\sum_{k=1}^{\infty} P(B_k B'_{k+1}) = P(B_1) \leq 1,$$

the right hand sum in (\*\*), being the remainder after  $n$  terms of a convergent series tends to zero as  $n \rightarrow \infty$ .

Hence

$$\lim_{n \rightarrow \infty} P(B_n) = \lim_{n \rightarrow \infty} \sum_{k=n}^{\infty} P(B_k B'_{k+1}) = 0$$

Thus the extended axiom of addition implies the axiom of continuity.

(b) Conversely (2)  $\Rightarrow$  (1), i.e., the extended axiom of addition follows from the axiom of continuity.

Let  $\{A_n\}$  be a countable sequence of pairwise disjoint events and let

$$\begin{aligned} A &= \bigcup_{i=1}^{\infty} A_i \\ &= (\bigcup_{i=1}^n A_i) \cup (\bigcup_{i=n+1}^{\infty} A_i) \end{aligned} \quad \dots(3)$$

Let us define a countable sequence  $\{B_n\}$  of events by

$$B_n = \bigcup_{i=n}^{\infty} A_i \quad \dots(4)$$

Obviously  $B_n$  is a decreasing sequence of events, i.e.,

$$B_1 \supset B_2 \supset \dots \supset B_n \supset B_{n+1} \supset \dots \quad \dots(5)$$

Also we have

$$A = (\bigcup_{i=1}^n A_i) \cup B_{n+1} \quad \dots(6)$$

Since  $A_i$ 's are pairwise disjoint, we get

$$A_i \cap B_{n+1} = \emptyset, \quad (i = 1, 2, \dots, n) \quad \dots(6a)$$

From (4) we see that if the event  $B_n$  has occurred it implies the occurrence of any one of the events  $A_{n+1}, A_{n+2}, \dots$ . Without loss of generality let us assume that this event is  $A_i$  ( $i = n+1, n+2, \dots$ ). Further since  $A_i$ 's are pairwise disjoint, the occurrence of  $A_i$  implies that events  $A_{i+1}, A_{i+2}, \dots$  do not occur leading to the conclusion that  $B_{i+1}, B_{i+2}, \dots$  will not occur.

$$\Rightarrow \bigcap_{i=n}^{\infty} B_i = \emptyset \quad \dots(7)$$

From (5) and (7), we observe that both the conditions of axiom of continuity are satisfied and hence we get

$$\lim_{n \rightarrow \infty} P(B_n) = 0 \quad \dots(8)$$

From (6), we get

$$\begin{aligned} P(A) &= P[(\bigcup_{i=1}^n A_i) \cup B_{n+1}] \\ &= \sum_{i=1}^n P(A_i) + P(B_{n+1}) \end{aligned}$$

(By axiom of Additivity)

$$\Rightarrow P(\bigcup_{i=1}^{\infty} A_i) = \lim_{n \rightarrow \infty} \sum_{i=1}^n P(A_i) + \lim_{n \rightarrow \infty} P(B_{n+1})$$

$$= \sum_{i=1}^{\infty} P(A_i), \quad [\text{From (8)}]$$

which is the extended axiom of addition.

### THEOREMS ON PROBABILITIES OF EVENTS

**Theorem 4-2.** *Probability of the impossible event is zero, i.e.,  $P(\phi) = 0$ .*

**Proof.** Impossible event contains no sample point and hence the certain event  $S$  and the impossible event  $\phi$  are mutually exclusive.

Hence  $S \cup \phi = S$

$$\therefore P(S \cup \phi) = P(S)$$

$$\Rightarrow P(S) + P(\phi) = P(S)$$

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

[By Axiom 3]

**Remark.** It may be noted  $P(A)=0$ , does not imply that  $A$  is necessarily an empty set. In practice, probability '0' is assigned to the events which are so rare that they happen only once in a lifetime. For example, if a person who does not know typing is asked to type the manuscript of a book, the probability of the event that he will type it correctly without any mistake is 0.

As another illustration, let us consider the random tossing of a coin. The event that the coin will stand erect on its edge, is assigned the probability 0.

The study of continuous random variable provides another illustration to the fact that  $P(A)=0$ , does not imply  $A=\phi$ , because in case of continuous random variable  $X$ , the probability at a point is always zero, i.e.,  $P(X=c)=0$  [See Chapter 5].

**Theorem 4-3.** Probability of the complementary event  $\bar{A}$  of  $A$  is given by

$$P(\bar{A}) = 1 - P(A)$$

**Proof.**  $A$  and  $\bar{A}$  are disjoint events.

Moreover,  $A \cup \bar{A} = S$

From axioms 2 and 3 of probability, we have

$$P(A \cup \bar{A}) = P(A) + P(\bar{A}) = P(S) = 1$$

$$\Rightarrow P(\bar{A}) = 1 - P(A)$$

**Cor. 1.** We have  $P(A) = 1 - P(\bar{A})$

$$\Rightarrow P(A) \leq 1 \quad (\because P(\bar{A}) \geq 0)$$

**Cor. 2.**  $P(\phi) = 0$ , since  $\phi = S$

$$\text{and } P(\phi) = P(S) = 1 - P(S) = 1 - 1 = 0.$$

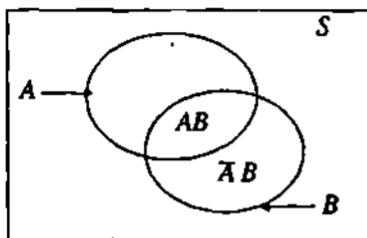
**Theorem 4-4.** For any two events  $A$  and  $B$ ,

$$P(\bar{A} \cap B) = P(B) - P(A \cap B) \quad [\text{Mysore Univ. B.Sc., 1992}]$$

**Proof.**

$\bar{A} \cap B$  and  $A \cap B$  are disjoint events and

$$(A \cap B) \cup (\bar{A} \cap B) = B$$



Hence by axiom 3, we get

$$\begin{aligned} P(B) &= P(A \cap B) + P(\bar{A} \cap B) \\ \Rightarrow P(\bar{A} \cap B) &= P(B) - P(A \cap B) \end{aligned}$$

**Remark.** Similarly, we shall get  
 $P(A \cap \bar{B}) = P(A) - P(A \cap B)$

**Theorem 4.5.** Probability of the union of any two events  $A$  and  $B$  is given by

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

**Proof.**  $A \cup B$  can be written as the union of the two mutually disjoint events,  $A$  and  $B \cap \bar{A}$ .

$$\therefore P(A \cup B) = P[A \cup (B \cap \bar{A})] = P(A) + P(B \cap \bar{A}) \\ = P(A) + P(B) - P(A \cap B) \quad (\text{c.f. Theorem 4.4})$$

**Theorem 4.6.** If  $B \subset A$ , then

$$(i) P(A \cap \bar{B}) = P(A) - P(B),$$

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

**Proof.** (i) When  $B \subset A$ ,  $B$  and  $A \cap \bar{B}$  are mutually exclusive events and their union is  $A$

Therefore

$$\begin{aligned} P(A) &= P[B \cup (A \cap \bar{B})] \\ &= P(B) + P(A \cap \bar{B}) \quad [\text{By axiom 3}] \\ \Rightarrow P(A \cap \bar{B}) &= P(A) - P(B) \end{aligned}$$

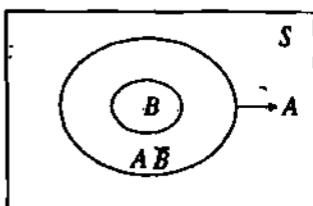
(ii) Using axiom 1,

$$P(A \cap \bar{B}) \geq 0 \Rightarrow P(A) - P(B) \geq 0$$

$$\text{Hence } P(B) \leq P(A)$$

Cor. Since  $(A \cap B) \subset A$  and  $(A \cap B) \subset B$ ,

$$P(A \cap B) \leq P(A) \text{ and } P(A \cap B) \leq P(B)$$

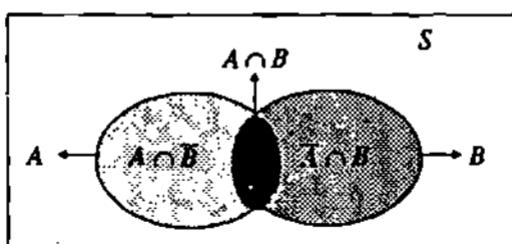


#### 4.6.2. Law of Addition of Probabilities

**Statement.** If  $A$  and  $B$  are any two events [subsets of sample space  $S$ ] and are not disjoint, then

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \quad \dots(4.5)$$

**Proof.**



We have

$$A \cup B = A \cup (\bar{A} \cap B)$$

Since  $A$  and  $(\bar{A} \cap B)$  are disjoint,

$$\begin{aligned} P(A \cup B) &= P(A) + P(\bar{A} \cap B) \\ &= P(A) + [P(\bar{A} \cap B) + P(A \cap B)] - P(A \cap B) \\ &= P(A) + P[(\bar{A} \cap B) \cup (A \cap B)] - P(A \cap B) \end{aligned}$$

[ $\because (\bar{A} \cap B)$  and  $(A \cap B)$  are disjoint]

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

**Remark.** An alternative proof is provided by Theorems 4.4 and 4.5.

**4.6.3. Extention of General Law of Addition of Probabilities.** For  $n$  events  $A_1, A_2, \dots, A_n$ , we have

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i) - \sum_{1 \leq i < j \leq n} P(A_i \cap A_j) + \sum_{1 \leq i < j < k \leq n} P(A_i \cap A_j \cap A_k) - \dots + (-1)^{n-1} P(A_1 \cap A_2 \cap \dots \cap A_n) \quad \dots(4.6)$$

**Proof.** For two events  $A_1$  and  $A_2$ , we have

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2) \quad \dots(*)$$

Hence (4.6) is true for  $n = 2$ .

Let us now suppose that (4.6) is true for  $n = r$ , (say). Then

$$P\left(\bigcup_{i=1}^r A_i\right) = \sum_{i=1}^r P(A_i) - \sum_{1 \leq i < j \leq r} P(A_i \cap A_j) + \dots + (-1)^{r-1} P(A_1 \cap A_2 \cap \dots \cap A_r) \quad \dots(**)$$

Now

$$\begin{aligned} P\left(\bigcup_{i=1}^{r+1} A_i\right) &= P\left[\left(\bigcup_{i=1}^r A_i\right) \cup A_{r+1}\right] \\ &= P\left(\bigcup_{i=1}^r A_i\right) + P(A_{r+1}) - P\left[\left(\bigcup_{i=1}^r A_i\right) \cap A_{r+1}\right] \quad \dots[\text{Using } (*)] \\ &= P\left(\bigcup_{i=1}^r A_i\right) + P(A_{r+1}) - P\left[\bigcup_{i=1}^r (A_i \cap A_{r+1})\right] \quad (\text{Distributive Law}) \\ &= \sum_{i=1}^r P(A_i) - \sum_{1 \leq i < j \leq r} P(A_i \cap A_j) + \dots \\ &\quad \dots + (-1)^{r-1} P(A_1 \cap A_2 \cap \dots \cap A_r) + P(A_{r+1}) \\ &\quad - P\left[\bigcup_{i=1}^r (A_i \cap A_{r+1})\right] \quad \dots[\text{From } (**)] \\ &= \sum_{i=1}^{r+1} P(A_i) - \sum_{1 \leq i < j \leq r} P(A_i \cap A_j) + \dots \\ &\quad + (-1)^{r-1} P(A_1 \cap A_2 \cap \dots \cap A_r) \end{aligned}$$

$$\begin{aligned}
 & - \left[ \sum_{i=1}^r P(A_i \cap A_{i+1}) - \sum_{1 \leq i < j \leq r} P(A_i \cap A_j \cap A_{j+1}) \right. \\
 & \quad \left. + \dots + (-1)^{r-1} P(A_1 \cap A_2 \cap \dots \cap A_r \cap A_{r+1}) \right] \quad \dots [\text{From } (**)] \\
 \Rightarrow P(\bigcup_{i=1}^{r+1} A_i) &= \sum_{i=1}^{r+1} P(A_i) - \left[ \sum_{1 \leq i < j \leq r} P(A_i \cap A_j) + \sum_{i=1}^r P(A_i \cap A_{i+1}) \right] \\
 & \quad + \dots + (-1)^r P(A_1 \cap A_2 \cap \dots \cap A_{r+1}) \\
 &= \sum_{i=1}^{r+1} P(A_i) - \sum_{1 \leq i < j \leq (r+1)} P(A_i \cap A_j) \\
 & \quad + \dots + (-1)^r P(A_1 \cap A_2 \cap \dots \cap A_{r+1})
 \end{aligned}$$

Hence if (4.6) is true for  $n=r$ , it is also true for  $n=(r+1)$ . But we have proved in (\*) that (4.6) is true for  $n=2$ . Hence by the principle of mathematical induction, it follows that (4.6) is true for all positive integral values of  $n$ .

**Remarks.** 1. If we write

$$P(A_i) = p_i, P(A_i \cap A_j) = p_{ij}, P(A_i \cap A_j \cap A_k) = p_{ijk}$$

and so on and

$$\begin{aligned}
 S_1 &= \sum_{i=1}^n p_i = \sum_{i=1}^n P(A_i) \\
 S_2 &= \sum_{1 \leq i < j \leq n} p_{ij} = \sum_{1 \leq i < j \leq n} P(A_i \cap A_j) \\
 S_3 &= \sum_{1 \leq i < j < k \leq n} p_{ijk} \quad \text{and so on,}
 \end{aligned}$$

then

$$P(\bigcup_{i=1}^n A_i) = S_1 - S_2 + S_3 - \dots + (-1)^{n-1} S_n \quad \dots (4.6a)$$

2. If all the events  $A_i$ , ( $i = 1, 2, \dots, n$ ) are mutually disjoint then (4.6) gives

$$P(\bigcup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i)$$

3. From practical point of view the theorem can be restated in a slightly different form. Let us suppose that an event  $A$  can materialise in several mutually exclusive forms, viz.,  $A_1, A_2, \dots, A_n$  which may be regarded as that many mutually exclusive events. If  $A$  happens then any one of the events  $A_i$ , ( $i = 1, 2, \dots, n$ ) must happen and conversely if any one of the events  $A_i$ , ( $i = 1, 2, \dots, n$ ) happens, then  $A$  happens. Hence the probability of happening of  $A$  is the same as the probability of happening of any one of its (unspecified) mutually exclusive forms. From this point of view, the total probability theorem can be restated as follows:

*The probability of happening of an event  $A$  is the sum of the probabilities of happening of its mutually exclusive forms  $A_1, A_2, \dots, A_n$ . Symbolically,*

$$P(A) = P(A_1) + P(A_2) + \dots + P(A_n) \quad \dots (4.6b)$$

The probabilities  $P(A_1), P(A_2), \dots, P(A_n)$  of the mutually exclusive forms of  $A$  are known as the *partial probabilities*. Since  $P(\bar{A})$  is their sum, it may be called the *total probability* of  $A$ . Hence the name of the theorem.

**Theorem 4.7. (Boole's inequality).** For  $n$  events  $\bar{A}_1, \bar{A}_2, \dots, \bar{A}_n$ , we have

$$(a) \quad P\left(\bigcap_{i=1}^n A_i\right) \geq \sum_{i=1}^n P(A_i) - (n-1) \quad \dots(4.7)$$

$$(b) \quad P\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n P(A_i) \quad \dots(4.7a)$$

[Delhi Univ. B.Sc. (Stat Hons.), 1992, 1989]

$$\text{Proof. } (a) \quad P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2) \leq 1$$

$$\Rightarrow P(A_1 \cap A_2) \geq P(A_1) + P(A_2) - 1 \quad (*)$$

Hence (4.7) is true for  $n=2$ .

Let us now suppose that (4.7) is true for  $n=r$  (say), such that

$$P\left(\bigcap_{i=1}^r A_i\right) \geq \sum_{i=1}^r P(A_i) - (r-1) \quad (**)$$

Then

$$\begin{aligned} P\left(\bigcap_{i=1}^{r+1} A_i\right) &= P\left(\bigcap_{i=1}^r A_i \cap A_{r+1}\right) \\ &\geq P\left(\bigcap_{i=1}^r A_i\right) + P(A_{r+1}) - 1 \quad [\text{From } (*)] \\ &\geq \sum_{i=1}^r P(A_i) - (r-1) + P(A_{r+1}) - 1 \quad [\text{From } (**)] \end{aligned}$$

$$\Rightarrow P\left(\bigcap_{i=1}^{r+1} A_i\right) \geq \sum_{i=1}^{r+1} P(A_i) - r$$

$\Rightarrow$  (4.7) is true for  $n=r+1$  also.

The result now follows by the principle of mathematical induction.

(b) Applying the inequality (4.7) to the events  $\bar{A}_1, \bar{A}_2, \dots, \bar{A}_n$ , we get

$$\begin{aligned} P(\bar{A}_1 \cap \bar{A}_2 \cap \dots \cap \bar{A}_n) &\geq [P(\bar{A}_1) + P(\bar{A}_2) + \dots + P(\bar{A}_n)] - (n-1) \\ &= [1 - P(A_1)] + [1 - P(A_2)] + \dots + [1 - P(A_n)] - (n-1) \\ &= 1 - P(A_1) - P(A_2) - \dots - P(A_n) \\ \Rightarrow P(A_1) + P(A_2) + \dots + P(A_n) &\geq 1 - P(\bar{A}_1 \cap \bar{A}_2 \cap \dots \cap \bar{A}_n) \\ &= 1 - P(\bar{A}_1 \cup \bar{A}_2 \cup \dots \cup \bar{A}_n) \\ &= P(A_1 \cup A_2 \cup \dots \cup A_n) \\ \Rightarrow P(A_1 \cup A_2 \cup \dots \cup A_n) &\leq P(A_1) + P(A_2) + \dots + P(A_n) \end{aligned}$$

as desired.

Aliter for (b) i.e., (4.7a). We have

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$$

$$\leq P(A_1) + P(A_2) \quad [\because P(A_1 \cap A_2) \geq 0] \quad \dots (***)$$

Hence (4.7a) is true for  $n=2$ .

Let us now suppose that (4.7a) is true for  $n=r$ , (say), so that

$$P\left(\bigcup_{i=1}^r A_i\right) \leq \sum_{i=1}^r P(A_i) \quad \dots (****)$$

Now

$$P\left(\bigcup_{i=1}^{r+1} A_i\right) = P\left(\bigcup_{i=1}^r A_i \cup A_{r+1}\right)$$

$$\leq P\left(\bigcup_{i=1}^r A_i\right) + P(A_{r+1}) \quad [\text{Using } (***)]$$

$$\leq \sum_{i=1}^r P(A_i) + P(A_{r+1}) \quad [\text{Using } (****)]$$

$$\Rightarrow P\left(\bigcup_{i=1}^{r+1} A_i\right) \leq \sum_{i=1}^{r+1} P(A_i)$$

Hence if (4.7a) is true for  $n=r$ , then it is also true for  $n=r+1$ . But we have proved in (\*\*\* ) that (4.7a) is true for  $n=2$ . Hence by mathematical induction we conclude that (4.7a) is true for all positive integral values of  $n$ .

**Theorem 4.8.** For  $n$  events  $A_1, A_2, \dots, A_n$ ,

$$P\left[\bigcup_{i=1}^n A_i\right] \geq \sum_{i=1}^n P(A_i) - \sum_{1 \leq i < j \leq n} P(A_i \cap A_j)$$

[Delhi Univ. B.Sc. (Stat Hons.), 1986]

**Proof.** We shall prove this theorem by the method of induction.

We know that

$$P(A_1 \cup A_2 \cup A_3) = P(A_1) + P(A_2) + P(A_3) - [P(A_1 \cap A_2) + P(A_2 \cap A_3) + P(A_3 \cap A_1)] + P(A_1 \cap A_2 \cap A_3)$$

$$\Rightarrow P\left(\bigcup_{i=1}^3 A_i\right) \geq \sum_{i=1}^3 P(A_i) - \sum_{1 \leq i < j \leq 3} P(A_i \cap A_j)$$

Thus the result is true for  $n=3$ . Let us now suppose that the result is true for  $n=r$  (say), so that

$$P\left(\bigcup_{i=1}^r A_i\right) \geq \sum_{i=1}^r P(A_i) - \sum_{1 \leq i < j \leq r} P(A_i \cap A_j) \quad \dots (*)$$

Now

$$P\left(\bigcup_{i=1}^{r+1} A_i\right) = P\left(\bigcup_{i=1}^r A_i \cup A_{r+1}\right)$$

$$= P\left(\bigcup_{i=1}^r A_i\right) + P(A_{r+1}) - P\left[\left(\bigcup_{i=1}^r A_i\right) \cap A_{r+1}\right]$$

$$\begin{aligned}
 &= P\left(\bigcup_{i=1}^r A_i\right) + P(A_{r+1}) - P\left[\bigcup_{i=1}^r (A_i \cap A_{r+1})\right] \\
 &\geq \left[ \sum_{i=1}^r P(A_i) - \sum_{1 \leq i < j < r} P(A_i \cap A_j) \right] \\
 &\quad + P(A_{r+1}) - P\left[\bigcup_{i=1}^r (A_i \cap A_{r+1})\right]
 \end{aligned} \quad \dots (**)$$

[From (\*)]

From Boole's inequality (c.f. Theorem 4.7 page 4-33), we get

$$\begin{aligned}
 P\left[\bigcup_{i=1}^r (A_i \cap A_{r+1})\right] &\leq \sum_{i=1}^r P(A_i \cap A_{r+1}) \\
 \Rightarrow -P\left[\bigcup_{i=1}^r (A_i \cap A_{r+1})\right] &\geq -\sum_{i=1}^r P(A_i \cap A_{r+1})
 \end{aligned}$$

∴ From (\*\*), we get

$$\begin{aligned}
 P\left(\bigcup_{i=1}^{r+1} A_i\right) &\geq \sum_{i=1}^{r+1} P(A_i) - \sum_{1 \leq i < j \leq r} P(A_i \cap A_j) - \sum_{i=1}^r P(A_i \cap A_{r+1}) \\
 \Rightarrow P\left(\bigcup_{i=1}^{r+1} A_i\right) &\geq \sum_{i=1}^{r+1} P(A_i) - \sum_{1 \leq i < j \leq r+1} P(A_i \cap A_j)
 \end{aligned}$$

Hence, if the theorem is true for  $n = r$ , it is also true for  $n = r + 1$ . But we have seen that the result is true for  $n = 3$ . Hence by mathematical induction, the result is true for all positive integral values of  $n$ .

#### 4.7. Multiplication Law of Probability and Conditional Probability

**Theorem 4.8.** For two events  $A$  and  $B$

$$\begin{aligned}
 P(A \cap B) &= P(A) \cdot P(B | A), \quad P(A) > 0 \\
 &= P(B) \cdot P(A | B), \quad P(B) > 0
 \end{aligned} \quad \dots (4.8)$$

where  $P(B | A)$  represents the conditional probability of occurrence of  $B$  when the event  $A$  has already happened and  $P(A | B)$  is the conditional probability of happening of  $A$ , given that  $B$  has already happened.

**Proof.**

$$P(A) = \frac{n(A)}{n(S)} ; \quad P(B) = \frac{n(B)}{n(S)} \quad \text{and} \quad P(A \cap B) = \frac{n(A \cap B)}{n(S)} \quad (*)$$

For the conditional event  $A | B$ , the favourable outcomes must be one of the sample points of  $B$ , i.e., for the event  $A | B$ , the sample space is  $B$  and out of the  $n(B)$  sample points,  $n(A \cap B)$  pertain to the occurrence of the event  $A$ . Hence

$$P(A | B) = \frac{n(A \cap B)}{n(B)}$$

Rewriting (\*), we get

$$P(A \cap B) = \frac{n(B)}{n(S)} \cdot \frac{n(A \cap B)}{n(B)} = P(B) \cdot P(A | B)$$

Similarly we can prove :

$$P(A \cap B) = \frac{n(A)}{n(S)} \cdot \frac{n(A \cap B)}{n(A)} = P(A) \cdot P(B|A)$$

$$\text{Remarks. 1. } P(B|A) = \frac{P(A \cap B)}{P(A)} \text{ and } P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Thus the conditional probabilities  $P(B|A)$  and  $P(A|B)$  are defined if and only if  $P(A) \neq 0$  and  $P(B) \neq 0$ , respectively.

2. (i) For  $P(B) > 0$ ,  $P(A|B) \leq P(A)$

(ii) The conditional probability  $P(A|B)$  is not defined if  $P(B) = 0$ .

(iii)  $P(B|B) = 1$ .

3. **Multiplication Law of Probability for Independent Events.** If  $A$  and  $B$  are independent then

$$P(A|B) = P(A) \text{ and } P(B|A) = P(B)$$

Hence (4.8) gives :

$$P(A \cap B) = P(A)P(B) \quad \dots(4.8a)$$

provided  $A$  and  $B$  are independent.

**4.7.1. Extension of Multiplication Law of Probability.** For  $n$  events  $A_1, A_2, \dots, A_n$ , we have

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \dots \times P(A_n|A_1 \cap A_2 \cap \dots \cap A_{n-1}) \quad \dots(4.8b)$$

where  $P(A_i|A_1 \cap A_2 \cap \dots \cap A_{i-1})$  represents the conditional probability of the event  $A_i$  given that the events  $A_j, A_k, \dots, A_1$  have already happened.

**Proof.** We have for three events  $A_1, A_2$ , and  $A_3$ ,

$$\begin{aligned} P(A_1 \cap A_2 \cap A_3) &= P[A_1 \cap (A_2 \cap A_3)] \\ &= P(A_1)P(A_2 \cap A_3|A_1) \\ &= P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \end{aligned}$$

Thus we find that (4.8b) is true for  $n=2$  and  $n=3$ . Let us suppose that (4.8b) is true for  $n=k$ , so that

$$P(A_1 \cap A_2 \cap \dots \cap A_k) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \dots P(A_k|A_1 \cap A_2 \cap \dots \cap A_{k-1})$$

Now

$$\begin{aligned} P[(A_1 \cap A_2 \cap \dots \cap A_k) \cap A_{k+1}] &= P(A_1 \cap A_2 \cap \dots \cap A_k) \\ &\quad \times P(A_{k+1}|A_1 \cap A_2 \cap \dots \cap A_k) \\ &= P(A_1)P(A_2|A_1) \dots P(A_k|A_1 \cap A_2 \cap \dots \cap A_{k-1}) \\ &\quad \times P(A_{k+1}|A_1 \cap A_2 \cap \dots \cap A_k) \end{aligned}$$

Thus (4.8b) is true for  $n=k+1$  also. Since (4.8b) is true for  $n=2$  and  $n=3$ , by the principle of mathematical induction, it follows that (4.8b) is true for all positive integral values of  $n$ .

**Remark.** If  $A_1, A_2, \dots, A_n$  are independent events then

$$\begin{aligned} P(A_2|A_1) &= P(A_2), \quad P(A_3|A_1 \cap A_2) = P(A_3) \\ \dots P(A_n|A_1 \cap A_2 \cap \dots \cap A_{n-1}) &= P(A_n) \end{aligned}$$

Hence (4.8b) gives :

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1) P(A_2) \dots \cap P(A_n), \quad \dots(4.8c)$$

provided  $A_1, A_2, \dots, A_n$  are independent.

**Remark.** **Mutually Exclusive (Disjoint) Events and Independent Events.**

Let  $A$  and  $B$  be mutually exclusive (disjoint) events with positive probabilities ( $P(A) > 0, P(B) > 0$ ), i.e., both  $A$  and  $B$  are possible events such that

$$A \cap B = \emptyset \Rightarrow P(A \cap B) = P(\emptyset) = 0 \quad \dots(i)$$

Further, by compound probability theorem we have

$$P(A \cap B) = P(A) \cdot P(B|A) = P(B) \cdot P(A|B) \quad \dots(ii)$$

Since  $P(A) \neq 0, P(B) \neq 0$ , from (i) and (ii) we get

$$P(A|B) = 0 \neq P(A), \quad P(B|A) = 0 \neq P(B) \quad \dots(iii)$$

$\Rightarrow A$  and  $B$  are dependent events.

Hence two possible mutually disjoint events are always dependent (not independent) events.

However, if  $A$  and  $B$  are independent events with  $P(A) > 0$  and  $P(B) > 0$ , then

$$P(A \cap B) = P(A) P(B) \neq 0$$

$\Rightarrow A$  and  $B$  cannot be mutually exclusive.

Hence two independent events (both of which are possible events), cannot be mutually disjoint.

**4.7.2.** Given  $n$  independent events  $A_i$ , ( $i=1, 2, \dots, n$ ) with respective probabilities of occurrence  $p_i$ , to find the probability of occurrence of at least one of them.

We have

$$P(A_i) = p_i \Rightarrow P(\bar{A}_i) = 1 - p_i; \quad i = 1, 2, \dots, n$$

[ $\because (\bar{A}_1 \cup \bar{A}_2 \cup \dots \cup \bar{A}_n) = (\bar{A}_1 \cap \bar{A}_2 \cap \dots \cap \bar{A}_n)$  (De-Morgan's Law)]

Hence the probability of happening of at least one of the events is given by

$$P(A_1 \cup A_2 \cup \dots \cup A_n) = 1 - P(\bar{A}_1 \cup \bar{A}_2 \cup \dots \cup \bar{A}_n) \quad \dots(*)$$

$$= 1 - P(\bar{A}_1 \cap \bar{A}_2 \cap \dots \cap \bar{A}_n)$$

$$= 1 - P(\bar{A}_1) P(\bar{A}_2) \dots P(\bar{A}_n) \quad \dots(**)$$

[c.f. Theorem 4.14 page 4.41]

$$= 1 - [(1 - p_1)(1 - p_2) \dots (1 - p_n)]$$

$$= \left[ \sum_{i=1}^n p_i - \sum_{\substack{i,j=1 \\ i < j}}^n (p_i p_j) + \sum_{\substack{i,j,k=1 \\ i < j < k}}^n (p_i p_j p_k) \right.$$

$$\left. \dots + (-1)^{n-1} (p_1 p_2 \dots p_n) \right]$$

**Remark.** The results in (\*) and (\*\*) are very important and are used quite often in numerical problems. Result (\*) stated in words gives:

$$P[\text{happening of at least one of the events } A_1, A_2, \dots, A_n]$$

$$= 1 - P(\text{none of the events } A_1, A_2, \dots, A_n \text{ happens})$$

or equivalently,

$$\begin{aligned} P \{ \text{none of the given events happens} \} \\ = 1 - P \{ \text{at least one of them happens} \}. \end{aligned}$$

**Theorem 4-9.** For any three events  $A, B$  and  $C$

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

**Proof.** We have

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\ \Rightarrow P[(A \cap C) \cup (B \cap C)] &= P(A \cap C) + P(B \cap C) - P(A \cap B \cap C) \end{aligned}$$

Dividing both sides by  $P(C)$ , we get

$$\begin{aligned} \frac{P[(A \cap C) \cup (B \cap C)]}{P(C)} &= \frac{P(A \cap C) + P(B \cap C) - P(A \cap B \cap C)}{P(C)}, P(C) > 0 \\ &= \frac{P(A \cap C)}{P(C)} + \frac{P(B \cap C)}{P(C)} - \frac{P(A \cap B \cap C)}{P(C)} \\ \Rightarrow \frac{P[(A \cup B) \cap C]}{P(C)} &= P(A | C) + P(B | C) - P(A \cap B | C) \\ \Rightarrow P[(A \cup B) | C] &= P(A | C) + P(B | C) - P(A \cap B | C) \end{aligned}$$

**Theorem 4-10.** For any three events  $A, B$  and  $C$

$$P(A \cap \bar{B} | C) + P(A \cap B | C) = P(A | C)$$

$$\begin{aligned} \text{Proof. } P(A \cap \bar{B} | C) + P(A \cap B | C) &= \frac{P(A \cap \bar{B} \cap C)}{P(C)} + \frac{P(A \cap B \cap C)}{P(C)} \\ &= \frac{P(A \cap \bar{B} \cap C) + P(A \cap B \cap C)}{P(C)} \\ &= \frac{P(A \cap C)}{P(C)} = P(A | C) \end{aligned}$$

**Theorem 4-11.** For a fixed  $B$  with  $P(B) > 0$ ,  $P(A | B)$  is a probability function. [Delhi Univ. B.Sc. (Stat. Hons.), 1991; (Maths Hons.), 1992]

**Proof.**

$$(i) \quad P(A | B) = \frac{P(A \cap B)}{P(B)} \geq 0$$

$$(ii) \quad P(S | B) = \frac{P(S \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$$

(iii) If  $\{A_n\}$  is any finite or infinite sequences of disjoint events, then

$$\begin{aligned} P[\bigcup_n A_n | B] &= \frac{P[(\bigcup_n A_n) \cap B]}{P(B)} = \frac{P[(\bigcup_n A_n \cdot B)]}{P(B)} \\ &= \frac{\sum_n P(A_n \cdot B)}{P(B)} = \sum_n \left[ \frac{P(A_n \cdot B)}{P(B)} \right] = \sum_n P(A_n | B) \end{aligned}$$

Hence the theorem.

**Remark.** For given  $B$  satisfying  $P(B) > 0$ , the conditional probability  $P[\cdot | B]$  also enjoys the same properties as the unconditional probability.

For example, in the usual notations, we have:

- (i)  $P[\emptyset | B] = 0$
- (ii)  $P[\bar{A} | B] = 1 - P[A | B]$
- (iii)  $P\left[\bigcup_{i=1}^n A_i | B\right] = \sum_{i=1}^n P[A_i | B],$

where  $A_1, A_2, \dots, A_n$  are mutually disjoint events.

- (iv)  $P(A_1 \cup A_2 | B) = P(A_1 | B) + P(A_2 | B) - P(A_1 A_2 | B)$
- (v) If  $E \subset F$ , then  $P(E | B) \leq P(F | B)$

and so on.

The proofs of results (iv) and (v) are given in theorems 4-9 and 4-13 respectively. Others are left as exercises to the reader.

**Theorem 4-12.** For any three events,  $A$ ,  $B$  and  $C$  defined on the sample space  $S$  such that  $B \subset C$  and  $P(A) > 0$ ,

$$P(B | A) \leq P(C | A)$$

$$\begin{aligned}\text{Proof. } P(C | A) &= \frac{P(C \cap A)}{P(A)} && \text{(By definition)} \\ &= \frac{P[B \cap C \cap A] + (\bar{B} \cap C \cap A)}{P(A)} \\ &= \frac{P[B \cap C \cap A]}{P(A)} + \frac{P(\bar{B} \cap C \cap A)}{P(A)} && \text{(Using axiom 3)} \\ &\approx P[(B \cap C | A) + (\bar{B} \cap C \cap A)]\end{aligned}$$

$$\text{Now } B \subset C \Rightarrow B \cap C = B$$

$$\therefore P(C | A) = P(B | A) + P(\bar{B} \cap C | A)$$

$$\Rightarrow P(C | A) \geq P(B | A)$$

**4-7-3. Independent Events.** An event  $B$  is said to be independent (or statistically independent) of event  $A$ , if the conditional probability of  $B$  given  $A$  i.e.,  $P(B | A)$  is equal to the unconditional probability of  $B$ , i.e., if

$$P(B | A) = P(B)$$

Since

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

and since  $P(B | A) = P(B)$  when  $B$  is independent of  $A$ , we must have  $P(A | B) = P(A)$  or it follows that  $A$  is also independent of  $B$ . Hence the events  $A$  and  $B$  are independent if and only if

$$P(A \cap B) = P(A) P(B) \quad \dots(4-9)$$

**4-7-4. Pairwise Independent Events**

**Definition.** A set of events  $A_1, A_2, \dots, A_n$  are said to be pair-wise independent if

$$P(A_i \cap A_j) = P(A_i) P(A_j) \quad \forall i \neq j \quad \dots(4-10)$$

**4.7.5. Conditions for Mutual Independence of  $n$  Events.** Let  $S$  denote the sample space for a number of events. The events in  $S$  are said to be mutually independent if the probability of the simultaneous occurrence of (any) finite number of them is equal to the product of their separate probabilities.

If  $A_1, A_2, \dots, A_n$  are  $n$  events, then for their mutual independence, we should have

$$(i) \quad P(A_i \cap A_j) = P(A_i)P(A_j), \quad (i \neq j; i, j = 1, 2, \dots, n)$$

$$(ii) \quad P(A_i \cap A_j \cap A_k) = P(A_i)P(A_j)P(A_k), \quad (i \neq j \neq k; i, j, k = 1, 2, \dots, n)$$

$$\vdots \qquad \vdots$$

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1)P(A_2) \dots P(A_n)$$

It is interesting to note that the above equations give respectively  ${}^n C_2, {}^n C_3, \dots, {}^n C_n$  conditions to be satisfied by  $A_1, A_2, \dots, A_n$ .

Hence the total number of conditions for the mutual independence of  $A_1, A_2, \dots, A_n$  is  ${}^n C_2 + {}^n C_3 + \dots + {}^n C_n$ .

Since  ${}^n C_0 + {}^n C_1 + {}^n C_2 + \dots + {}^n C_n = 2^n$ , we get the required number of conditions as  $(2^n - 1 - n)$ .

In particular for three events  $A_1, A_2$  and  $A_3$ , ( $n=3$ ), we have the following  $2^3 - 1 - 3 = 4$ , conditions for their mutual independence.

$$P(A_1 \cap A_2) = P(A_1)P(A_2)$$

$$P(A_2 \cap A_3) = P(A_2)P(A_3)$$

$$P(A_1 \cap A_3) = P(A_1)P(A_3)$$

$$P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3) \quad \dots(4.11)$$

**Remarks.** 1. It may be observed that pairwise or mutual independence of events  $A_1, A_2, \dots, A_n$ , is defined only when  $P(A_i) \neq 0$ , for  $i=1, 2, \dots, n$ .

2. If the events  $A$  and  $B$  are such that  $P(A_i) \neq 0$ ,  $P(B) \neq 0$  and  $A$  is independent of  $B$ , then  $B$  is independent of  $A$ .

**Proof.** We are given that

$$P(A|B) = P(A)$$

$$\Rightarrow \frac{P(A \cap B)}{P(B)} = P(A)$$

$$\Rightarrow P(A \cap B) = P(A)P(B)$$

$$\Rightarrow \frac{P(B \cap A)}{P(A)} = P(B) \quad [\because P(A) \neq 0 \text{ and } A \cap B = B \cap A]$$

$$\Rightarrow P(B|A) = P(B),$$

which by definition of independent events, means that  $B$  is independent of  $A$ .

3. It may be noted that pairwise independence of events does not imply their mutual independence. For illustrations, see Examples 4.50 and 4.51.

**Theorem 4.13.** If  $A$  and  $B$  are independent events then  $A$  and  $\bar{B}$  are also independent events.

**Proof.** By theorem 4.4, we have

$$\begin{aligned} P(A \cap \bar{B}) &= P(A) - P(A \cap B) \\ &= P(A) - P(A)P(B) \quad [\because A \text{ and } B \text{ are independent}] \\ &= P(A)[1 - P(B)] \\ &= P(A)P(\bar{B}) \end{aligned}$$

$\Rightarrow A$  and  $\bar{B}$  are independent events.

**Aliter.**  $P(A \cap B) = P(A)P(B) = P(A)P(B|A) = P(B)P(A|B)$

i.e.,  $P(B|A) = P(B) \Rightarrow B$  is independent of  $A$ .

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

Also  $P(B|A) + P(\bar{B}|A) = 1 \Rightarrow P(B) + P(\bar{B}|A) = 1$

or  $P(\bar{B}|A) = 1 - P(B) = P(\bar{B})$

$\therefore \bar{B}$  is independent of  $A$  and by symmetry we say that  $A$  is independent of  $\bar{B}$ . Thus  $A$  and  $\bar{B}$  are independent events.

**Remark.** Similarly, we can prove that if  $A$  and  $B$  are independent events then  $\bar{A}$  and  $\bar{B}$  are also independent events.

**Theorem 4.14.** If  $A$  and  $B$  are independent events then  $\bar{A}$  and  $\bar{B}$  are also independent events.

**Proof.** We are given  $P(A \cap B) = P(A)P(B)$

$$\begin{aligned} \text{Now } P(\bar{A} \cap \bar{B}) &\doteq P(\bar{A} \cup \bar{B}) = 1 - P(A \cup B) \\ &= 1 - [P(A) + P(B) - P(A \cap B)] \\ &= 1 - [P(A) + P(B) - P(A)P(B)] \\ &= 1 - P(A) - P(B) + P(A)P(B) \\ &= [1 - P(B)] - P(A)[1 - P(B)] \\ &= [1 - P(A)][1 - P(B)] = P(\bar{A})P(\bar{B}) \end{aligned}$$

$\therefore \bar{A}$  and  $\bar{B}$  are independent events.

**Aliter.** We know

$$\begin{aligned} P(\bar{A}|\bar{B}) + P(A|\bar{B}) &= 1 \\ \Rightarrow P(\bar{A}|\bar{B}) + P(A) &= 1 \quad (\text{c.f. Theorem 4.13}) \\ \Rightarrow P(\bar{A}|\bar{B}) &= 1 - P(A) = P(\bar{A}) \end{aligned}$$

$\therefore \bar{A}$  and  $\bar{B}$  are independent events.

**Theorem 4.15.** If  $A, B, C$  are mutually independent events then  $A \cup B$  and  $C$  are also independent.

**Proof.** We are required to prove:

$$\begin{aligned} P[(A \cup B) \cap C] &= P(\bar{A} \cup \bar{B})P(C) \\ \text{L.H.S.} &= P[(A \cap C) \cup (B \cap C)] \quad [\text{Distributive Law}] \\ &= P(A \cap C) + P(B \cap C) - P(A \cap B \cap C) \\ &= P(A)P(C) + P(B)P(C) - P(A)P(B)P(C) \\ &\quad [\because A, B \text{ and } C \text{ are mutually independent}] \\ &= P(C)[P(A) + P(B) - P(A \cap B)] \end{aligned}$$

$$= P(C) P(A \cup B) = \text{R.H.S.}$$

Hence  $(A \cup B)$  and  $C$  are independent.

**Theorem 4-16.** If  $A, B$  and  $C$  are random events in a sample space and if  $A, B$  and  $C$  are pairwise independent and  $A$  is independent of  $(B \cup C)$ , then  $A, B$  and  $C$  are mutually independent.

**Proof.** We are given

$$\left. \begin{aligned} P(A \cap B) &= P(A)P(B) \\ P(B \cap C) &= P(B)P(C) \\ P(A \cap C) &= P(A)P(C) \\ P[A \cap (B \cup C)] &= P(A)P(B \cup C) \end{aligned} \right\} \quad \dots (*)$$

$$\begin{aligned} \text{Now } P[A \cap (B \cup C)] &= P[(A \cap B) \cup (A \cap C)] \\ &= P(A \cap B) + P(A \cap C) - P[A \cap B] \cap (A \cap C)] \\ &= P(A) \cdot P(B) + P(A) \cdot P(C) - P(A \cap B \cap C) \quad \dots (***) \\ \text{and } P(A)P(B \cup C) &= P(A)[P(B) + P(C) - P(B \cap C)] \\ &= P(A) \cdot P(B) + P(A)P(C) - P(A)P(B \cap C) \quad \dots (****) \end{aligned}$$

From  $(**)$  and  $(****)$ , on using  $(*)$ , we get

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

Hence  $A, B, C$  are mutually independent.

**Theorem 4-17.** For any two events  $A$  and  $B$ ,

$$P(A \cap B) \leq P(A) \leq P(A \cup B) \leq P(A) + P(B)$$

[Patna Univ. B.A.(Stat. Hons.), 1992; Delhi Univ. B.Sc.(Stat. Hons.), 1989]

**Proof.** We have

$$A = (A \cap \bar{B}) \cup (A \cap B)$$

Using axiom 3, we have

$$P(A) = P[(A \cap \bar{B}) \cup (A \cap B)] = P(A \cap \bar{B}) + P(A \cap B)$$

$$\text{Now } P[(A \cap \bar{B})] \geq 0 \quad (\text{From axiom 1})$$

$$\therefore P(A) \geq P(A \cap B) \quad \dots (*)$$

$$\text{Similarly } P(B) \geq P(A \cap B) \quad \dots (**) \quad (\text{From axiom 1})$$

$$\Rightarrow P(B) - P(A \cap B) \geq 0$$

$$\text{Now } P(A \cup B) = P(A) + [P(B) - P(A \cap B)] \quad \dots (**)$$

$$\therefore P(A \cup B) \geq P(A) \Rightarrow P(A) \leq P(A \cup B) \quad \dots (***)$$

$$\text{Also } P(A \cup B) \leq P(A) + P(B) \quad \text{[From (**)]}$$

Hence from  $(*)$ ,  $(**)$  and  $(***)$ , we get

$$P(A \cap B) \leq P(A) \leq P(A \cup B) \leq P(A) + P(B)$$

Aliter. Since  $A \cap B \subset A$ , by Theorem 4-6 (ii) page 4-30, we get

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

$$\text{Also } A \subset (A \cup B) \Rightarrow P(A) \leq P(A \cup B)$$

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\ &\leq P(A) + P(B) \quad (\because P(A \cap B) \geq 0) \end{aligned}$$

Combining the above results, we get

$$P(A \cap B) \leq P(A) \leq P(A \cup B) \leq P(A) + P(B)$$

**Example 4-12.** Two dice, one green and the other red, are thrown. Let A be the event that the sum of the points on the faces shown is odd, and B be the event of at least one ace (number '1').

(a) Describe the (i) complete sample space, (ii) events  $A$ ,  $B$ ,  $\bar{B}$ ,  $A \cap B$ ,  $A \cup B$ , and  $A \cap \bar{B}$  and find their probabilities assuming that all the 36 sample points have equal probabilities.

(b) Find the probabilities of the events :

(i)  $(\bar{A} \cup \bar{B})$  (ii)  $(\bar{A} \cap \bar{B})$  (iii)  $(A \cap \bar{B})$  (iv)  $(\bar{A} \cap B)$  (v)  $(\bar{A} \cap \bar{B})$  (vi)  $(\bar{A} \cup B)$  (vii)  $(\bar{A} \cup \bar{B})$  (viii)  $\bar{A} \cap (A \cup B)$  (ix)  $A \cup (\bar{A} \cap B)$  (x)  $(A \mid B)$  and  $(B \mid A)$ , and (xi)  $(\bar{A} \mid \bar{B})$  and  $(\bar{B} \mid \bar{A})$  ..

**Solution.** (a) The sample space consists of the 36 elementary events .

- (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)

where, for example, the ordered pair (4, 5) refers to the elementary event that the green die shows 4 and the red die shows 5.

$A$  = The event that the sum of the numbers shown by the two dice is odd.

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

$$P(A) = \frac{n(A)}{n(S)} = \frac{18}{36}$$

$B$  = The event that at least one face is 1,

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

$$P(B) = \frac{n(B)}{n(S)} = \frac{11}{36}$$

$\bar{B}$  = The event that each of the face obtained is not an ace.

$$= \{(2,2); (2,3); (2,4); (2,5); (2,6); (3,2); (3,3); (3,4); (3,5); (3,6); (4,2); (4,3); (4,4); (4,5); (4,6); (5,2); (5,3); (5,4); (5,5); (5,6); (6,2); (6,3); (6,4); (6,5); (6,6)\} \text{ and therefore}$$

$$P(\bar{B}) = \frac{n(\bar{B})}{n(S)} = \frac{25}{36}$$

$A \cap B$  = The event that sum is odd and at least one face is an ace.

$$= \{(1,2); (2,1); (1,4); (4,1); (1,6); (6,1)\}$$

$$\therefore P(A \cap B) = \frac{n(A \cap B)}{n(S)} = \frac{6}{36} = \frac{1}{6}$$

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

$$\therefore P(A \cup B) = \frac{n(A \cup B)}{n(S)} = \frac{23}{36}$$

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

$$P(A \cap \bar{B}) = \frac{n(A \cap \bar{B})}{n(S)} = \frac{12}{36} = \frac{1}{3}$$

$$(b) (i) P(\bar{A} \cup \bar{B}) = P(\bar{A} \cap \bar{B}) = 1 - P(A \cap B) = 1 - \frac{1}{6} = \frac{5}{6}$$

$$(ii) P(\bar{A} \cap \bar{B}) = P(\bar{A} \cup \bar{B}) = 1 - P(A \cap B) = 1 - \frac{23}{36} = \frac{13}{36}$$

$$(iii) P(A \cap \bar{B}) = P(A) - P(A \cap B) = \frac{18}{36} - \frac{6}{36} = \frac{12}{36} = \frac{1}{3}$$

$$(iv) P(\bar{A} \cap B) = P(B) - P(A \cap B) = \frac{11}{36} - \frac{6}{36} = \frac{5}{36}$$

$$(v) P(\bar{A} \cap \bar{B}) = 1 - P(A \cap B) = 1 - \frac{1}{6} = \frac{5}{6}$$

$$(vi) P(\bar{A} \cup B) = P(\bar{A}) + P(B) - P(\bar{A} \cap B) \\ = \left(1 - \frac{18}{36}\right) + \frac{11}{36} - \frac{5}{36} = \frac{2}{3}$$

$$(vii) P(\bar{A} \cup \bar{B}) = 1 - P(A \cap B) = 1 - \frac{23}{36} = \frac{13}{36}$$

$$(viii) P[\bar{A} \cap (A \cup B)] = P[(A \cap \bar{A}) \cup (\bar{A} \cap B)] \\ = P(\bar{A} \cap B) = \frac{5}{36} \quad [\because A \cap \bar{A} = \emptyset]$$

$$(ix) P[A \cup (\bar{A} \cap B)] = P(A) + P(\bar{A} \cap B) - P(A \cap \bar{A} \cap B) \\ = P(A) + P(\bar{A} \cap B) = \frac{18}{36} + \frac{5}{36} = \frac{23}{36}$$

$$(x) P(A | B) = \frac{P(A \cap B)}{P(B)} = \frac{6/36}{11/36} = \frac{6}{11}$$

$$P(B | A) = \frac{P(A \cap B)}{P(A)} = \frac{6/36}{18/36} = \frac{6}{18} = \frac{1}{3}$$

$$(xi) P(\bar{A} | \bar{B}) = \frac{P(\bar{A} \cap \bar{B})}{P(\bar{B})} = \frac{13/36}{25/36} = \frac{13}{25}$$

$$P(\bar{B} | \bar{A}) = \frac{P(\bar{A} \cap \bar{B})}{P(\bar{A})} = \frac{13/36}{18/36} = \frac{13}{18}$$

**Example 4.13.** If two dice are thrown, what is the probability that the sum is (a) greater than 8, and (b) neither 7 nor 11?

**Solution.** (a) If  $S$  denotes the sum on the two dice, then we want  $P(S > 8)$ . The required event can happen in the following mutually exclusive ways:

- (i)  $S = 9$     (ii)  $S = 10$     (iii)  $S = 11$     (iv)  $S = 12$ .

Hence by addition theorem of probability

$$P(S > 8) = P(S = 9) + P(S = 10) + P(S = 11) + P(S = 12)$$

In a throw of two dice, the sample space contains  $6^2 = 36$  points.

The number of favourable cases can be enumerated as follows:

$S = 9$  : (3, 6), (6, 3), (4, 5), (5, 4), i.e., 4 sample points.

$$\therefore P(S = 9) = \frac{4}{36}$$

$S = 10$  : (4, 6), (6, 4), (5, 5), i.e., 3 sample points.

$$\therefore P(S = 10) = \frac{3}{36}$$

$S = 11$  : (5, 6), (6, 5), i.e., 2 sample points.

$$\therefore P(S = 11) = \frac{2}{36}$$

$S = 12$  : (6, 6), i.e., 1 sample point.

$$\therefore P(S = 12) = \frac{1}{36}$$

$$\therefore P(S > 8) = \frac{4}{36} + \frac{3}{36} + \frac{2}{36} + \frac{1}{36} = \frac{10}{36} = \frac{5}{18}$$

(b) Let  $A$  denote the event of getting the sum of 7 and  $B$  denote the event of getting the sum of 11 with a pair of dice.

$S = 7$  : (1, 6), (6, 1), (2, 5), (5, 2), (3, 4), (4, 3), i.e., 6 distinct sample points.

$$\therefore P(A) = P(S = 7) = \frac{6}{36} = \frac{1}{6}$$

$$S = 11 : (5, 6), (6, 5), P(B) = P(S = 11) = \frac{2}{36} = \frac{1}{18}$$

$$\therefore \text{Required probability} = P(\bar{A} \cap \bar{B}) = 1 - P(A \cup B)$$

$$= 1 - [P(A) + P(B)]$$

( $\because A$  and  $B$  are disjoint events)

$$= 1 - \frac{1}{6} - \frac{1}{18} = \frac{7}{9}$$

**Example 4-14.** An urn contains 4 tickets numbered 1, 2, 3, 4 and another contains 6 tickets numbered 2, 4, 6, 7, 8, 9. If one of the two urns is chosen at random and a ticket is drawn at random from the chosen urn, find the probabilities that the ticket drawn bears the number (i) 2 or 4, (ii) 3, (iii) 1 or 9

[Calicut Univ. B.Sc., 1992]

**Solution.** (i) Required event can happen in the following mutually exclusive ways:

(I) First urn is chosen and then a ticket is drawn.

(II) Second urn is chosen and then a ticket is drawn.

Since the probability of choosing any urn is  $\frac{1}{2}$ , the required probability ' $p$ ' is given by

$$p = P(I) + P(II)$$

$$= \frac{1}{2} \times \frac{2}{4} + \frac{1}{2} \times \frac{2}{6} = \frac{5}{12}$$

$$(ii) \text{ Required probability} = \frac{1}{2} \times \frac{1}{4} + \frac{1}{2} \times 0 = \frac{1}{8}$$

( $\because$  in the 2nd urn there is no ticket with number 3)

$$(iii) \text{ Required probability} = \frac{1}{2} \times \frac{1}{4} + \frac{1}{2} \times \frac{1}{6} = \frac{5}{24}$$

**Example 4-15.** A card is drawn from a well-shuffled pack of playing cards. What is the probability that it is either a spade or an ace?

**Solution.** The equiprobable sample space  $S$  of drawing a card from a well-shuffled pack of playing cards consists of 52 sample points.

If  $A$  and  $B$  denote the events of drawing a 'spade card' and 'an ace' respectively then  $A$  consists of 13 sample points and  $B$  consists of 4 sample points so that,

$$P(A) = \frac{13}{52} \text{ and } P(B) = \frac{4}{52}$$

The compound event  $A \cap B$  consists of only one sample point, viz., ace of spade so that,

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

The probability that the card drawn is either a spade or an ace is given by

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\ &= \frac{13}{52} + \frac{4}{52} - \frac{1}{52} = \frac{4}{13} \end{aligned}$$

**Example 4-16.** A box contains 6 red, 4 white and 5 black balls. A person draws 4 balls from the box at random. Find the probability that among the balls drawn there is at least one ball of each colour. (Nagpur Univ. B.Sc., 1992)

**Solution.** The required event  $E$  that 'in a draw of 4 balls from the box at random there is at least one ball of each colour', can materialise in the following mutually disjoint ways :

- (i) 1 Red, 1 White, 2 Black balls
- (ii) 2 Red, 1 White, 1 Black balls
- (iii) 1 Red, 2 White, 1 Black balls.

Hence by the addition theorem of probability, the required probability is given by

$$\begin{aligned} P(E) &= P(i) + P(ii) + P(iii) \\ &= \frac{{}^6C_1 \times {}^4C_1 \times {}^5C_2}{{}^{15}C_4} + \frac{{}^6C_2 \times {}^4C_1 \times {}^5C_1}{{}^{15}C_4} + \frac{{}^6C_1 \times {}^4C_2 \times {}^5C_1}{{}^{15}C_4} \\ &= \frac{1}{{}^{15}C_4} [6 \times 4 \times 10 + 15 \times 4 \times 5 + 6 \times 6 \times 5] \\ &= \frac{4!}{15 \times 14 \times 13 \times 12} [240 + 300 + 180] \\ &= \frac{24 \times 720}{15 \times 14 \times 13 \times 12} = 0.5275 \end{aligned}$$

**Example 4-17.** Why does it pay to bet consistently on seeing 6 at least once in 4 throws of a die, but not on seeing a double six at least once in 24 throws with two dice? (de Mere's Problem).

**Solution.** The probability of getting a '6' in a throw of die =  $1/6$ .

∴ The probability of not getting a '6' in a throw of die

$$= 1 - 1/6 = 5/6.$$

By compound probability theorem, the probability that in 4 throws of a die no '6' is obtained =  $(5/6)^4$

Hence the probability of obtaining '6' at least once in 4 throws of a die =  $1 - (5/6)^4 = 0.516$

Now, if a trial consists of throwing two dice at a time, then the probability of getting a 'double' of '6' in a trial =  $1/36$ .

Thus the probability of not getting a 'double of 6' in a trial =  $35/36$ .

The probability that in 24 throws, with two dice each, no 'double of 6' is obtained =  $(35/36)^{24}$

Hence the probability of getting a 'double of 6' at least once in 24 throws =  $1 - (35/36)^{24} = 0.491$ .

Since the probability in the first case is greater than the probability in the second case, the result follows.

**Example 4-18.** A problem in Statistics is given to the three students A, B and C whose chances of solving it are  $1/2, 3/4$ , and  $1/4$  respectively.

What is the probability that the problem will be solved if all of them try independently? [Madurai Kamraj Univ. B.Sc., 1986; Delhi Univ. B.A., 1991]

**Solution.** Let A, B, C denote the events that the problem is solved by the students A, B, C respectively. Then

$$P(A) = \frac{1}{2}, P(B) = \frac{3}{4} \text{ and } P(C) = \frac{1}{4}$$

The problem will be solved if at least one of them solves the problem. Thus we have to calculate the probability of occurrence of at least one of the three events A, B, C, i.e.,  $P(A \cup B \cup C)$ .

$$\begin{aligned} P(A \cup B \cup C) &= P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) \\ &\quad - P(B \cap C) + P(A \cap B \cap C) \\ &= P(A) + P(B) + P(C) - P(A)P(B) - P(A)P(C) \\ &\quad - P(B)P(C) + P(A)P(B)P(C) \\ &\quad (\because A, B, C \text{ are independent events.}) \\ &= \frac{1}{2} + \frac{3}{4} + \frac{1}{4} - \frac{1}{2} \cdot \frac{3}{4} - \frac{3}{4} \cdot \frac{1}{4} \\ &\quad - \frac{1}{2} \cdot \frac{1}{4} + \frac{1}{2} \cdot \frac{3}{4} \cdot \frac{1}{4} \\ &= \frac{29}{32} \end{aligned}$$

$$\begin{aligned}
 \text{Aliter. } P(A \cup B \cup C) &= 1 - P(\overline{A \cup B \cup C}) \\
 &= 1 - P(\overline{A} \cap \overline{B} \cap \overline{C}) \\
 &= 1 - P(\overline{A})P(\overline{B})P(\overline{C}) \\
 &= 1 - \left(1 - \frac{1}{2}\right)\left(1 - \frac{3}{4}\right)\left(1 - \frac{1}{4}\right) \\
 &= \frac{29}{32}
 \end{aligned}$$

**Example 4.19.** If  $A \cap B = \phi$ , then show that

$$P(A) \leq P(\overline{B})$$

[Delhi Univ. B.Sc. (Maths Hons.) 1987]

**Solution.** We have

$$\begin{aligned}
 A &= (A \cap B) \cup (A \cap \overline{B}) \\
 &= \phi \cup (A \cap \overline{B}) \quad [\text{Using } *] \\
 &= A \cap \overline{B} \\
 \Rightarrow A &\subseteq \overline{B} \\
 \Rightarrow P(A) &\leq P(\overline{B})
 \end{aligned}$$

as desired.

**Aliter.** Since  $A \cap B = \phi$ , we have  $A \subset \overline{B}$ , which implies that  $P(A) \leq P(\overline{B})$ .

**Example 4.20.** Let  $A$  and  $B$  be two events such that

$$P(A) = \frac{3}{4} \text{ and } P(B) = \frac{5}{8}$$

show that

$$\begin{aligned}
 (a) \quad P(A \cup B) &\geq \frac{3}{4} \\
 (b) \quad \frac{3}{8} &\leq P(A \cap B) \leq \frac{5}{8}
 \end{aligned}$$

[Delhi Univ. B.Sc. Stat (Hons.) 1986, 1988]

**Solution.** (i) We have

$$\begin{aligned}
 A &\subset (A \cup B) \\
 \Rightarrow P(A) &\leq P(A \cup B) \\
 \Rightarrow \frac{3}{4} &\leq P(A \cup B) \\
 \Rightarrow P(A \cup B) &\geq \frac{3}{4}
 \end{aligned}$$

$$(ii) \quad A \cap B \subseteq B$$

$$\Rightarrow P(A \cap B) \leq P(B) = \frac{5}{8} \quad ... (i)$$

$$\text{Also } P(A \cup B) = P(A) + P(B) - P(A \cap B) \leq 1$$

$$\Rightarrow \frac{3}{4} + \frac{5}{8} - 1 \leq P(A \cap B)$$

$$\Rightarrow \frac{6+5-8}{8} \leq P(A \cap B)$$

$$\Rightarrow \frac{3}{8} \leq P(A \cap B) \quad \dots(ii)$$

From (i) and (ii) we get

$$\frac{3}{8} \leq P(A \cap B) \leq \frac{5}{8}$$

**Example 4.21.** (Chebychev's Problem). What is the chance that two numbers, chosen at random, will be prime to each other?

**Solution.** If any number 'a' is divided by a prime number 'r', then the possible remainders are 0, 1, 2, ..., r-1. Hence the chance that 'a' is divisible by r is 1/r (because the only case favourable to this is remainder being 0). Similarly, the probability that any number 'b' chosen at random is divisible by r is 1/r. Since the numbers a and b are chosen at random, the probability that none of them is divisible by 'r' is given (by compound probability theorem) by :

$$\left(1 - \frac{1}{r}\right) \times \left(1 - \frac{1}{r}\right) = \left(1 - \frac{1}{r}\right)^2; \quad r = 2, 3, 5, 7, \dots$$

Hence the required probability that the two numbers chosen at random are prime to each other is given by

$$P = \prod_r \left(1 - \frac{1}{r}\right)^2, \quad \text{where } r \text{ is a prime number.}$$

$$= \frac{6}{\pi^2} \quad (\text{From trigonometry})$$

**Example 4.22.** A bag contains 10 gold and 8 silver coins. Two successive drawings of 4 coins are made such that : (i) coins are replaced before the second trial, (ii) the coins are not replaced before the second trial. Find the probability that the first drawing will give 4 gold and the second 4 silver coins.

[Allahabad Univ. B.Sc., 1987]

**Solution.** Let A denote the event of drawing 4 gold coins in the first draw and B denote the event of drawing 4 silver coins in the second draw. Then we have to find the probability of  $P(A \cap B)$ .

(i) *Draws with replacement.* If the coins drawn in the first draw are replaced back in the bag before the second draw then the events A and B are independent and the required probability is given (using the multiplication rule of probability) by the expression

$$P(A \cap B) = P(A) \cdot P(B) \quad \dots(*)$$

*1st draw.* Four coins can be drawn out of 10+8=18 coins in  ${}^{18}C_4$  ways, which gives the exhaustive number of cases. In order that all these coins are of gold, they must be drawn out of the 10 gold coins and this can be done in  ${}^{10}C_4$  ways. Hence

$$P(A) = {}^{10}C_4 / {}^{18}C_4$$

*2nd draw.* When the coins drawn in the first draw are replaced before the 2nd draw, the bag contains 18 coins. The probability of drawing 4 silver coins in the 2nd draw is given by  $P(B) = {}^8C_4 / {}^{18}C_4$ .

Substituting in (\*), we have

$$P(A \cap B) = \frac{{}^{10}C_4}{{}^{18}C_4} \times \frac{{}^8C_4}{{}^{18}C_4}$$

*(ii) Draws without replacement.* If the coins drawn are not replaced back before the second draw, then the events A and B are not independent and the required probability is given by .

$$P(A \cap B) = P(A) \cdot P(B | A) \quad \dots (**)$$

$$\text{As discussed in part (i), } P(A) = {}^{10}C_4 / {}^{18}C_4.$$

Now, if the 4 gold coins which were drawn in the first draw are not replaced back, there are  $18 - 4 = 14$  coins left in the bag and  $P(B | A)$  is the probability of drawing 4 silver coins from the bag containing 14 coins out of which 6 are gold coins and 8 are silver coins.

$$\text{Hence } P(B | A) = {}^8C_4 / {}^{14}C_4$$

Substituting in (\*\*) we get

$$P(A \cap B) = \frac{{}^{10}C_4}{{}^{18}C_4} \times \frac{{}^8C_4}{{}^{14}C_4}$$

**Example 4.23.** A consignment of 15 record players contains 4 defectives. The record players are selected at random, one by one, and examined. Those examined are not put back. What is the probability that the 9th one examined is the last defective?

**Solution.** Let A be the event of getting exactly 3 defectives in examination of 8 record players and let B the event that the 9th piece examined is a defective one.

Since it is a problem of sampling without replacement and since there are 4 defectives out of 15 record players, we have

$$P(A) = \frac{\binom{4}{3} \times \binom{11}{5}}{\binom{15}{8}}$$

$P(B | A)$  = Probability that the 9th examined record player is defective given that there were 3 defectives in the first 8 pieces examined.

$$= 1/7,$$

since there is only one defective piece left among the remaining  $15 - 8 = 7$  record players.

Hence the required probability is

$$P(A \cap B) = P(A) \cdot P(B | A)$$

$$= \frac{\binom{4}{3} \times \binom{11}{5}}{\binom{15}{8}} \times \frac{1}{7} = \frac{8}{195}$$

**Example 4-24.**  $p$  is the probability that a man aged  $x$  years will die in a year. Find the probability that out of  $n$  men  $A_1, A_2, \dots, A_n$  each aged  $x$ ,  $A_1$  will die in a year and will be the first to die. [Delhi Univ. B.Sc., 1985]

**Solution.** Let  $E_i$ , ( $i = 1, 2, \dots, n$ ) denote the event that  $A_i$  dies in a year. Then

$$P(E_i) = p, (i = 1, 2, \dots, n) \text{ and } P(\bar{E}_i) = 1 - p.$$

The probability that none of  $n$  men  $A_1, A_2, \dots, A_n$  dies in a year

$$= P(\bar{E}_1 \cap \bar{E}_2 \cap \dots \cap \bar{E}_n) = P(\bar{E}_1) P(\bar{E}_2) \dots P(\bar{E}_n)$$

(By compound probability theorem)

$$= (1 - p)^n$$

∴ The probability that at least one of  $A_1, A_2, \dots, A_n$ , dies in a year

$$= 1 - P(\bar{E}_1 \cap \bar{E}_2 \cap \dots \cap \bar{E}_n) = 1 - (1 - p)^n$$

The probability that among  $n$  men,  $A_1$  is the first to die is  $1/n$  and since this event is independent of the event that at least one man dies in a year, required probability is

$$\frac{1}{n} [1 - (1 - p)^n]$$

**Example 4-25.** The odds against Manager X settling the wage dispute with the workers are 8:6 and odds in favour of manager Y settling the same dispute are 14:16.

(i) What is the chance that neither settles the dispute, if they both try, independently of each other?

(ii) What is the probability that the dispute will be settled?

**Solution.** Let  $A$  be the event that the manager X will settle the dispute and  $B$  be the event that the Manager Y will settle the dispute. Then clearly

$$P(\bar{A}) = \frac{8}{8+6} = \frac{4}{7} \Rightarrow P(A) = 1 - P(\bar{A}) = \frac{6}{14} = \frac{3}{7}$$

$$P(\bar{B}) = \frac{14}{14+16} = \frac{7}{15} \Rightarrow P(B) = 1 - P(\bar{B}) = \frac{16}{14+16} = \frac{8}{15}$$

The required probability that neither settles the dispute is given by :

$$P(\bar{A} \cap \bar{B}) = P(\bar{A}) \times P(\bar{B}) = \frac{4}{7} \times \frac{8}{15} = \frac{32}{105}$$

[Since  $A$  and  $B$  are independent  $\Rightarrow \bar{A}$  and  $\bar{B}$  are also independent]

(ii) The dispute will be settled if at least one of the managers X and Y settles the dispute. Hence the required probability is given by:

$$P(A \cup B) = \text{Prob. [At least one of X and Y settles the dispute]}$$

$$= 1 - \text{Prob. [None settles the dispute]}$$

$$= 1 - P(\bar{A} \cap \bar{B}) = 1 - \frac{32}{105} = \frac{73}{105}$$

**Example 4-26.** The odds that person X speaks the truth are 3:2 and the odds that person Y speaks the truth are 5:3. In what percentage of cases are they likely to contradict each other on an identical point.

**Solution.** Let us define the events:

$$A : X \text{ speaks the truth}, \quad B : Y \text{ speaks the truth}$$

Then  $\bar{A}$  and  $\bar{B}$  represent the complementary events that X and Y tell a lie respectively. We are given:

$$P(A) = \frac{3}{3+2} = \frac{3}{5} \Rightarrow P(\bar{A}) = 1 - \frac{3}{5} = \frac{2}{5}$$

$$\text{and } P(B) = \frac{5}{5+3} = \frac{5}{8} \Rightarrow P(\bar{B}) = 1 - \frac{5}{8} = \frac{3}{8}$$

The event  $E$  that X and Y contradict each other on an identical point can happen in the following mutually exclusive ways:

(i) X speaks the truth and Y tells a lie, i.e., the event  $A \cap \bar{B}$  happens,

(ii) X tells a lie and Y speaks the truth, i.e., the event  $\bar{A} \cap B$  happens.

Hence by addition theorem of probability the required probability is given by:

$$P(E) = P(i) + P(ii) = P(A \cap \bar{B}) + P(\bar{A} \cap B)$$

$$= P(A) \cdot P(\bar{B}) + P(\bar{A}) \cdot P(B),$$

[Since A and B are independent]

$$= \frac{3}{5} \times \frac{3}{8} + \frac{2}{5} \times \frac{5}{8} = \frac{19}{40} = 0.475$$

Hence A and B are likely to contradict each other on an identical point in 47.5% of the cases.

**Example 4-27.** A special dice is prepared such that the probabilities of throwing 1, 2, 3, 4, 5 and 6 points are :

$$\frac{1-k}{6}, \frac{1+2k}{6}, \frac{1-k}{6}; \quad \frac{1+k}{6}, \frac{1-2k}{6}, \text{ and } \frac{1+k}{6}$$

respectively. If two such dice are thrown, find the probability of getting a sum equal to 9. [Delhi Univ. B.Sc. (Stat. Hons.), 1988]

**Solution.** Let  $(x, y)$  denote the numbers obtained in a throw of two dice,  $x$  denoting the number on the first dice and  $y$  denoting the number on the second dice. The sum  $S = x+y = 9$ , can be obtained in the following mutually disjoint ways:

(i) (3, 6), (ii) (6, 3), (iii) (4, 5), (iv) (5, 4)

Hence by addition theorem of probability:

$$P(S=9) = P(3, 6) + P(6, 3) + P(4, 5) + P(5, 4)$$

$$= P(x=3)P(y=6) + P(x=6)P(y=3) + P(x=4)P(y=5) \\ + P(x=5)P(y=4),$$

since the number on one dice is independent of the number on the other dice.

$$\begin{aligned} \therefore P(S=9) &= \frac{(1-k)}{6} \cdot \frac{(1+k)}{6} + \frac{(1+k)}{6} \cdot \frac{(1-k)}{6} + \frac{(1+k)}{6} \cdot \frac{(1-2k)}{6} \\ &\quad + \frac{(1-2k)}{6} \cdot \frac{(1+k)}{6} \\ &= 2 \left( \frac{1+k}{36} \right) |(1-k) + (1-2k)| \\ &= \frac{1}{18} (1+k)(2-3k) \end{aligned}$$

**Example 4.28.** (a) A and B alternately cut a pack of cards and the pack is shuffled after each cut. If A starts and the game is continued until one cuts a diamond, what are the respective chances of A and B first cutting a diamond?

(b) One shot is fired from each of the three guns.  $E_1, E_2, E_3$  denote the events that the target is hit by the first, second and third gun respectively. If  $P(E_1) = 0.5$ ,  $P(E_2) = 0.6$  and  $P(E_3) = 0.8$  and  $E_1, E_2, E_3$  are independent events, find the probability that (a) exactly one hit is registered, (b) at least two hits are registered.

**Solution.** (a) Let  $E_1$  and  $E_2$  denote the events of A and B cutting a diamond respectively. Then

$$P(E_1) = P(E_2) = \frac{13}{52} = \frac{1}{4} \Rightarrow P(\bar{E}_1) = P(\bar{E}_2) = \frac{3}{4}$$

If A starts the game, he can first cut the diamond in the following mutually exclusive ways:

(i)  $E_1$  happens, (ii)  $\bar{E}_1 \cap E_2 \cap E_1$  happens, (iii)  $\bar{E}_1 \cap \bar{E}_2 \cap \bar{E}_1 \cap \bar{E}_2 \cap E_1$  happens, and so on. Hence by addition theorem of probability, the probability 'p' that A first wins is given by

$$\begin{aligned} p &= P(i) + P(ii) + P(iii) + \dots \dots \\ &= P(E_1) + P(\bar{E}_1 \cap E_2 \cap E_1) + P(\bar{E}_1 \cap \bar{E}_2 \cap \bar{E}_1 \cap \bar{E}_2 \cap E_1) + \dots \\ &= P(E_1) + P(\bar{E}_1) P(E_2) P(E_1) + P(\bar{E}_1) P(\bar{E}_2) P(\bar{E}_1) P(\bar{E}_2) P(E_1) + \dots \\ &\quad \text{(By Compound Probability Theorem)} \\ &= \frac{1}{4} + \frac{3}{4} \cdot \frac{3}{4} \cdot \frac{1}{4} + \frac{3}{4} \cdot \frac{3}{4} \cdot \frac{3}{4} \cdot \frac{1}{4} + \dots \dots \\ &= \frac{\frac{1}{4}}{1 - \frac{9}{16}} = \frac{4}{7} \end{aligned}$$

The probability that B first cuts a diamond

$$= 1 - p = 1 - \frac{4}{7} = \frac{3}{7}$$

(b) We are given

$$P(\bar{E}_1) = 0.5, P(\bar{E}_2) = 0.4 \text{ and } P(\bar{E}_3) = 0.2$$

(a) Exactly one hit can be registered in the following mutually exclusive ways:

(i)  $E_1 \cap \overline{E_2} \cap \overline{E_3}$  happens, (ii)  $\overline{E_1} \cap E_2 \cap \overline{E_3}$  happens, (iii)  $\overline{E_1} \cap \overline{E_2} \cap E_3$  happens.

Hence by addition probability theorem, the required probability 'p' is given by :

$$P = P(E_1 \cap \overline{E_2} \cap \overline{E_3}) + P(\overline{E_1} \cap E_2 \cap \overline{E_3}) + P(\overline{E_1} \cap \overline{E_2} \cap E_3)$$

$$= P(E_1)P(\overline{E_2})P(\overline{E_3}) + P(\overline{E_1})P(E_2)P(\overline{E_3}) + P(\overline{E_1})P(\overline{E_2})P(E_3)$$

(Since  $E_1, E_2$  and  $E_3$  are independent)

$$= 0.5 \times 0.4 \times 0.2 + 0.5 \times 0.6 \times 0.2 + 0.5 \times 0.4 \times 0.8 = 0.26.$$

(b) At least two hits can be registered in the following mutually exclusive ways:

(i)  $E_1 \cap E_2 \cap \overline{E_3}$  happens (ii)  $E_1 \cap \overline{E_2} \cap E_3$  happens, (iii)  $\overline{E_1} \cap E_2 \cap E_3$  happens. (iv)  $E_1 \cap E_2 \cap E_3$  happens.

Required probability

$$= P(E_1 \cap E_2 \cap \overline{E_3}) + P(E_1 \cap \overline{E_2} \cap E_3) + P(\overline{E_1} \cap E_2 \cap E_3) + P(E_1 \cap E_2 \cap E_3)$$

$$= 0.5 \times 0.6 \times 0.2 + 0.5 \times 0.4 \times 0.8 + 0.5 \times 0.6 \times 0.8 + 0.5 \times 0.6 \times 0.8$$

$$= 0.06 + 0.16 + 0.24 + 0.24 = 0.70$$

**Example 4.29.** Three groups of children contain respectively 3 girls and 1 boy, 2 girls and 2 boys, and 1 girl and 3 boys. One child is selected at random from each group. Show that the chance that the three selected consist of 1 girl and 2 boys is  $13/32$ . [Madurai Univ. B.Sc., 1988; Nagpur Univ. B.Sc., 1991]

**Solution.** The required event of getting 1 girl and 2 boys among the three selected children can materialise in the following three mutually disjoint cases:

Group No. →	I	II	III
(i)	Girl	Boy	Boy
(ii)	Boy	Girl	Boy
(iii)	Boy	Boy	Girl

Hence by addition theorem of probability,

$$\text{Required probability} = P(i) + P(ii) + P(iii) \quad \dots (*)$$

Since the probability of selecting a girl from the first group is  $3/4$ , of selecting a boy from the second is  $2/4$ , and of selecting a boy from the third group is  $3/4$ , and since these three events of selecting children from three groups are independent of each other, by compound probability theorem, we have

$$P(i) = \frac{3}{4} \times \frac{2}{4} \times \frac{3}{4} = \frac{9}{32}$$

Similarly, we have

$$P(ii) = \frac{1}{4} \times \frac{2}{4} \times \frac{3}{4} = \frac{3}{32}$$

$$\text{and } P(iii) = \frac{1}{4} \times \frac{2}{4} \times \frac{1}{4} = \frac{1}{32}$$

Substituting in (\*), we get

$$\text{Required probability} = \frac{9}{32} + \frac{3}{32} + \frac{1}{32} = \frac{13}{32}$$

### EXERCISE 4 (b)

1. (a) Which function defines a probability space on  $S = (e_1, e_2, e_3)$

$$(i) P(e_1) = \frac{1}{4}, P(e_2) = \frac{1}{3}, P(e_3) = \frac{1}{2}$$

$$(ii) P(e_1) = \frac{2}{3}, P(e_2) = -\frac{1}{3}, P(e_3) = \frac{2}{3}$$

$$(iii) P(e_1) = \frac{1}{4}, P(e_2) = \frac{1}{3}, P(e_3) = \frac{1}{2}, \text{ and}$$

$$(iv) P(e_1) = 0, P(e_2) = \frac{1}{3}, P(e_3) = \frac{2}{3}$$

**Ans.** (i) No, (ii) No, (iii) No, and (iv) Yes

(b) Let  $S = (e_1, e_2, e_3, e_4)$ , and let  $P$  be a probability function on  $S$ .

$$(i) \text{ Find } P(e_1), \text{ if } P(e_2) = \frac{1}{3}, P(e_3) = \frac{1}{6}, P(e_4) = \frac{1}{9},$$

$$(ii) \text{ Find } P(e_1) \text{ and } P(e_2) \text{ if } P(e_3) = P(e_4) = \frac{1}{4} \text{ and } P(e_1) = 2P(e_2), \text{ and}$$

$$(iii) \text{ Find } P(e_1) \text{ if } P\{(e_2, e_3)\} = \frac{2}{3}, P\{(e_2, e_4)\} = \frac{1}{2} \text{ and } P(e_2) = \frac{1}{3}.$$

$$\text{Ans. (i)} P(e_1) = \frac{7}{18}, \text{ (ii)} P(e_1) = \frac{1}{3}, P(e_2) = \frac{1}{6}, \text{ and (iii)} P(e_1) = \frac{1}{6}$$

2. (a) With usual notations, prove that

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

Deduce a similar result for  $P(A \cup B \cup C)$ , where  $C$  is one more event.

(b) For any event :  $E_i, P(E_i) = p_i, (i = 1, 2, 3); P(E_i \cap E_j) = p_{ij}, (i, j = 1, 2, 3)$  and  $P(E_1 \cap E_2 \cap E_3) = p_{123}$ , find the probability that of the three events, (i) at least one, and (ii) exactly one happens.

(c) Discuss briefly the axiomatic approach to probability, illustrating by examples how it meets the deficiencies of the classical approach.

(d) If  $A$  and  $B$  are any two events, state the results giving

$$(i) P(A \cup B) \text{ and (ii)} P(A \cap B).$$

$A$  and  $B$  are mutually exclusive events and  $P(A) = \frac{1}{2}, P(B) = \frac{1}{3}$ . Find  $P(A \cup B)$  and  $P(A \cap B)$ .

3. Let  $S = \left\{1, \frac{1}{2}, \left(\frac{1}{2}\right)^2, \dots, \left(\frac{1}{2}\right)^n\right\}$ , be a classical event space and  $A, B$  be events given by

$$A = \left\{ 1, \frac{1}{2} \right\}, \quad B = \left\{ \left( \frac{1}{2} \right)^k \mid k \text{ is an even positive integer} \right\}$$

Find  $P(\bar{A} \cap \bar{B})$

[Calcutta Univ. B.Sc. (Stat Hons.), 1986]

4. What is a 'probability space'? State (i) the 'law of total probability' and (ii) Boole's inequality for events not necessarily mutually exclusive.

5. (a) Explain the following with examples:

- (i) random experiment, (ii) an event, (iii) an event space. State the axioms of probability and explain their frequency interpretations.

A man forgets the last digit of a telephone number, and dials the last digit at random. What is the probability of calling no more than three wrong numbers?

(b) Define conditional probability and give its frequency interpretation. Show that conditional probabilities satisfy the axioms of probability.

6. Prove the following laws, in each case assuming the conditional probabilities being defined.

$$(a) P(E|E) = 1, \quad (b) P(\phi|F) = 0$$

$$(c) \text{ If } E_1 \subseteq E_2, \text{ then } P(E_1|F) < P(E_2|F)$$

$$(d) P(\bar{E}|F) = 1 - P(E|F)$$

$$(e) P(E_1 \cup E_2|F) = P(E_1|F) + P(E_2|F) - P(P(E_1 \cap E_2|F))$$

$$(f) \text{ If } P(F) = 1 \text{ then } P(E|F) = P(E)$$

$$(g) P(E - F) = P(E) - P(E \cap F)$$

$$(h) \text{ If } P(F) > 0, \text{ and } E \text{ and } F \text{ are mutually exclusive then } P(E|F) = 0$$

$$(i) \text{ If } P(E|F) = P(E), \text{ then } P(E|\bar{F}) = P(E) \text{ and } P(\bar{E}|F) = P(\bar{E})$$

7. (a) If  $P(\bar{A}) = a$ ,  $P(\bar{B}) = b$ , then prove that  $P(A \cap B) \geq 1 - a - b$ .

(b) If  $P(A) = \alpha$ ,  $P(B) = \beta$ , then prove that  $P(A|B) \geq (\alpha + \beta - 1)/\beta$ .

**Hint.** In each case use  $P(A \cup B) \leq 1$

8. Prove or disprove:

(a) (i) If  $P(A|B) \geq P(A)$ , then  $P(B|A) \geq P(B)$

(ii) If  $P(A) = P(\bar{B})$ , then  $A = \bar{B}$ .

[Delhi Univ. B.Sc. (Maths Hons.), 1988]

(b) If  $P(A) = 0$ , then  $A = \phi$

[Delhi Univ. B.Sc. (Maths Hons.), 1990]

**Ans. Wrong.**

(c) For possible events  $A, B, C$ ,

(i) If  $P(A) > P(B)$ , then  $P(A|C) > P(B|C)$

(ii) If  $P(A|C) \geq P(B|C)$  and  $P(A|\bar{C}) \geq P(B|\bar{C})$ ,  
then  $P(A) \geq P(B)$ . [Delhi Univ. B.Sc. (Maths Hons.), 1989]

(d) If  $P(A) = 0$ , then  $P(A \cap B) = 0$ .

[Delhi Univ. B.Sc. (Maths Hons.), 1986]

(e) (i) If  $P(A) = P(B) = p$ , then  $P(A \cap B) \leq p^2$

(ii) If  $P(B|\bar{A}) = P(B|A)$ , then  $A$  and  $B$  are independent.

[Delhi Univ. B.Sc. (Maths Hons.), 1990]

(f) If  $P(A) > 0$ ,  $P(B) > 0$  and  $P(A|B) = P(B|A)$ ,  
then  $P(A) = P(B)$ .

9. (a) Let  $A$  and  $B$  be two events, neither of which has probability zero. Then if  $A$  and  $B$  are disjoint,  $A$  and  $B$  are independent.

[Delhi Univ. B.Sc.(Stat. Hons.), 1986]

(b) Under what conditions does the following equality hold?

$$P(A) = P(A|B) + P(A|\bar{B})$$

[Punjab Univ. B.Sc. (Maths Hons.), 1992]

Ans.  $B = S$  or  $\bar{B} = S$

10. (a) If  $A$  and  $B$  are two events and the probability  $P(B) \neq 1$ , prove that

$$P(A|\bar{B}) = \frac{[P(A) - P(A \cap B)]}{[1 - P(B)]}$$

where  $\bar{B}$  denotes the event complementary to  $B$  and hence deduce that

$$P(A \cap B) \geq P(A) + P(B) - 1$$

[Delhi Univ. B.Sc. (Stat. Hons.), 1989]

Also show that  $P(A) >$  or  $< P(A|\bar{B})$  according as

$$P(A|\bar{B}) > \text{or} < P(A).$$

[Sri Venkat. Univ. B.Sc. 1992 ; Karnataka Univ. B.Sc. 1991]

**Hint.** (i)

$$P(A|\bar{B}) = \frac{P(A \cap \bar{B})}{P(\bar{B})} = \frac{[P(A) - P(A \cap B)]}{[1 - P(B)]}$$

(ii) Since  $P(A|\bar{B}) \leq 1$ ,  $P(A) - P(A \cap B) \leq 1 - P(B)$

$$\Rightarrow P(A) + P(B) - 1 \leq P(A \cap B)$$

$$(iii) \quad \frac{P(A|\bar{B})}{P(A)} = \frac{P(\bar{B}|A)}{P(\bar{B})} = \frac{1 - P(B|A)}{1 - P(B)}$$

Now  $P(A|\bar{B}) > P(A)$  if  $\{1 - P(B|A)\} > \{1 - P(B)\}$

i.e., if  $P(B|A) < P(B)$

$$\text{i.e., if } \frac{P(B|A)}{P(B)} < 1$$

$$\text{i.e., if } \frac{P(A|B)}{P(A)} < 1 \text{ i.e., if } P(A) > P(A|B)$$

(b) If  $A$  and  $B$  are two mutually exclusive events show that

$$P(A|\bar{B}) = P(A)/[1 - P(B)]$$

[Delhi Univ. B.Sc. (Stat. Hons.), 1987]

(c) If  $A$  and  $B$  are two mutually exclusive events and  $P(A \cup B) \neq 0$ , then

$$P(A|A \cup B) = \frac{P(A)}{P(A) + P(B)} \quad [\text{Guahati Univ. B.Sc. 1991}]$$

(d) If  $A$  and  $B$  are two independent events show that

$$P(A \cup B) = 1 - P(\bar{A})P(\bar{B})$$

(e) If  $\bar{A}$  denotes the non-occurrence of  $A$ , then prove that

$$P(\bar{A}_1 \cup \bar{A}_2 \cup \bar{A}_3) = 1 - P(A_1) P(A_2 | A_1) P(A_3 | A_1 \cap A_2)$$

[Agra Univ. B.Sc., 1987]

11. If  $A$ ,  $B$  and  $C$  are three arbitrary events and

$$S_1 = P(A) + P(B) + P(C)$$

$$S_2 = P(A \cap B) + P(B \cap C) + P(C \cap A)$$

$$S_3 = P(A \cap B \cap C).$$

Prove that the probability that exactly one of the three events occurs is given by  $S_1 - 2S_2 + 3S_3$ .

12. (a) For the events  $A_1, A_2, \dots, A_n$  assuming

$$P\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n P(A_i), \text{ prove that}$$

$$(i) P\left(\bigcap_{i=1}^n A_i\right) \geq 1 - \sum_{i=1}^n P(\bar{A}_i) \text{ and that}$$

$$(ii) P\left(\bigcap_{i=1}^n A_i\right) \geq \sum_{i=1}^n P(A_i) - (n-1)$$

[Sardar Patel Univ. B.Sc. Nov.1992]

(b) Let  $A$ ,  $B$  and  $C$  denote events. If  $P(A | C) \geq P(B | C)$  and

$$P(A | \bar{C}) \geq P(B | \bar{C}), \text{ then show that } P(A) \geq P(B).$$

[Calcutta Univ. B.Sc. (Maths Hons.), 1992]

13. (a) If  $A$  and  $B$  are independent events defined on a given probability space  $(\Omega, \mathcal{A}, P(\cdot))$ , then prove that  $A$  and  $\bar{B}$  are independent,  $\bar{A}$  and  $B$  are independent.

[Delhi Univ. B.Sc. (Maths Hons.), 1988]

(b)  $A$ ,  $B$  and  $C$  are three events such that  $A$  and  $B$  are independent,  $P(C) = 0$ . Show that  $A$ ,  $B$  and  $C$  are independent.

(c) An event  $A$  is known to be independent of the events  $B, B \cup C$  and  $B \cap C$ . Show that it is also independent of  $C$ .

[Nagpur Univ. B.Sc. 1992]

(d) Show that if an event  $C$  is independent of two mutually exclusive events  $A$  and  $B$ , then  $C$  is also independent of  $A \cup B$ .

(e) The outcome of an experiment is equally likely to be one of the four points in three-dimensional space with rectangular coordinates  $(1, 0, 0)$ ,  $(0, 1, 0)$ ,  $(0, 0, 1)$  and  $(1, 1, 1)$ . Let  $E$ ,  $F$  and  $G$  be the events :  $x$ -coordinate = 1,  $y$ -coordinate = 1 and  $z$ -coordinate = 1, respectively. Check if the events  $E$ ,  $F$  and  $G$  are independent.

[Calcutta Univ. B.Sc., 1988]

14. Explain what is meant by "Probability Space". You fire at a target with each of the three guns;  $A$ ,  $B$  and  $C$  denote respectively the event — hit the target with the first, second and third gun. Assuming that the events are independent and have probabilities  $P(A) = a$ ,  $P(B) = b$  and  $P(C) = c$ , express in terms of  $A$ ,  $B$  and  $C$  the following events:

(i) You will not hit the target at all.

(ii) You will hit the target at least twice. Find also the probabilities of these events. [Sardar Patel Univ. B.Sc., 1990]

15. (a) Suppose  $A$  and  $B$  are any two events and that  $P(A) = p_1$ ,  $P(B) = p_2$  and  $P(A \cap B) = p_3$ . Show that the formulae of each of the following probabilities in terms of  $p_1$ ,  $p_2$  and  $p_3$  can be expressed as follows :

$$(i) P(\overline{A} \cup \overline{B}) = 1 - p_2 \quad (ii) P(\overline{A} \cap \overline{B}) = 1 - p_1 - p_2 + p_3$$

$$(iii) P(A \cap \overline{B}) = p_1 - p_3 \quad (iv) P(\overline{A} \cap B) = p_2 - p_3$$

$$(v) P(\overline{A} \cap \overline{B}) = 1 - p_3 \quad (vi) P(\overline{A} \cup B) = 1 - p_1 + p_3$$

$$(vii) P(\overline{A} \cup \overline{B}) = 1 - p_1 - p_2 + p_3 \quad (viii) P[\overline{A} \cap (A \cup B)] = p_2 - p_3$$

$$(ix) P[A \cup (\overline{A} \cap B)] = p_1 + p_2 - p_3$$

$$(x) P(A|B) = \frac{p_3}{p_2} \text{ and } P(B|A) = \frac{p_3}{p_1}$$

$$(xi) P(\overline{A} \mid \overline{B}) = \frac{1 - p_1 - p_2 + p_3}{1 - p_2} \text{ and } P(\overline{B} \mid \overline{A}) = \frac{1 - p_1 - p_2 + p_3}{1 - p_1}$$

[Allahabad Univ. B.Sc. (Stat.), 1991]

(b) If  $P(A) = 1/3$ ,  $P(B) = 3/4$  and  $P(A \cup B) = 11/12$ , find  $P(A|B)$  and  $P(B|A)$ .

(c) Let  $P(A) = p$ ,  $P(A|B) = q$ ,  $P(B|A) = r$ . Find the relation between the numbers  $p$ ,  $q$  and  $r$  such that  $\overline{A}$  and  $\overline{B}$  are mutually exclusive.

[Delhi Univ. B.Sc. (Maths Hons.), 1985]

**Hint.**  $P(AB) = P(A)P(B|A) = P(B)P(A|B)$

$$\Rightarrow P(AB) = pr = P(B).q \Rightarrow P(B) = pr/q$$

If  $\overline{A}$  and  $\overline{B}$  are mutually disjoint, then  $P(\overline{A} \cap \overline{B}) = 0$ .

$$\Rightarrow 1 - P(A \cup B) = 0 \Rightarrow 1 - [p + (pr/q) - pr] = 0$$

16. (a) In terms of probabilities,  $p_1 = P(A)$ ,  $p_2 = P(B)$  and  $p_3 = P(A \cap B)$ ;

Express (i)  $P(A \cup B)$ , (ii)  $P(A|B)$ , (iii)  $P(\overline{A} \cap B)$  under the condition that (i)  $A$  and  $B$  are mutually exclusive, (ii)  $A$  and  $B$  are mutually independent.

(b) Let  $A$  and  $B$  be the possible outcomes of an experiment and suppose

$$P(A) = 0.4, P(A \cup B) = 0.7 \text{ and } P(B) = p$$

(i) For what choice of  $p$  are  $A$  and  $B$  mutually exclusive?

(ii) For what choice of  $p$  are  $A$  and  $B$  independent?

[Aligarh Univ. B.Sc., 1988 ; Guwahati Univ. B.Sc., 1991]

**Ans.** (i) 0.3, (ii) 0.5

(c) Let  $A_1, A_2, A_3, A_4$  be four independent events for which  $P(A_1) = p$ ,  $P(A_2) = q$ ,  $P(A_3) = r$  and  $P(A_4) = s$ . Find the probability that

(i) at least one of the events occurs, (ii) exactly two of the events occur, and (iii) at most three of the events occur. [Civil Services (Main), 1985]

17. (a) Two six-faced unbiased dice are thrown. Find the probability that the sum of the numbers shown is 7 or their product is 12.

**Ans.** 2/9

(b) Defects are classified as  $A$ ,  $B$  or  $C$ , and the following probabilities have been determined from available production data :

$P(A) = 0.20$ ,  $P(B) = 0.16$ ,  $P(C) = 0.14$ ,  $P(A \cap B) = 0.08$ ,  $P(A \cap C) = 0.05$ ,  
 $P(B \cap C) = 0.04$ , and  $P(A \cap B \cap C) = 0.02$ .

What is the probability that a randomly selected item of product will exhibit at least one type of defect ? What is the probability that it exhibits both  $A$  and  $B$  defects but is free from type  $C$  defect ? [Bombay Univ. B.Sc., 1991]

(c) A language class has only three students  $A$ ,  $B$ ,  $C$  and they independently attend the class. The probabilities of attendance of  $A$ ,  $B$  and  $C$  on any given day are  $1/2$ ,  $2/3$  and  $3/4$  respectively. Find the probability that the total number of attendances in two consecutive days is exactly three.

[Lucknow Univ. B.Sc. 1990; Calcutta Univ. B.Sc.(Maths Hons.), 1986]

18. (a) Cards are drawn one by one from a full deck. What is the probability that exactly 10 cards will precede the first ace. [Delhi Univ. B.Sc., 1988]

$$\text{Ans. } \left( \frac{48}{52} \times \frac{47}{51} \times \frac{46}{50} \times \dots \times \frac{39}{43} \right) \times \frac{4}{42} = \frac{164}{4165}$$

(b) Each of two persons tosses three fair coins. What is the probability that they obtain the same number of heads.

$$\text{Ans. } \left( \frac{1}{8} \right)^2 + \left( \frac{3}{8} \right)^2 + \left( \frac{3}{8} \right)^2 + \left( \frac{1}{8} \right)^2 = \frac{5}{16}.$$

19. (a) Given that  $A$ ,  $B$  and  $C$  are mutually exclusive events, explain why each of the following is not a permissible assignment of probabilities.

$$(i) P(A) = 0.24, \quad P(B) = 0.4 \quad \text{and} \quad P(A \cup C) = 0.2,$$

$$(ii) P(A) = 0.7, \quad P(B) = 0.1 \quad \text{and} \quad P(B \cap C) = 0.3$$

$$(iii) P(A) = 0.6, \quad P(A \cap B) = 0.5$$

(b) Prove that for  $n$  arbitrary independent events  $A_1, A_2, \dots, A_n$ ,

$$P(A_1 \cup A_2 \cup A_3 \cup \dots \cup A_n) + P(\bar{A}_1) P(\bar{A}_2) \dots P(\bar{A}_n) = 1.$$

(c)  $A_1, A_2, \dots, A_n$  are  $n$  independent events with

$$P(A_i) = 1 - \frac{1}{\alpha^i}, \quad i = 1, 2, \dots, n.$$

Find the value of  $P(A_1 \cup A_2 \cup A_3 \cup \dots \cup A_n)$ . (Nagpur Univ. B.Sc., 1987)

$$\text{Ans. } 1 - \frac{1}{\alpha^{n(n+1)/2}}$$

(d) Suppose the events  $A_1, A_2, \dots, A_n$  are independent and that

$P(A_i) = \frac{1}{i+1}$  for  $1 \leq i \leq n$ . Find the probability that none of the  $n$  events occurs, justifying each step in your calculations.

$$\text{Ans. } 1/(n+1)$$

20. (a)  $A$  denotes getting a heart card,  $B$  denotes getting a face card (King, Queen or Jack),  $\bar{A}$  and  $\bar{B}$  denote the complementary events. A card is drawn at

random from a full deck. Compute the following probabilities.

- (i)  $P(A)$ , (ii)  $P(A \cap \bar{B})$ , (iii)  $P(A \cup B)$ , (iv)  $P(A \cap B)$ ,
- (v)  $P(\bar{A} \cup B)$ .

Assume natural assignment of probabilities.

**Ans.** (i)  $1/4$ , (ii)  $5/26$ , (iii)  $11/26$ , (iv)  $3/5$ , (v)  $21/26$ .

(b) A town has two doctors  $X$  and  $Y$  operating independently. If the probability that doctor  $X$  is available is  $0.9$  and that for  $Y$  is  $0.8$ , what is the probability that at least one doctor is available when needed? [Gorakhpur Univ. B.Sc., 1988]

**Ans.**  $0.98$

21. (a) The odds that a book will be favourably reviewed by 3 independent critics are 5 to 2, 4 to 3 and 3 to 4 respectively. What is the probability that, of the three reviews, a majority will be favourable? [Gauhati Univ. B.Sc., 1987]

**Ans.**  $209/343$ .

(b)  $A$ ,  $B$  and  $C$  are independent witnesses of an event which is known to have occurred.  $A$  speaks the truth three times out of four,  $B$  four times out of five and  $C$  five times out of six. What is the probability that the occurrence will be reported truthfully by majority of three witnesses?

**Ans.**  $31/60$ .

(c) A man seeks advice regarding one of two possible courses of action from three advisers who arrived at their recommendations independently. He follows the recommendation of the majority. The probability that the individual advisers are wrong are  $0.1$ ,  $0.05$  and  $0.05$  respectively. What is the probability that the man takes incorrect advise? [Gujarat Univ. B.Sc., 1987]

22. (a) The odds against a certain event are 5 to 2 and odds in favour of another (independent) event are 6 to 5. Find the chance that at least one of the events will happen. [Madras Univ. B.Sc., 1987]

**Ans.**  $52/77$ .

(b) A person takes four tests in succession. The probability of his passing the first test is  $p$ , that of his passing each succeeding test is  $p$  or  $p/2$  according as he passes or fails the preceding one. He qualifies provided he passes at least three tests. What is his chance of qualifying? [Gauhati Univ. B.Sc. (Hons.) 1988]

23. (a) The probability that a 50-years old man will be alive at 60 is  $0.83$  and the probability that a 45-years old woman will be alive at 55 is  $0.87$ . What is the probability that a man who is 50 and his wife who is 45 will both be alive 10 years hence?

**Ans.**  $0.7221$ .

(b) It is 8:5 against a husband who is 55 years old living till he is 75 and 4:3 against his wife who is now 48, living till she is 68. Find the probability that (i) the couple will be alive 20 years hence, and (ii) at least one of them will be alive 20 years hence.

**Ans.** (i)  $15/91$ , (ii)  $59/91$ .

(c) A husband and wife appear in an interview for two vacancies in the same post. The probability of husband's selection is  $1/7$  and that of wife's selection is  $1/5$ . What is the probability that only one of them will be selected?

Ans. 2/7

[Delhi Univ. B.Sc., 1986]

24. (a) The chances of winning of two race-horses are  $1/3$  and  $1/6$  respectively. What is the probability that at least one will win when the horses are running (a) in different races, and (b) in the same race?

Ans. (a)  $8/18$  (b)  $1/2$

(b) A problem in statistics is given to three students whose chances of solving it are  $1/2$ ,  $1/3$  and  $1/4$ . What is the probability that the problem will be solved?

Ans. 3/4

[Meerut Univ. B.Sc., 1990]

25. (a) Ten pairs of shoes are in a closet. Four shoes are selected at random. Find the probability that there will be at least one pair among the four shoes selected?

$$\text{Ans. } 1 - \frac{{}^{10}C_4 \times 2^4}{{}^{20}C_4}$$

(b) From 100 tickets numbered 1, 2, ..., 100 four are drawn at random. What is the probability that 3 of them will bear number from 1 to 20 and the fourth will bear any number from 21 to 100?

$$\text{Ans. } \frac{{}^{20}C_3 \times {}^{80}C_1}{{}^{100}C_4}$$

26. A six faced die is so biased that it is twice as likely to show an even number as an odd number when thrown. It is thrown twice. What is the probability that the sum of the two numbers thrown is odd?

Ans. 4/9

27. From a group of 8 children, 5 boys and 3 girls, three children are selected at random. Calculate the probabilities that selected group contains (i) no girl, (ii) only one girl, (iii) one particular girl, (iv) at least one girl, and (v) more girls than boys.

Ans. (i)  $5/28$ , (ii)  $15/28$ , (iii)  $5/28$ , (iv)  $23/28$ , (v)  $2/7$ .

28. If three persons, selected at random, are stopped on a street, what are the probabilities that :

- (a) all were born on a Friday;
- (b) two were born on a Friday and the other on a Tuesday;
- (c) none was born on a Monday.

Ans. (a)  $1/343$ , (b)  $3/343$ , (c)  $216/343$ .

29. (a) A and B toss a coin alternately on the understanding that the first who obtains the head wins. If A starts, show that their respective chances of winning are  $2/3$  and  $1/3$ .

(b) A, B and C, in order, toss a coin. The first one who throws a head wins. If A starts, find their respective chances of winning. (Assume that the game may

continue indefinitely.)

**Ans.**  $4/7, 2/7, 1/7$ .

(c) A man alternately tosses a coin and throws a die, beginning with coin. What is the probability that he will get a head before he gets a '5 or 6' on die?

**Ans.**  $3/4$ .

**30. (a)** Two ordinary six-sided dice are tossed.

(i) What is the probability that both the dice show the number 5.

(ii) What is the probability that both the dice show the same number.

(iii) Given that the sum of two numbers shown is 8, find the conditional probability that the number noted on the first dice is larger than the number noted on the second dice.

(b) Six dice are thrown simultaneously. What is the probability that all will show different faces?

**31. (a)** A bag contains 10 balls, two of which are red, three blue and five black. Three balls are drawn at random from the bag, that is every ball has an equal chance of being included in the three. What is the probability that

(i) the three balls are of different colours,

(ii) two balls are of the same colour, and

(iii) the balls are all of the same colour?

**Ans.** (i)  $30/120$ , (ii)  $79/120$ , (iii)  $11/120$ .

(b) A is one of six horses entered for a race and is to be ridden by one of the two jockeys B and C. It is 2 to 1 that B rides A, in which case all the horses are equally likely to win, with rider C, A's chance is trebled.

(i) Find the probability that A wins.

(ii) What are odds against A's winning?

[Shivaji Univ. B.Sc. (Stat. Hons.), 1992]

**Hint.** Probability of A's winning

$$= P(B \text{ rides } A \text{ and } A \text{ wins}) + P(C \text{ rides } A \text{ and } A \text{ wins})$$

$$= \frac{2}{3} \times \frac{1}{6} + \frac{1}{3} \times \frac{3}{6} = \frac{5}{18}$$

∴ Probability of A's losing  $= 1 - 5/18 = 13/18$ .

Hence odds against A's winning are :  $13/18 : 5/18$ , i.e.,  $13 : 5$ .

**32. (a)** Two-third of the students in a class are boys and the rest girls. It is known that the probability of a girl getting a first class is 0.25 and that of boy getting a first class is 0.28. Find the probability that a student chosen at random will get first class marks in the subject.

**Ans.** 0.27

(b) You need four eggs to make omelettes for breakfast. You find a dozen eggs in the refrigerator but do not realise that two of these are rotten. What is the probability that of the four eggs you choose at random

(i) none is rotten,

(ii) exactly one is rotten?

Ans. (i) 625/1296 : (ii) 500/1296.

(c) The probability of occurrence of an event  $A$  is 0.7, the probability of non-occurrence of another event  $B$  is 0.5 and that of at least one of  $A$  or  $B$  not occurring is 0.6. Find the probability that at least one of  $A$  or  $B$  occurs.

[Mysore Univ. B.Sc., 1991]

33. (a) The odds against  $A$  solving a certain problem are 4 to 3 and odds in favour of  $B$  solving the same problem are 7 to 5. What is the probability that the problem is solved if they both try independently? [Gujarat Univ. B.Sc., 1987]

Ans. 16/21

(b) A certain drug manufactured by a company is tested chemically for its toxic nature. Let the event 'the drug is toxic' be denoted by  $E$  and the event 'the chemical test reveals that the drug is toxic' be denoted by  $F$ . Let  $P(E) = \theta$ ,  $P(F | E) = P(\bar{F} | \bar{E}) = 1 - \theta$ . Then show that probability that the drug is not toxic given that the chemical test reveals that it is toxic is free from  $\theta$ .

Ans. 1/2

[M.S. Baroda Univ. B.Sc., 1992]

34. A bag contains 6 white and 9 black balls. Four balls are drawn at a time. Find the probability for the first draw to give 4 white and the second draw to give 4 black balls in each of the following cases :

(i) The balls are replaced before the second draw.

(ii) The balls are not replaced before the second draw.

[Jammu Univ. B.Sc., 1992]

$$\text{Ans. (i)} \frac{{}^6C_4}{{}^{15}C_4} \times \frac{{}^9C_4}{{}^{15}C_4} \quad \text{(ii)} \frac{{}^6C_4}{{}^{15}C_4} \times \frac{{}^9C_4}{{}^{11}C_4}$$

35. The chances that doctor  $A$  will diagnose a disease  $X$  correctly is 60%. The chances that a patient will die by his treatment after correct diagnosis is 40% and the chance of death by wrong diagnosis is 70%. A patient of doctor  $A$ , who had disease  $X$ , died. What is the chance that his disease was diagnosed correctly?

**Hint.** Let us define the following events:

$E_1$  : Disease  $X$  is diagnosed correctly by doctor  $A$ .

$E_2$  : A patient (of doctor  $A$ ) who has disease  $X$  dies.

Then we are given :

$$P(E_1) = 0.6 \Rightarrow P(\bar{E}_1) = 1 - 0.6 = 0.4$$

$$\text{and } P(E_2 | E_1) = 0.4 \quad \text{and } P(E_2 | \bar{E}_1) = 0.7$$

$$\text{We want } P(E_1 | E_2) = \frac{P(E_1 \cap E_2)}{P(E_2)} = \frac{P(E_1 \cap E_2)}{P(E_1 \cap E_2) + P(\bar{E}_1 \cap E_2)} = \frac{6}{13}$$

36. The probability that at least 2 of 3 people  $A$ ,  $B$  and  $C$  will survive for 10 years is  $247/315$ . The probability that  $A$  alone will survive for 10 years is  $4/105$  and the probability that  $C$  alone will die within 10 years is  $2/21$ . Assuming that the events of the survival of  $A$ ,  $B$  and  $C$  can be regarded as independent, calculate the

probability of surviving 10 years for each person.

Ans. 3/5, 5/7, 7/9.

37. A and B throw alternately a pair of unbiased dice, A beginning. A wins if he throws 7 before B throws 6, and B wins if he throws 6 before A throws 7. If A and B respectively denote the events that A wins and B wins the series, and  $a$  and  $b$  respectively denote the events that it is A's and B's turn to throw the dice, show that

$$(i) P(A|a) = \frac{1}{6} + \frac{5}{6} P(A|b), \quad (ii) P(A|b) = \frac{31}{36} P(A|a),$$

$$(iii) P(B|a) = \frac{5}{6} P(B|b), \text{ and } (iv) P(B|b) = \frac{5}{36} + \frac{13}{36} P(B|a).$$

Hence or otherwise, find  $P(A|a)$  and  $P(B|a)$ . Also comment on the result that  $P(A|a) + P(B|a) = 1$ .

38. A bag contains an assortment of blue and red balls. If two balls are drawn at random, the probability of drawing two red balls is five times the probability of drawing two blue balls. Furthermore, the probability of drawing one ball of each colour is six times the probability of drawing two blue balls. How many red and blue balls are there in the bag?

**Hint.** Let number of red and blue balls in the bag be  $r$  and  $b$  respectively. Then

$$p_1 = \text{Prob. of drawing two red balls} = \frac{r(r-1)}{(r+b)(r+b-1)}$$

$$p_2 = \text{Prob. of drawing two blue balls} = \frac{b(b-1)}{(r+b)(r+b-1)}$$

$$p_3 = \text{Prob. of drawing one red and one blue ball} = \left[ \frac{2br}{(r+b)(r+b-1)} \right]$$

$$\text{Now } p_1 = 5p_2 \text{ and } p_3 = 6p_2$$

$$\therefore r(r-1) = 5b(b-1) \text{ and } 2br = 6b(b-1)$$

$$\text{Hence } b = 3 \text{ and } r = 6.$$

39. Three newspapers A, B and C are published in a certain city. It is estimated from a survey that 20% read A; 16% read B, 14% read C, 8% read A and B, 5% read A and C, 4% read B and C and 2% read all the three newspapers. What is the probability that a normally chosen person

(i) does not read any paper, (ii) does not read C

(iii) reads A but not B, (iv) reads only one of these papers, and

(v) reads only two of these papers.

Ans. (i) 0.65, (ii) 0.86, (iii) 0.12, (iv) 0.22, (v) 0.11.

40. (a) A die is thrown twice, the event space  $S$  consisting of the 36 possible pairs of outcomes  $(a,b)$  each assigned probability  $1/36$ . Let A, B and C denote the following events :

$A = \{(a,b) | a \text{ is odd}\}, B = \{(a,b) | b \text{ is odd}\}, C = \{(a,b) | a+b \text{ is odd}\}$

Check whether A, B and C are independent or independent in pairs only.

[Calcutta Univ. B.Sc. Hons., 1985]

(b) Eight tickets numbered 111, 121, 122, 122, 211, 212, 212, 221 are placed in a hat and stirred. One of them is then drawn at random. Show that the event  $A$  : "the first digit on the ticket drawn will be 1",  $B$  : "the second digit on the ticket drawn will be 1" and  $C$  : "the third digit on the ticket drawn will be 1", are not pairwise independent although

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

41. (a) Four identical marbles marked 1, 2, 3 and 123 respectively are put in an urn and one is drawn at random. Let  $A_i$ , ( $i = 1, 2, 3$ ), denote the event that the number  $i$  appears on the drawn marble. Prove that the events  $A_1, A_2$  and  $A_3$  are pairwise independent but not mutually independent.

[Gauhati Univ. B.Sc. (Hons.), 1988]

Hint.  $P(A_1) = \frac{1}{2} = P(A_2) = P(A_3)$ ;  $P(A_1 A_2) = P(A_1 A_3) = P(A_2 A_3) = \frac{1}{4}$

$$P(A_1 A_2 A_3) = \frac{1}{4}.$$

(b) Two fair dice are thrown independently. Define the following events :

$A$  : Even number on the first dice

$B$  : Even number on the second dice.

$C$  : Same number on both dice.

Discuss the independence of the events  $A, B$  and  $C$ .

(c) A die is of the shape of a regular tetrahedron whose faces bear the numbers 111, 112, 121, 122.  $A_1, A_2, A_3$  are respectively the events that the first two, the last two and the extreme two digits are the same, when the die is tossed at random. Find whether or not the events  $A_1, A_2, A_3$  are (i) pairwise independent, (ii) mutually (i.e. completely) independent. Determine  $P(A_1 | A_2 A_3)$  and explain its value by argument.

[Civil Services (Main), 1983]

42. (a) For two events  $A$  and  $B$  we have the following probabilities:

$$P(A) = P(A | B) = \frac{1}{4} \text{ and } P(B | A) = \frac{1}{2}.$$

Check whether the following statements are true or false :

(i)  $A$  and  $B$  are mutually exclusive, (ii)  $A$  and  $B$  are independent, (iii)  $A$  is a subevent of  $B$ , and (iv)  $P(\bar{A} | B) = \frac{3}{4}$

Ans. (i) False, (ii) True, (iii) False, and (iv) True.

(b) Consider two events  $A$  and  $B$  such that  $P(A) = 1/4, P(B | A) = 1/2, P(A | B) = 1/4$ . For each of the following statements, ascertain whether it is true or false :

(i)  $A$  is a sub-event of  $B$ , (ii)  $P(\bar{A} | \bar{B}) = 3/4$ ,

(iii)  $P(A | B) + P(A | \bar{B}) = 1$

43. (a) Let  $A$  and  $B$  be two events such that  $P(A) = \frac{3}{4}$  and  $P(B) = \frac{5}{8}$ .

Show that

$$(i) P(A \cup B) \geq \frac{3}{4}, \quad (ii) \frac{3}{8} \leq P(A \cap B) \leq \frac{5}{8}, \text{ and } (iii) \frac{1}{8} \leq P(A \cap \bar{B}) \leq \frac{3}{8}.$$

[Coimbatore Univ. B.E., Nov. 1990; Delhi Univ. B.Sc.(Stat. Hons.), 1986]

(b) Given two events  $A$  and  $B$ . If the odds against  $A$  are 2 to 1 and those in favour of  $A \cup B$  are 3 to 1, show that

$$\frac{5}{12} \leq P(B) \leq \frac{3}{4}$$

Give an example in which  $P(B) = 3/4$  and one in which  $P(B) = 5/12$ .

44. Let  $A$  and  $B$  be events, neither of which has probability zero. Prove or disprove the following events :

- (i) If  $A$  and  $B$  are disjoint,  $A$  and  $B$  are independent,
- (ii) If  $A$  and  $B$  are independent,  $A$  and  $B$  are disjoint.

45. (a) It is given that  $P(A_1 \cup A_2) = \frac{5}{6}$ ,  $P(A_1 \cap A_2) = \frac{1}{3}$  and  $P(\bar{A}_2) = \frac{1}{2}$ ,

where  $P(\bar{A}_2)$  stands for the probability that  $A_2$  does not happen. Determine  $P(A_1)$  and  $P(A_2)$ .

Hence show that  $A_1$  and  $A_2$  are independent.

$$\text{Ans. } P(A_1) = \frac{2}{3}, \quad P(A_2) = \frac{1}{2}$$

(b)  $A$  and  $B$  are events such that

$$P(A \cup B) = \frac{3}{4}, \quad P(A \cap B) = \frac{1}{4}, \text{ and } P(\bar{A}) = \frac{2}{3}.$$

Find (i)  $P(A)$ , (ii)  $P(B)$  and (iii)  $P(A \cap \bar{B})$ .

[Madras Univ. B.E., 1989]

Ans. (i)  $1/3$ , (ii)  $2/3$  (iii)  $1/12$ .

46. A thief has a bunch of  $n$  keys, exactly one of which fits a lock. If the thief tries to open the lock by trying the keys at random, what is the probability that he requires exactly  $k$  attempts, if he rejects the keys already tried? Find the same probability if he does not reject the keys already tried.

[Aligarh Univ. B.Sc., 1991]

$$\text{Ans. (i) } \frac{1}{n}, \quad \text{(ii) } \left(\frac{n-1}{n}\right)^{k-1} \cdot \frac{1}{n}$$

(b) There are  $M$  urns numbered 1 to  $M$  and  $M$  balls numbered 1 to  $M$ . The balls are inserted randomly in the urns with one ball in each urn. If a ball is put into the urn bearing the same number as the ball, a match is said to have occurred. Find the probability that no match has occurred. [Civil Services (Main), 1984]

**Hint.** See Example 4-54 page 4-97.

47. If  $n$  letters are placed at random in  $n$  correctly addressed envelopes, find the probability that

- (i) none of the letters is placed in the correct envelope,

- (ii) At least one letter goes to the correct envelope,  
 (iii) All letters go to the correct envelopes.

[Delhi Univ. B.Sc. (Stat Hons.), 1987, 1984]

48. An urn contains  $n$  white and  $m$  black balls, a second urn contains  $N$  white and  $M$  black balls. A ball is randomly transferred from the first to the second urn and then from the second to the first urn. If a ball is now selected randomly from the first urn, prove that the probability that it is white is

$$\frac{n}{n+m} + \frac{mN - nM}{(n+m)^2(N+M+1)}$$

[Delhi Univ. B.Sc. (Stat.Hons.) 1986]

**Hint.** Let us define the following events :

$B_i$  : Drawing of a black ball from the  $i$ th urn,  $i = 1, 2$ .

$W_i$  : Drawing of a white ball from the  $i$ th urn ,  $i = 1, 2$ .

The four distinct possibilities for the first two exchanges are  $B_1 W_2, B_1 B_2, W_1 B_2, W_1 W_2$ . Hence if  $E$  denotes the event of drawing a white ball from the first urn after the exchanges, then

$$P(E) = P(B_1 W_2 E) + P(B_1 B_2 E) + P(W_1 B_2 E) + P(W_1 W_2 E) \quad \dots (*)$$

We have :

$$P(B_1 W_2 E) = P(B_1) \cdot P(W_2 | B_1) P(E | B_1 W_2) = \frac{m}{m+n} \times \frac{N}{M+N+1} \times \frac{n+1}{m+n}$$

$$P(B_1 B_2 E) = P(B_1) \cdot P(B_2 | B_1) \cdot P(E | B_1 B_2) = \frac{m}{m+n} \times \frac{M+1}{M+N+1} \times \frac{n}{m+n}$$

$$P(W_1 B_2 E) = P(W_1) \cdot P(B_2 | W_1) \cdot P(E | W_1 B_2) = \frac{n}{m+n} \times \frac{M}{M+N+1} \times \frac{n-1}{m+n}$$

$$P(W_1 W_2 E) = P(W_1) \cdot P(W_2 | W_1) \cdot P(E | W_1 W_2) = \frac{n}{m+n} \times \frac{N+1}{M+N+1} \times \frac{n}{m+n}$$

Substituting in (\*) and simplifying we get the result.

49. A particular machine is prone to three similar types of faults  $A_1, A_2$  and  $A_3$ . Past records on breakdowns of the machine show the following : the probability of a breakdown (*i.e.*, at least one fault) is 0.1; for each  $i$ , the probability that fault  $A_i$  occurs and the others do not is 0.02 ; for each pair  $i, j$  the probability that  $A_i$  and  $A_j$  occur but the third fault does not is 0.012. Determine the probabilities of

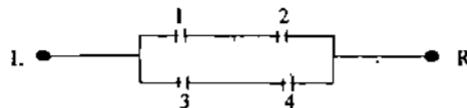
(a) the fault of type  $A_1$  occurring irrespective of whether the other faults occur or not,

(b) a fault of type  $A_1$  given that  $A_2$  has occurred,

(c) faults of type  $A_1$  and  $A_2$  given that  $A_3$  has occurred.

[London U. B.Sc. 1976]

50. The probability of the closing of each relay of the circuit shown below is given by  $p$ . If all the relays function independently, what is the probability that a circuit exists between the terminals L and R?



$$\text{Ans. } p^2(2-p^2).$$

**4.9. Bayes Theorem.** If  $E_1, E_2, \dots, E_n$  are mutually disjoint events with  $P(E_i) \neq 0$ , ( $i = 1, 2, \dots, n$ ) then for any arbitrary event A which is a subset of  $\bigcup_{i=1}^n E_i$ , such that  $P(A) > 0$ , we have

$$P(E_i | A) = \frac{P(E_i) P(A | E_i)}{\sum_{i=1}^n P(E_i) P(A | E_i)}, \quad i = 1, 2, \dots, n. \quad \dots(4.12)$$

**Proof.** Since  $A \subset \bigcup_{i=1}^n E_i$ , we have

$$A = A \cap (\bigcup_{i=1}^n E_i) = \bigcup_{i=1}^n (A \cap E_i) \quad [\text{By distributive law}]$$

Since  $(A \cap E_i) \subset E_i$ , ( $i = 1, 2, \dots, n$ ) are mutually disjoint events, we have by addition theorem of probability (or Axiom 3 of probability)

$$P(A) = P\left[\bigcup_{i=1}^n (A \cap E_i)\right] = \sum_{i=1}^n P(A \cap E_i) = \sum_{i=1}^n P(E_i) P(A | E_i), \quad \dots(*)$$

by compound theorem of probability.

Also we have

$$P(A \cap E_i) = P(A) P(E_i | A)$$

$$P(E_i | A) = \frac{P(A \cap E_i)}{P(A)} = \frac{P(E_i) P(A | E_i)}{\sum_{i=1}^n P(E_i) P(A | E_i)} \quad [\text{From } (*)]$$

**Remarks.** 1. The probabilities  $P(E_1), P(E_2), \dots, P(E_n)$  are termed as the '*a priori probabilities*' because they exist before we gain any information from the experiment itself.

2. The probabilities  $P(A | E_i)$ ,  $i = 1, 2, \dots, n$  are called '*likelihoods*' because they indicate how likely the event A under consideration is to occur, given each and every *a priori* probability.

3. The probabilities  $P(E_i | A)$ ,  $i = 1, 2, \dots, n$  are called '*posterior probabilities*' because they are determined after the results of the experiment are known.

4. From (\*) we get the following important result:

"If the events  $E_1, E_2, \dots, E_n$  constitute a partition of the sample space S and  $P(E_i) \neq 0$ ,  $i = 1, 2, \dots, n$ , then for any event A in S we have

$$P(A) = \sum_{i=1}^n P(A \cap E_i) = \sum_{i=1}^n P(E_i) P(A | E_i) \quad \dots(4.12a)$$

**Cor. (Bayes theorem for future events)**

The probability of the materialisation of another event  $C$ , given

$P(C | A \cap E_1), P(C | A \cap E_2), \dots, P(C | A \cap E_n)$  is

$$P(C | A) = \frac{\sum_{i=1}^n P(E_i) P(A | E_i) P(C | E_i \cap A)}{\sum_{i=1}^n P(E_i) P(A | E_i)} \quad \dots(4.12b)$$

**Proof.** Since the occurrence of event  $A$  implies the occurrence of one and only one of the events  $E_1, E_2, \dots, E_n$ , the event  $C$  (granted that  $A$  has occurred) can occur in the following mutually exclusive ways:

$$\begin{aligned} & C \cap E_1, C \cap E_2, \dots, C \cap E_n \\ \text{i.e., } & C = (C \cap E_1) \cup (C \cap E_2) \cup \dots \cup (C \cap E_n) \\ \Rightarrow & C | A = [(C \cap E_1) | A] \cup [(C \cap E_2) | A] \cup \dots \cup [(C \cap E_n) | A] \\ \therefore P(C | A) &= P[(C \cap E_1) | A] + P[(C \cap E_2) | A] + \dots + P[(C \cap E_n) | A] \\ &= \sum_{i=1}^n P[(C \cap E_i) | A] \\ &= \sum_{i=1}^n P(E_i | A) P[C | (E_i \cap A)] \end{aligned}$$

Substituting the value of  $P(E_i | A)$  from (\*), we get

$$P(C | A) = \frac{\sum_{i=1}^n P(E_i) P(A | E_i) P(C | E_i \cap A)}{\sum_{i=1}^n P(E_i) P(A | E_i)}$$

**Remark.** It may happen that the materialisation of the event  $E_i$  makes  $C$  independent of  $A$ , then we have

$$P(C | E_i \cap A) = P(C | E_i),$$

and the above formula reduces to

$$P(C | A) = \frac{\sum_{i=1}^n P(E_i) P(A | E_i) P(C | E_i)}{\sum_{i=1}^n P(E_i) P(A | E_i)} \quad \dots(4.12c)$$

The event  $C$  can be considered in regard to  $A$  as Future Event.

**Example 4-30.** In 1989 there were three candidates for the position of principal - Mr. Chatterji, Mr. Ayangar and Dr. Singh - whose chances of getting the appointment are in the proportion 4:2:3 respectively. The probability that Mr. Chatterji if selected would introduce co-education in the college is 0.3. The probabilities of Mr. Ayangar and Dr. Singh doing the same are respectively 0.5 and 0.8. What is the probability that there was co-education in the college in 1990?

[Delhi Univ. B.Sc.(Stat. Hons.), 1992; Gorakhpur Univ. B.Sc., 1992]

**Solution.** Let the events and probabilities be defined as follows:

$A$  : Introduction of co-education

$E_1$  : Mr. Chatterji is selected as principal

$E_2$  : Mr. Ayangar is selected as principal

$E_3$  : Dr. Singh is selected as principal.

Then

$$P(E_1) = \frac{4}{9}, \quad P(E_2) = \frac{2}{9} \quad \text{and} \quad P(E_3) = \frac{3}{9}$$

$$P(A | E_1) = \frac{3}{10}, \quad P(A | E_2) = \frac{5}{10} \quad \text{and} \quad P(A | E_3) = \frac{8}{10}$$

$$\begin{aligned} \therefore P(A) &= P[(A \cap E_1) \cup (A \cap E_2) \cup (A \cap E_3)] \\ &= P(A \cap E_1) + P(A \cap E_2) + P(A \cap E_3) \\ &= P(E_1) P(A | E_1) + P(E_2) P(A | E_2) + P(E_3) P(A | E_3) \\ &= \frac{4}{9} \cdot \frac{3}{10} + \frac{2}{9} \cdot \frac{5}{10} + \frac{3}{9} \cdot \frac{8}{10} = \frac{23}{45} \end{aligned}$$

**Example 4-31.** The contents of urns I, II and III are as follows:

1 white, 2 black and 3 red balls,

2 white, 1 black and 1 red balls, and

4 white, 5 black and 3 red balls.

One urn is chosen at random and two balls drawn. They happen to be white and red. What is the probability that they come from urns I, II or III?

[Delhi Univ. B.Sc. (Stat. Hons.), 1988]

**Solution.** Let  $E_1$ ,  $E_2$ , and  $E_3$  denote the events that the urn I, II and III is chosen, respectively, and let  $A$  be the event that the two balls taken from the selected urn are white and red. Then

$$P(E_1) = P(E_2) = P(E_3) = \frac{1}{3}$$

$$P(A | E_1) = \frac{1 \times 3}{6C_2} = \frac{1}{5}, \quad P(A | E_2) = \frac{2 \times 1}{4C_2} = \frac{1}{3},$$

$$\text{and} \quad P(A | E_3) = \frac{4 \times 3}{12C_2} = \frac{2}{11}$$

Hence .

$$\begin{aligned} P(E_2 | A) &= \frac{P(E_2) P(A | E_2)}{\sum_{i=1}^3 P(E_i) P(A | E_i)} \\ &= \frac{\frac{1}{3} \times \frac{1}{3}}{\frac{1}{3} \times \frac{1}{5} + \frac{1}{3} \times \frac{1}{3} + \frac{1}{3} \times \frac{2}{11}} = \frac{55}{118} \end{aligned}$$

Similarly

$$\begin{aligned} P(E_3 | A) &= \frac{\frac{1}{3} \times \frac{2}{11}}{\frac{1}{3} \times \frac{1}{5} + \frac{1}{3} \times \frac{1}{3} + \frac{1}{3} \times \frac{1}{11}} = \frac{30}{118} \\ \therefore P(E_1 | A) &= 1 - \frac{55}{118} - \frac{30}{118} = \frac{33}{118} \end{aligned}$$

**Example 4.32.** In answering a question on a multiple choice test a student either knows the answer or he guesses. Let  $p$  be the probability that he knows the answer and  $1-p$  the probability that he guesses. Assume that a student who guesses at the answer will be correct with probability  $1/5$ , where  $5$  is the number of multiple-choice alternatives. What is the conditional probability that a student knew the answer to a question given that he answered it correctly?

[Delhi Univ. B.Sc. (Maths Hons.), 1985]

**Solution.** Let us define the following events:

$E_1$  : The student knew the right answer.

$E_2$  : The student guesses the right answer.

$A$  : The student gets the right answer.

Then we are given

$$P(E_1) = p, \quad P(E_2) = 1-p, \quad P(A | E_2) = 1/5$$

$$P(A | E_1) = P[\text{student gets the right answer given that he knew the right answer}] = 1$$

We want  $P(E_1 | A)$ .

Using Bayes' rule, we get :

$$P(E_1 | A) = \frac{P(E_1) \cdot P(A | E_1)}{P(E_1) P(A | E_1) + P(E_2) P(A | E_2)} = \frac{p \times 1}{p \times 1 + (1-p) \times \frac{1}{5}} = \frac{5p}{4p+1}$$

**Example 4.33.** In a bolt factory machines  $A$ ,  $B$  and  $C$  manufacture respectively 25%, 35% and 40% of the total. Of their output 5, 4, 2 per cent are defective bolts. A bolt is drawn at random from the product and is found to be defective. What are the probabilities that it was manufactured by machines  $A$ ,  $B$  and  $C$ ?

**Solution.** Let  $E_1, E_2$  and  $E_3$  denote the events that a bolt selected at random is manufactured by the machines  $A, B$  and  $C$  respectively and let  $E$  denote the event of its being defective. Then we have

$$P(E_1) = 0.25, P(E_2) = 0.35, P(E_3) = 0.40$$

The probability of drawing a defective bolt manufactured by machine  $A$  is  $P(E | E_1) = 0.05$ .

Similarly, we have

$$P(E | E_2) = 0.04, \text{ and } P(E | E_3) = 0.02$$

Hence the probability that a defective bolt selected at random is manufactured by machine  $A$  is given by

$$\begin{aligned} P(E_1 | E) &= \frac{P(E_1) P(E | E_1)}{\sum_{i=1}^3 P(E_i) P(E | E_i)} \\ &= \frac{0.25 \times 0.05}{0.25 \times 0.05 + 0.35 \times 0.04 + 0.40 \times 0.02} = \frac{125}{345} = \frac{25}{69} \end{aligned}$$

Similarly

$$P(E_2 | E) = \frac{0.35 \times 0.04}{0.25 \times 0.05 + 0.35 \times 0.04 + 0.40 \times 0.02} = \frac{140}{345} = \frac{28}{69}$$

and

$$P(E_3 | E) = 1 - [P(E_1 | E) + P(E_2 | E)] = 1 - \frac{25}{69} - \frac{28}{69} = \frac{16}{69}$$

This example illustrates one of the chief applications of Bayes Theorem.

#### EXERCISE 4 (d)

1. (a) State and prove Baye's Theorem.

(b) The set of events  $A_k$ , ( $k = 1, 2, \dots, n$ ) are (i) exhaustive and (ii) pairwise mutually exclusive. If for all  $k$  the probabilities  $P(A_k)$  and  $P(E | A_k)$  are known, calculate  $P(A_k | E)$ , where  $E$  is an arbitrary event. Indicate where conditions (i) and (ii) are used.

(c) The events  $E_1, E_2, \dots, E_n$  are mutually exclusive and  $E = E_1 \cup E_2 \cup \dots \cup E_n$ . Show that if  $P(A | E_i) = P(B | E_i)$ ;  $i = 1, 2, \dots, n$ , then  $P(A | E) = P(B | E)$ . Is this conclusion true if the events  $E_i$  are not mutually exclusive?

[Calcutta Univ. B.Sc. (Maths Hons.), 1990]

(d) What are the criticisms against the use of Bayes theorem in probability theory.  
[Sri. Venkateswara Univ. B.Sc., 1991]

(e) Using the fundamental addition and multiplication rules of probability, show that

$$P(B | A) = \frac{P(B) P(A | B)}{P(B) P(A | B) + P(\bar{B}) P(\bar{A} | \bar{B})}$$

where  $\bar{B}$  is the event complementary to the event  $B$ .

[Delhi Univ. M.A. (Econ.), 1987]

2. (a) Two groups are competing for the positions on the Board of Directors of a corporation. The probabilities that the first and second groups will win are 0.6 and 0.4 respectively. Furthermore, if the first group wins the probability of introducing a new product is 0.8 and the corresponding probability if the second group wins is 0.3. What is the probability that the new product will be introduced?

$$\text{Ans. } 0.6 \times 0.8 + 0.4 \times 0.3 = 0.6$$

(b) The chances of  $X, Y, Z$  becoming managers of a certain company are 4:2:3. The probabilities that bonus scheme will be introduced if  $X, Y, Z$  become managers, are 0.3, 0.5 and 0.8 respectively. If the bonus scheme has been introduced, what is the probability that  $X$  is appointed as the manager.

$$\text{Ans. } 0.51$$

(c) A restaurant serves two special dishes,  $A$  and  $B$  to its customers consisting of 60% men and 40% women. 80% of men order dish  $A$  and the rest  $B$ . 70% of women order dish  $B$  and the rest  $A$ . In what ratio of  $A$  to  $B$  should the restaurant prepare the two dishes? (Bangalore Univ. B.Sc., 1991)

$$\text{Ans. } P(A) = P[(A \cap M) \cup (A \cap W)] = 0.6 \times 0.8 + 0.4 \times 0.3 = 0.6$$

$$\text{Similarly } P(B) = 0.4. \text{ Required ratio} = 0.6 : 0.4 = 3 : 2.$$

3. (a) There are three urns having the following compositions of black and white balls.

Urn 1 : 7 white, 3 black balls

Urn 2 : 4 white, 6 black balls

Urn 3 : 2 white, 8 black balls.

One of these urns is chosen at random with probabilities 0.20, 0.60 and 0.20 respectively. From the chosen urn two balls are drawn at random without replacement. Calculate the probability that both these balls are white.

$$\text{Ans. } 8/45.$$

(Madurai Univ. B.Sc., 1991)

(b) Bowl I contain 3 red chips and 7 blue chips, bowl II contain 6 red chips and 4 blue chips. A bowl is selected at random and then 1 chip is drawn from this bowl. (i) Compute the probability that this chip is red, (ii) Relative to the hypothesis that the chip is red, find the conditional probability that it is drawn from bowl II.

[Delhi Univ. B.Sc. (Maths Hons.) 1987]

(c) In a factory machines  $A$  and  $B$  are producing springs of the same type. Of this production, machines  $A$  and  $B$  produce 5% and 10% defective springs, respectively. Machines  $A$  and  $B$  produce 40% and 60% of the total output of the factory. One spring is selected at random and it is found to be defective. What is the possibility that this defective spring was produced by machine  $A$ ?

[Delhi Univ. M.A. (Econ.), 1986]

(d) Urn  $A$  contains 2 white, 1 black and 3 red balls, urn  $B$  contains 3 white, 2 black and 4 red balls and urn  $C$  contains 4 white, 3 black and 2 red balls. One urn is chosen at random and 2 balls are drawn. They happen to be red and black. What

is the probability that both balls came from urn 'B' ?

[Madras U. B.Sc. April; 1989]

(e) Urn  $X_1, X_2, X_3$ , each contains 5 red and 3 white balls. Urns  $Y_1, Y_2$ , each contain 2 red and 4 white balls. An urn is selected at random and a ball is drawn. It is found to be red. Find the probability that the ball comes out of the urns of the first type.

[Bombay U. B.Sc., April 1992]

(f) Two shipments of parts are received. The first shipment contains 1000 parts with 10% defectives and the second shipment contains 2000 parts with 5% defectives. One shipment is selected at random. Two parts are tested and found good. Find the probability (*a posterior*) that the tested parts were selected from the first shipment.

[Burdwan Univ. B.Sc. (Hons.), 1988]

(g) There are three machines producing 10,000 ; 20,000 and 30,000 bullets per hour respectively. These machines are known to produce 5%, 4% and 2% defective bullets respectively. One bullet is taken at random from an hour's production of the three machines. What is the probability that it is defective? If the drawn bullet is defective, what is the probability that this was produced by the second machine?

[Delhi Univ. B.Sc. (Stat. Hons.), 1991]

4. (a) Three urns are given each containing red and white chips as indicated.

Urn 1 : 6 red and 4 white.

Urn 2 : 2 red and 6 white.

Urn 3 : 1 red and 8 white.

(i) An urn is chosen at random and a ball is drawn from this urn. The ball is red. Find the probability that the urn chosen was urn I.

(ii) An urn is chosen at random and two balls are drawn without replacement from this urn. If both balls are red, find the probability that urn I was chosen. Under these conditions, what is the probability that urn III was chosen.

Ans. 108/173, 112/12, 0 [Gauhati Univ. B.Sc., 1990]

(b) There are ten urns of which each of three contains 1 white and 9 black balls, each of other three contains 9 white and 1 black ball, and of the remaining four, each contains 5 white and 5 black balls. One of the urns is selected at random and a ball taken blindly from it turns out to be white. What is the probability that an urn containing 1 white and 9 black balls was selected? (Agra Univ. B.Sc., 1991)

Hint :  $P(E_1) = \frac{3}{10}$ ,  $P(E_2) = \frac{3}{10}$  and  $P(E_3) = \frac{4}{10}$ .

Let A be the event of drawing a white ball.

$$P(A) = \frac{3}{10} \times \frac{1}{10} + \frac{3}{10} \times \frac{9}{10} + \frac{4}{10} \times \frac{5}{10} = \frac{1}{2}$$

$$P(A | E_1) = \frac{1}{10} \text{ and } P(E_1 | A) = \frac{3}{50}$$

(c) It is known that an urn containing altogether 10 balls was filled in the following manner: A coin was tossed 10 times, and according as it showed heads or tails, one white or one black ball was put into the urn. Balls are drawn from this

urn one at a time, 10 times in succession (with replacement) and every one turns out to be white. Find the chance that the urn contains nothing but white balls.

**Ans.** 0.0702.

**5. (a)** From a vessel containing 3 white and 5 black balls, 4 balls are transferred into an empty vessel. From this vessel a ball is drawn and is found to be white. What is the probability that out of four balls transferred, 3 are white and 1 black. [Delhi Univ. B.Sc. (Stat. Hons.), 1985]

**Hint.** Let the five mutually exclusive events for the four balls transferred be  $E_0, E_1, E_2, E_3$ , and  $E_4$ , where  $E_i$  denotes the event that  $i$  white balls are transferred and let  $A$  be the event of drawing a white ball from the new vessel.

$$\text{Then } P(E_0) = \frac{^5C_4}{^8C_4}, P(E_1) = \frac{^3C_1 \times ^5C_3}{^8C_4}, P(E_2) = \frac{^3C_2 \times ^5C_2}{^8C_4}$$

$$P(E_3) = \frac{^3C_3 \times ^5C_1}{^8C_4} \text{ and } P(E_4) = 0$$

$$\text{Also } P(A | E_0) = 0, P(A | E_1) = \frac{1}{4}, P(A | E_2) = \frac{2}{4}, (A | E_3) = \frac{3}{4},$$

$$\text{and } P(A | E_4) = 1. \text{ Hence } P(E_3 | A) = \frac{1}{7}.$$

**(b)** The contents of the urns 1 and 2 are as follows :

Urn 1 : 4 white and 5 black balls.

Urn 2 : 3 white and 6 black balls.

One urn is chosen at random and a ball is drawn and its colour noted and replaced back to the urn. Again a ball is drawn from the same urn, colour noted and replaced. The process is repeated 4 times and as a result one ball of white colour and three balls of black colour are obtained. What is the probability that the urn chosen was the urn 1 ? [Poona Univ. B.E., 1989]

**Hint.**  $P(E_1) = P(E_2) = 1/2,$

$$P(A | E_1) = 4/9, \quad 1 - P(A | E_1) = 5/9$$

$$P(A | E_2) = 1/3, \quad 1 - P(A | E_2) = 2/3$$

The probability that the urn chosen was the urn 1

$$= \frac{\frac{1}{2} \cdot \frac{4}{9} \left(\frac{5}{9}\right)^3}{\frac{1}{2} \cdot \frac{4}{9} \cdot \left(\frac{5}{9}\right)^3 + \frac{1}{2} \cdot \frac{1}{3} \cdot \left(\frac{2}{3}\right)^3}$$

**(c)** There are five urns numbered 1 to 5. Each urn contains 10 balls. The  $i$ th urn has  $i$  defective balls and  $10 - i$  non-defective balls;  $i = 1, 2, \dots, 5$ . An urn is chosen at random and then a ball is selected at random from that urn. (i) What is the probability that a defective ball is selected ?

(ii) If the selected ball is defective, find the probability that it came from urn  $i$ , ( $i = 1, 2, \dots, 5$ ). [Delhi Univ. B.Sc. (Maths Hons.), 1987]

**Hint.:** Define the following events :

$E_i$  :  $i$ th urn is selected at random.

$A$  : Defective ball is selected.

$$P(E_i) = 1/5; i = 1, 2, \dots, 5.$$

$$P(A | E_i) = P[\text{Defective ball from } i\text{th urn}] = i/10, (i = 1, 2, \dots, 5)$$

$$P(E_i) \cdot P(A | E_i) = \frac{1}{5} \times \frac{i}{10} = \frac{i}{50}, (i = 1, 2, \dots, 5).$$

$$(i) \quad P(A) = \sum_{i=1}^5 P(E_i) P(A | E_i) = \sum_{i=1}^5 \left( \frac{i}{50} \right) = \frac{1+2+3+4+5}{50} = \frac{3}{10}$$

$$(ii) \quad P(E_i | A) = \frac{P(E_i) P(A | E_i)}{\sum P(E_i) P(A | E_i)} = \frac{i/50}{3/10} = \frac{i}{15}; i = 1, 2, \dots, 5.$$

For example, the probability that the defective ball came from 5th urn  
 $= (5/15) = 1/3$ .

6. (a) A bag contains six balls of different colours and a ball is drawn from it at random.  $A$  speaks truth thrice out of 4 times and  $B$  speaks truth 7 times out of 10 times. If both  $A$  and  $B$  say that a red ball was drawn, find the probability of their joint statement being true.

[Delhi Univ. B.Sc. (Stat. Hons.), 1987; Kerala Univ. B.Sc., 1988]

(b)  $A$  and  $B$  are two very weak students of Statistics and their chances of solving a problem correctly are  $1/8$  and  $1/12$  respectively. If the probability of their making a common mistake is  $1/100$  and they obtain the same answer, find the chance that their answer is correct. [Poona Univ. B.Sc., 1989]

$$\text{Ans. Reqd. Probability} = \frac{\frac{1}{8} \times \frac{1}{12}}{\frac{1}{8} \times \frac{1}{12} + (1 - \frac{1}{8}) \cdot (1 - \frac{1}{12}) \cdot \frac{1}{100}} = \frac{13}{14}$$

7. (a) Three boxes, practically indistinguishable in appearance, have two drawers each. Box I contains a gold coin in one and a silver coin in the other drawer, box II contains a gold coin in each drawer and box III contains a silver coin in each drawer. One box is chosen at random and one of its drawers is opened at random and a gold coin found. What is the probability that the other drawer contains a coin of silver? [Gujarat Univ. B.Sc., 1992]

$$\text{Ans. } 1/3, 1/3.$$

(b) Two cannons No. 1 and 2 fire at the same target. Cannon No. 1 gives on an average 9 shots in the time in which Cannon No. 2 fires 10 projectiles. But on an average 8 out of 10 projectiles from Cannon No. 1 and 7 out of 10 from Cannon No. 2 strike the target. In the course of shooting, the target is struck by one projectile. What is the probability of a projectile which has struck the target belonging to Cannon No. 2? [Lucknow Univ. B.Sc., 1991]

$$\text{Ans. } 0.493$$

(c) Suppose 5 men out of 100 and 25 women out of 10,000 are colour blind. A colour blind person is chosen at random. What is the probability of his being male? (Assume males and females to be in equal number.)

Hint.  $E_1$  = Person is a male,  $E_2$  = Person is a female.

$A =$  Person is colour blind.

Then  $P(E_1) = P(E_2) = \frac{1}{2}$ ,  $P(A | E_1) = 0.05$ ,  $P(A | E_2) = 0.0025$ .

Hence find  $P(E_1 | A)$ .

8. (a) Three machines  $X, Y, Z$  with capacities proportional to 2:3:4 are producing bullets. The probabilities that the machines produce defective are 0.1, 0.2 and 0.1 respectively. A bullet is taken from a day's production and found to be defective. What is the probability that it came from machine  $X$ ?

[Madras Univ. B.Sc., 1988]

- (b) In a factory 2 machines  $M_1$  and  $M_2$  are used for manufacturing screws which may be uniquely classified as good or bad.  $M_1$  produces per day  $n_1$  boxes of screws, of which on the average,  $p_1\%$  are bad while the corresponding numbers for  $M_2$  are  $n_2$  and  $p_2$ . From the total production of both  $M_1$  and  $M_2$  for a certain day, a box is chosen at random, a screw taken out of it and it is found to be bad. Find the chance that the selected box is manufactured (i) by  $M_1$ , (ii)  $M_2$ .

Ans. (i)  $n_1 p_1 / (n_1 p_1 + n_2 p_2)$ , (ii)  $n_2 p_2 / (n_1 p_1 + n_2 p_2)$ .

9. (a) A man is equally likely to choose any one of three routes  $A, B, C$  from his house to the railway station, and his choice of route is not influenced by the weather. If the weather is dry, the probabilities of missing the train by routes  $A, B, C$  are respectively  $1/20, 1/10, 1/5$ . He sets out on a dry day and misses the train. What is the probability that the route chosen was  $C$ ?

On a wet day, the respective probabilities of missing the train by routes  $A, B, C$  are  $1/20, 1/5, 1/2$  respectively. On the average, one day in four is wet. If he misses the train, what is the probability that the day was wet?

[Allahabad Univ. B.Sc., 1991]

- (b) A doctor is to visit the patient and from past experience it is known that the probabilities that he will come by train, bus or scooter are respectively  $3/10, 1/5$ , and  $1/10$ , the probability that he will use some other means of transport being, therefore,  $2/5$ . If he comes by train, the probability that he will be late is  $1/4$ , if by bus  $1/3$  and if by scooter  $1/12$ , if he uses some other means of transport it can be assumed that he will not be late. When he arrives he is late. What is the probability that (i) he comes by train (ii) he is not late?

[Burdwan Univ. B.Sc. (Hons.), 1990]

Ans. (i)  $1/2$ , (ii)  $9/34$

10. State and prove Bayes rule and explain why, in spite of its easy deductibility from the postulates of probability, it has been the subject of such extensive controversy.

In the chest X-ray tests, it is found that the probability of detection when a person has actually T.B. is 0.95 and probability of diagnosing incorrectly as having T.B. is 0.002. In a certain city 0.1% of the adult population is suspected to be suffering from T.B. If an adult is selected at random and is diagnosed as having

T.B. on the basis of the X-ray test, what is the probability of his actually having a T.B.? (Nagpur Univ. B.E., 1991)

**Ans.** 0.97

11. A certain transistor is manufactured at three factories at Barnsley, Bradford and Bristol. It is known that the Barnsley factory produces twice as many transistors as the Bradford one, which produces the same number as the Bristol one (during the same period). Experience also shows that 0.2% of the transistors produced at Barnsley and Bradford are faulty and so are 0.4% of those produced at Bristol.

A service engineer, while maintaining an electronic equipment, finds a defective transistor. What is the probability that the Bradford factory is to blame?

(Bangalore Univ. B.E., Oct. 1992)

12. The sample space consists of integers from 1 to  $2n$  which are assigned probabilities proportional to their logarithms. Find the probabilities and show that the conditional probability of the integer 2, given that an even integer occurs, is

$$\frac{\log 2}{[n \log 2 + \log(n!)])} \quad (\text{Lucknow Univ. M.A., 1992})$$

[Hint. Let  $E_i$  : the event that the integer  $2i$  is drawn, ( $i = 1, 2, 3, \dots, n$ ).]

$A$  : the event of drawing an even integer.

$$\Rightarrow A = E_1 \cup E_2 \cup \dots \cup E_n \Rightarrow P(A) = \sum_{i=1}^n P(E_i)$$

But  $P(E_i) = k \log(2i)$  (Given)

$$\therefore P(A) = k \sum_{i=1}^n \log(2i) = k \log \prod_{i=1}^n (2i) = k [n \log 2 + \log(n!)]$$

$$\therefore P(E_1 | A) = \frac{\log(2)}{[n \log 2 + \log(n!)])}$$

13. In answering a question on a multiple choice test, an examinee either knows the answer (with probability  $p$ ), or he guesses (with probability  $1 - p$ ). Assume that the probability of answering a question correctly is unity for an examinee who knows the answer and  $1/m$  for the examinee who guesses, where  $m$  is the number of multiple choice alternatives. Supposing an examinee answers a question correctly, what is the probability that he really knows the answer?

(Delhi Univ. M.C.A., 1990; M.Sc. (Stat.), 1989)

**Hint.** Let  $E_1$  = The examinee knows the answer,

$\bar{E}_1$  = The examinee guesses the answer,

and  $A$  = The examinee answers correctly.

Then  $P(E_1) = p$ ,  $P(\bar{E}_1) = 1 - p$ ,  $P(A | E_1) = 1$  and  $P(A | \bar{E}_1) = 1/m$

Now use Bayes theorem to prove

$$P(E_1 | A) = \frac{mp}{1 + (m-1)p}$$

14. Die  $A$  has four red and two white faces whereas die  $B$  has two red and four white faces. A biased coin is flipped once. If it falls heads, the game continues by

throwing die A, if it falls tails die B is to be used.

- Show that the probability of getting a red face at any throw is 1/2.
- If the first two throws resulted in red faces, what is the probability of red face at the 3rd throw?
- If red face turns up at the first  $n$  throws, what is the probability that die A is being used?

Ans. (ii) 3/5 (iii)  $\frac{2^n}{2^n + 1}$

15. A manufacturing firm produces steel pipes in three plants with daily production volumes of 500, 1,000 and 2,000 units respectively. According to past experience it is known that the fraction of defective outputs produced by the three plants are respectively 0.005, 0.008 and 0.010. If a pipe is selected at random from a day's total production and found to be defective, from which plant does that pipe come?

Ans. Third plant.

16. A piece of mechanism consists of 11 components, 5 of type A, 3 of type B, 2 of type C and 1 of type D. The probability that any particular component will function for a period of 24 hours from the commencement of operations without breaking down is independent of whether or not any other component breaks down during that period and can be obtained from the following table:

*Component type: ABCD*

*Probability: 0.60 0.70 0.30 0.2*

(i) Calculate the probability that 2 components chosen at random from the 11 components will both function for a period of 24 hours from the commencement of operations without breaking down.

(ii) If at the end of 24 hours of operations neither of the 2 components chosen in (i) has broken down, what is the probability that they are both type C components.

*Hint.*

$$\begin{aligned} \text{(i) Required probability} &= \frac{1}{^{11}C_2} [{}^5C_2 \times (0.6)^2 + {}^3C_2 (0.7)^2 + {}^2C_2 (0.3)^2 \\ &\quad + {}^5C_1 \times {}^3C_1 \times 0.6 \times 0.7 + {}^5C_1 \times {}^2C_1 \times (0.6) \times (0.3) \\ &\quad + {}^5C_1 \times {}^1C_1 \times (0.6) \times (0.2) + {}^3C_1 \times {}^2C_1 \times 0.7 \times 0.3 \\ &\quad + {}^3C_1 \times {}^1C_1 \times 0.7 \times 0.2 + {}^2C_1 \times {}^1C_1 \times 0.3 \times 0.2] \\ &= p \text{ (Say).} \end{aligned}$$

(ii) Required probability (By Bayes theorem)

$$= \frac{{}^2C_2 \times (0.3)^2}{p} = \frac{0.09}{p}$$

**4.10. Geometric probability.** In remark 3, § 4.3.1 it was pointed out that the classical definition of probability fails if the total number of outcomes of an experiment is infinite. Thus, for example, if we are interested in finding the

probability that a point selected at random in a given region will lie in a specified part of it, the classical definition of probability is modified and extended to what is called *geometrical probability or probability in continuum*. In this case, the general expression for probability 'p' is given by

$$p = \frac{\text{Measure of specified part of the region}}{\text{Measure of the whole region}}$$

where 'measure' refers to the length, area or volume of the region if we are dealing with one, two or three dimensional space respectively.

**Example 4.34.** Two points are taken at random on the given straight line of length  $a$ . Prove that the probability of their distance exceeding a given length  $c$  ( $c < a$ ) is equal to  $\left(1 - \frac{c}{a}\right)^2$ .

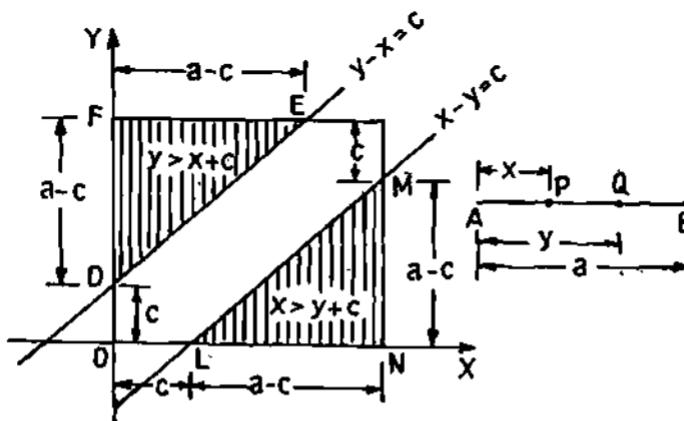
[Burdwan Univ. B.Sc. (Hons.), 1992; Delhi Univ. M.A. (Econ.), 1987]

**Solution.** Let  $P$  and  $Q$  be any two points taken at random on the given straight line  $AB$  of length ' $a$ '. Let  $AP = x$  and  $AQ = y$ ,

$$(0 \leq x \leq a, 0 \leq y \leq a).$$

Then we want  $P(|x - y| > c)$ .

The probability can be easily calculated geometrically. Plotting the lines  $x - y = c$  and  $y - x = c$  along the co-ordinate axes, we get the following diagram:



Since  $0 \leq x \leq a, 0 \leq y \leq a$ , total area  $= a \cdot a = a^2$ .

Area favourable to the event  $|x - y| > c$  is given by

$$\begin{aligned} \Delta LMN + \Delta DEF &= \frac{1}{2} LN \cdot MN + \frac{1}{2} EF \cdot DF \\ &= \frac{1}{2} (a - c)^2 + \frac{1}{2} (a - c)^2 = (a - c)^2 \end{aligned}$$

$$P(|x - y| > c) = \frac{(a - c)^2}{a^2} = \left(1 - \frac{c}{a}\right)^2$$

**Example 4-35. (Bertrand's Problem).** If a chord is taken at random in a circle, what is the chance that its length  $l$  is not less than ' $a$ ', the radius of the circle?

**Solution.** Let the chord  $AB$  make an angle  $\theta$  with the diameter  $OA'$  of the circle with centre  $O$  and radius  $OA=a$ . Obviously  $\theta$  lies between  $-\pi/2$  and  $\pi/2$ . Since all the positions of the chord  $AB$  and consequently all the values of  $\theta$  are equally likely,  $\theta$  may be regarded as a random variable which is uniformly distributed c.f. § 8-1 over  $(-\pi/2, \pi/2)$  with probability density function

$$f(\theta) = \frac{1}{\pi}; -\pi/2 < \theta \leq \pi/2$$

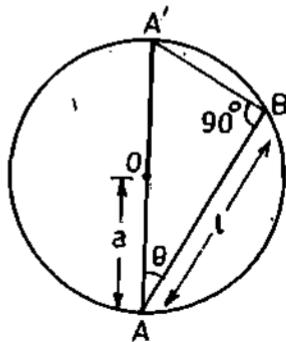
$\angle ABA'$ , being the angle in a semi-circle, is a right angle. From  $\triangle ABA'$  we have

$$\frac{AB}{AA'} = \cos \theta$$

$$\Rightarrow l = 2a \cos \theta$$

The required probability ' $p$ ' is given by

$$\begin{aligned} p &= P(l \geq a) = P(2a \cos \theta \geq a) \\ &= P(\cos \theta \geq 1/2) = P(|\theta| \leq \pi/3) \\ &= \int_{-\pi/3}^{\pi/3} f(\theta) d\theta = \frac{1}{\pi} \int_{-\pi/3}^{\pi/3} d\theta = \frac{2}{3} \end{aligned}$$



**Example 4-36.** A rod of length ' $a$ ' is broken into three parts at random. What is the probability that a triangle can be formed from these parts?

**Solution.** Let the lengths of the three parts of the rod be  $x$ ,  $y$  and  $a - (x + y)$ . Obviously, we have

$$x > 0; y > 0 \text{ and } x + y < a \Rightarrow y < a - x \quad \dots(*)$$

In order that these three parts form the sides of a triangle, we should have

$$\left. \begin{array}{l} x + y > a - (x + y) \Rightarrow y > \frac{a}{2} - x \\ x + a - (x + y) > y \Rightarrow y < \frac{a}{2} \\ y + a - (x + y) > x \Rightarrow y < \frac{a}{2} \end{array} \right\} \quad \dots(**)$$

since in a triangle, the sum of any two sides is greater than the third. Equivalently,  $(**)$  can be written as

$$\frac{a}{2} - x < y < \frac{a}{2} \quad \wedge \quad 0 < x < \frac{a}{2} \quad \dots(***)$$

Hence, on using  $(*)$  and  $(***)$ , the required probability is given by

$$\begin{aligned} \frac{\int_0^{a/2} \int_{(a/2)-x}^{a/2} dy dx}{\int_0^a \int_0^{a-x} dy dx} &= \frac{\int_0^{a/2} \left[ \frac{a}{2} - \left( \frac{a}{2} - x \right) \right] dx}{\int_0^a (a-x) dx} \\ &= \frac{\left| \frac{x^2}{2} \right|_0^{a/2}}{\left| -(a-x)^2 \right|_0^a} = \frac{\frac{a^2/8}{a^2/2}}{\frac{1}{4}} = \frac{1}{4} \end{aligned}$$

**Example 4-37. (Buffon's Needle Problem).** A vertical board is ruled with horizontal parallel lines at constant distance 'a' apart. A needle of length  $l$  ( $< a$ ) is thrown at random on the table. Find the probability that it will intersect one of the lines.

**Solution.** Let  $y$  denote the distance from the centre of the needle to the nearest parallel and  $\phi$  be angle formed by the needle with this parallel. The quantities  $y$  and  $\phi$  fully determine the position of the needle. Obviously  $y$  ranges from 0 to  $a/2$  (since  $l < a$ ) and  $\phi$  from 0 to  $\pi$ .

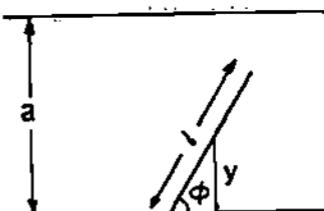
Since the needle is dropped randomly, all possible values of  $y$  and  $\phi$  may be regarded as equally likely and consequently the joint probability density function  $f(y, \phi)$  of  $y$  and  $\phi$  is given by the uniform distribution. (c.f. § 8.1) by

$$f(y, \phi) = k; \quad 0 \leq \phi \leq \pi, \\ 0 \leq y \leq a/2, \quad \dots (*)$$

where  $k$  is a constant.

The needle will intersect one of the lines if the distance of its centre from the line is less than  $\frac{1}{2} l \sin \phi$ , i.e., the required event can be represented by the inequality

$0 < y < \frac{1}{2} l \sin \phi$ . Hence the required probability  $p$  is given by



$$\begin{aligned} p &= \frac{\int_0^{\pi} \int_0^{(l \sin \phi)/2} f(y, \phi) dy d\phi}{\int_0^{\pi} \int_0^{a/2} f(y, \phi) dy d\phi} \\ &= \frac{\frac{l}{2} \int_0^{\pi} \sin \phi d\phi}{(a/2) \cdot \pi} \\ &= \frac{l}{a \pi} \left| -\cos \phi \right|_0^{\pi} = \frac{2l}{a \pi} \end{aligned}$$

### EXERCISE 4 (e)

1. Two points are selected at random in a line  $AC$  of length ' $a$ ' so as to lie on the opposite sides of its mid-point  $O$ . Find the probability that the distance between them is less than  $a/3$ .

2. (a) Two points are selected at random on a line of length  $a$ . What is the probability that none of three sections in which the line is thus divided is less than  $a/4$ ?

Ans. 1/16.

(b) A rectilinear segment  $AB$  is divided by a point  $C$  into two parts  $AC=a$ ,  $CB=b$ . Points  $X$  and  $Y$  are taken at random on  $AC$  and  $CB$  respectively. What is the probability that  $AX$ ,  $XY$  and  $BY$  can form a triangle?

(c)  $ABG$  is a straight line such that  $AB$  is 6 inches and  $BG$  is 5 inches. A point  $Y$  is chosen at random on the  $BG$  part of the line. If  $C$  lies between  $B$  and  $G$  in such a way that  $AC=t$  inches, find

(i) the probability that  $Y$  will lie in  $BC$ .

(ii) the probability that  $Y$  will lie in  $CG$ .

What can you say about the sum of these probabilities?

(d) The sides of a rectangle are taken at random each less than  $a$  and all lengths are equally likely. Find the chance that the diagonal is less than  $a$ .

3. (a) Three points are taken at random on the circumference of a circle. Find the chance that they lie on the same semi-circle.

(b) A chord is drawn at random in a given circle. What is the probability that it is greater than the side of an equilateral triangle inscribed in that circle?

(c) Show that the probability of choosing two points randomly from a line segment of length 2 inches and their being at a distance of at least 1 inch from each other is  $1/4$ . [Delhi Univ. M.A. (Econ.), 1985]

4. A point is selected at random inside a circle. Find the probability that the point is closer to the centre of the circle than to its circumference.

5. One takes at random two points  $P$  and  $Q$  on a segment  $AB$  of length  $a$

(i) What is the probability for the distance  $PQ$  being less than  $b$  ( $< a$ )?

(ii) Find the chance that the distance between them is greater than a given length  $b$ .

6. Two persons  $A$  and  $B$ , make an appointment to meet on a certain day at a certain place, but without fixing the time further than that it is to be between 2 p.m. and 3 p.m. and that each is to wait not longer than ten minutes for the other. Assuming that each is independently equally likely to arrive at any time during the hour, find the probability that they meet.

Third person  $C$ , is to be at the same place from 2.10 p.m. until 2.40 p.m. on the same day. Find the probabilities of  $C$  being present when  $A$  and  $B$  are there together (i) When  $A$  and  $B$  remain after they meet, (ii) When  $A$  and  $B$  leave as soon as they meet.

**Hint.** Denote the times of arrival of  $A$  by  $x$  and of  $B$  by  $y$ . For the meeting to take place it is necessary and sufficient that

$$|x - y| < 10$$

We depict  $x$  and  $y$  as Cartesian coordinates in the plane; for the scale unit we take one minute. All possible outcomes can be described as points of a square with side 60. We shall finally get [c.f. Example 4.34, with  $a = 60$ ,  $c = 10$ ]

$$P(|x - y| < 10) = 1 - (5/6)^2 = 11/36$$

7. The outcome of an experiment are represented by points in the square bounded by  $x = 0$ ,  $x = 2$  and  $y = 2$  in the  $xy$ -plane. If the probability is distributed uniformly, determine the probability that  $x^2 + y^2 > 1$

**Hint.**

$$\text{Required probability } P(E) = \int_E \frac{1}{4} dx dy = 1 - \int_{E'} \frac{1}{4} dx dy$$

where  $E$  is the region for which  $x^2 + y^2 > 1$  and  $E'$  is the region for which  $x^2 + y^2 \leq 1$ .

$$\therefore 4P(E) = 4 - \int_0^1 \int_0^1 dx dy = 3 \quad \Rightarrow \quad P(E) = \frac{3}{4}$$

8. A floor is paved with tiles, each tile being a parallelogram such that the distance between pairs of opposite sides are  $a$  and  $b$  respectively, the length of the diagonal being  $l$ . A stick of length  $c$  falls on the floor parallel to the diagonal. Show that the probability that it will lie entirely on one tile is

$$\left(1 - \frac{c}{l}\right)^2$$

If a circle of diameter  $d$  is thrown on the floor, show that the probability that it will lie on one tile is

$$\left(1 - \frac{d}{a}\right) \left(1 - \frac{d}{b}\right)$$

9. Circular discs of radius  $r$  are thrown at random on to a plane circular table of radius  $R$  which is surrounded by a border of uniform width  $r$  lying in the same plane as the table. If the discs are thrown independently and at random, and  $N$  stay on the table, show that the probability that a fixed point on the table but not on the border, will be covered is

$$1 - \left(1 - \frac{r^2}{(R+r)^2}\right)^N$$

### SOME MISCELLANEOUS EXAMPLES

**Example 4.38.** A die is loaded in such a manner that for  $n=1, 2, 3, 4, 5, 6$ , the probability of the face marked  $n$ , landing on top when the die is rolled is proportional to  $n$ . Find the probability that an odd number will appear on tossing the die.

[Madras Univ. B.Sc. (Stat. Main), 1987]

**Solution.** Here we are given

$P(n) \propto n$  or  $P(n) = kn$ , where  $k$  is the constant of proportionality.  
Also  $P(1) + P(2) + \dots + P(6) = 1 \Rightarrow k(1+2+3+4+5+6) = 1$  or  $k = 1/21$

$$\text{Required Probability} = P(1) + P(3) + P(5) = \frac{1+3+5}{21} = \frac{3}{7}$$

**Example 4.39.** In terms of probability :

$$p_1 = P(A), p_2 = P(B), p_3 = P(A \cap B), (p_1, p_2, p_3 > 0)$$

Express the following in terms of  $p_1, p_2, p_3$ .

$$(a) P(\overline{A} \cup \overline{B}), (b) P(\overline{A} \cup \overline{B}), (c) P(\overline{A} \cap B), (d) P(\overline{A} \cup B), (e) P(\overline{A} \cap \overline{B})$$

$$(f) P(A \cap \overline{B}), (g) P(A|B), (h) P(B|\overline{A}), (i) P[\overline{A} \cap (A \cup B)]$$

**Solution.**

$$(a) P(\overline{A} \cup \overline{B}) = 1 - P(A \cup B) = 1 - [P(A) + P(B) - P(AB)] \\ = 1 - p_1 - p_2 + p_3.$$

$$(b) P(\overline{A} \cup \overline{B}) = P(\overline{A \cap B}) = 1 - P(A \cap B) = 1 - p_3$$

$$(c) P(\overline{A} \cap B) = P(B - AB) = P(B) - P(A \cap B) = p_2 - p_3$$

$$(d) P(\overline{A} \cup B) = P(\overline{A}) + P(B) - P(\overline{A} \cap B) = 1 - p_1 + p_2 - (p_2 - p_3) \\ = 1 - p_1 + p_3$$

$$(e) P(\overline{A} \cap \overline{B}) = P(\overline{A \cup B}) = 1 - p_1 - p_2 + p_3. \quad [\text{Part (a)}]$$

$$(f) P(A \cap \overline{B}) = P(A - A \cap B) = P(A) - P(A \cap B) = p_1 - p_3$$

$$(g) P(A|B) = P(A \cap B)/P(B) = p_3/p_2$$

$$(h) P(B|\overline{A}) = P(\overline{A} \cap B)/P(\overline{A}) = (p_2 - p_3)/(1 - p_1)$$

$$(i) P[\overline{A} \cap (A \cup B)] = P[(\overline{A} \cap A) \cup (\overline{A} \cap B)] \\ = P(\overline{A} \cap B) = p_2 - p_3. \quad [\because A \cap \overline{A} = \emptyset]$$

**Example 4.40.** Let  $P(A) = p, P(A|B) = q, P(B|A) = r$ . Find relations between the numbers  $p, q, r$  for the following cases :

(a) Events  $A$  and  $B$  are mutually exclusive.

(b)  $A$  and  $B$  are mutually exclusive and collectively exhaustive.

(c)  $A$  is a subevent of  $B$ ;  $B$  is a superevent of  $A$ .

(d)  $\overline{A}$  and  $\overline{B}$  are mutually exclusive.

[Delhi Univ. B.Sc. (Maths Hons.) 1985]

**Solution.** From given data :  $P(A) = p, P(A \cap B) = P(\overline{A})P(B|A) = rp$

$$\therefore P(B) = \frac{P(A \cap B)}{P(A|B)} = \frac{rp}{q}$$

$$(a) P(A \cap B) = 0 \Rightarrow rp = 0.$$

$$(b) P(A \cap B) = 0 \text{ and } P(A) + P(B) = 1$$

$$\Rightarrow p(q+r) = q; rp = 0 \Rightarrow pq = q \Rightarrow p = 1 \vee q = 0.$$

$$(c) A \subseteq B \Rightarrow A \cap B = A \text{ or } P(A \cap B) = P(A) \Rightarrow rp = p \Rightarrow r = 1 \vee p = 0.$$

$$B \subseteq A \Rightarrow A \cap B = B \text{ or } P(A \cap B) = P(B)$$

$$\Rightarrow rp = (rp/q) \text{ or } rp(q-1) = 0 \Rightarrow q = 1$$

$$(d) P(\overline{A} \cap \overline{B}) = 1 - P(A \cup B) \Rightarrow 0 = 1 - [P(A) + P(B) - P(A \cap B)]$$

$$\text{So } P(A) + P(B) = 1 + P(A \cap B) \Rightarrow p[1 + (r/q)] = 1 + rp \\ \therefore p(q+r) = q(1+pr).$$

**Example 4-41.** (a) Twelve balls are distributed at random among three boxes. What is the probability that the first box will contain 3 balls?

(b) If  $n$  biscuits be distributed among  $N$  persons, find the chance that a particular person receives  $r$  ( $< n$ ) biscuits. [Marathwada Univ. B.Sc. 1992]

**Solution.** (a) Since each ball can go to any one of the three boxes, there are 3 ways in which a ball can go to any one of the three boxes. Hence there are  $3^{12}$  ways in which 12 balls can be placed in the three boxes.

Number of ways in which 3 balls out of 12 can go to the first box is  ${}^{12}C_3$ . Now the remaining 9 balls are to be placed in 2 boxes and this can be done in  $2^9$  ways. Hence the total number of favourable cases  $= {}^{12}C_3 \times 2^9$ .

$$\therefore \text{Required probability} = \frac{{}^{12}C_3 \times 2^9}{3^{12}}$$

(b) Take any one biscuit. This can be given to any one of the  $N$  beggars so that there are  $N$  ways of distributing any one biscuit. Hence the total number of ways in which  $n$  biscuit can be distributed at random among  $N$  beggars

$$= N \cdot N \dots N (\text{ }n \text{ times}) = N^n.$$

' $r$ ' biscuits can be given to any particular beggar in  ${}^N C_r$  ways. Now we are left with  $(n-r)$  biscuits which are to be distributed among the remaining  $(N-1)$  beggars and this can be done in  $(N-1)^{n-r}$  ways.

$$\therefore \text{Number of favourable cases} = {}^N C_r \cdot (N-1)^{n-r}$$

$$\text{Hence, required probability} = \frac{{}^N C_r (N-1)^{n-r}}{N^n}$$

**Example 4-42.** A car is parked among  $N$  cars in a row, not at either end. On his return the owner finds that exactly  $r$  of the  $N$  places are still occupied. What is the probability that both neighbouring places are empty?

**Solution.** Since the owner finds on return that exactly  $r$  of the  $N$  places (including owner's car) are occupied, the exhaustive number of cases for such an arrangement is  ${}^{N-1}C_{r-1}$  [since the remaining  $r-1$  cars are to be parked in the remaining  $N-1$  places and this can be done in  ${}^{N-1}C_{r-1}$  ways].

Let  $A$  denote the event that both the neighbouring places to owner's car are empty. This requires the remaining  $(r-1)$  cars to be parked in the remaining  $N-3$  places and hence the number of cases favourable to  $A$  is  ${}^{N-3}C_{r-1}$ . Hence

$$P(A) = \frac{{}^{N-3}C_{r-1}}{{}^{N-1}C_{r-1}} = \frac{(N-r)(N-r-1)}{(N-1)(N-2)}$$

**Example 4-43.** What is the probability that at least two out of  $n$  people have the same birthday? Assume 365 days in a year and that all days are equally likely.

**Solution.** Since the birthday of any person can fall on any one of the 365 days, the exhaustive number of cases for the birthdays of  $n$  persons is  $365^n$ .

If the birthdays of all the  $n$  persons fall on different days, then the number of favourable cases is

$$365 \cdot (365 - 1) \cdot (365 - 2) \dots [365 - (n - 1)],$$

because in this case the birthday of the first person can fall on any one of 365 days, the birthday of the second person can fall on any one of the remaining 364 days and so on.

Hence the probability ( $p$ ) that birthdays of all the  $n$  persons are different is given by :

$$\begin{aligned} p &= \frac{365 \cdot (365 - 1) \cdot (365 - 2) \dots [365 - (n - 1)]}{365^n} \\ &= \left(1 - \frac{1}{365}\right) \left(1 - \frac{2}{365}\right) \left(1 - \frac{3}{365}\right) \dots \left(1 - \frac{n-1}{365}\right) \end{aligned}$$

Hence the required probability that at least two persons have the same birthday is

$$1 - p = 1 - \left(1 - \frac{1}{365}\right) \left(1 - \frac{2}{365}\right) \left(1 - \frac{3}{365}\right) \dots \left(1 - \frac{n-1}{365}\right)$$

**Example 4.44.** A five-figure number is formed by the digits 0, 1, 2, 3, 4 (without repetition). Find the probability that the number formed is divisible by 4.

[Delhi Univ. B.Sc. (Stat. Hons.), 1990]

**Solution.** The total number of ways in which the five digits 0, 1, 2, 3, 4 can be arranged among themselves is  $5!$ . Out of these, the number of arrangements which begin with 0 (and, therefore, will give only 4-digit numbers) is  $4!$ . Hence the total number of five digit numbers that can be formed from the digits 0, 1, 2, 3, 4 is

$$5! - 4! = 120 - 24 = 96$$

The number formed will be divisible by 4 if the number formed by the two digits on extreme right (i.e., the digits in the unit and tens places) is divisible by 4. Such numbers are :

$$04, 12, 20, 24, 32, \text{ and } 40$$

If the numbers end in 04, the remaining three digits, viz., 1, 2 and 3 can be arranged among themselves in  $3!$  ways. Similarly, the number of arrangements of the numbers ending with 20 and 40 is  $3!$  in each case.

If the numbers end with 12, the remaining three digits 0, 3, 4 can be arranged in  $3!$  ways. Out of these we shall reject those numbers which start with 0 (i.e., have 0 as the first digit). There are  $(3 - 1)! = 2!$  such cases. Hence, the number of five digit numbers ending with 12 is

$$3! - 2! = 6 - 2 = 4$$

Similarly the number of 5-digit numbers ending with 24 and 32 each is 4. Hence the total number of favourable cases is

$$3 \times 3! + 3 \times 4 = 18 + 12 = 30$$

$$\text{Hence required probability} = \frac{30}{96} = \frac{5}{16}$$

**Example 4.45.** (*Huyghen's problem*). A and B throw alternately with a pair of ordinary dice. A wins if he throws 6 before B throws 7, and B wins if he throws 7 before A throws 6. If A begins, show that his chance of winning is  $\frac{30}{61}$

[Delhi Univ. B.Sc. (Stat. Hons.), 1991; Delhi Univ. B.Sc., 1987]

**Solution.** Let  $E_1$  denote the event of A's throwing '6' and  $E_2$  the event of B's throwing '7' with a pair of dice. Then  $\bar{E}_1$  and  $\bar{E}_2$  are the complementary events.

'6' can be obtained with two dice in the following ways:

(1, 5), (5, 1), (2, 4), (4, 2), (3, 3), i.e., in 5 distinct ways.

$$\therefore P(E_1) = \frac{5}{36} \text{ and } P(\bar{E}_1) = 1 - \frac{5}{36} = \frac{31}{36}$$

'7' can be obtained with two dice as follows:

(1, 6), (6, 1), (2, 5), (5, 2), (3, 4), (4, 3), i.e., in 6 distinct ways.

$$\therefore P(E_2) = \frac{6}{36} = \frac{1}{6} \text{ and } P(\bar{E}_2) = 1 - \frac{1}{6} = \frac{5}{6}$$

If A starts the game, he will win in the following mutually exclusive ways:

(i)  $E_1$  happens (ii)  $\bar{E}_1 \cap \bar{E}_2 \cap E_1$  happens

(iii)  $\bar{E}_1 \cap \bar{E}_2 \cap \bar{E}_1 \cap \bar{E}_2 \cap E_1$  happens, and so on.

Hence by addition theorem of probability, the required probability of A's winning, (say),  $P(A)$  is given by

$$\begin{aligned} P(A) &= P(i) + P(ii) + P(iii) + \dots \\ &= P(E_1) + P(\bar{E}_1 \cap \bar{E}_2 \cap E_1) + P(\bar{E}_1 \cap \bar{E}_2 \cap \bar{E}_1 \cap \bar{E}_2 \cap E_1) + \dots \\ &= P(E_1) + P(\bar{E}_1) P(\bar{E}_2) P(E_1) + P(\bar{E}_1) P(\bar{E}_2) P(\bar{E}_1) P(\bar{E}_2) P(E_1) + \dots \\ &\quad \text{(By compound probability theorem)} \\ &= \frac{5}{36} + \frac{31}{36} \times \frac{5}{6} \times \frac{5}{36} + \frac{31}{36} \times \frac{5}{6} \times \frac{31}{36} \times \frac{5}{6} \times \frac{5}{36} + \dots \\ &= \frac{5/36}{1 - \frac{31}{36} \times \frac{5}{6}} = \frac{30}{61} \end{aligned}$$

**Example 4.46.** A player tosses a coin and is to score one point for every head and two points for every tail turned up. He is to play on until his score reaches or passes  $n$ . If  $p_n$  is the chance of attaining exactly  $n$  score, show that

$$p_n = \frac{1}{2} [p_{n-1} + p_{n-2}],$$

and hence find the value of  $p_n$ .

[Delhi Univ. B.Sc. (Stat. Hons.), 1992]

**Solution.** The score  $n$  can be reached in the following two mutually exclusive ways:

(i) By throwing a tail when score is  $(n - 2)$ , and

(ii) By throwing a head when score is  $(n - 1)$ .

Hence by addition theorem of probability, we get

$$p_n = P(i) + P(ii) = \frac{1}{2} \cdot p_{n-2} + \frac{1}{2} \cdot p_{n-1} = \frac{1}{2} (p_{n-1} + p_{n-2}) \quad \dots(*)$$

To find  $p_n$  explicitly, (\*) may be re-written as

$$p_n + \frac{1}{2} p_{n-1} = p_{n-1} + \frac{1}{2} p_{n-2}$$

$$= p_{n-2} + \frac{1}{2} p_{n-3}$$

...      ...

...      ...

$$= p_2 + \frac{1}{2} p_1$$

$\dots(**)$

Since the score 2 can be obtained as

(i) Head in first throw and head in 2nd throw,

(ii) Tail in the first throw, we have

$$p_2 = \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} = \frac{1}{4} + \frac{1}{2} = \frac{3}{4} \text{ and obviously } p_1 = \frac{1}{2}$$

Hence, from (\*\*), we get

$$\begin{aligned} p_n + \frac{1}{2} p_{n-1} &= \frac{3}{4} + \frac{1}{2} \cdot \frac{1}{2} = 1 = \frac{2}{3} + \frac{1}{3} = \frac{2}{3} + \frac{1}{2} \cdot \frac{2}{3} \\ p_n - \frac{2}{3} &= (-\frac{1}{2}) (p_{n-1} - \frac{2}{3}) \\ p_{n-1} - \frac{2}{3} &= (-\frac{1}{2}) (p_{n-2} - \frac{2}{3}) \\ &\vdots && \vdots \\ p_2 - \frac{2}{3} &= (-\frac{1}{2}) (p_1 - \frac{2}{3}) \end{aligned}$$

Multiplying all the above equations, we get

$$\begin{aligned} p_n - \frac{2}{3} &= (-\frac{1}{2})^{n-1} (p_1 - \frac{2}{3}) \\ &= (-\frac{1}{2})^{n-1} (\frac{1}{2} - \frac{2}{3}) = (-1)^n \cdot \frac{1}{2^n} \cdot \frac{1}{3} \\ \Rightarrow p_n &= \frac{2}{3} + (-1)^n \frac{1}{2^n} \cdot \frac{1}{3} \\ &= \frac{1}{3} \left[ 2 + (-1)^n \frac{1}{2^n} \right] \end{aligned}$$

**Example 4.47.** A coin is tossed  $(m+n)$  times,  $(mn)$ . Show that the probability of at least  $m$  consecutive heads is  $\frac{n+2}{2^{m+1}}$ .

[Kurukshetra Univ. M.Sc. 1990; Calcutta Univ. B.Sc.(Hons.), 1986]

**Solution.** Since  $m > n$ , only one sequence of  $m$  consecutive heads is possible. This sequence may start either with the first toss or second toss or third toss, and so on, the last one will be starting with  $(n+1)$ th toss.

Let  $E_i$  denote the event that the sequence of  $m$  consecutive heads starts with  $i$ th toss. Then the required probability is

$$P(E_1) + P(E_2) + \dots + P(E_{n+1}) \quad \dots(*)$$

Now  $P(E_1) = P[\text{Consecutive heads in first } m \text{ tosses and head or tail in the rest}]$

$$= \left(\frac{1}{2}\right)^m$$

$P(E_2) = P[\text{Tail in the first toss, followed by } m \text{ consecutive heads and head or tail in the next}]$

$$= \frac{1}{2} \left(\frac{1}{2}\right)^m = \frac{1}{2^{m+1}}$$

In general,

$P(E_r) = P[\text{tail in the } (r-1)\text{th trial followed by } m \text{ consecutive heads and head or tail in the next}]$

$$= \frac{1}{2} \left(\frac{1}{2}\right)^m = \frac{1}{2^{m+1}}, \quad \forall r = 2, 3, \dots, n+1.$$

Substituting in (\*),

$$\text{Required probability} = \frac{1}{2^m} + \frac{n}{2^{m+1}} = \frac{2+n}{2^{m+1}}$$

**Example 4.48.** Cards are dealt one by one from a well-shuffled pack until an ace appears. Show that the probability that exactly  $n$  cards are dealt before the first ace appears is

$$\frac{4(51-n)(50-n)(49-n)}{52.51.50.49}$$

[Delhi Univ. B.Sc. 1992]

**Solution.** Let  $E_i$  denote the event that an ace appears when the  $i$ th card is dealt. Then the required probability ' $p$ ' is given by

$$\begin{aligned} p &= P[\text{Exactly } n \text{ cards are dealt before the first ace appears}] \\ &= P[\text{The first ace appears at the } (n+1)\text{th dealing}] \\ &= P(\bar{E}_1 \cap \bar{E}_2 \cap \bar{E}_3 \cap \dots \cap \bar{E}_{n-1} \cap \bar{E}_n \cap E_{n+1}) \\ &= P(\bar{E}_1) P(\bar{E}_2 | \bar{E}_1) P(\bar{E}_3 | \bar{E}_1 \cap \bar{E}_2) \dots \\ &\quad \times P(\bar{E}_n | \bar{E}_1 \cap \bar{E}_2 \cap \dots \cap \bar{E}_{n-1}) \times P(E_{n+1} | \bar{E}_1 \cap \bar{E}_2 \cap \dots \cap \bar{E}_n) \end{aligned} \quad \dots(*)$$

Now

$$P(E_1) = \frac{4}{52} \Rightarrow P(\bar{E}_1) = \frac{48}{52}$$

$$P(E_2 | \bar{E}_1) = \frac{4}{51} \Rightarrow P(\bar{E}_2 | \bar{E}_1) = \frac{47}{51}$$

$$P(E_3 | \bar{E}_1 \cap \bar{E}_2) = \frac{4}{50} \quad \Rightarrow \quad P(\bar{E}_3 | \bar{E}_1 \cap \bar{E}_2) = \frac{46}{50}$$

⋮

$$P(E_{n-1} | \bar{E}_1 \cap \bar{E}_2 \cap \dots \cap \bar{E}_{n-2}) = \frac{4}{52 - (n-2)}$$

$$\therefore P(\bar{E}_{n-1} | \bar{E}_1 \cap \bar{E}_2 \cap \dots \cap \bar{E}_{n-2}) = \frac{50-n}{52-(n-2)}$$

$$P(E_n | \bar{E}_1 \cap \bar{E}_2 \cap \dots \cap \bar{E}_{n-1}) = \frac{4}{52-(n-1)}$$

$$\therefore P(\bar{E}_n | \bar{E}_1 \cap \bar{E}_2 \cap \dots \cap \bar{E}_{n-1}) = \frac{49-n}{52-(n-1)}$$

and,  $P(E_{n+1} | \bar{E}_1 \cap \bar{E}_2 \cap \dots \cap \bar{E}_n) = \frac{4}{52-n}$

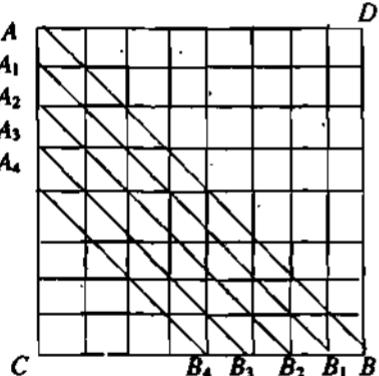
Hence, from (\*) we get

$$P = \left[ \frac{48}{52} \times \frac{47}{51} \times \frac{46}{50} \times \frac{45}{49} \times \frac{44}{48} \times \frac{43}{47} \times \dots \times \frac{52-n}{52-(n-4)} \right. \\ \left. \times \frac{51-n}{52-(n-3)} \times \frac{50-n}{52-(n-2)} \times \frac{49-n}{52-(n-1)} \times \frac{4}{52-n} \right] \\ = \frac{(51-n)(50-n)(49-n)4}{52 \times 51 \times 50 \times 49}$$

**Example 4.49.** If four squares are chosen at random on a chess-board, find the chance that they should be in a diagonal line.

[Delhi Univ. B.Sc. (Stat. Hons.), 1988]

**Solution.** In a chess-board there are  $8 \times 8 = 64$  squares as shown in the following diagram.



Let us consider the number of ways in which the 4 squares selected at random are in a diagonal line parallel to  $AB$ . Consider the  $\Delta ABC$ . Number of ways in which 4 selected squares are along the lines  $A_4B_4$ ,  $A_3B_3$ ,  $A_2B_2$ ,  $A_1B_1$  and  $AB$  are  ${}^4C_4$ ,  ${}^5C_4$ ,  ${}^6C_4$ ,  ${}^7C_4$  and  ${}^8C_4$  respectively.

Similarly, in  $\Delta ABD$  there are an equal number of ways of selecting 4 squares in a diagonal line parallel to  $AB$ .

Hence, total number of ways in which the 4 selected squares are in a diagonal line parallel to  $AB$  are  $2({}^4C_4 + {}^5C_4 + {}^6C_4 + {}^7C_4) + {}^8C_4$ .

Since there is an equal number of ways in which 4 selected squares are in a diagonal line parallel to  $CD$ , the required number of favourable cases is given by

$$2 [ 2( {}^4C_4 + {}^5C_4 + {}^6C_4 + {}^7C_4 ) + {}^8C_4 ]$$

Since 4 squares can be selected out of 64 in  ${}^64C_4$  ways, the required probability is

$$\begin{aligned} &= \frac{2 [ 2( {}^4C_4 + {}^5C_4 + {}^6C_4 + {}^7C_4 ) + {}^8C_4 ]}{{}^64C_4} \\ &= \frac{[ 4 ( 1 + 5 + 15 + 35 ) + 140 ] \times 4!}{64 \times 63 \times 62 \times 61} = \frac{91}{158844} \end{aligned}$$

**Example 4.50.** An urn contains four tickets marked with numbers 112, 121, 211, 222 and one ticket is drawn at random. Let  $A_i$ , ( $i=1, 2, 3$ ) be the event that  $i$ th digit of the number of the ticket drawn is 1. Discuss the independence of the events  $A_1, A_2$  and  $A_3$ . [Delhi Univ. B.Sc.(Stat. Hons.), 1987; Poona Univ. B.Sc., 1986]

**Solution.** We have

$$P(A_1) = \frac{2}{4} = \frac{1}{2} = P(A_2) = P(A_3)$$

$A_1 \cap A_2$  is the event that the first two digits in the number which the selected ticket bears are each equal to unity and the only favourable case is ticket with number 112.

$$\therefore P(A_1 \cap A_2) = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = P(A_1) P(A_2)$$

Similarly,

$$P(A_2 \cap A_3) = \frac{1}{4} = P(A_2) P(A_3)$$

$$\text{and } P(A_3 \cap A_1) = \frac{1}{4} = P(A_3) P(A_1)$$

Thus we conclude that the events  $A_1, A_2$  and  $A_3$  are pairwise independent.

Now  $P(A_1 \cap A_2 \cap A_3) = P\{\text{all the three digits in the number are 1's}\}$

$$\begin{aligned} &= P(\emptyset) \\ &= 0 \neq P(A_1) P(A_2) P(A_3) \end{aligned}$$

Hence  $A_1, A_2$  and  $A_3$  though pairwise independent are not mutually independent.

**Example 4.51.** Two fair dice are thrown independently. Three events  $A, B$  and  $C$  are defined as follows:

$A$  : Odd face with first dice

$B$  : Odd face with second dice

$C$  : Sum of points on two dice is odd.

Are the events  $A, B$  and  $C$  mutually independent?

[Delhi Univ. B.Sc. (Stat. Hons.) 1983; M.S. Baroda Univ. B.Sc. 1987]

**Solution.** Since each of the two dice can show any one of the six faces 1, 2, 3, 4, 5, 6, we get :

$$P(A) = \frac{3 \times 6}{36} = \frac{1}{2} \quad [\because A = \{1, 3, 5\} \times \{1, 2, 3, 4, 5, 6\}]$$

$$P(B) = \frac{3 \times 6}{36} = \frac{1}{2} \quad [\because B = \{1, 2, 3, 4, 5, 6\} \times \{1, 3, 5\}]$$

The sum of points on two dice will be odd if one shows odd number and the other shows even number. Hence favourable cases for C are :

$$(1, 2), (1, 4), (1, 6); \quad (4, 1), (4, 3), (4, 5)$$

$$(2, 1), (2, 3), (2, 5); \quad (5, 2), (5, 4), (5, 6)$$

$$(3, 2), (3, 4), (3, 6); \quad (6, 1), (6, 3), (6, 5)$$

i.e., 18 cases in all.

$$\text{Hence } P(C) = \frac{18}{36} = \frac{1}{2}$$

Cases favourable to the events  $A \cap B$ ,  $A \cap C$ ,  $B \cap C$  and  $A \cap B \cap C$  are given below :

Event	Favourable cases
$A \cap B$	(1,1), (1, 3), (1, 5), (3, 1), (3, 3), (3, 5), (5, 1) (5, 3) (5, 5), i.e., 9 in all.
$A \cap C$	(1, 2), (1, 4), (1, 6), (3, 2), (3, 4), (3, 6), (5, 2), (5, 4) (5, 6), i.e., 9 in all.
$B \cap C$	(2, 1), (4, 1), (6, 1) (2, 3), (4, 3), (6, 3), (2, 5), (4, 5), (6, 5), i.e., 9 in all
$A \cap B \cap C$	Nil, because $A \cap B$ implies that sum of points on two dice is even and hence $(A \cap B) \cap C = \emptyset$

$$\therefore P(A \cap B) = \frac{9}{36} = \frac{1}{4} = P(A) \cdot P(B)$$

$$P(A \cap C) = \frac{9}{36} = \frac{1}{4} = P(A) \cdot P(C)$$

$$P(B \cap C) = \frac{9}{36} = \frac{1}{4} = P(B) \cdot P(C)$$

and  $P(A \cap B \cap C) = P(\emptyset) = 0 \neq P(A) \cdot P(B) \cdot P(C)$

Hence the events A, B and C are pairwise independent but not mutually independent.

**Example 4-52.** Let  $A_1, A_2, \dots, A_n$  be independent events and  $P(A_k) = p_k$ . Further, let  $p$  be the probability that none of the events occurs; then show that

$$p \leq e^{-\sum p_k}$$

[Agra Univ. M.Sc., 1987]

**Solution.** We have

$$\begin{aligned}
 p &= P(\bar{A}_1 \cap \bar{A}_2 \cap \dots \cap \bar{A}_n) \\
 &= \prod_{i=1}^n P(\bar{A}_i) = \prod_{i=1}^n [1 - P(A_i)] = \prod_{i=1}^n (1 - p_i) \\
 &\quad [\text{since } A_i \text{'s are independent}] \\
 &\leq \prod_{i=1}^n e^{-p_i} \quad [\because 1-x \leq e^{-x} \text{ for } 0 \leq x \leq 1 \\
 &\quad \text{and } 0 \leq p_i \leq 1] \\
 \Rightarrow p &\leq \exp \left[ - \sum_{i=1}^n p_i \right],
 \end{aligned}$$

as desired.

**Remark.** We have

$$1-x \leq e^{-x} \text{ for } 0 \leq x \leq 1 \quad \dots(*)$$

**Proof.** The inequality (\*) is obvious for  $x=0$  and  $x=1$ . Consider  $0 < x < 1$ . Then

$$\begin{aligned}
 \log(1-x)^{-1} &= -\log(1-x) \\
 &= \left[ x + \frac{x^2}{2} + \frac{x^3}{3} + \frac{x^4}{4} + \dots \right], \quad \dots(**)
 \end{aligned}$$

the expansion being valid since  $0 < x < 1$ . Further since  $x > 0$ , we get from (\*\*)

$$\begin{aligned}
 \log(1-x)^{-1} &> x \\
 \Rightarrow -\log(1-x) &> x \\
 \Rightarrow \log(1-x) &< -x \\
 \Rightarrow 1-x &< e^{-x},
 \end{aligned}$$

as desired.

**Example 4.53.** In the following Fig.(a) and (b) assume that the probability of a relay being closed is  $P$  and that a relay is open or closed independently of any other. In each case find the probability that current flows from  $L$  to  $R$ .

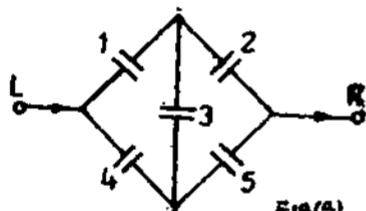


Fig.(a)

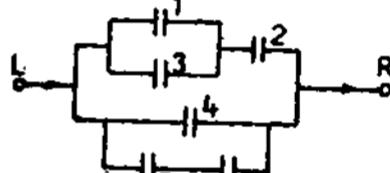


Fig.(b)

**Solution.** Let  $A_i$  denote the event that the relay  $i$ , ( $i = 1, 2, \dots, 6$ ) is closed. Let  $E$  be the event that current flows from  $L$  to  $R$ .

In Fig. (a) the current will flow from  $L$  to  $R$  if at least one of the circuits from  $L$  to  $R$  is closed. Thus for the current to flow from  $L$  to  $R$  we have the following favourable cases:

$$(i) A_1 \cap A_2 = B_1, \quad (ii) A_4 \cap A_5 = B_2,$$

$$(iii) A_1 \cap A_3 \cap A_5 = B_3, \quad (iv) A_4 \cap A_3 \cap A_2 = B_4.$$

The probability  $p_1$  that current flows from  $L$  to  $R$  is given by

$$p_1 = P(B_1 \cup B_2 \cup B_3 \cup B_4) = \sum_i P(B_i) - \sum_{i < j} P(B_i \cap B_j) + \sum_{i < j < k} P(B_i \cap B_j \cap B_k) \\ - P(B_1 \cap B_2 \cap B_3 \cap B_4) \quad \dots (*)$$

Since the relays operate independently of each other, we have

$$P(B_1) = P(A_1 \cap A_2) = P(A_1) \cdot P(A_2) = p \cdot p = p^2$$

$$P(B_2) = P(A_4 \cap A_5) = P(A_4) \cdot P(A_5) = p \cdot p = p^2$$

$$P(B_3) = P(A_1) P(A_3) P(A_5) = p^3$$

$$P(B_4) = P(A_4) P(A_3) P(A_2) = p^3$$

Similarly

$$P(B_1 \cap B_2) = P(A_1 \cap A_2 \cap A_4 \cap A_5) = P(A_1) P(A_2) P(A_4) P(A_5) = p^4$$

$$P(B_1 \cap B_2 \cap B_3) = P(A_1 \cap A_2 \cap A_3 \cap A_4 \cap A_5) = p^5$$

and so on. Finally, substituting in (\*), we get

$$p_1 = (p^2 + p^2 + p^3 + p^3) - (p^4 + p^4 + p^4 + p^4 + p^3) \\ + (p^5 + p^5 + p^5 + p^5) - p^5 \\ = 2p^2 + 2p^3 - 5p^4 + 2p^5$$

In Fig. (b). Arguing as in the above case, the required probability  $p_2$  that the current flows from  $L$  to  $R$  is given by

$$p_2 = P(E_1 \cup E_2 \cup E_3 \cup E_4)$$

where

$$E_1 = A_1 \cap A_2, E_2 = A_3 \cap A_2, E_3 = A_4, E_4 = A_5 \cap A_6$$

$$\therefore p_2 = \sum_i P(E_i) - \sum_{i < j} P(E_i \cap E_j) + \sum_{i < j < k} P(E_i \cap E_j \cap E_k) \\ - P(E_1 \cap E_2 \cap E_3 \cap E_4)$$

$$= (p^2 + p^2 + p + p^2) - (p^3 + p^3 + p^4 + p^3 + p^3) \\ + (p^4 + p^5 + p^5 + p^5) - p^6$$

$$= p + 3p^2 - 4p^3 - p^4 + 3p^5 - p^6$$

**Matching Problem.** Let us have  $n$  letters corresponding to which there exist  $n$  envelopes bearing different addresses. Considering various letters being put in various envelopes, a *match* is said to occur if a letter goes into the right envelope. (Alternatively, if in a party there are  $n$  persons with  $n$  different hats, a *match* is said to occur if in the process of selecting hats at random, the  $i$ th person rightly gets the  $i$ th hat.)

A match at the  $k$ th position for  $k=1, 2, \dots, n$ . Let us first consider the event  $A_k$  when a match occurs at the  $k$ th place. For better understanding let us put the envelopes bearing numbers 1, 2, ...,  $n$  in ascending order. When  $A_k$  occurs,  $k$  th

letter goes to the  $k$ th envelope but  $(n - 1)$  letters can go to the remaining  $(n - 1)$  envelopes in  $(n - 1)!$  ways.

$$\text{Hence } P(A_k) = \frac{(n-1)!}{n!} = \frac{1}{n},$$

where  $P(A_k)$  denotes the probability of the  $k$ th match. It is interesting to see that  $P(A_k)$  does not depend on  $k$ .

**Example 4.54.** (a) 'n' different objects 1, 2, ..., n are distributed at random in n places marked 1, 2, ..., n. Find the probability that none of the objects occupies the place corresponding to its number. [Calcutta Univ. B.A.(Stat.Hons.) 1986;

Delhi Univ. B.Sc.(Maths Hons.), 1990; B.Sc.(Stat.Hons.) 1988]

(b) If n letters are randomly placed in correctly addressed envelopes, prove that the probability that exactly r letters are placed in correct envelopes is given by

$$\frac{1}{r!} \sum_{k=0}^{n-r} (-1)^k \frac{1}{k!}; \quad r = 1, 2, \dots, n$$

[Bangalore Univ. B.Sc., 1987]

**Solution (Probability of no match).** Let  $E_i$ , ( $i = 1, 2, \dots, n$ ) denote the event that the  $i$ th object occupies the place corresponding to its number so that  $\bar{E}_i$  is the complementary event. Then the probability 'p' that none of the objects occupies the place corresponding to its number is given by

$$\begin{aligned} p &= P(\bar{E}_1 \cap \bar{E}_2 \cap \bar{E}_3 \cap \dots \cap \bar{E}_n) \\ &= 1 - P(\text{at least one of the objects occupies the place corresponding to its number}) \\ &= 1 - P(E_1 \cup E_2 \cup E_3 \cup \dots \cup E_n) \\ &= 1 - \left[ \sum_{i=1}^n P(E_i) - \sum_{\substack{i,j=1 \\ i < j}}^n P(E_i \cap E_j) + \sum_{\substack{i,j,k=1 \\ i < j < k}}^n P(E_i \cap E_j \cap E_k) - \dots \right. \\ &\quad \left. + (-1)^{n-1} P(E_1 \cap E_2 \cap \dots \cap E_n) \right] \quad \dots(*) \end{aligned}$$

$$\text{Now } P(E_i) = \frac{1}{n}, \quad \forall \quad i$$

$$\begin{aligned} P(E_i \cap E_j) &= P(E_i) P(E_j | E_i) \\ &= \frac{1}{n} \cdot \frac{1}{n-1}, \quad \forall \quad i, j \quad (i < j) \end{aligned}$$

$$\begin{aligned} P(E_i \cap E_j \cap E_k) &= P(E_i) P(E_j | E_i) P(E_k | E_i \cap E_j) \\ &= \frac{1}{n} \cdot \frac{1}{n-1} \cdot \frac{1}{n-2}, \quad \forall \quad i, j, k \quad (i < j < k) \end{aligned}$$

and so on. Finally,

$$P(E_1 \cap E_2 \cap E_3 \cap \dots \cap E_n) = \frac{1}{n} \cdot \frac{1}{n-1} \cdot \frac{1}{n-2} \cdots \frac{1}{2} \cdot 1$$

Substituting in (\*), we get

$$\begin{aligned}
 p &= 1 - \left[ {}^n C_1 \frac{1}{n} - {}^n C_2 \frac{1}{n(n-1)} + {}^n C_3 \frac{1}{n(n-1)(n-2)} - \dots \right. \\
 &\quad \left. + (-1)^{n-1} \frac{1}{n(n-1)\dots 3 \cdot 2 \cdot 1} \right] \\
 &= 1 - \left[ 1 - \frac{1}{2!} + \frac{1}{3!} - \dots + (-1)^{n-1} \frac{1}{n!} \right] \\
 &= \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \dots + (-1)^n \frac{1}{n!} \\
 &= \sum_{k=0}^n \frac{(-1)^k}{k!}
 \end{aligned}$$

**Remark.** For large  $n$ ,

$$\begin{aligned}
 p &= 1 - 1 + \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \dots \\
 &= e^{-1} = 0.36787
 \end{aligned}$$

Hence the probability of at least one match is

$$\begin{aligned}
 1-p &= 1 - \frac{1}{2!} + \frac{1}{3!} - \dots + \frac{(-1)^n}{n!} \\
 &= 1 - \frac{1}{e}, \text{ (for large } n\text{)}
 \end{aligned}$$

(b) [Probability of exactly  $r$  matches  $\{r \leq (n-2)\}$ ] Let  $A_i$ , ( $i = 1, 2, \dots, n$ ) denote the event that  $i$ th letter goes to the correct envelope. Then the probability that none of the  $n$  letters goes to the correct envelope is

$$P(\bar{A}_1 \cap \bar{A}_2 \cap \dots \cap \bar{A}_n) = \sum_{k=0}^n \frac{(-1)^k}{k!} \quad \dots (**) \text{ [(c.f. part (a)]}$$

The probability that each of the ' $r$ ' letters is in the right envelope is

$\frac{1}{n(n-1)(n-2)\dots(n-r+1)}$ , and the probability that none of the remaining  $(n-r)$  letters goes in the correct envelope is obtained by replacing  $n$  by  $(n-r)$  in  $(**)$  and is thus given by  $\sum_{k=0}^{n-r} \frac{(-1)^k}{k!}$ . Hence by compound probability theorem,

the probability that out of  $n$  letters exactly  $r$  letters go to correct envelopes, (in a specified order), is

$$\frac{1}{n(n-1)(n-2)\dots(n-r+1)} \sum_{k=0}^{n-r} \frac{(-1)^k}{k!}; \quad r \leq n-2.$$

Since  $r$  letters can go to  $n$  envelopes in  ${}^n C_r$  mutually exclusive ways, the required probability of exactly  $r$  letters going to correct envelopes, (in any order, whatsoever), is given by

$${}^n C_r \times \frac{1}{n(n-1)(n-2) \dots (n-r+1)} \sum_{k=0}^{n-r} \frac{(-1)^k}{k!} = \frac{1}{r!} \sum_{k=0}^{n-r} (-1)^k \frac{1}{k!}$$

**Example 4.55.** Each of the  $n$  urns contains ' $a$ ' white balls and ' $b$ ' black balls. One ball is transferred from the first urn to the second, then one ball from the latter into the third, and so on. If  $p_k$  is the probability of drawing a white ball from the  $k$ th urn, show that

$$p_{k+1} = \frac{a+1}{a+b+1} p_k + \frac{a}{a+b+1} (1-p_k)$$

Hence for the last urn, prove that

$$p_n = \frac{a}{a+b} \quad [\text{Punjab Univ., B.Sc.(Maths Hons.), 1988}]$$

**Solution.** The event of drawing a white ball from the  $k$ th urn can materialise in the following two ways:

- (i) The ball transferred from the  $(k-1)$ th urn is white and then a white ball is drawn from the  $k$ th urn.
- (ii) The ball transferred from the  $(k-1)$ th urn is black and then a white ball is drawn from the  $k$ th urn.

The probability of case (i) is  $p_{k-1} \times \frac{a+1}{a+b+1}$ ,

since the probability of drawing a white ball from the  $(k-1)$ th urn is  $p_{k-1}$  and then the probability of drawing white ball from the  $k$ th urn is

$$\frac{a+1}{a+b+1}.$$

Since the probability of drawing a black ball from the  $(k-1)$ th urn is  $[1-p_{k-1}]$  and then the probability of drawing a white ball from the  $k$ th urn is

$$\frac{a}{a+b+1},$$

the probability of case (ii) is given by

$$\frac{a}{a+b+1} [1-p_{k-1}]$$

Since the cases (i) and (ii) are mutually exclusive, we have by addition theorem of probability

$$p_k = \frac{a+1}{a+b+1} p_{k-1} + \frac{a}{a+b+1} [1-p_{k-1}] \quad \dots (*)$$

$$\therefore p_k = \frac{1}{a+b+1} p_{k-1} + \frac{a}{a+b+1} \quad \dots (1)$$

Replacing  $k$  by  $k+1$  in (\*) we get the required result.

Changing  $k$  to  $k-1, k-2, \dots$  and so on, we get

$$p_{k+1} = \frac{1}{a+b+1} p_{k-2} + \frac{a}{a+b+1} \quad \dots (2)$$

$$p_{k-2} = \frac{1}{a+b+1} p_{k-3} + \frac{a}{a+b+1} \quad \dots(3)$$

$$\begin{array}{ccccccc} & \cdot & & \cdot & & \cdot & \\ & \cdot & & \cdot & & \cdot & \\ & \cdot & & \cdot & & \cdot & \\ & \cdot & & \cdot & & \cdot & \\ p_2 = \frac{1}{a+b+1} p_1 + \frac{a}{a+b+1} & & & & & & \dots(k-1) \end{array}$$

But  $p_1$  = Probability of drawing a white ball from the first urn =  $\frac{a}{a+b}$ .

Multiplying (1) by 1, (2) by  $\frac{1}{a+b+1}$ , (3) by  $\left(\frac{1}{a+b+1}\right)^2$ , ..., and  $(k-1)$ th equation by  $\left(\frac{1}{a+b+1}\right)^{k-2}$  and adding, we get

$$\begin{aligned} p_k &= \left(\frac{1}{a+b+1}\right)^{k-1} p_1 + \frac{a}{a+b+1} \left[ 1 + \frac{1}{a+b+1} + \frac{1}{(a+b+1)^2} + \dots \right. \\ &\quad \left. + \left(\frac{1}{a+b+1}\right)^{k-2} \right] \\ &= \left(\frac{1}{a+b+1}\right)^{k-1} \times \frac{a}{(a+b)} + \frac{a}{a+b+1} \left[ \frac{1 - \left(\frac{1}{a+b+1}\right)^{k-1}}{\left(1 - \frac{1}{a+b+1}\right)} \right] \\ &= \frac{a}{a+b} \left(\frac{1}{a+b+1}\right)^{k-1} + \frac{a}{a+b} \left[ 1 - \left(\frac{1}{a+b+1}\right)^{k-1} \right] \\ &= \frac{a}{a+b} \left[ \left(\frac{1}{a+b+1}\right)^{k-1} + \left\{ 1 - \left(\frac{1}{a+b+1}\right)^{k-1} \right\} \right] \\ &= \frac{a}{a+b}, \quad (k=1, 2, \dots, n) \end{aligned}$$

Since the probability of drawing a white ball from the  $k$ th urn is independent of  $k$ , we have

$$p_n = \frac{a}{a+b}.$$

**Example 4-56.** (i) Let the probability  $p_n$  that a family has exactly  $n$  children be  $\alpha p^n$  when  $n \geq 1$  and  $p_0 = 1 - \alpha p (1 + p + p^2 + \dots)$ . Suppose that all sex distributions of  $n$  children have the same probability. Show that for  $k \geq 1$ , the probability that a family contains exactly  $k$  boys is  $2 \alpha \cdot p^k / (2-p)^{k+1}$ .

(ii) Given that a family includes at least one boy, show that the probability that there are two or more boys is  $p/(2-p)$ .

**Solution.** We are given

$$\begin{aligned} p_n &= P[\text{that a family has exactly } n \text{ children}] \\ &= \alpha p^n, \quad n \geq 1. \end{aligned}$$

and  $p_0 = 1 - \alpha p(1 + p + p^2 + \dots)$

Let  $E_j$  be the event that the number of children in a family is  $j$  and let  $A$  be the event that a family contains exactly  $k$  boys. Then

$$P(E_j) = p_j, \quad j = 0, 1, 2, \dots$$

Now, since each child can have any of the two sex distributions (either boy or girl), the total number of possible distributions for a family to have ' $j$ ' children is  $2^j$ .

$$\therefore P(A|E_j) = \frac{jC_k}{2^j}, \quad j \geq k$$

$$\begin{aligned} \text{and } P(A) &= \sum_{j=k}^{\infty} P(E_j) P(A|E_j) = \sum_{j=k}^{\infty} p_j P(A|E_j) \\ &= \sum_{j=k}^{\infty} \alpha p^j \left[ \frac{jC_k}{2^j} \right], \quad j \geq k \geq 1 \\ &= \alpha \sum_{j=k}^{\infty} \left( \frac{p}{2} \right)^j jC_k \\ &= \alpha \sum_{r=0}^{\infty} {}^{k+r}C_k \left( \frac{p}{2} \right)^{k+r} \quad [\text{Put } j-k=r] \\ &= \alpha \left( \frac{p}{2} \right)^k \sum_{r=0}^{\infty} {}^{k+r}C_r \left( \frac{p}{2} \right)^r \quad [{}^nC_r = {}^nC_{n-r}] \end{aligned}$$

We know that

$$\begin{aligned} {}^nC_r &= (-1)^r \cdot {}^{n+r-1}C_r \Rightarrow (-1)^r \cdot {}^nC_r = {}^{n+r-1}C_r, \\ &\therefore (-1)^r \cdot {}^{-(k+1)}C_r = {}^{k+r}C_r. \end{aligned}$$

Hence

$$\begin{aligned} P(A) &= \alpha \left( \frac{p}{2} \right)^k \sum_{r=0}^{\infty} (-1)^r \cdot {}^{-(k+1)}C_r \cdot \left( \frac{p}{2} \right)^r \\ &= \alpha \left( \frac{p}{2} \right)^k \sum_{r=0}^{\infty} {}^{-(k+1)}C_r \left( \frac{p}{2} \right)^r \\ &= \alpha \left( \frac{p}{2} \right)^k \left( 1 - \frac{p}{2} \right)^{(k+1)} \\ &= \alpha \left( \frac{p}{2} \right)^k \frac{2^{k+1}}{(2-p)^{k+1}} = \frac{2 \alpha p^k}{(2-p)^{k+1}}. \end{aligned}$$

(b) Let  $B$  denote the event that a family includes at least one boy and  $C$  denote the event that a family has two or more boys. Then

$$\begin{aligned}
 P(B) &= \sum_{k=1}^{\infty} P[\text{family has exactly } k \text{ boys}] \\
 &= \sum_{k=1}^{\infty} \frac{2\alpha p^k}{(2-p)^{k+1}} = \frac{2\alpha}{2-p} \sum_{k=1}^{\infty} \left(\frac{p}{2-p}\right)^k \\
 &= \frac{2\alpha}{2-p} \times \frac{p/(2-p)}{1-[p/(2-p)]} = \frac{\alpha p}{(1-p)(2-p)}
 \end{aligned}$$

$$\begin{aligned}
 P(C) &= \sum_{k=2}^{\infty} P[\text{family has exactly } k \text{ boys}] \\
 &= \sum_{k=2}^{\infty} \frac{2\alpha p^k}{(2-p)^{k+1}} = \frac{2\alpha}{2-p} \sum_{k=2}^{\infty} \left(\frac{p}{2-p}\right)^k \\
 &= \frac{2\alpha}{2-p} \cdot \frac{[p/(2-p)]^2}{1-[p/(2-p)]} = \frac{\alpha p^2}{(2-p)^2(1-p)}
 \end{aligned}$$

Since  $C \subset B$  and  $B \cap C = C$ ,  $P(B \cap C) = P(C) \Rightarrow P(B)P(C|B) = P(C)$   
Therefore,

$$P(C|B) = \frac{P(C)}{P(B)} = \frac{\alpha p^2}{(2-p)^2(1-p)} \times \frac{(1-p)(2-p)}{\alpha p} = \frac{p}{2-p}$$

**Example 4.57.** A slip of paper is given to person A who marks it either with a plus sign or a minus sign; the probability of his writing a plus sign is  $1/3$ . A passes the slip to B, who may either leave it alone or change the sign before passing it to C. Next C passes the slip to D after perhaps changing the sign. Finally D passes it to a referee after perhaps changing the sign. The referee sees a plus sign on the slip. It is known that B, C and D each change the sign with probability  $2/3$ . Find the probability that A originally wrote a plus.

**Solution.** Let us define the following events :

$E_1$  : A wrote a plus sign;  $E_2$  : A wrote a minus sign

$E$  : The referee observes a plus sign on the slip.

We are given :  $P(E_1) = 1/3$ ,  $P(E_2) = 1 - 1/3 = 2/3$

We want  $P(E_1|E)$ , which by Bayes' rule is given by :

$$P(E_1|E) = \frac{P(E_1)P(E|E_1)}{P(E_1)P(E|E_1) + P(E_2)P(E|E_2)} \quad \dots(i)$$

$P(E|E_1) = P[\text{Referee observes the plus sign given that 'A' wrote the plus sign on the slip}]$

$= P[(\text{Plus sign was not changed at all}) \cup (\text{Plus sign was changed exactly twice in passing from 'A' to referee through B, C and D})]$

$= P(\hat{E}_3 \cup \hat{E}_4)$ , (say).

$$= P(E_3) + P(E_4) \quad \dots(ii)$$

Let  $A_1, A_2$  and  $A_3$  respectively denote the events that  $B, C$  and  $D$  change the sign on the slip. Then we are given

$$P(A_1) = P(A_2) = P(A_3) = 2/3 ; \quad P(\bar{A}_1) = P(\bar{A}_2) = P(\bar{A}_3) = 1/3$$

We have

$$\begin{aligned} P(E_3) &= P(\bar{A}_1 \cap \bar{A}_2 \cap \bar{A}_3) = P(\bar{A}_1) P(\bar{A}_2) P(\bar{A}_3) = (1/3)^3 = 1/27 \\ P(E_4) &= P[(A_1 A_2 \bar{A}_3) \cup (A_1 \bar{A}_2 A_3) \cup (\bar{A}_1 A_2 A_3)] \\ &= P(A_1 A_2 \bar{A}_3) + P(A_1 \bar{A}_2 A_3) + P(\bar{A}_1 A_2 A_3) \\ &= P(A_1) P(A_2) P(\bar{A}_3) + P(A_1) P(\bar{A}_2) P(A_3) + P(\bar{A}_1) P(A_2) P(A_3) \\ &= \frac{2}{3} \cdot \frac{2}{3} \cdot \frac{1}{3} + \frac{2}{3} \cdot \frac{1}{3} \cdot \frac{2}{3} + \frac{1}{3} \cdot \frac{2}{3} \cdot \frac{2}{3} = \frac{4}{9} \end{aligned}$$

Substituting in (ii) we get

$$P(E | E_1) = \frac{1}{27} + \frac{4}{9} = \frac{13}{27} \quad \dots(iii)$$

Similarly,

$P(E | E_2) = P$  [Referee observes the plus sign given that 'A' wrote minus sign on the slip]

$$\begin{aligned} &= P[(\text{Minus sign was changed exactly once}) \\ &\quad \cup (\text{Minus sign was changed thrice})] \end{aligned}$$

$$= P(E_5 \cup E_6), \text{(say)},$$

$$= P(E_5) + P(E_6) \quad \dots(iv)$$

$$\begin{aligned} P(E_5) &= P[(A_1 \bar{A}_2 \bar{A}_3) \cup (\bar{A}_1 A_2 \bar{A}_3) \cup (\bar{A}_1 \bar{A}_2 A_3)] \\ &= P(A_1) P(\bar{A}_2) P(\bar{A}_3) + P(\bar{A}_1) P(A_2) P(\bar{A}_3) + P(\bar{A}_1) P(\bar{A}_2) P(A_3) \\ &= \frac{2}{3} \cdot \frac{1}{3} \cdot \frac{1}{3} + \frac{1}{3} \cdot \frac{2}{3} \cdot \frac{1}{3} + \frac{1}{3} \cdot \frac{1}{3} \cdot \frac{2}{3} = \frac{2}{9} \end{aligned}$$

$$P(E_6) = P(A_1 A_2 A_3) = P(A_1) P(A_2) P(A_3) = \frac{2}{3} \cdot \frac{2}{3} \cdot \frac{2}{3} = \frac{8}{27}$$

Substituting in (iv) we get :

$$P(E | E_2) = \frac{2}{9} + \frac{8}{27} = \frac{14}{27} \quad \dots(v)$$

Substituting from (iii) and (v) in (i) we get :

$$P(E_1 | E) = \frac{\frac{1}{3} \times \frac{13}{27}}{\frac{1}{3} \times \frac{13}{27} + \frac{2}{3} \times \frac{14}{27}} = \frac{13}{13 + 28} = \frac{13}{41}$$

**Example 4-58.** Three urns of the same appearance have the following proportion of balls.

First urn	:	2 black	1 white
Second Urn	:	1 black	2 white
Third urn	:	2 black	2 white

One of the urns is selected and one ball is drawn. It turns out to be white. What is the probability of drawing a white ball again, the first one not having been returned?

**Solution.** Let us define the events:

$E_i$  = The event of selection of  $i$ th urn, ( $i = 1, 2, 3$ )

and  $A$  = The event of drawing a white ball.

Then

$$P(E_1) = P(E_2) = P(E_3) = 1/3$$

$$\text{and } P(A|E_1) = 1/3, P(A|E_2) = 2/3 \text{ and } P(A|E_3) = 1/2$$

Let  $C$  denote the future event of drawing another white ball from the urns.

Then

$$\begin{aligned} P(C|E_1 \cap A) &= 0, P(C|E_2 \cap A) = 1/2, \text{ and } P(C|E_3 \cap A) = 1/3 \\ &\therefore P(C|A) = \frac{\sum_{i=1}^3 P(E_i) P(A|E_i) P(C|E_i \cap A)}{\sum_{i=1}^3 P(E_i) P(A|E_i)} \\ &= \frac{\frac{1}{3} \cdot \frac{1}{3} \cdot 0 + \frac{1}{3} \cdot \frac{2}{3} \cdot \frac{1}{2} + \frac{1}{3} \cdot \frac{1}{2} \cdot \frac{1}{3}}{\frac{1}{3} \cdot \frac{1}{3} + \frac{1}{3} \cdot \frac{2}{3} + \frac{1}{3} \cdot \frac{1}{2}} = \frac{1}{3} \end{aligned}$$

### MISCELLANEOUS EXERCISE ON CHAPTER IV

1. Probabilities of occurrence of  $n$  independent events  $E_1, E_2, \dots, E_n$  are  $p_1, p_2, \dots, p_n$  respectively. Find the probability of occurrence of the compound event in which  $E_1, E_2, \dots, E_r$  occur and  $E_{r+1}, E_{r+2}, \dots, E_n$  do not occur.

$$\text{Ans. } \prod_{i=1}^r p_i \times \prod_{i=r+1}^n (1-p_i)$$

2. Prove that for any integer  $m \geq 1$ ,

$$(a) P\left(\bigcap_{i=1}^m A_i\right) \leq P(A_i) \leq P\left(\bigcup_{i=1}^m A_i\right) \leq \sum_{i=1}^m P(A_i)$$

$$(b) P\left(\bigcap_{i=1}^m A_i\right) \geq 1 - \sum_{i=1}^m P(\bar{A}_i)$$

3. Establish the inequalities :

$$P(A \cap B \cap C) \leq P(A \cap B) \leq P(A \cup B) \leq P(A \cup B \cup C) \leq P(A) + P(B) + P(C)$$

4. Let  $A_1, A_2, \dots, A_n$  be mutually independent events with  $P(A_k) = p_k$ ,  $k = 1, 2, \dots, n$ .

Let  $p$  be the probability that none of the events  $A_1, A_2, \dots, A_n$  occurs. Show that

$$p = \prod_{k=1}^n (1-p_k) \leq \exp \left\{ - \sum_{k=1}^n p_k \right\}$$

Use the above relation to compute the probability that in six tosses of a fair die, no "aces are obtained". Compare this with the upper bound given above. Show that if each  $p_k$  is small compared with  $n$ , the upper bound is a good approximation.

5. A and B play a match, the winner being the one who first wins two games in succession, no games being drawn. Their respective chances of winning a particular game are  $p : q$ . Find

- (i) A's initial chance of winning.
- (ii) A's chance of winning after having won the first game.

6. A carpenter has a tool chest with two compartments, each one having a lock. He has two keys for each lock, and he keeps all four keys in the same ring. His habitual procedure in opening a compartment is to select a key at random and try it. If it fails, he selects one of the remaining three and tries it and so on. Show that the probability that he succeeds on the first, second and third try is  $1/2, 1/3, 1/6$  respectively.

(Lucknow Univ. B.Sc., 1990)

7. Three players A, B and C agree to play a series of games observing the following rules : two players participate in each game, while third is idle, and the game is to be won by one of them. The loser in each game quits and his place in the next game is taken by the player who was idle. The player who succeeds in winning over both of his opponents without interruption wins the whole series of games.

Supposing the probability for each player to win a single game is  $1/2$ , and that the first game is played by A and B, find the probability for A, B and C respectively to win the whole series if the number of games is unlimited.

Ans.  $5/14, 5/14, 2/7$

8. In a certain group of mathematicians, 60 per cent have insufficient background of modern Algebra, 50 per cent have inadequate knowledge of Mathematical Statistics and 80 per cent are in either one or both of the two categories. What is the percentage of people who know Mathematical Statistics among those who have a sufficient background of Modern Algebra? (Ans. 0.50)

9. (a) If A has  $(n+1)$  and B has  $n$  fair coins, which they flip, show that the probability that A gets more heads than B is  $\frac{1}{2}$ .

(b) A student is given a column of 10 dates and column of 10 events and is asked to match the correct date to each event. He is not allowed to use any item more than once. Consider the case where the student knows how to match four of the items but he is very doubtful of the remaining six. He decides to match these at random. Find the probabilities that he will correctly match (i) all the items, (ii) at-least seven of the items, and (iii) at least five.

Ans. (a)  $\frac{1}{6!}$ , (b)  $\frac{10}{6!}$ , (c)  $1 - \frac{1}{6!}$

10. An astrologer claims that he can predict before birth the sex of a baby just to be born. Suppose that the astrologer has no real power but he tosses a coin just

once before every birth and if the head turns up he predicts a boy for that birth and if the tail turns up he predicts a girl. Let  $p$  be the probability of the event that at a certain birth a male child is born, and  $p'$  the probability of a head turning up in a single toss with astrologer's coin. Find the probability of a correct prediction and that of at least one correct prediction in  $n$  predictions.

11. From a pack of 52 cards an even number of cards is drawn. Show that the probability of half of these cards being red is

$$\frac{[52!/(26!)^2 - 1]}{(2^{51} - 1)}$$

12. A sportsman's chance of shooting an animal at a distance  $r (> a)$  is  $a^2/r^2$ . He fires when  $r = 2a$ , and if he misses he reloads and fires when  $r = 3a, 4a, \dots$  If he misses at distance  $na$ , the animal escapes. Find the odds against the sportsman.

Ans.  $n+1 : n-1$

$$\text{Hint. } P[\text{Sportsman shoots at a distance } ia] = \frac{a^2}{(ia)^2} = \frac{1}{i^2}$$

$$\Rightarrow P[\text{Sportsman misses the shot at a distance } ia] = 1 - \frac{1}{i^2}$$

$$\begin{aligned}\therefore P[\text{Animal escapes}] &= \prod_{i=2}^n \left(1 - \frac{1}{i^2}\right) = \prod_{i=2}^n \left[\left(\frac{i-1}{i}\right)\left(\frac{i+1}{i}\right)\right] \\ &= \prod_{i=2}^n \left(\frac{i-1}{i}\right) \prod_{i=2}^n \left(\frac{i+1}{i}\right) = \frac{n+1}{2n}\end{aligned}$$

$$\text{Required ratio} = \frac{n+1}{2n} : \left(1 - \frac{n+1}{2n}\right) = (n+1) : (n-1)$$

13. (a) Pataudi, the captain of the Indian team, is reported to have observed the rule of calling 'heads' every time the toss was made during the five matches of the Test series with the Australian team. What is the probability of his winning the toss in all the five matches?

Ans.  $(1/2)^5$

How will the probability be affected if

(i) he had made a rule of tossing a coin privately to decide whether to call "heads" or "tails" on each occasion.

(ii) the factors determining his choice were not predetermined but he called out whatever occurred to him on the spur of the moment?

(b) A lot contains 50 defective and 50 non-defective bulbs. Two bulbs are drawn at random one at a time, with replacement. The events  $A, B, C$  are defined as

$A = \{\text{The first bulb is defective}\}$

$B = \{\text{The second bulb is non-defective}\}$

$C = \{\text{The two bulbs are both defective or both non-defective}\}$

Determine whether

(i)  $A, B, C$  are pairwise independent,

(ii)  $A, B, C$  are independent.

14.  $A, B$  and  $C$  are three urns which contain 2 white, 1 black, 3 white, 2 black and 2 white and 2 black balls, respectively. One ball is drawn from urn  $A$  and put into the urn  $B$ ; then a ball is drawn from urn  $B$  and put into the urn  $C$ . Then a ball is drawn from urn  $C$ . Find the probability that the ball drawn is white.

Ans. 4/15.

15. An urn contains  $a$  white and  $b$  black balls and a series of drawings of one ball at a time is made, the ball removed being returned to the urn immediately after the next drawing is made. If  $p_n$  denotes the probability that the  $n$ th ball drawn is black, show that

$$p_n = (b - p_{n-1}) / (a + b - 1).$$

Hence find  $p_n$ .

16. A person is to be tested to see whether he can differentiate between the taste of two brands of cigarettes. If he cannot differentiate, it is assumed that the probability is one-half that he will identify a cigarette correctly. Under which of the following two procedures is there less chance that he will make all correct identifications when he actually cannot differentiate between the two brands?

(i) The subject is given four pairs each containing both brands of cigarettes (this is known to the subject), he must identify for each pair which cigarette represents each brand.

(ii) The subject is given eight cigarettes and is told that the first four are of one brand and the last four of the other brand.

How do you explain the difference in results despite the fact that eight cigarettes are tested in each case?

Ans. (i) 1/16 (ii) 1/2

17. (Sampling with replacement). A sample of size  $r$  is taken from a population of  $n$  people. Find the probability  $U_r$  that  $N$  given people will be included in the sample.

$$\text{Ans. } U_r = \sum_{m=0}^N (-1)^m \binom{N}{m} \left(1 - \frac{m}{n}\right)^r$$

18. In a lottery  $m$  tickets are drawn at a time out of the total number of  $n$  tickets, and returned before the next drawing is made. Show that the chance that in  $k$  drawings, each of the numbers 1, 2, 3, ...,  $n$  will appear at least once is given by

$$P_k = 1 - \binom{n}{1} \left(1 - \frac{m}{n}\right)^k + \binom{n}{2} \left(1 - \frac{m}{n}\right)^k \left(1 - \frac{m}{n-1}\right)^k - \dots$$

[Nagpur Univ. M.Sc. 1987]

19. In a certain book of  $N$  pages, no page contains more than four errors,  $n_1$  of them contain one error,  $n_2$  contain two errors,  $n_3$  contain three errors and  $n_4$  contain four errors. Two copies of the book are opened at any two given pages. Show that the probability that the number of errors in these two pages shall not exceed five is

$$1 - \frac{1}{N^2} (n_1^2 + n_4^2 + 2n_1 n_4 + 2n_3 n_4)$$

**Hint.** Let  $E_i$  I : the event that a page of first book contains  $i$  errors.  
and  $E_i$  II : the event that a page of second book contains  $i$  errors.

$P$  (No. of errors in the two pages shall not exceed 5)

$$= 1 - P [E_2 \text{ I } E_4 \text{ II} + E_3 \text{ I } E_4 \text{ II} + E_4 \text{ I } E_2 \text{ II} + E_3 \text{ I } E_3 \text{ II} + E_4 \text{ I } E_2 \text{ II}]$$

20. (a) Of three independent events, the chance that the first only should happen is  $a$ , the chance of the second only is  $b$  and the chance of the third only is  $c$ . Show that the independent chances of the three events are respectively

$$\frac{a}{a+x}, \frac{b}{b+x}, \frac{c}{c+x}$$

where  $x$  is the root of the equation

$$(a+x)(b+x)(c+x) = x^2$$

**Hint.**  $P(E_1 \cap \bar{E}_2 \cap \bar{E}_3) = P(E_1)[1 - P(E_2)][1 - P(E_3)] = a$  ... (\*)

$P(\bar{E}_1 \cap E_2 \cap \bar{E}_3) = [1 - P(E_1)]P(E_2)[1 - P(E_3)] = b$  ... (\*\*)

$P(\bar{E}_1 \cap \bar{E}_2 \cap E_3) = [1 - P(E_1)][1 - P(E_2)]P(E_3) = c$  ... (\*\*\*)

Multiplying (\*), (\*\*) and (\*\*\*) we get

$$P(E_1)P(E_2)P(E_3)x^2 = abc,$$

where  $x = [1 - P(E_1)][1 - P(E_2)][1 - P(E_3)]$

Multiplying (\*) by  $[1 - P(E_1)]$ , we get

$$P(E_1) = \frac{a}{a+x}, \text{ and so on.}$$

(b) Of three independent events, the probability that the first only should happen is  $1/4$ , the probability that the second only should happen is  $1/8$ , and the probability that the third only should happen is  $1/12$ . Obtain the unconditional probabilities of the three events.

Ans.  $1/2, 1/3, 1/4$ .

(c) A total of  $n$  shells are fired at a target. The probability of the  $i$ th shell hitting the target is  $p_i$ ;  $i = 1, 2, 3, \dots, n$ . Assuming that the  $n$  firings are  $n$  mutually independent events, find the probability that at least two shells out of  $n$  hit the target. [Calcutta Univ. B.Sc.(Maths Hons.), 1988]

(d) An urn contains  $M$  balls numbered 1 to  $M$ , where the first  $K$  balls are defective and the remaining  $M-K$  are non-defective. A sample of  $n$  balls is drawn from the urn. Let  $A_k$  be the event that the sample of  $n$  balls contains exactly  $k$  defectives. Find  $P(A_k)$  when the sample is drawn (i) with replacement and, (ii) without replacement. [Delhi Univ. B.Sc. (Maths Hons.), 1989]

21. For three independent events  $A$ ,  $B$  and  $C$ , the probability for  $A$  to occur is  $a$ , the probability that  $A$ ,  $B$  and  $C$  will not occur is  $b$ , and the probability that at least one of the three events will not occur is  $c$ . If  $p$  denotes the probability that  $C$  occurs but neither  $A$  nor  $B$  occurs, prove that  $p$  satisfies the quadratic equation

$$ap^2 + [ab - (1-a)(a+c-1)]p + b(1-a)(1-c) = 0$$

and hence deduce that  $c > \frac{(1-a)^2 + ab}{(1-a)}$

Further show that the probability of occurrence of  $C$  is  $p/(p+b)$ , and that of  $B$ 's happening is  $(1-c)(p+b)/ap$ .

**Hint.** Let  $P(A) = x$ ,  $P(B) = y$  and  $P(C) = z$

$$\text{Then } x = a, (1-x)(1-y)(1-z) = b, 1 - xyz = c$$

$$\text{and } p = z(1-x)(1-y)$$

Elimination of  $x$ ,  $y$  and  $z$  gives quadratic equation in  $p$ .

22. (a) The chance of success in each trial is  $p$ . If  $p_k$  is the probability that there are even number of successes in  $k$  trials, prove that

$$p_k = p + p_{k-1}(1-2p)$$

$$\text{Deduce that } p_k = \frac{1}{2}[1 + (1-2p)^k]$$

(b) If a day is dry, the conditional probability that the following day will also be dry is  $p$ ; if a day is wet, the conditional probability that the following day will be dry is  $p'$ . If  $u_n$  is the probability that the  $n$ th day will be dry, prove that

$$u_n - (p-p')u_{n-1} - p' = 0 ; n \geq 2$$

If the first day is dry,  $p = 3/4$  and  $p' = 1/4$ , find  $u_n$ .

23. There are  $n$  similar biased dice such that the probability of obtaining a 6 with each one of them is the same and equal to  $p$ . If all the dice are rolled once, show that  $p_n$ , the probability that an odd number of 6's is obtained satisfies the difference equation

$$p_n + (2p-1)p_{n-1} = p$$

and hence derive an explicit expression for  $p_n$ .

$$\text{Ans. } p_n = \frac{1}{2}[1 + (1-2p)^n]$$

24. Suppose that each day the weather can be uniquely classified as 'fine' or 'bad'. Suppose further that the probability of having fine weather on the last day of a certain year is  $P_0$  and we have the probability  $p$  that the weather on an arbitrary day will be of the same kind as on the preceding day. Let the probability of having fine weather on the  $n$ th day of the following year be  $P_n$ . Show that

$$P_n = (2p-1)P_{n-1} + (1-p)$$

Deduce that

$$P_3 = (2p-1)^3 \left( P_0 - \frac{1}{2} \right) + \frac{1}{2}$$

25. A closet contains  $n$  pairs of shoes. If  $2r$  shoes are chosen at random (with  $2r < n$ ), what is the probability that there will be (i) no complete pair,

(ii) exactly one complete pair, (iii) exactly two complete pairs among them?

Hint. (i)  $P(\text{no complete pair}) = \binom{n}{2r} 2^{2r} \div \binom{2n}{2r}$

(ii)  $P(\text{exactly one complete pair}) = n \binom{n-1}{2r-2} 2^{2r-2} + \binom{2n}{2r}$

and (iii)  $P(\text{exactly two complete pairs}) = \binom{n}{2} \binom{n-2}{2r-4} 2^{2r-4} + \binom{2n}{2r}$

26. Show that the probability of getting no right pair out of  $n$ , when the left foot shoes are paired randomly with the right foot shoes, is the sum of the first  $(n+1)$  terms in the expansion of  $e^{-1}$ .

27. (a) In a town consisting of  $(n+1)$  inhabitants, a person narrates a rumour to a second person, who in turn narrates it to a third person, and so on. At each step the recipient of the rumour is chosen at random from the  $n$  available persons, excluding the narrator himself. Find the probability that the rumour will be told  $r$  times without:

(i) returning to the originator,

(ii) being narrated to any person more than once.

(b) Do the above problem when, at each step the rumour is told by one person to a gathering of  $N$  randomly chosen people.

Ans. (a) (i)  $\frac{n(n-1)^{r-1}}{n^r} = \left(1 - \frac{1}{n}\right)^{r-1}$ ; (ii)  $\frac{n(n-1)(n-2)\dots(n-r+1)}{n^r}$

(b) (i)  $\left(1 - \frac{N}{n}\right)^{r-1}$ ; (ii)  $\left[\left(\frac{n}{N}\right)\right]^r$

28. What is the probability that (i) the birthdays of twelve people will fall in twelve different calendar months (assume equal probabilities for the twelve months) and (ii) the birthdays of six people will fall in exactly two calendar months?

Hint. (i) The birthday of the first person, for instance, can fall in 12 different ways and so for the second, and so on.

$\therefore$  The total number of cases =  $12^{12}$ .

Now there are 12 months in which the birthday of one person can fall and 11 months in which the birthday of the second person can fall and 10 months for another third person, and so on.

$\therefore$  The total number of favourable cases =  $12.11.10\dots.3.2.1$

Hence the required probability =  $\frac{12!}{12^{12}}$

(ii) The total number of ways in which the birthdays of 6 persons can fall in any of the month =  $12^6$ .

$\therefore$  The required probability =  $\frac{\binom{12}{2}(2^6 - 2)}{12^6}$

29. An elevator starts with 7 passengers and stops at 10 floors. What is the probability  $p$  that no two passengers leave at the same floor?

[Delhi Univ. M.C.A., 1988]

30. A bridge player knows that his two opponents have exactly five hearts between two of them. Each opponent has thirteen cards. What is the probability that there is three-two split on the hearts (that is one player has three hearts and the other two)?

[Delhi Univ. B.Sc.(Maths Hons.), 1988]

31. An urn contains 2 white and 2 black balls. A ball is drawn at random. If it is white, it is not replaced into the urn. Otherwise it is replaced along with another ball of the same colour. The process is repeated. Find the probability that the third ball drawn is black.

[Burdwan Univ. B.Sc. (Hons.), 1990]

Ans.  $\frac{23}{30}$

32. There is a series of  $n$  urns. In the  $i$ th urn there are  $i$  white and  $(n-i)$  black balls,  $i=1, 2, 3, \dots, k$ . One urn is chosen at random and 2 balls are drawn from it. Both turn out to be white. What is the probability that the  $j$ th urn was chosen, where  $j$  is a particular number between 3 and  $n$ .

**Hint.** Let  $E_j$  denote the event of selection of  $j$ th urn,  $j=3, 4, \dots, n$  and  $A$  denote the event of drawing of 2 white balls, then

$$P(A|E_j) = \left(\frac{j}{n}\right)\left(\frac{j-1}{n-1}\right), \quad P(E_j) = \frac{1}{n}, \quad P(A) = \sum_{i=1}^n \frac{1}{n} \left(\frac{i}{n}\right)\left(\frac{i-1}{n-1}\right)$$

$$\therefore P(E_j|A) = \frac{\frac{1}{n} \left(\frac{j}{n}\right)\left(\frac{j-1}{n-1}\right)}{\sum_{i=1}^n \left(\frac{1}{n}\right)\left(\frac{i}{n}\right)\left(\frac{i-1}{n-1}\right)}$$

33. There are  $(N+1)$  identical urns marked 0, 1, 2, ...,  $N$  each of which contains  $N$  white and red balls. The  $k$ th urn contains  $k$  red and  $N-k$  white balls, ( $k=0, 1, 2, \dots, N$ ). An urn is chosen at random and  $n$  random drawings of a ball are made from it, the ball drawn being replaced after each draw. If the balls drawn are all red, show that the probability that the next drawing will also yield a red ball is approximately  $(n+1)/(n+2)$  when  $N$  is large.

34. A printing machine can print  $n$  letters, say  $\alpha_1, \alpha_2, \dots, \alpha_n$ . It is operated by electrical impulses, each letter being produced by a different impulse. Assume that  $p$  is the constant probability of printing the correct letter and the impulses are independent. One of the  $n$  impulses, chosen at random, was fed into the machine twice and both times the letter  $\alpha_i$  was printed. Compute the probability that the impulse chosen was meant to print  $\alpha_i$ .

[Delhi Univ. M.Sc.(Stat.), 1981]

Ans.  $(n-1)p^2/(n p^2 - 2p + 1)$

35. Two players  $A$  and  $B$  agree to contest a match consisting of a series of games, the match to be won by the player who first wins three games, with the provision that if the players win two games each, the match is to continue until it

is won by one player winning two games more than his opponent. The probability of  $A$  winning any given game is  $p$ , and the games cannot be drawn.

(i) Prove that  $f(p)$ , the initial probability of  $A$  winning the match is given by:

$$f(p) = p^3(4 - 5p + 2p^2)/(1 - 2p + 2p^2)$$

(ii) Show that the equation  $f(p) = p$  has five real roots, of which three are admissible values of  $p$ . Find these three roots and explain their significance.

[Civil Services (Main), 1986]

36. Two players  $A$  and  $B$  start playing a series of games with Rs.  $a$  and  $b$  respectively. The stake is Re. 1 on a game and no game can be drawn. If the probability of  $A$  winning any game is a constant  $p$ , find the initial probability of  $A$  exhausting the funds of  $B$  or his own. Also show that if the resources of  $B$  are unlimited then

(i)  $A$  is certain to be ruined if  $p = \frac{1}{2}$ , and

(ii)  $A$  has an even chance of escaping ruin if  $p = 2^{1/a}/(1 + 2^{1/a})$ .

**Hint.** Let  $u_n$  be the probability of  $A$ 's final win when he has Rs.  $a$ .

Then  $u_n = pu_{n+1} + (1-p)u_{n-1}$  where  $u_0 = 0$  and  $u_{a+b} = 1$

$$\therefore u_{n+1} - u_n = \left( \frac{1-p}{p} \right) (u_n - u_{n-1})$$

$$\text{Hence } u_{n+1} - u_n = \left( \frac{1-p}{p} \right)^n u_1, \text{ by repeated application,}$$

$$\text{so that } u_n = u_1 \left[ 1 - \left( \frac{1-p}{p} \right)^n \right] / \left[ 1 - \left( \frac{1-p}{p} \right)^{a+b} \right]$$

$$\text{Hence using } u_{a+b} = 1, u_n = \left[ 1 - \left( \frac{1-p}{p} \right)^n \right] / \left[ 1 - \left( \frac{1-p}{p} \right)^{a+b} \right]$$

$$\therefore \text{Initial probability of } A\text{'s win is } u_a = \frac{p^a - (1-p)^a}{p^{a+b} - (1-p)^{a+b}} \cdot p^b$$

$$\text{Probability of } A\text{'s ruin} = 1 - u_a.$$

For  $p = \frac{1}{2}$ ,  $u_a = \frac{a}{a+b} \rightarrow 0$  as  $b \rightarrow \infty$  and for  $p \neq \frac{1}{2}$ ,  $u_a = \frac{1}{2}$  if  $p = 2^{1/a}/(1 + 2^{1/a})$ .

37. In a game of skill a player has probability  $1/3$ ,  $5/12$  and  $1/4$  of scoring 0, 1 and 2 points respectively at each trial, the game terminating on the first realization of a zero score at a trial. Assuming that the trials are independent, prove that the probability of the player obtaining a total score of  $n$  points is

$$u_n = \frac{3}{13} \left( \frac{3}{4} \right)^n + \frac{4}{39} \left( -\frac{1}{3} \right)^n$$

**Hint.** Event can materialize in the two mutually exclusive ways:

(i) at the  $(n-1)$ th trial, a score of  $(n-1)$  points is obtained and a score of 1 point is obtained at the  $n$ th trial.

(ii) at the  $(n - 2)$ th trial, a score of  $(n - 2)$  points is obtained and a score of 2 points is obtained at the last two trials.

$$\text{Hence } u_n = \frac{5}{12} u_{n-1} + \frac{1}{4} u_{n-2} \text{ where } u_0 = \frac{1}{3}, u_1 = \frac{1}{3} \cdot \frac{5}{12} = \frac{5}{36}$$

$$\text{Also } u_n = \left( \frac{3}{4} - \frac{1}{3} \right) u_{n-1} + \frac{1}{4} u_{n-2} \Rightarrow u_n + \frac{1}{3} u_{n-1} = \frac{3}{4} \left( u_{n-1} + \frac{1}{3} u_{n-2} \right)$$

This equation can be solved as a homogeneous difference equation of second order with the initial conditions

$$u_0 = \frac{1}{3}, u_1 = \frac{1}{3} \cdot \frac{5}{12} = \frac{5}{36}$$

38. The following weather forecasting is used by an amateur forecaster. Each day is classified as 'dry' or 'wet' and the probability that any given day is same as the preceding one is assumed to be a constant  $p$ , ( $0 < p < 1$ ). Based on past records, it is supposed that January 1 has a probability  $\beta$  of being dry. Letting

$\beta_n$  = Probability that  $n$ th day of the year is dry, obtain an expression for  $\beta_n$  in terms of  $\beta$  and  $p$ . Also evaluate  $\lim_{n \rightarrow \infty} \beta_n$ .

$$\text{Hint. } \beta_n = p \cdot \beta_{n-1} + (1-p)(1-\beta_{n-1})$$

$$\Rightarrow \beta_n = (2p-1)\beta_{n-1} + (1-p); n = 2, 3, 4, \dots$$

$$\text{Ans. } \beta_n = (2p-1)^{n-1} (\beta - \frac{1}{2}) + \frac{1}{2}; \lim_{n \rightarrow \infty} \beta_n = \frac{1}{2}$$

39. Two urns contain respectively ' $a$  white and  $b$  black' and ' $b$  white and  $a$  black' balls. A series of drawings is made according to the following rules:

(i) Each time only one ball is drawn and immediately returned to the same urn it came from.

(ii) If the ball drawn is white, the next drawing is made from the first urn.

(iii) If it is black, the next drawing is made from the second urn.

(iv) The first ball drawn comes from the first urn.

What is the probability that  $n$ th ball drawn will be white?

Hint.  $p_r = P$  [Drawing a white ball at the  $r$ th draw].

$$p_r = \frac{a}{a+b} p_{r-1} + \frac{b}{a+b} (1-p_{r-1})$$

$$\Rightarrow p_r = \frac{a-b}{a+b} \cdot p_{r-1} + \frac{b}{a+b}$$

$$\text{Ans. } p_n = \frac{1}{2} + \frac{1}{2} \left( \frac{a-b}{a+b} \right)^n$$

40. If a coin is tossed repeatedly, show that the probability of getting  $m$  heads before  $n$  tails is :

$$\frac{1}{2^{m+n-1}} \sum_{i=m}^{m+n-1} {}^{m+n-1} C_i \quad [\text{Burdwan Univ. (Maths Hons.), 1991}]$$

### OBJECTIVE TYPE QUESTIONS

**I.** Find out the correct answer from group Y for each item of group X:

- | Group X   | Group Y  |
|---|--|
| (a) At least one of the events A or B occurs.     | (i) $(\bar{A} \cap B) \cup (A \cap \bar{B}) \cup (\bar{A} \cap \bar{B})$ |
| (b) Neither A nor B occurs.                       | (ii) $(A \cup B) - (A \cap B)$   |
| (c) Exactly one of the events A or B occurs.      | (iii) $A \subset B$  |
| (d) If event A occurs, so does B.                 | (iv) $B \subset A$   |
| (e) Not more than one of the events A or B occur. | (v) $[A - (A \cap B)] \cup [B - (A \cap B)]$                             |
|   | (vi) $A \cap \bar{B}$  |
|   | (vii) $\bar{1} - (A \cup \bar{B})$                                       |
|   | (viii) $A \cup B$  |
|   | (ix) $\bar{1} - (A \cap B)$  |

**II.** Match the correct expression of probabilities on the left:

- |  |                                     |
|--|-------------------------------------|
| (a) $P(\phi)$ , where $\phi$ is null set | (i) $1 - P(A)$                      |
| (b) $P(A B)P(B)$                         | (ii) $P(A \cap B)$                  |
| (c) $P(\bar{A})$                         | (iii) $P(A) - P(A \cap B)$          |
| (d) $P(\bar{A} \cap \bar{B})$            | (iv) 0                              |
| (e) $P(A \sim B)$                        | (v) $1 - P(A) - P(B) + P(A \cap B)$ |
|  | (vi) $P(A) + P(B) - P(A \cap B)$    |

**III.** Given that A, B and C are mutually exclusive events, explain why the following are not permissible assignments of probabilities:

- (i)  $P(A) = 0.24$ ,  $P(B) = 0.4$  and  $P(A \cup C) = 0.2$
- (ii)  $P(A) = 0.4$ ,  $P(B) = 0.61$
- (iii)  $P(A) = 0.6$ ,  $P(A \cap \bar{B}) = 0.5$

**IV.** In each of the following, indicate whether events A and B are :

(i) independent, (ii) mutually exclusive, (iii) dependent but not mutually exclusive.

- |  |  |
|--|--|
| (a) $P(A \cap B) = 0$                                      | (b) $P(A \cap B) = 0.3$ , $P(A) = 0.45$                |
| (c) $P(A \cup B) = 0.85$ , $P(A) = 0.3$ , $P(B) = 0.6$     | (d) $P(A \cup B) = 0.70$ , $P(A) = 0.5$ , $P(B) = 0.4$ |
| (e) $P(A \cup B) = 0.90$ , $P(A B) = 0.8$ , $P(B) = 0.5$ . |  |

**V.** Give the correct label as answer like a or b etc., for the following questions:

- (i) The probability of drawing any one spade card from a pack of cards is  
 (a)  $\frac{1}{52}$       (b)  $\frac{1}{13}$       (c)  $\frac{4}{13}$       (d)  $\frac{1}{4}$

(ii) The probability of drawing one white ball from a bag containing 6 red, 8 black, 10 yellow and 1 green balls is

- |                    |       |       |                     |                     |
|--------------------|-------|-------|---------------------|---------------------|
| (a) $\frac{1}{25}$ | (b) 0 | (c) 1 | (d) $\frac{24}{25}$ | (e) $\frac{15}{20}$ |
|--------------------|-------|-------|---------------------|---------------------|

(iii) A coin is tossed three times in succession, the number of sample points in sample space is

- (a) 6      (b) 8      (c) 3

(iv) In the simultaneous tossing of two perfect coins, the probability of having at least one head is

- (a)  $\frac{1}{2}$       (b)  $\frac{1}{4}$       (c)  $\frac{3}{4}$       (d) 1

(v) In the simultaneous tossing of two perfect dice, the probability of obtaining 4 as the sum of the resultant faces is

- (a)  $\frac{4}{12}$       (b)  $\frac{1}{12}$       (c)  $\frac{3}{12}$       (d)  $\frac{2}{12}$

(vi) A single letter is selected at random from the word 'probability'. The probability that it is a vowel is

- (a)  $\frac{3}{11}$       (b)  $\frac{2}{11}$       (c)  $\frac{4}{11}$       (d) 0

(vii) An urn contains 9 balls, two of which are red, three blue and four black. Three balls are drawn at random. The chance that they are of the same colour is

- (a)  $\frac{5}{84}$       (b)  $\frac{3}{9}$       (c)  $\frac{3}{7}$       (d)  $\frac{7}{17}$

(viii) A number is chosen at random among the first 120 natural numbers. The probability of the number chosen being a multiple of 5 or 15 is

- (a)  $\frac{1}{5}$       (b)  $\frac{1}{8}$       (c)  $\frac{1}{16}$

(ix) If  $A$  and  $B$  are mutually exclusive events, then

- (a)  $P(A \cup B) = P(A) \cdot P(B)$   
 (b)  $P(A \cup B) = P(A) + P(B)$ ,      (c)  $P(A \cup B) = 0$ .

(x) If  $A$  and  $B$  are two independent events, the probability that both  $A$  and  $B$  occur is  $\frac{1}{8}$  and the probability that neither of them occurs is  $\frac{3}{8}$ . The probability of the occurrence of  $A$  is :

- (a)  $\frac{1}{2}$ ,      (b)  $\frac{1}{3}$ ,      (c)  $\frac{1}{4}$ ,      (d)  $\frac{1}{5}$ .

#### VI. Fill in the blanks :

(i) Two events are said to be equally likely if .....

(ii) A set of events is said to be independent if .....

(iii) If  $P(A) \cdot P(B) \cdot P(C) = P(A \cap B \cap C)$ , then the events  $A, B, C$  are .....

(iv) Two events  $A$  and  $B$  are mutually exclusive if  $P(A \cap B) = \dots$  and are independent if  $P(A \cap B) = \dots$

(v) The probability of getting a multiple of 2 in a throw of a dice is  $1/2$  and of getting a multiple of 3 is  $1/3$ . Hence probability of getting a multiple of 2 or 3 is .....

(vi) Let  $A$  and  $B$  be independent events and suppose the event  $C$  has probability 0 or 1. Then  $A, B$  and  $C$  are ..... events.

(vii) If  $A, B, C$  are pairwise independent and  $A$  is independent of  $B \cup C$ , then  $A, B, C$  are ..... independent.

- (viii) A man has tossed 2 fair dice. The conditional probability that he has tossed two sixes, given that he has tossed at least one six is .....  
 (ix) Let  $A$  and  $B$  be two events such that  $P(A) = 0.3$  and  $P(A \cup B) = 0.8$ . If  $A$  and  $B$  are independent events then  $P(B) = \dots$

VII. Each of following statements is either true or false. If it is true prove it, otherwise, give a counter example to show that it is false.

- (i) The probability of occurrence of at least one of two events is the sum of the probability of each of the two events.
- (ii) Mutually exclusive events are independent.
- (iii) For any two events  $A$  and  $B$ ,  $P(A \cap B)$  cannot be less than either  $P(A)$  or  $P(B)$ .
- (iv) The conditional probability of  $A$  given  $B$  is always greater than  $P(A)$ .
- (v) If the occurrence of an event  $A$  implies the occurrence of another event  $B$  then  $P(A)$  cannot exceed  $P(B)$ .
- (vi) For any two events  $A$  and  $B$ ,  $P(A \cup B)$  cannot be greater than either  $P(A)$  or  $P(B)$ .
- (vii) Mutually exclusive events are not independent.
- (viii) Pairwise independence does not necessarily imply mutual independence.
- (ix) Let  $A$  and  $B$  be events neither of which has probability zero. Then if  $A$  and  $B$  are disjoint,  $A$  and  $B$  are independent.
- (x) The probability of any event is always a proper fraction.
- (xi) If  $0 < P(B) < 1$  so that  $P(A|B)$  and  $P(A|\bar{B})$  are both defined, then  $P(A) = P(B)P(A|B) + P(\bar{B})P(A|\bar{B})$ .
- (xii) For two events  $A$  and  $B$  if  
 $P(A) = P(A|B) = 1/4$  and  $P(A|\bar{B}) = 1/2$ , then
  - (a)  $A$  and  $B$  are mutually exclusive.
  - (b)  $A$  and  $B$  are independent.
  - (c)  $A$  is a sub-event of  $B$ .
  - (d)  $P(\bar{A}|B) = 3/4$ . [Delhi Univ. B.Sc.(Stat. Hons.), 1992]
- (xiii) Two events can be independent and mutually exclusive simultaneously.
- (xiv) Let  $A$  and  $B$  be events, neither of which has probability zero. Prove or disprove the following :
  - (a) If  $A$  and  $B$  are disjoint,  $A$  and  $B$  are independent.
  - (b) If  $A$  and  $B$  are independent,  $A$  and  $B$  are disjoint.
  - (xv) If  $P(A) = 0$ , then  $A = \emptyset$ .

**Random Variables — Distribution Functions**

**5.1. Random Variable.** Intuitively by a *random variable* (r.v.) we mean a real number  $X$  connected with the outcome of a random experiment  $E$ . For example, if  $E$  consists of two tosses of a coin, we may consider the random variable which is the number of heads (0, 1 or 2).

Outcome :	HII	HT	TH	TT
Value of $X$ :	2	1	1	0

Thus to each outcome  $\omega$ , there corresponds a real number  $X(\omega)$ . Since the points of the sample space  $S$  correspond to outcomes, this means that a real number, which we denote by  $X(\omega)$ , is defined for each  $\omega \in S$ . From this standpoint, we define random variable to be a real function on  $S$  as follows:

"Let  $S$  be the sample space associated with a given random experiment. A real-valued function defined on  $S$  and taking values in  $R(-\infty, \infty)$  is called a one-dimensional random variable. If the function values are ordered pairs of real numbers (i.e., vectors in two-space) the function is said to be a two-dimensional random variable. More generally, an  $n$ -dimensional random variable is simply a function whose domain is  $S$  and whose range is a collection of  $n$ -tuples of real numbers (vectors in  $n$ -space)."

For a mathematical and rigorous definition of the random variable, let us consider the probability space, the triplet  $(S, B, P)$ , where  $S$  is the sample space, viz., space of outcomes,  $B$  is the  $\sigma$ -field of subsets in  $S$ , and  $P$  is a probability function on  $B$ .

**Def.** A random variable (r.v.) is a function  $X(\omega)$  with domain  $S$  and range  $(-\infty, \infty)$  such that for every real number  $a$ , the event  $\{\omega : X(\omega) \leq a\} \in B$ .

**Remarks:** 1. The refinement above is the same as saying that the function  $X(\omega)$  is measurable real function on  $(S, B)$ .

2. We shall need to make probability statements about a random variable  $X$  such as  $P(X \leq a)$ . For the simple example given above we should write  $P(X \leq 1) = P(HH, HT, TH) = 3/4$ . That is,  $P(X \leq a)$  is simply the probability of the set of outcomes  $\omega$  for which  $X(\omega) \leq a$  or

$$P(X \leq a) = P(\{\omega : X(\omega) \leq a\})$$

Since  $P$  is a measure on  $(S, B)$  i.e.,  $P$  is defined on subsets of  $B$ , the above probability will be defined only if  $\{\omega : X(\omega) \leq a\} \in B$ , which implies that  $X(\omega)$  is a measurable function on  $(S, B)$ .

3. One-dimensional random variables will be denoted by capital letters,  $X, Y, Z, \dots$  etc. A typical outcome of the experiment (i.e., a typical element of the sample space) will be denoted by  $\omega$  or  $e$ . Thus  $X(\omega)$  represents the real number which the random variable  $X$  associates with the outcome  $\omega$ . The values which  $X, Y, Z, \dots$  etc., can assume are denoted by lower case letters viz.,  $x, y, z, \dots$  etc.

4. *Notations.* If  $x$  is a real number, the set of all  $\omega$  in  $S$  such that  $X(\omega) = x$  is denoted briefly by writing  $X = x$ . Thus

$$P(X = x) = P\{\omega : X(\omega) = x\}$$

Similarly  $P(X \leq a) = P\{\omega : X(\omega) \in [-\infty, a]\}$

and  $P[a < X \leq b] = P\{\omega : X(\omega) \in (a, b]\}$

Analogous meanings are given to

$$P(X = a \text{ or } X = b) = P\{(X = a) \cup (X = b)\},$$

$$P(X = a \text{ and } X = b) = P\{(X = a) \cap (X = b)\}, \text{ etc.}$$

**Illustrations :** 1. If a coin is tossed, then

$$S = \{\omega_1, \omega_2\} \text{ where } \omega_1 = H, \omega_2 = T$$

$$X(\omega) = \begin{cases} 1, & \text{if } \omega = H \\ 0, & \text{if } \omega = T \end{cases}$$

$X(\omega)$  is a Bernoulli random variable. Here  $X(\omega)$  takes only two values. A random variable which takes only a finite number of values is called *single*.

2. An experiment consists of rolling a die and reading the number of points on the upturned face. The most natural random variable  $X$  to consider is

$$X(\omega) = \omega ; \omega = 1, 2, \dots, 6$$

If we are interested in whether the number of points is even or odd, we consider a random variable  $Y$  defined as follows :

$$Y(\omega) = \begin{cases} 0, & \text{if } \omega \text{ is even} \\ 1, & \text{if } \omega \text{ is odd} \end{cases}$$

3. If a dart is thrown at a circular target, the sample space  $S$  is the set of all points  $\omega$  on the target. By imagining a coordinate system placed on the target with the origin at the centre, we can assign various random variables to this experiment. A natural one is the two dimensional random variable which assigns to the point  $\omega$ , its rectangular coordinates  $(x, y)$ . Another is that which assigns  $\omega$  its polar coordinates  $(r, \theta)$ . A one dimensional random variable assigns to each  $\omega$  only one of the coordinates  $x$  or  $y$  (for cartesian system),  $r$  or  $\theta$  (for polar system). The event  $E$ , "that the dart will land in the first quadrant" can be described by a random variable which assigns to each point  $\omega$  its polar coordinate  $\theta$  so that  $X(\omega) = \theta$  and then  $E = \{\omega : 0 \leq X(\omega) \leq \pi/2\}$ .

4. If a pair of fair dice is tossed then  $S = \{1, 2, 3, 4, 5, 6\} \times \{1, 2, 3, 4, 5, 6\}$  and  $n(S) = 36$ . Let  $X$  be a random variable with image set

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

$$P(X = 1) = P\{1, 1\} = 1/36$$

$$P(X = 2) = P\{(2, 1), (2, 2), (1, 2)\} = 3/36$$

$$P(X = 3) = P\{(3, 1), (3, 2), (3, 3), (2, 3), (1, 3)\} = 5/36$$

$$P(X=4) = P\{(4,1), (4,2), (4,3), (4,4), (3,4), (2,4), (1,4)\} = 7/36$$

$$\text{Similarly } P(X=5) = 9/36 \text{ and } P(X=6) = 11/36$$

**Some theorems on Random Variables.** Here we shall state (without proof) some of the fundamental results and theorems on random variables.

**Theorem 5-1.** A function  $X(\omega)$  from  $S$  to  $R$  ( $-\infty, \infty$ ) is a random variable if and only if

$$\{\omega : X(\omega) < a\} \in \mathcal{B}$$

**Theorem 5-2.** If  $X_1$  and  $X_2$  are random variables and  $C$  is a constant then  $CX_1, X_1 + X_2, X_1 X_2$  are also random variables.

**Remark.** It will follow that  $C_1 X_1 + C_2 X_2$  is a random variable for constants  $C_1$  and  $C_2$ . In particular  $X_1 - X_2$  is a r.v.

**Theorem 5-3.** If  $\{X_n(\omega), n \geq 1\}$  are random variables then

$\sup_n X_n(\omega), \inf_n X_n(\omega), \limsup_{n \rightarrow \infty} X_n(\omega)$  and  $\liminf_{n \rightarrow \infty} X_n(\omega)$  are all random variables, whenever they are finite for all  $\omega$ .

**Theorem 5-4.** If  $X$  is a random variable then

- (i)  $\frac{1}{X}$  where  $\left(\frac{1}{X}\right)(\omega) = \infty$  if  $X(\omega) = 0$
- (ii)  $X_+(\omega) = \max [0, X(\omega)]$
- (iii)  $X_-(\omega) = -\min [0, X(\omega)]$
- (iv)  $|X|$

are random variables.

**Theorem 5-5.** If  $X_1$  and  $X_2$  are random variables then

(i)  $\max[X_1, X_2]$  and (ii)  $\min[X_1, X_2]$  are also random variables.

**Theorem 5-6.** If  $X$  is a r.v. and  $f(\cdot)$  is a continuous function, then  $f(X)$  is a r.v.

**Theorem 5-7.** If  $X$  is a r.v. and  $f(\cdot)$  is an increasing function, then  $f(X)$  is a r.v.

**Corollary.** If  $f$  is a function of bounded variations on every finite interval  $[a,b]$ , and  $X$  is a r.v. then  $f(X)$  is a r.v.

(proofs of the above theorems are beyond the scope of this book)

### EXERCISE 5 (a)

1. Let  $X$  be a one dimensional random variable. (i) If  $a < b$ , show that the two events  $a < X \leq b$  and  $X \leq a$  are disjoint, (ii) Determine the union of the two events in part (i), (iii) show that  $P(a < X \leq b) = P(X \leq b) - P(X \leq a)$ .

2. Let a sample space  $S$  consist of three elements  $\omega_1, \omega_2$ , and  $\omega_3$ . Let  $P(\omega_1) = 1/4, P(\omega_2) = 1/2$  and  $P(\omega_3) = 1/4$ . If  $X$  is a random variable defined on  $S$  by  $X(\omega_1) = 10, X(\omega_2) = -3, X(\omega_3) = 15$ , find  $P(-2 \leq X \leq 2)$ .

3. Let  $S = \{e_1, e_2, \dots, e_n\}$  be the sample space of some experiment and let  $E \subseteq S$  be some event associated with the experiment.

Define  $\psi_E$ , the *characteristic random variable* of  $E$  as follows :

$$\psi_E(e_i) = \begin{cases} 1 & \text{if } e_i \in E \\ 0 & \text{if } e_i \notin E. \end{cases}$$

In other words,  $\psi_E$  is equal to 1 if  $E$  occurs, and  $\psi_E$  is equal to 0 if  $E$  does not occur.

Verify the following properties of characteristic random variables :

- (i)  $\psi_\emptyset$  is identically zero , i.e.,  $\psi_\emptyset(e_i) = 0 ; i = 1, 2, \dots, n$
- (ii)  $\psi_S$  is identically one , i.e.,  $\psi_S(e_i) = 1 ; i = 1, 2, \dots, n$
- (iii)  $E = F \Rightarrow \psi_E(e_i) = \psi_F(e_i) ; i = 1, 2, \dots, n$  and conversely
- (iv) If  $E \subseteq F$  then  $\psi_E(e_i) \leq \psi_F(e_i) ; i = 1, 2, \dots, n$
- (v)  $\psi_E(e_i) + \psi_F(e_i)$  is identically 1 :  $i = 1, 2, \dots, n$
- (vi)  $\psi_{E \cap F}(e_i) = \psi_E(e_i)\psi_F(e_i) ; i = 1, 2, \dots, n$
- (vii)  $\psi_{E \cup F}(e_i) = \psi_E(e_i) + \psi_F(e_i) - \psi_E(e_i)\psi_F(e_i)$ , for  $i = 1, 2, \dots, n$ .

**5.2. Distribution Function.** Let  $X$  be a r.v. on  $(S, \mathcal{B}, P)$ . Then the function :

$$F_X(x) = P(X \leq x) = P\{\omega : X(\omega) \leq x\}, -\infty < x < \infty$$

is called the distribution function (d.f.) of  $X$ .

If clarity permits, we may write  $F(x)$  instead of  $F_X(x)$ . ....(5.1)

**5.2.1. Properties of Distribution Function.** We now proceed to derive a number of properties common to all distribution functions.

**Property 1.** If  $F$  is the d.f. of the r.v.  $X$  and if  $a < b$ , then

$$P(a < X \leq b) = F(b) - F(a)$$

**Proof.** The events ' $a < X \leq b$ ' and ' $X \leq a$ ' are disjoint and their union is the event ' $X \leq b$ '. Hence by addition theorem of probability

$$\begin{aligned} P(a < X \leq b) + P(X \leq a) &= P(X \leq b) \\ \Rightarrow P(a < X \leq b) &= P(X \leq b) - P(X \leq a) = F(b) - F(a) \quad \dots(5.2) \end{aligned}$$

**Cor. 1.**

$$\begin{aligned} P(a \leq X \leq b) &= P\{(X = a) \cup (a < X \leq b)\} \\ &= P(X = a) + P(a < X \leq b) \\ &\quad \text{(using additive property of } P\text{)} \\ &= P(X = a) + [F(b) - F(a)] \quad \dots(5.2a) \end{aligned}$$

Similarly, we get

$$\begin{aligned} P(a < X < b) &= P(a < X \leq b) - P(X = b) \\ &= F(b) - F(a) - P(X = b) \quad \dots(5.2b) \end{aligned}$$

$$P(a \leq X < b) = P(a < X < b) + P(X = a)$$

$$= F(b) - F(a) - P(X=b) + P(X=a) \quad \dots(5.2.c)$$

**Remark.** When  $P(X=a)=0$  and  $P(X=b)=0$ , all four events  $a \leq X \leq b$ ,  $a < X < b$ ,  $a \leq X < b$  and  $a < X \leq b$  have the same probability  $F(b) - F(a)$ .

**Property 2.** If  $F$  is the d.f. of one-dimensional r.v.  $X$ , then  
(i)  $0 \leq F(x) \leq 1$ , (ii)  $F(x) \leq F(y)$  if  $x < y$ .

In other words, all distribution functions are monotonically non-decreasing and lie between 0 and 1.

**Proof.** Using the axioms of certainty and non-negativity for the probability function  $P$ , part (i) follows triviality from the definition of  $F(x)$ .

For part (ii), we have for  $x < y$ ,

$$F(y) - F(x) = P(x < X \leq y) \geq 0 \quad (\text{Property 1})$$

$$\Rightarrow F(y) \geq F(x)$$

$$\Rightarrow F(x) \leq F(y) \text{ when } x < y \quad \dots(5.3)$$

**Property 3.** If  $F$  is d.f. of one-dimensional r.v.  $X$ , then

$$F(-\infty) = \lim_{x \rightarrow -\infty} F(x) = 0$$

$$\text{and} \quad F(\infty) = \lim_{x \rightarrow \infty} F(x) = 1$$

**Proof.** Let us express the whole sample space  $S$  as a countable union of disjoint events as follows :

$$S = [\bigcup_{n=-\infty}^{\infty} (-n < X \leq -n+1)] \cup [\bigcup_{n=0}^{\infty} (n < X \leq n+1)]$$

$$\Rightarrow P(S) = \sum_{n=-\infty}^{\infty} P(-n < X \leq -n+1) + \sum_{n=0}^{\infty} P(n < X \leq n+1)$$

( $\because P$  is additive)

$$\begin{aligned} \Rightarrow 1 &= \lim_{a \rightarrow \infty} \sum_{n=1}^a [F(-n+1) - F(-n)] \\ &\quad + \lim_{b \rightarrow \infty} \sum_{n=0}^b [F(n+1) - F(n)] \\ &= \lim_{a \rightarrow \infty} [F(0) - F(-a)] + \lim_{b \rightarrow \infty} [F(b+1) - F(0)] \\ &= [F(0) - F(-\infty)] + [F(\infty) - F(0)] \\ \therefore 1 &= F(\infty) - F(-\infty) \quad \dots(*) \end{aligned}$$

Since  $-\infty < \infty$ ,  $F(-\infty) \leq F(\infty)$ . Also

$$F(-\infty) \geq 0 \text{ and } F(\infty) \leq 1$$

(Property 2)

$$\therefore 0 \leq F(-\infty) \leq F(\infty) \leq 1 \quad (**)$$

(\*) and (\*\*) give  $F(-\infty) = 0$  and  $F(\infty) = 1$ .

**Remarks.** 1. Discontinuities of  $F(x)$  are at most countable.

$$2. \quad F(a) - F(a-0) = \lim_{h \rightarrow 0} P(a-h \leq X \leq a), \quad h > 0$$

$$\therefore F(a) - F(a-0) = P(X=a)$$

$$\text{and } F(a+0) - F(a) = \lim_{h \rightarrow 0} P(a \leq X \leq a+h) = 0, \quad h > 0$$

$$\Rightarrow F(a+0) = F(a)$$

**5.3. Discrete Random Variable.** If a random variable takes at most a countable number of values, it is called a discrete random variable. In other words, a real valued function defined on a discrete sample space is called a discrete random variable.

**5.3.1. Probability Mass Function (and probability distribution of a discrete random variable).**

Suppose  $X$  is a one-dimensional discrete random variable taking at most a countably infinite number of values  $x_1, x_2, \dots$ . With each possible outcome  $x_i$ , we associate a number  $p_i = P(X = x_i) = p(x_i)$ , called the probability of  $x_i$ . The numbers  $p(x_i)$ ;  $i = 1, 2, \dots$  must satisfy the following conditions :

$$(i) \quad p(x_i) \geq 0 \quad \forall i, \quad (ii) \quad \sum_{i=1}^{\infty} p(x_i) = 1$$

This function  $p$  is called the probability mass function of the random variable  $X$  and the set  $\{x_i, p(x_i)\}$  is called the probability distribution (p.d.) of the r.v.  $X$ .

**Remarks:** 1. The set of values which  $X$  takes is called the spectrum of the random variable.

2. For discrete random variable, a knowledge of the probability mass function enables us to compute probabilities of arbitrary events. In fact, if  $E$  is a set of real numbers, we have

$$P(X \in E) = \sum_{x \in E \cap S} p(x), \quad \text{where } S \text{ is the sample space.}$$

**Illustration.** Toss of coin,  $S = \{H, T\}$ . Let  $X$  be the random variable defined by

$$X(H) = 1, \text{ i.e., } X = 1, \text{ if 'Head' occurs.}$$

$$X(T) = 0, \text{ i.e., } X = 0, \text{ if 'Tail' occurs.}$$

If the coin is 'fair' the probability function is given by

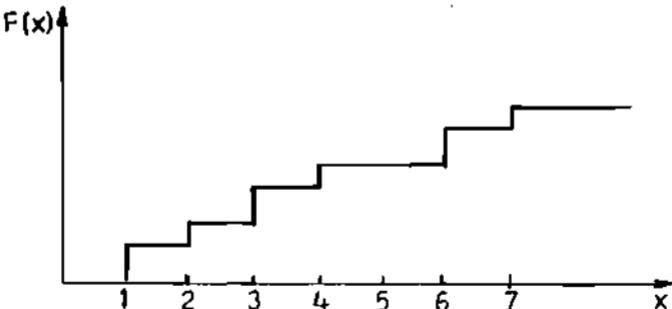
$$P(\{H\}) = P(\{T\}) = \frac{1}{2}$$

and we can speak of the probability distribution of the random variable  $X$  as

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

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

**5.3.2. Discrete Distribution Function.** In this case there are a countable number of points  $x_1, x_2, x_3, \dots$  and numbers  $p_i \geq 0$ ,  $\sum_{i=1}^{\infty} p_i = 1$  such that  $F(X) = \sum_{\{i : x_i \leq X\}} p_i$ . For example if  $x_i$  is just the integer  $i$ ,  $F(x)$  is a "step function" having jump  $p_i$  at  $i$ , and being constant between each pair of integers.



**Theorem 5.5.**  $p(x_j) = P(X = x_j) = F(x_j) - F(x_{j-1})$ , where  $F$  is the d.f. of  $X$ .

**Proof.** Let  $x_1 < x_2 < \dots$  We have

$$\begin{aligned} F(x_j) &= P(X \leq x_j) \\ &= \sum_{i=1}^j P(X = x_i) = \sum_{i=1}^j p(x_i) \end{aligned}$$

and

$$F(x_{j-1}) = P(X \leq x_{j-1}) = \sum_{i=1}^{j-1} p(x_i)$$

$$\therefore F(x_j) - F(x_{j-1}) = p(x_j) \quad \dots(5.5)$$

Thus, given the distribution function of discrete random variable, we can compute its probability mass function.

**Example 5.1.** An experiment consists of three independent tosses of a fair coin. Let

$X$  = The number of heads

$Y$  = The number of head runs,

$Z$  = The length of head runs,

a head run being defined as consecutive occurrence of at least two heads, its length then being the number of heads occurring together in three tosses of the coin.

Find the probability function of (i)  $X$ , (ii)  $Y$ , (iii)  $Z$ , (iv)  $X+Y$  and (v)  $XY$  and construct probability tables and draw their probability charts.

**Solution.**

**Table 1**

S. No.	Elementary event	Random Variables				
		X	Y	Z	X+Y	XY
1	HHH	3	1	3	4	3
2	HHT	2	1	2	3	2
3	HTH	2	0	0	2	0
4	HTT	1	0	0	1	0
5	THH	2	1	2	3	2
6	THT	1	0	0	1	0
7	TTH	1	0	0	1	0
8	TTT	0	0	0	0	0

Here sample space is.

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

(i) Obviously  $X$  is a r.v. which can take the values 0, 1, 2, and 3

$$P(3) = P(HHH) = (1/2)^3 = 1/8$$

$$P(2) = P[HHT \cup HTH \cup THH]$$

$$= P(HHT) + P(HTH) + P(THH) = 1/8 + 1/8 + 1/8 = 3/8$$

Similarly  $P(1) = 3/8$  and  $P(0) = 1/8$ .

These probabilities could also be obtained directly from the above table 1.

**Table 2**  
**Probability table of X**

Values of X (x)	0	1	2	3
p(x)	1/8	3/8	3/8	1/8

**Table 3****(ii) Probability Table of Y**

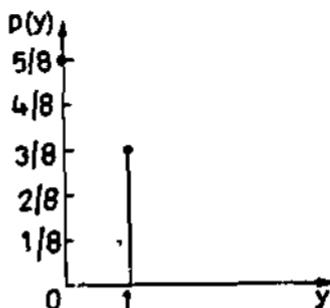
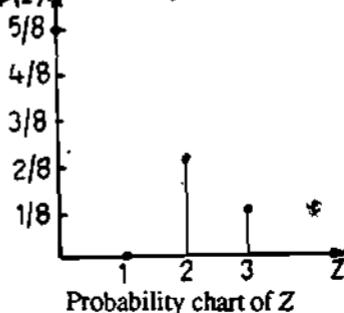
Values of $Y$ , (y)	0    1
$p(y)$	$5/8 \quad 3/8$

This is obvious from table 1.

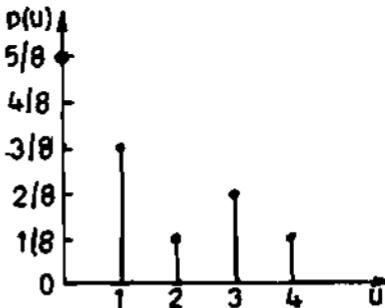
(iii) From table 1 , we have

**Table 4****Probability Table of**

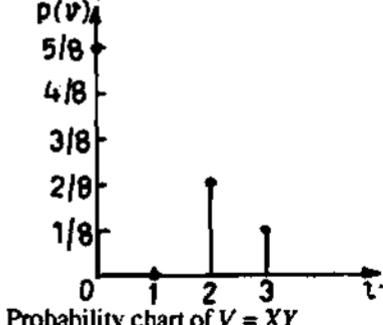
Values of $Z$ , (z)	0    1    2    3
$p(z)$	$5/8 \quad 0 \quad 2/8 \quad 1/8$

Probability chart of  $Y$ Probability chart of  $Z$ (iv) Let  $U = X + Y$ . From table 1, we get**Table 5****Probability Table of U**

Values of $U$ , (u)	0    1    2    3    4
$p(u)$	$1/8 \quad 3/8 \quad 1/8 \quad 2/8 \quad 1/8$

Probability chart of  $U = X+Y$ (v) Let  $V = XY$ **Table 6****Probability Table of V**

Values of $V$ , (v)	0    1    2    3
$p(v)$	$5/8 \quad 0 \quad 2/8 \quad 1/8$

Probability chart of  $V = XY$

**Example 5.2.** A random variable  $X$  has the following probability distribution :

$x:$	0	1	2	3	4	5	6	7
$p(x):$	0	$k$	$2k$	$2k$	$3k$	$k^2$	$2k^2$	$7k^2 + k$

(i) Find  $k$ , (ii) Evaluate  $P(X < 6)$ ,  $P(X \geq 6)$ , and  $P(0 < X < 5)$ , (iii) If  $P(X \leq c) > \frac{1}{2}$ , find the minimum value of  $c$ , and (iv) Determine the distribution function of  $X$ .

[Madurai Univ. B.Sc., Oct. 1988]

**Solution.** Since  $\sum_{x=0}^7 p(x) = 1$ , we have

$$\Rightarrow k + 2k + 2k + 3k + k^2 + 2k^2 + 7k^2 + k = 1$$

$$\Rightarrow 10k^2 + 9k - 1 = 0$$

$$\Rightarrow (10k - 1)(k + 1) = 0 \Rightarrow k = 1/10$$

[ $\because k = -1$ , is rejected, since probability cannot be negative.]

$$(ii) P(X < 6) = P(X = 0) + P(X = 1) + \dots + P(X = 5)$$

$$= \frac{1}{10} + \frac{2}{10} + \frac{2}{10} + \frac{3}{10} + \frac{1}{100} = \frac{81}{100}$$

$$P(X \geq 6) = 1 - P(X < 6) = \frac{19}{100}$$

$$P(0 < X < 5) = P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) = 8k = 4/5$$

$$(iii) P(X \leq c) > \frac{1}{2}. \text{ By trial, we get } c = 4.$$

(iv)	$X$	$F_X(x) = P(X \leq x)$
	0	0
	1	$k = 1/10$
	2	$3k = 3/10$
	3	$5k = 5/10$
	4	$8k = 4/5$
	5	$8k + k^2 = 81/100$
	6	$8k + 3k^2 = 83/100$
	7	$9k + 10k^2 = 1$

### EXERCISE 5 (b)

1. (a) A student is to match three historical events (Mahatma Gandhi's Birthday, India's freedom, and First World War) with three years (1947, 1914, 1896). If he guesses with no knowledge of the correct answers, what is the probability distribution of the number of answers he gets correctly?

(b) From a lot of 10 items containing 3 defectives, a sample of 4 items is drawn at random. Let the random variable  $X$  denote the number of defective items in the sample. Answer the following when the sample is drawn without replacement.

- (i) Find the probability distribution of  $X$ ,  
(ii) Find  $P(X \leq 1)$ ,  $P(X < 1)$  and  $P(0 < X < 2)$

**Ans. (a)**

$x$	0	1	2	3
$p(x)$	$\frac{1}{3}$	$\frac{1}{2}$	0	$\frac{1}{6}$

**(b) (i)**

$x$	0	1	2	3
$p(x)$	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{3}{10}$	$\frac{1}{30}$

**(ii)**  $2/3, 5/6, 1/2$

2. (a) A random variable  $X$  can take all non-negative integral values, and the probability that  $X$  takes the value  $r$  is proportional to  $\alpha^r$  ( $0 < \alpha < 1$ ). Find  $P(X = 0)$ . [Calcutta Univ. B.Sc. 1987]

**Ans.**  $P(X = r) = A\alpha^r$ ;  $r = 0, 1, 2, \dots$ ;  $A = 1 - \alpha$ ;  $P(X = 0) = A = 1 - \alpha$

(b) Suppose that the random variable  $X$  has possible values 1, 2, 3, ... and  $P(X = j) = \frac{1}{2^j}$ ,  $j = 1, 2, \dots$  (i) Compute  $P(X \text{ is even})$ , (ii) Compute  $P(X \geq 5)$ , and (iii) Compute  $P(X \text{ is divisible by } 3)$ .

**Ans.** (i)  $1/3$ , (ii)  $1/16$ , and (iii)  $1/7$ .

3. (a) Let  $X$  be a random variable such that

$$P(X = -2) = P(X = -1), P(X = 2) = P(X = 1) \text{ and}$$

$$P(X > 0) = P(X < 0) = P(X = 0).$$

Obtain the probability mass function of  $X$  and its distribution function.

**Ans.**

$X$	:	-2	-1	0	1	2
$p(x)$	:	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{6}$
$F(x)$	:	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{4}{6}$	$\frac{5}{6}$	1

(b) A random variable  $X$  assumes the values  $-3, -2, -1, 0, 1, 2, 3$  such that

$$P(X = -3) = P(X = -2) = P(X = -1),$$

$$P(X = 1) = P(X = 2) = P(X = 3),$$

and  $P(X = 0) = P(X > 0) = P(X < 0)$ ,

Obtain the probability mass function of  $X$  and its distribution function, and find further the probability mass function of  $Y = 2X^2 + 3X + 4$ .

[Poona Univ. B.Sc., March 1991]

**Ans.**

$X$	:	-3	-2	-1	0	1	2	3
$p(x)$	:	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{3}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$
$Y$	:	13	6	3	4	9	18	31
$p(y)$	:	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{3}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$

4. (a) A random variable  $X$  has the following probability function :

Values of $X, x$ :	-2	-1	0	1	2	3
$p(x)$ :	0.1	$k$	0.2	$2k$	0.3	$k$

(i) Find the value of  $k$ , and calculate mean and variance.

(ii) Construct the c.d.f.  $F(X)$  and draw its graph.

Ans. (i) 0.1, 0.8 and 2.16, (ii)  $F(X) = 0.1, 0.2, 0.4, 0.6, 0.9, 1.0$

(b) Given the probability function

$x$	0	1	2	3
$p(x)$	0.1	0.3	0.5	0.1

Let  $Y = X^2 + 2X$ , then find (i) the probability function of  $Y$ , (ii) mean and variance of  $Y$ .

Ans. (i)	$y$	0	3	8	15			
	$p(y)$	0.1	0.3	0.5	0.1			

5. A random variable  $X$  has the following probability distribution :

Values of $X, x$	0	1	2	3	4	.5	6.	7	8
$p(x)$	$a$	$3a$	$5a$	$7a$	$9a$	$11a$	$13a$	$15a$	$17a$

(i) Determine the value of  $a$ .

(ii) Find  $P(X < 3)$ ,  $P(X \geq 3)$ ,  $P(0 < X < 5)$ .

(iii) What is the smallest value of  $x$  for which  $P(X \leq x) > 0.5$ ?

(iv) Find out the distribution function of  $X$ ?

Ans. (i)  $a = 1/81$ , (ii)  $9/81, 72/81, 24/81$ , (iii) 6

(iv) x	0	1	2	3	4	5	6	7	8
$F(x)$	$a$	$4a$	$9a$	$16a$	$25a$	$36a$	$49a$	$64a$	$81a$

6. (a) Let  $p(x)$  be the probability function of a discrete random variable  $X$  which assumes the values  $x_1, x_2, x_3, x_4$ , such that  $2p(x_1) = 3p(x_2) = p(x_3) = 5p(x_4)$ . Find probability distribution and cumulative probability distribution of  $X$ . (Sardar Patel Univ. B.Sc. 1987)

Ans.	$x$	$x_1$	$x_2$	$x_3$	$x_4$
	$p(x)$	$15/16$	$10/16$	$30/16$	$6/16$

(b) The following is the distribution function of a discrete random variable  $X$  :

$x$	-3	-1	0	1	2	3	5	8
$f(x)$	0.10	0.30	0.45	0.5	0.75	0.90	0.95	1.00

(i) Find the probability distribution of  $X$ .

(ii) Find  $P(X \text{ is even})$  and  $P(1 \leq X \leq 8)$ .

(iii) Find  $P(X = -3 | X < 0)$  and  $P(X \geq 3 | X > 0)$ .

[Ans. (ii) 0.30, 0.55, (iii) 1/3, 5/11].

7. If  $p(x) = \frac{x}{15}$ ;  $x = 1, 2, 3, 4, 5$   
 $\approx 0$ , elsewhere

Find (i)  $P(X = 1 \text{ or } 2)$ , and (ii)  $P\left\{\frac{1}{2} < X < \frac{5}{2} \mid X > 1\right\}$

[Allahabad Univ. B.Sc., April 1992]

Hint. (i)  $P\{X = 1 \text{ or } 2\} = P(X = 1) + P(X = 2) = \frac{1}{15} + \frac{2}{15} = \frac{1}{5}$

(ii)  $P\left\{\frac{1}{2} < X < \frac{5}{2} \mid X > 1\right\} = \frac{P\left\{\left(\frac{1}{2} < X < \frac{5}{2}\right) \cap X > 1\right\}}{P(X > 1)}$

$$= \frac{P\{(X = 1 \text{ or } 2) \cap X > 1\}}{P(X > 1)} = \frac{P(X = 2)}{1 - P(X = 1)} = \frac{\frac{2}{15}}{1 - (\frac{1}{15})} = \frac{1}{7}$$

8. The probability mass function of a random variable  $X$  is zero except at the points  $x = 0, 1, 2$ . At these points it has the values  $p(0) = 3c^3$ ,  $p(1) = 4c - 10c^2$  and  $p(2) = 5c - 1$  for some  $c > 0$ .

(i) Determine the value of  $c$ .

(ii) Compute the following probabilities,  $P(X < 2)$  and  $P(1 < X \leq 2)$ .

(iii) Describe the distribution function and draw its graph.

(iv) Find the largest  $x$  such that  $F(x) < \frac{1}{2}$ .

(v) Find the smallest  $x$  such that  $F(x) \geq \frac{1}{3}$ . [Poona Univ. B.Sc., 1987]

Ans. (i)  $\frac{1}{3}$ , (ii)  $\frac{1}{3}, \frac{2}{3}$ , (iv) 1, (v) 1.

9. (a) Suppose that the random variable  $X$  assumes three values 0, 1 and 2 with probabilities  $\frac{1}{3}, \frac{1}{6}$  and  $\frac{1}{2}$  respectively. Obtain the distribution function of  $X$ .  
[Gujarat Univ. B.Sc., 1992]

(b) Given that  $f(x) = k(1/2)^x$  is a probability distribution for a random variable which can take on the values  $x = 0, 1, 2, 3, 4, 5, 6$ , find  $k$  and find an expression for the corresponding cumulative probabilities  $F(x)$ .

[Nagpur Univ. B.Sc., 1987]

5.4. Continuous Random Variable. A random variable  $X$  is said to be continuous if it can take all possible values between certain limits. In other words, a random variable is said to be continuous when its different values cannot be put in 1-1 correspondence with a set of positive integers.

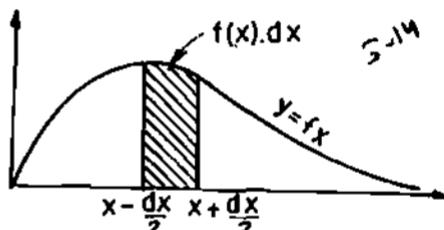
A continuous random variable is a random variable that (at least conceptually) can be measured to any desired degree of accuracy. Examples of continuous random variables are age, height, weight etc.

5.4.1. Probability Density Function (Concept and Definition). Consider the small interval  $(x, x + dx)$  of length  $dx$  round the point  $x$ . Let  $f(x)$  be any continuous

function of  $x$  so that  $f(x) dx$  represents the probability that  $X$  falls in the infinitesimal interval  $(x, x + dx)$ . Symbolically

$$P(x \leq X \leq x + dx) = f_X(x) dx \quad \dots (5.5)$$

In the figure,  $f(x) dx$  represents the area bounded by the curve  $y = f(x)$ ,  $x$ -axis and the ordinates at the points  $x$  and  $x + dx$ . The function  $f_X(x)$  so defined is known as *probability density function or simply density function of random variable  $X$*  and is usually abbreviated as



*p.d.f.* The expression,  $f(x) dx$ , usually written as  $dF(x)$ , is known as the *probability differential* and the curve  $y = f(x)$  is known as the *probability density curve or simply probability curve*.

**Definition.** p.d.f.  $f_X(x)$  of the r.v.  $X$  is defined as :

$$f_X(x) = \lim_{\delta x \rightarrow 0} \frac{P(x \leq X \leq x + \delta x)}{\delta x} \quad \dots (5.5a)$$

The probability for a variate value to lie in the interval  $dx$  is  $f(x) dx$  and hence the probability for a variate value to fall in the finite interval  $[\alpha, \beta]$  is :

$$P(\alpha \leq X \leq \beta) = \int_{\alpha}^{\beta} f(x) dx \quad \dots (5.5b)$$

which represents the area between the curve  $y = f(x)$ ,  $x$ -axis and the ordinates at  $x = \alpha$  and  $x = \beta$ . Further since total probability is unity, we have  $\int_a^b f(x) dx = 1$ , where  $[a, b]$  is the range of the random variable  $X$ . The range of the variable may be finite or infinite.

The probability density function (*p.d.f.*) of a random variable (*r.v.*)  $X$  usually denoted by  $f_X(x)$  or simply by  $f(x)$  has the following obvious properties

$$(i) f(x) \geq 0, -\infty < x < \infty \quad \dots (5.5c)$$

$$(ii) \int_{-\infty}^{\infty} f(x) dx = 1 \quad \dots (5.5d)$$

(iii) The probability  $P(E)$  given by

$$P(E) = \int_E f(x) dx \quad \dots (5.5e)$$

is well defined for any event  $E$ .

**Important Remark.** In case of discrete random variable, the probability at a point, i.e.,  $P(x = c)$  is not zero for some fixed  $c$ . However, in case of continuous random variables the probability at a point is always zero, i.e.,  $P(x = c) = 0$  for all possible values of  $c$ . This follows directly from (5.5b) by taking  $\alpha = \beta = c$ .

This also agrees with our discussion earlier that  $P(E) = 0$  does not imply that the event  $E$  is null or impossible event. This property of continuous r.v., viz.,

$$P(X = c) = 0, \quad \forall c \quad \dots (5.5f)$$

leads us to the following important result :

$P(\alpha \leq X \leq \beta) = P(\alpha \leq X < \beta) = P(\alpha < X \leq \beta) = P(\alpha < X < \beta) \dots (5.5g)$   
i.e., in case of continuous r.v., it does matter whether we include the end points of the interval from  $\alpha$  to  $\beta$ .

However, this result is in general not true for discrete random variables.

**5.4.2. Various Measures of Central Tendency, Dispersion, Skewness, and Kurtosis for Continuous Probability Distribution.** The formulae for these measures in case of discrete frequency distribution can be easily extended to the case of continuous probability distribution by simply replacing  $p_i = f_i/N$  by  $f(x) dx$ ,  $x$  by  $x$  and the summation over 'i' by integration over the specified range of the variable  $X$ .

Let  $f_X(x)$  or  $f(x)$  be the p.d.f. of a random variable  $X$  where  $X$  is defined from  $a$  to  $b$ . Then

$$(i) \quad \text{Arithmetic mean} = \int_a^b x f(x) dx \quad \dots (5.6)$$

(ii) *Harmonic mean.* Harmonic mean  $H$  is given by

$$\frac{1}{H} = \int_a^b \left( \frac{1}{x} \right) f(x) dx \quad \dots (5.6a)$$

(iii) *Geometric mean.* Geometric mean  $G$  is given by

$$\log G = \int_a^b \log x f(x) dx \quad \dots (5.6b)$$

$$(iv) \quad \mu'_1 \text{ (about origin)} = \int_a^b x' f(x) dx \quad \dots (5.7)$$

$$\mu'_r \text{ (about the point } x = A) = \int_a^b (x - A)^r f(x) dx \quad \dots (5.7a)$$

$$\text{and } \mu_r \text{ (about mean)} = \int_a^b (x - \text{mean})^r f(x) dx \quad \dots (5.7b)$$

In particular, from (5.7), we have

$$\mu'_1 \text{ (about origin)} = \text{Mean} = \int_a^b x f(x) dx$$

$$\text{and } \mu'_2 = \int_a^b x^2 f(x) dx$$

$$\text{Hence } \mu_2 = \mu'_2 - \mu_1'^2 = \int_a^b x^2 f(x) dx - \left( \int_a^b x f(x) dx \right)^2 \quad \dots (5.7c)$$

From (5.7), on putting  $r=3$  and  $4$  respectively, we get the values of  $\mu_3'$  and  $\mu_4'$  and consequently the moments about mean can be obtained by using the relations :

$$\text{and } \begin{aligned} \mu_3 &= \mu_3' - 3\mu_2'\mu_1' + 2\mu_1'^3 \\ \mu_4 &= \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'\mu_1'^2 - 3\mu_1'^4 \end{aligned} \quad \left. \right\} \quad \dots (5.7 d)$$

and hence  $\beta_1$  and  $\beta_2$  can be computed.

(v) *Median.* Median is the point which divides the entire distribution in two equal parts. In case of continuous distribution, median is the point which divides the total area into two equal parts. Thus if  $M$  is the median, then

$$\int_a^M f(x) dx = \int_M^b f(x) dx = \frac{1}{2} \quad \dots (5.8)$$

Thus solving

$$\int_a^M f(x) dx = \frac{1}{2} \quad \text{or} \quad \int_M^b f(x) dx = \frac{1}{2} \quad \dots (5.8 a)$$

for  $M$ , we get the value of median.

(vi) *Mean Deviation.* Mean deviation about the mean  $\mu_1'$  is given by

$$M.D. = \int_a^b |x - \text{mean}| f(x) dx \quad \dots (5.9)$$

(vii) *Quartiles and Deciles.*  $Q_1$  and  $Q_3$  are given by the equations

$$\int_a^{Q_1} f(x) dx = \frac{1}{4} \quad \text{and} \quad \int_a^{Q_3} f(x) dx = \frac{3}{4} \quad \dots (5.10)$$

$D_i$ ,  $i$ th decile is given by

$$\int_a^{D_i} f(x) dx = \frac{i}{10} \quad \dots (5.10 a)$$

(viii) *Mode.* Mode is the value of  $x$  for which  $f(x)$  is maximum. Mode is thus the solution of

$$f'(x) = 0 \quad \text{and} \quad f''(x) < 0 \quad \dots (5.11)$$

provided it lies in  $[a, b]$ .

**Example 5.3.** The diameter of an electric cable, say  $X$ , is assumed to be a continuous random variable with p.d.f.  $f(x) = 6x(1-x)$ ,  $0 \leq x \leq 1$ .

(i) Check that above is p.d.f.,

(ii) Determine a number  $b$  such that  $P(X < b) = P(X > b)$

(Aligarh Univ. B.Sc.(Hons).1990)

**Solution.** Obviously, for  $0 \leq x \leq 1$ ,  $f(x) \geq 0$

$$\begin{aligned} \text{Now } \int_0^1 f(x) dx &= 6 \int_0^1 x(1-x) dx \\ &= 6 \int_0^1 (x - x^2) dx = 6 \left[ \frac{x^2}{2} - \frac{x^3}{3} \right]_0^1 = 1 \end{aligned}$$

Hence  $f(x)$  is the p.d.f. of r.v.  $X$

$$(ii) \quad P(X < b) = P(X > b) \quad \dots (*)$$

$$\begin{aligned}
 &\Rightarrow \int_0^b f(x) dx = \int_b^1 f(x) dx \\
 &\Rightarrow 6 \int_0^b x(1-x) dx = 6 \int_b^1 x(1-x) dx \\
 &\Rightarrow \left| \frac{x^2}{2} - \frac{x^3}{3} \right|_0^b = \left| \frac{x^2}{2} - \frac{x^3}{3} \right|_b^1 \\
 &\Rightarrow \left( \frac{b^2}{2} - \frac{b^3}{3} \right) = \left[ \left( \frac{1}{2} - \frac{1}{3} \right) - \left( \frac{b^2}{2} - \frac{b^3}{3} \right) \right] \\
 &\Rightarrow 3b^2 - 2b^3 = [1 - 3b^2 + 2b^3] \\
 &\Rightarrow 4b^3 - 6b^2 + 1 = 0 \\
 &\quad (2b-1)(2b^2-2b+1)=0 \\
 &\Rightarrow 2b-1=0 \quad \text{or} \quad 2b^2-2b+1=0
 \end{aligned}$$

Hence  $b = 1/2$  is the only real value lying between 0 and 1 and satisfying (\*).

**Example 5-4.** A continuous random variable  $X$  has a p.d.f.

$f(x) = 3x^2$ ,  $0 \leq x \leq 1$ . Find  $a$  and  $b$  such that

- (i)  $P\{X \leq a\} = P\{X > a\}$ , and  
(ii)  $P\{X > b\} = 0.05$ . [Calicut Univ. B.Sc., Sept. 1988]

**Solution.** (i) Since  $P\{X \leq a\} = P\{X > a\}$ ,

each must be equal to  $1/2$ , because total probability is always one.

$$\begin{aligned}
 &\therefore P\{X \leq a\} = \frac{1}{2} \Rightarrow \int_0^a f(x) dx = \frac{1}{2} \\
 &\Rightarrow 3 \int_0^a x^2 dx = \frac{1}{2} \Rightarrow 3 \left| \frac{x^3}{3} \right|_0^a = \frac{1}{2} \\
 &\Rightarrow a^3 = \frac{1}{2} \Rightarrow a = \left( \frac{1}{2} \right)^{\frac{1}{3}} \\
 &(ii) \quad P\{X > b\} = 0.05 \Rightarrow \int_b^1 f(x) dx = 0.05 \\
 &\Rightarrow 3 \left| \frac{x^3}{3} \right|_b^1 = \frac{1}{20} \Rightarrow 1 - b^3 = \frac{1}{20} \\
 &\Rightarrow b^3 = \frac{19}{20} \Rightarrow b = \left( \frac{19}{20} \right)^{\frac{1}{3}}.
 \end{aligned}$$

**Example 5-5.** Let  $X$  be a continuous random variate with p.d.f.

$$\begin{aligned}
 f(x) &= ax, \quad 0 \leq x \leq 1 \\
 &= a, \quad 1 \leq x \leq 2 \\
 &= -ax + 3a, \quad 2 \leq x \leq 3 \\
 &= 0, \quad \text{elsewhere}
 \end{aligned}$$

(i) Determine the constant  $a$ .

(ii) Compute  $P(X \leq 1.5)$ . [Sardar Patel Univ. B.Sc., Nov.1988]

**Solution.** (i) Constant 'a' is determined from the consideration that total probability is unity, i.e.,

$$\begin{aligned} & \int_{-\infty}^{\infty} f(x) dx = 1 \\ \Rightarrow & \int_{-\infty}^0 f(x) dx + \int_0^1 f(x) dx + \int_1^2 f(x) dx + \int_2^3 f(x) dx + \int_3^{\infty} f(x) dx = 1 \\ \Rightarrow & \int_0^1 ax dx + \int_1^2 a dx + \int_2^3 (-ax + 3a) dx = 1 \\ \Rightarrow & a \left[ \frac{x^2}{2} \Big|_0^1 + a \left[ x \Big|_1^2 + a \left[ -\frac{x^2}{2} + 3x \Big|_2^3 \right] \right] = 1 \\ \Rightarrow & \frac{a}{2} + a + a \left[ \left( -\frac{9}{2} + 9 \right) - (-2 + 6) \right] = 1 \\ \Rightarrow & \frac{a}{2} + a + \frac{a}{2} = 1 \quad \Rightarrow \quad 2a = 1 \quad \Rightarrow \quad a = \frac{1}{2} \end{aligned}$$

$$\begin{aligned} (ii) P(X \leq 1.5) &= \int_{-\infty}^{1.5} f(x) dx = \int_{-\infty}^0 f(x) dx + \int_0^1 f(x) dx + \int_1^{1.5} f(x) dx \\ &= a \int_0^1 x dx + \int_1^{1.5} a dx \\ &= a \left[ \frac{x^2}{2} \Big|_0^1 + a \left[ x \Big|_1^{1.5} \right] \right] = \frac{a}{2} + 0.5a \\ &= a = \frac{1}{2} \quad [\because a = \frac{1}{2}, \text{ Part (i)}] \end{aligned}$$

**Example 5.6.** A probability curve  $y = f(x)$  has a range from 0 to  $\infty$ . If  $f(x) = e^{-x}$ , find the mean and variance and the third moment about mean.

[Andhra Univ. B.Sc. 1988; Delhi Univ. B.Sc. Sept. 1987]

**Solution.**

$$\begin{aligned} \mu_r \text{ (rth moment about origin)} &= \int_0^{\infty} x^r f(x) dx \\ &= \int_0^{\infty} x^r e^{-x} dx = \Gamma(r+1) = r! \end{aligned}$$

(Using Gamma Integral)

Substituting  $r = 1, 2$  and  $3$  successively, we get

$$\text{Mean} = \mu_1' = 1! = 1, \mu_2' = 2! = 2, \mu_3' = 3! = 6$$

$$\text{Hence variance} = \mu_2 = \mu_2' - \mu_1'^2 = 2 - 1 = 1$$

$$\text{and } \mu_3 = \mu_3' - 3\mu_2' \mu_1' + 2\mu_1'^3 = 6 - 3 \times 2 + 2 = 2$$

**Example 5-7.** In a continuous distribution whose relative frequency density is given by

$$f(x) = y_0 \cdot x(2-x), \quad 0 \leq x \leq 2,$$

find mean, variance,  $\beta_1$ , and  $\beta_2$  and hence show that the distribution is symmetrical. Also (i) find mean deviation about mean and (ii) show that for this distribution  $\mu_{2r+1} = 0$ , (iii) find the mode, harmonic mean and median.

[Delhi Univ. B.Sc.(Stat. Hons.), 1992; B.Sc., Oct. 1992]

**Solution.** Since total probability is unity, we have

$$\begin{aligned} & \int_0^2 f(x) dx = 1 \\ \Rightarrow & y_0 \int_0^2 x(2-x) dx = 1 \Rightarrow y_0 = 3/4 \\ \therefore & f(x) = \frac{3}{4}x(2-x) \end{aligned}$$

$$\mu_r' = \int_0^2 x^r f(x) dx = \frac{3}{4} \int_0^2 x^{r+1}(2-x) dx = \frac{3 \cdot 2^{r+1}}{(r+2)(r+3)}$$

In particular

$$\text{Mean} = \mu_1' = \frac{3 \cdot 2^2}{3 \cdot 4} = 1, \quad \mu_2' = \frac{3 \cdot 2^3}{4 \cdot 5} = \frac{6}{5},$$

$$\mu_3' = \frac{3 \cdot 2^4}{5 \cdot 6} = \frac{8}{5}, \quad \text{and} \quad \mu_4' = \frac{3 \cdot 2^5}{6 \cdot 7} = \frac{16}{7}$$

$$\text{Hence variance} = \mu_2 = \mu_2' - \mu_1'^2 = \frac{6}{5} - 1 = \frac{1}{5}$$

$$\mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2\mu_1'^3 = \frac{8}{5} - 3 \cdot \frac{6}{5} \cdot 1 + 2 = 0$$

$$\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'\mu_1'^2 - 3\mu_1'^4 = \frac{16}{7} - 4 \cdot \frac{8}{5} \cdot 1 + 6 \cdot \frac{6}{5} \cdot 1 - 3 \cdot 1 = \frac{3}{35}$$

$$\therefore \beta_1 = \frac{\mu_3^2}{\mu_2^2} = 0 \quad \text{and} \quad \beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{3/35}{(1/5)^2} = \frac{15}{7}$$

Since  $\beta_1 = 0$ , the distribution is symmetrical.

Mean deviation about mean

$$\begin{aligned} &= \int_0^2 |x-1| f(x) dx \\ &= \int_0^1 |x-1| f(x) dx + \int_1^2 |x-1| f(x) dx \\ &= \frac{3}{4} \left[ \int_0^1 (1-x)x(2-x) dx + \int_1^2 (x-1)x(2-x) dx \right] \end{aligned}$$

$$\begin{aligned}
 &= \frac{3}{4} \left[ \int_0^1 (2x - 3x^2 + x^3) dx + \int_1^2 (3x^2 - x^3 - 2x) dx \right] \\
 &= \frac{3}{4} \left[ \left| x^2 - \frac{3x^3}{3} + \frac{x^4}{4} \right|_0^1 + \left| 3x^2 - \frac{x^3}{3} - \frac{2x^2}{2} \right|_1^2 \right] = \frac{3}{8}
 \end{aligned}$$

$$\begin{aligned}
 \mu_{2n+1} &= \int_0^2 (x - \text{mean})^{2n+1} f(x) dx \\
 &= \frac{3}{4} \int_0^2 (x-1)^{2n+1} x(2-x) dx \\
 &= \frac{3}{4} \int_{-1}^1 t^{2n+1} (t+1)(1-t) dt \quad (x-1=t) \\
 &= \frac{3}{4} \int_{-1}^1 t^{2n+1} (1-t^2) dt
 \end{aligned}$$

Since  $t^{2n+1}$  is an odd function of  $t$  and  $(1-t^2)$  is an even function of  $t$ , the integrand  $t^{2n+1}(1-t^2)$  is an odd function of  $t$ .

Hence  $\mu_{2n+1} = 0$ .

$$\text{Now } f'(x) = \frac{3}{4} (2-2x) = 0 \Rightarrow x = 1$$

$$\text{and } f''(x) = \frac{3}{4} (-2) = -\frac{3}{2} < 0$$

Hence mode = 1

Harmonic mean  $H$  is given by

$$\begin{aligned}
 \frac{1}{H} &= \int_0^2 \frac{1}{x} f(x) dx \\
 &= \frac{3}{4} \int_0^2 (2-x) dx = \frac{3}{2} \\
 \Rightarrow H &= \frac{2}{3}
 \end{aligned}$$

If  $M$  is the median, then

$$\begin{aligned}
 \int_0^M f(x) dx &= \frac{1}{2} \\
 \Rightarrow \frac{3}{4} \int_0^M x(2-x) dx &= \frac{1}{2} \\
 \Rightarrow \left| x^2 - \frac{x^3}{3} \right|_0^M &= \frac{2}{3} \\
 \Rightarrow 3M^2 - M^3 &= 2 \\
 \Rightarrow M^3 - 3M^2 + 2 &= 0 \\
 \Rightarrow (M-1)(M^2 - 2M - 2) &= 0
 \end{aligned}$$

The only value of  $M$  lying in  $[0, 2]$  is  $M = 1$ . Hence median is 1.

Aliter. Since we have proved that distribution is symmetrical,

$$\text{Mode} = \text{Median} = \text{Mean} = 1$$

**Example 5.8.** The elementary probability law of a continuous random variable  $X$  is

$$f(x) = y_0 e^{-b(x-a)}, \quad a \leq x < \infty, \quad b > 0$$

where  $a, b$  and  $y_0$  are constants.

Show that  $y_0 = b = \sqrt{\sigma}$  and  $a = m - \sigma$ , where  $m$  and  $\sigma$  are respectively the mean and standard deviation of the distribution. Show also that  $\beta_1 = 4$  and  $\beta_2 = 9$ . [Gauhati Univ. B.Sc., 1992]

**Solution.** Since total probability is unity,

$$\begin{aligned} \int_a^\infty f(x) dx &= 1 \Rightarrow y_0 \int_a^\infty e^{-b(x-a)} dx = 1 \\ \Rightarrow y_0 \left| \frac{e^{-b(x-a)}}{-b} \right|_a^\infty &= 1 \Rightarrow y_0 \frac{1}{b} = 1, \quad (b > 0) \\ \Rightarrow y_0 &= b \\ \mu'_r \quad (\text{rth moment about the point } 'x = a') &= \int_a^\infty (x-a)^r f(x) dx = b \int_a^\infty (x-a)^r e^{-b(x-a)} dx \\ &= b \int_0^\infty t^r e^{-bt} dt \quad [\text{On putting } x-a=t] \\ &= b \frac{\Gamma(r+1)}{b^{r+1}} = \frac{r!}{b^r} \quad [\text{Using Gamma Integral}] \end{aligned}$$

In particular

$$\mu'_1 = 1/b, \quad \mu'_2 = 2/b^2, \quad \mu'_3 = 6/b^3, \quad \mu'_4 = 24/b^4$$

$$\therefore m = \text{Mean} = a + \mu'_1 = a + (1/b)$$

$$\text{and} \quad \sigma^2 = \mu'_2 - \mu'_1^2 = 1/b^2$$

$$\Rightarrow \sigma = \frac{1}{b} \quad \text{and} \quad m = a + \frac{1}{b} = a + \sigma$$

$$\text{Hence} \quad y_0 = b = \frac{1}{\sigma} \quad \text{and} \quad a = m - \sigma$$

$$\text{Also} \quad \mu_3 = \mu'_3 - 3\mu'_2\mu'_1 + 2\mu'_1^3 = \frac{1}{b^3}(6 - 3 \cdot 2 + 2) = \frac{2}{b^3} = 2\sigma^3$$

$$\text{and} \quad \mu_4 = \mu'_4 - 4\mu'_3\mu'_1 + 6\mu'_2\mu'_1^2 - 3\mu'_1^4$$

$$= \frac{1}{b^4} (24 - 4.6.1 + 6.2.1 - 3) = \frac{9}{b^4} = 9\sigma^4$$

Hence  $\beta_1 = \mu_3/\mu_2^2 = 4\sigma^6/\sigma^6 = 4$  and  $\beta_2 = \mu_4/\mu_2^2 = 9\sigma^4/\sigma^4 = 9$

**Example 5.9.** For the following probability distribution

$$dF = y_o \cdot e^{-|x|} dx, \quad -\infty < x < \infty$$

show that  $y_o = \frac{1}{2}$ ,  $\mu'_1 = 0$ ,  $\sigma = \sqrt{2}$  and mean deviation about mean = 1.

**Solution:** We have  $\int_{-\infty}^{\infty} f(x) dx = 1$

$$\Rightarrow y_o \int_{-\infty}^{\infty} e^{-|x|} dx = 1 \Rightarrow 2y_o \int_0^{\infty} e^{-|x|} dx = 1,$$

(since  $e^{-|x|}$  is an even function of  $x$ )

$$\Rightarrow 2y_o \int_0^{\infty} e^{-x} dx = 1, \quad (\text{since in } 0 \leq x < \infty, |x| = x)$$

$$\Rightarrow 2y_o \left[ \frac{e^{-x}}{-1} \right]_0^{\infty} = 1 \Rightarrow 2y_o = 1, \text{ i.e., } y_o = \frac{1}{2}$$

$$\begin{aligned} \mu'_1 \text{ (about origin)} &= \int_{-\infty}^{\infty} x f(x) dx = \frac{1}{2} \int_{-\infty}^{\infty} x e^{-|x|} dx \\ &= 0, \end{aligned}$$

(since the integrand  $x \cdot e^{-|x|}$  is an odd function of  $x$ )

$$\mu'_2 = \int_{-\infty}^{\infty} x^2 f(x) dx = \frac{1}{2} \int_{-\infty}^{\infty} x^2 e^{-|x|} dx$$

$$= \frac{1}{2} 2 \int_0^{\infty} x^2 e^{-x} dx$$

(since the integrand  $x^2 e^{-x}$  is an even function of  $x$ )

$$\therefore \mu'_2 = \int_0^{\infty} x^2 e^{-x} dx = \Gamma(3) \quad (\text{on using Gamma Integral})$$

$$\Rightarrow \mu'_2 = 2! = 2$$

$$\text{Now } \sigma^2 = \mu_2 = \mu'_2 - \mu'_1^2 = 2$$

$$\text{M.D. about mean} = \int_{-\infty}^{\infty} |x - \text{mean}| f(x) dx$$

$$= \frac{1}{2} \int_{-\infty}^{\infty} |x| e^{-|x|} dx \quad (\because \text{Mean} = \mu'_1 = 0)$$

$$= \frac{1}{2} \cdot 2 \int_0^{\infty} |x| e^{-x} dx$$

$$= \int_0^{\infty} x e^{-x} dx = \Gamma(2) = 1$$

**Example 5-10.** A random variable  $X$  has the probability law :

$$dF(x) = \frac{x}{b^2} e^{-x^2/2b^2} dx, \quad 0 \leq x < \infty$$

Find the distance between the quartiles and show that the ratio of this distance to the standard deviation of  $X$  is independent of the parameter 'b'.

**Solution.** If  $Q_1$  and  $Q_3$  are the first and third quartiles respectively, we have

$$\int_0^{Q_1} f(x) dx = \frac{1}{4} \Rightarrow \frac{1}{b^2} \int_0^{Q_1} x e^{-x^2/2b^2} dx = \frac{1}{4}$$

$$\text{Put } y = \frac{x^2}{2b^2} \quad \text{then} \quad dy = \frac{x}{b^2} dx$$

$$\therefore \int_0^{Q_1^2/2b^2} e^{-y} dy = \frac{1}{4} \Rightarrow \left| \frac{e^{-y}}{-1} \right|_0^{Q_1^2/2b^2} = \frac{1}{4}$$

$$\Rightarrow 1 - e^{-Q_1^2/2b^2} = \frac{1}{4} \Rightarrow e^{-Q_1^2/2b^2} = \frac{3}{4}$$

$$\Rightarrow Q_1 = \sqrt{2b \sqrt{\log(4/3)}}$$

Again we have  $\int_0^{Q_3} f(x) dx = \frac{3}{4}$  which, on proceeding similarly, will give

$$1 - e^{-Q_3^2/2b^2} = 3/4 \Rightarrow e^{-Q_3^2/2b^2} = 1/4 \\ \Rightarrow Q_3 = \sqrt{2b \sqrt{\log(4)}}$$

The distance between the quartiles is given by

$$Q_3 - Q_1 = \sqrt{2b [\sqrt{\log 4} - \sqrt{\log(4/3)}]}$$

$$\mu'_1 = \int_0^\infty x f(x) dx = \int_0^\infty x \frac{x}{b^2} e^{-x^2/2b^2} dx \\ = \int_0^\infty \sqrt{2by^{1/2}} e^{-y} dy \quad \left( y = \frac{x^2}{2b^2} \right)$$

$$= \sqrt{2b} \int_0^\infty e^{-y} y^{3/2-1} dy \\ = \sqrt{2b} \Gamma\left(\frac{3}{2}\right) = \sqrt{2} \cdot b \cdot \frac{1}{2} \Gamma\left(\frac{1}{2}\right) = \sqrt{2b} \frac{\sqrt{\pi}}{2} = b \sqrt{(\pi/2)}$$

$$\mu'_2 = \int_0^\infty x^2 f(x) dx = \int_0^\infty x^2 \frac{x}{b^2} e^{-x^2/2b^2} dx \\ = 2b^2 \int_0^\infty y e^{-y} dy \quad \left( y = \frac{x^2}{2b^2} \right)$$

$$= 2b^2 \Gamma(2) = 2b^2 \cdot 1! = 2b^2$$

$$\therefore \sigma^2 = \mu_2 - \mu_1'^2 = 2b^2 - b^2 \cdot \frac{\pi}{2} = b^2 \left( 2 - \frac{\pi}{2} \right)$$

$$\Rightarrow \sigma = b \sqrt{2 - (\pi/2)}$$

$$\text{Hence } \frac{Q_3 - Q_1}{\sigma} = \frac{\sqrt{2} [\sqrt{\log 4} - \sqrt{\log (4/3)}]}{\sqrt{2 - (\pi/2)}},$$

which is independent of the parameter 'b'.

**Example 5.11.** Prove that the geometric mean  $G$  of the distribution

$$dF = 6(2-x)(x-1)dx, \quad 1 \leq x \leq 2$$

is given by  $6 \log (16G) = 19$ .

(Kanpur Univ. B.Sc., Oct. 1992)

**Solution.** By definition, we have

$$\begin{aligned} \log G &= \int_1^2 \log x f(x) dx = 6 \int_1^2 \log x (2-x)(x-1) dx \\ &= -6 \int_1^2 (x^2 - 3x + 2) \log x dx \end{aligned}$$

Integrating by parts, we get

$$\begin{aligned} \log G &= -6 \left[ \left| \left( \frac{x^3}{3} - \frac{3x^2}{2} + 2x \right) \log x \right|_1^2 \right. \\ &\quad \left. - \int_1^2 \left( \frac{x^3}{3} - \frac{3x^2}{2} + 2x \right) \frac{1}{x} dx \right] \\ &= -4 \log 2 + 6 \times \frac{19}{36} \quad \text{(on simplification)} \end{aligned}$$

$$\therefore \log G + 4 \log 2 = \frac{19}{6} \Rightarrow \log G + \log 2^4 = \frac{19}{6}$$

$$\Rightarrow \log G + \log 16 = \frac{19}{6} \Rightarrow \log (16G) = \frac{19}{6}$$

$$\Rightarrow 6 \log (16G) = 19$$

**Example 5.12.** The time one has to wait for a bus at a downtown bus stop is observed to be random phenomenon ( $X$ ) with the following probability density function :

$$\begin{aligned} f_X(x) &= 0, \quad \text{for } x < 0 \\ &= \frac{1}{9}(x+1), \quad \text{for } 0 \leq x < 1 \\ &= \frac{4}{9}(x - \frac{1}{2}), \quad \text{for } 1 \leq x < \frac{3}{2} \\ &= \frac{4}{9}(\frac{5}{2} - x), \quad \text{for } \frac{3}{2} \leq x < 2 \\ &= \frac{1}{9}(4 - x), \quad \text{for } 2 \leq x < 3 \\ &= \frac{1}{9}, \quad \text{for } 3 \leq x < 6 \end{aligned}$$

$$= 0, \quad \text{for } 6 \leq x,$$

Let the events A and B be defined as follows :

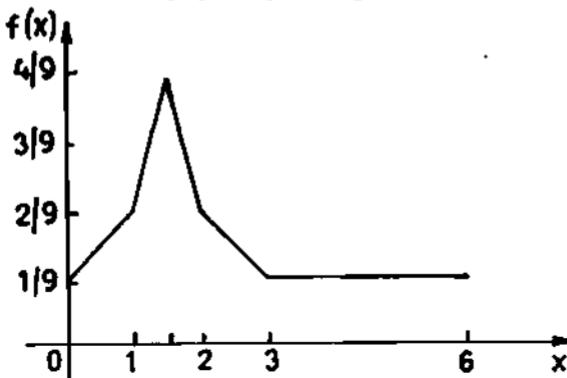
A : One waits between 0 to 2 minutes inclusive.

B : One waits between 0 to 3 minutes inclusive.

(i) Draw the graph of probability density function.

$$(ii) \text{ Show that (a) } P(B|A) = \frac{2}{3}, \text{ (b) } P(\bar{A} \cap \bar{B}) = \frac{1}{3}$$

**Solution.** (i) The graph of p.d.f. is given below.



$$\begin{aligned} (ii) (a) \quad P(A) &= \int_0^2 f(x) dx = \int_0^1 \frac{1}{9}(x+1) dx + \int_1^2 \frac{4}{9}\left(x - \frac{1}{2}\right) dx \\ &\quad + \int_2^3 \frac{4}{9}\left(\frac{5}{2} - x\right) dx \\ &= \frac{1}{2} \quad (\text{on simplification}) \end{aligned}$$

$$\begin{aligned} P(A \cap B) &= P(1 \leq X \leq 2) = \int_1^2 f(x) dx \\ &= \int_1^{3/2} \frac{4}{9}\left(x - \frac{1}{2}\right) dx + \int_{3/2}^2 \frac{4}{9}\left(\frac{5}{2} - x\right) dx \\ &= \frac{4}{9} \left[ \frac{x^2}{2} - \frac{x}{2} \right]_1^{3/2} + \frac{4}{9} \left[ \frac{5}{2}x - \frac{x^2}{2} \right]_{3/2}^2 = \frac{1}{3} \\ &\quad (\text{on simplification}) \end{aligned}$$

$$\therefore P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{1/3}{1/2} = \frac{2}{3}$$

(b)  $\bar{A} \cap \bar{B}$  means that waiting time is more than 3 minutes.

$$\begin{aligned} \therefore P(\bar{A} \cap \bar{B}) &= P(X > 3) = \int_3^\infty f(x) dx = \int_3^6 f(x) dx + \int_6^\infty f(x) dx \\ &= \int_3^6 \frac{1}{9} dx = \frac{1}{9} \left[ x \right]_3^6 = \frac{1}{3} \end{aligned}$$

**Example 5-13.** The amount of bread (in hundreds of pounds)  $X$  that a certain bakery is able to sell in a day is found to be a numerical valued random phenomenon, with a probability function specified by the probability density function  $f(x)$ , given by

$$\begin{aligned}f(x) &= A \cdot x, && \text{for } 0 \leq x < 5 \\&= A(10 - x), && \text{for } 5 \leq x < 10 \\&= 0, && \text{otherwise}\end{aligned}$$

- (a) Find the value of  $A$  such that  $f(x)$  is a probability density function.  
 (b) What is the probability that the number of pounds of bread that will be sold tomorrow is

(i) more than 500 pounds,

(ii) less than 500 pounds,

(iii) between 250 and 750 pounds? [Agra Univ. B.Sc., 1989]

(c) Denoting by  $A, B, C$  the events that the pounds of bread sold are as in (i), (ii) and (iii) respectively, find  $P(A|B), P(A|C)$ . Are (i)  $A$  and  $B$  independent events? (ii) Are  $A$  and  $C$  independent events?

**Solution.** (a) In order that  $f(x)$  should be a probability density function

$$\int_{-\infty}^{\infty} f(x) dx = 1$$

$$\text{i.e., } \int_0^5 A x dx + \int_5^{10} A(10 - x) dx = 1$$

$$\Rightarrow A = \frac{1}{25} \quad (\text{On simplification})$$

(b) (i) The probability that the number of pounds of bread that will be sold tomorrow is more than 500 pounds, i.e.,

$$\begin{aligned}P(5 \leq X \leq 10) &= \int_5^{10} \frac{1}{25} (10 - x) dx = \frac{1}{25} \left| 10x - \frac{x^2}{2} \right|_5^{10} \\&= \frac{1}{25} \left( \frac{25}{2} \right) = \frac{1}{2} = 0.5\end{aligned}$$

(ii) The probability that the number of pounds of bread that will be sold tomorrow is less than 500 pounds, i.e.,

$$P(0 \leq X \leq 5) = \int_0^5 \frac{1}{25} \cdot x dx = \frac{1}{25} \left| \frac{x^2}{2} \right|_0^5 = \frac{1}{2} = 0.5$$

(iii) The required probability is given by

$$P(2.5 \leq X \leq 7.5) = \int_{2.5}^5 \frac{1}{25} x dx + \int_5^{7.5} \frac{1}{25} (10 - x) dx = \frac{3}{4}$$

(c) The events  $A, B$  and  $C$  are given by

$$A : 5 < X \leq 10; \quad B : 0 \leq X < 5; \quad C : 2.5 < X < 7.5$$

Then from parts *b* (*i*), (*ii*) and (*iii*), we have

$$P(A) = 0.5, \quad P(B) = 0.5, \quad P(C) = \frac{3}{4}$$

The events  $A \cap B$  and  $A \cap C$  are given by

$$A \cap B = \emptyset \text{ and } A \cap C : 5 < X < 7.5$$

$$\therefore P(A \cap B) = P(\emptyset) = 0$$

$$\begin{aligned} \text{and } P(A \cap C) &= \int_5^{7.5} f(x) dx = \frac{1}{25} \int_5^{7.5} (10-x) dx \\ &= \frac{1}{25} \times \frac{75}{8} = \frac{3}{8} \\ P(A) \cdot P(C) &= \frac{1}{2} \times \frac{3}{4} = \frac{3}{8} = P(A \cap C) \end{aligned}$$

$\Rightarrow A$  and  $C$  are independent.

$$\text{Again } P(A) \cdot P(B) = \frac{1}{4} \neq P(A \cap B)$$

$\Rightarrow A$  and  $B$  are not independent.

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = 0$$

$$P(A|C) = \frac{P(A \cap C)}{P(C)} = \frac{3/8}{3/4} = \frac{1}{2}$$

**Example 5.14.** The mileage  $C$  in thousands of miles which car owners get with a certain kind of tyre is a random variable having probability density function

$$\begin{aligned} f(x) &= \frac{1}{20} e^{-x/20}, \text{ for } x > 0 \\ &= 0, \text{ for } x \leq 0 \end{aligned}$$

Find the probabilities that one of these tyres will last

- (i) at most 10,000 miles,
- (ii) anywhere from 16,000 to 24,000 miles.
- (iii) at least 30,000 miles.

(Bombay Univ. B.Sc. 1989)

**Solution.** Let r.v.  $X$  denote the mileage (in '000 miles) with a certain kind of tyre. Then required probability is given by:

$$\begin{aligned} (i) \quad P(X \leq 10) &= \int_0^{10} f(x) dx = \frac{1}{20} \int_0^{10} e^{-x/20} dx \\ &= \frac{1}{20} \left[ \frac{e^{-x/20}}{-1/20} \right]_0^{10} = 1 - e^{-1/2} \\ &= 1 - 0.6065 = 0.3935 \end{aligned}$$

$$(ii) P(16 \leq X \leq 24) = \frac{1}{20} \int_{16}^{24} \exp\left(-\frac{x}{20}\right) dx = \left| -e^{-x/20} \right|_{16}^{24} \\ = e^{-16/20} - e^{-24/20} = e^{-4/5} - e^{-6/5} \\ = 0.4493 - 0.3012 = 0.1481$$

$$(iii) P(X \geq 30) = \int_{30}^{\infty} f(x) dx = \frac{1}{20} \left| \frac{e^{-x/20}}{-1/20} \right|_{30}^{\infty} \\ = e^{-15} = 0.2231$$

### EXERCISE 5 (c)

1. (a) A continuous random variable  $X$  follows the probability law

$$f(x) = Ax^2, 0 \leq x \leq 1$$

Determine  $A$  and find the probability that (i)  $X$  lies between 0.2 and 0.5, (ii)  $X$  is less than 0.3, (iii)  $1/4 < X < 1/2$  and (iv)  $X > 3/4$  given  $X > 1/2$ .

Ans.  $A = 0.3$ , (i) 0.117, (ii) 0.027, (iii) 15/256 and (iv) 27/56.

(b) If a random variable  $X$  has the density function

$$f(x) = \begin{cases} 1/4, & -2 < x < 2 \\ 0, & \text{elsewhere} \end{cases}$$

Obtain (i)  $P(X < 1)$ , (ii)  $P(|X| > 1)$  (iii)  $P(2X + 3 > 5)$

(Kerala Univ. B.Sc., Sept. 1992)

Hint. (ii)  $P(|X| > 1) = P(X > 1 \text{ or } X < -1) = \int_{-2}^{-1} f(x) dx + \int_1^2 f(x) dx$

or  $P(|X| > 1) = 1 - P(|X| \leq 1) = 1 - P(-1 \leq X \leq 1)$

Ans. (i) 3/4, (ii) 1/2 (iii) 1/4.

2. Are any of the following probability mass or density functions?

Prove your answer in each case.

$$(a) f(x) = x; x = \frac{1}{16}, \frac{3}{16}, \frac{1}{4}, \frac{1}{2}$$

$$(b) f(x) = \lambda e^{-\lambda x}; x \geq 0; \lambda > 0$$

$$(c) f(x) = \begin{cases} 2x, & 0 < x < 1 \\ 4 - 2x, & 1 < x < 2 \\ 0, & \text{elsewhere} \end{cases}$$

(Calicut Univ. B. Sc., Oct. 1989)

Ans. (a) and (b) are p.m.f./p.d.f.'s, (c) is not.

3. If  $f_1$  and  $f_2$  are p.d.f.'s and  $\theta_1 + \theta_2 = 1$ , check if,

$$g(x) = \theta_1 f_1(x) + \theta_2 f_2(x), \text{ is a p.d.f.}$$

Ans.  $g(x)$  is a p.d.f. if  $0 \leq (\theta_1, \theta_2) \leq 1$  s.t.  $\theta_1 + \theta_2 = 1$ .

4. A continuous random variable  $X$  has the probability density function :

$$f(x) = A + Bx, \quad 0 \leq x \leq 1.$$

If the mean of the distribution is  $\frac{1}{2}$ , find  $A$  and  $B$ .

**Hint :** Solve  $\int_0^1 f(x) dx = 1$  and  $\int_0^1 x f(x) dx = \frac{1}{2}$ . Find  $A$  and  $B$ .

5. For the following density function

$$f(x) = c x^2 (1-x), \quad 0 < x < 1,$$

find (i) the constant  $c$ , and (ii) mean.

[Calicut Univ. B.Sc.(subs.), 1991]

**Ans.** (i)  $c = 12$ ; (ii) mean =  $3/5$ .

6. A continuous distribution of a variable  $X$  in the range  $(-3, 3)$  is defined by

$$\begin{aligned} f(x) &= \frac{1}{16} (3+x)^2, \quad -3 \leq x \leq -1 \\ &= \frac{1}{16} (6-2x^2), \quad -1 \leq x \leq 1 \\ &= \frac{1}{16} (3-x)^2, \quad 1 \leq x \leq 3 \end{aligned}$$

(i) Verify that the area under the curve is unity.

(ii) Find the mean and variance of the above distribution.

(Madras Univ. B.Sc., Oct. 1992; Gujarat Univ. B.Sc., Oct. 1986)

**Hint:**  $\int_{-3}^3 f(x) dx = \int_{-3}^{-1} f(x) dx + \int_{-1}^1 f(x) dx + \int_1^3 f(x) dx$

**Ans.** Mean=0, Variance=1

7. If the random variable  $X$  has the p.d.f.,

$$\begin{aligned} f(x) &= \frac{1}{2} (x+1), \quad -1 < x < 1 \\ &= 0, \text{ elsewhere,} \end{aligned}$$

Find the coefficient of skewness and kurtosis.

8. (a) A random variable  $X$  has the probability density function given by

$$f(x) = 6x(1-x), \quad 0 \leq x \leq 1$$

Find the mean  $\mu$ , mode and S.D.  $\sigma$ . Compute  $P(\mu - 2\sigma < X < \mu + 2\sigma)$ .

Find also the mean deviation about the median.

(Lucknow Univ. B.Sc., 1988)

(b) For the continuous distribution

$$dF = y_o(x - x^2) dx; \quad 0 \leq x \leq 1, \quad y_o \text{ being a constant.}$$

Find (i) arithmetic mean, (ii) harmonic mean, (iii) Median, (iv) Mode and (v)  $r$ th moment about mean. Hence find  $\beta_1$  and  $\beta_2$  and show that the distribution is symmetrical.

(Delhi Univ. B.Sc., 1992; Karnataka Univ. B.Sc., 1991)

**Ans.** Mean = Median = Mode =  $\frac{1}{2}$

(c) Find the mean, mode and median for the distribution,

$$dF(x) = \sin x dx, \quad 0 \leq x \leq \pi/2$$

Ans. 1,  $\pi/2$ ,  $\pi/3$

9. If the function  $f(x)$  is defined by

$$f(x) = c e^{-\alpha x}, \quad 0 \leq x < \infty, \quad \alpha > 0$$

(i) Find the value of constant  $c$ .

(ii) Evaluate the first four moments about mean.

[Gauhati Univ. B.Sc. 1987]

Ans. (i)  $c = \alpha$ , (ii) 0,  $1/\alpha^2$ ,  $2/\alpha^3$ ,  $9/\alpha^4$ .

10. (a) Show that for the exponential distribution

$$dP = y_0 e^{-x/\sigma} dx, \quad 0 \leq x < \infty, \quad \sigma > 0$$

the mean and S.D. are both equal to  $\sigma$  and that the interquartile range is  $\sigma \log_e 3$ . Also find  $\mu_r'$  and show that  $\beta_1 = 4$ ,  $\beta_2 = 9$ .

[Agra Univ. B.Sc., 1986 ; Madras Univ. B.Sc., 1987]

(b) Define the harmonic mean (H.M.) of variable  $X$  as the reciprocal of the expected value of  $1/X$ , show that the H.M. of variable which ranges from 0 to  $\infty$  with probability density  $\frac{1}{6} x^3 e^{-x}$  is 3.

11. (a) Find the mean, variance and the co-efficients  $\beta_1$ ,  $\beta_2$  of the distribution,

$$dF = k x^2 e^{-x} dx, \quad 0 < x < \infty.$$

Ans.  $k = 1/2$ ; 3, 3,  $4/3$  and 5.

(b) Calculate  $\beta_1$  for the distribution,

$$dF = k x e^{-x} dx, \quad 0 < x < \infty$$

Ans. 2

[Delhi Univ. B.Sc. (Hons. Subs.), 1988]

12. A continuous random variable  $X$  has a p.d.f. given by

$$f(x) = k x e^{-\lambda x}, \quad x \geq 0, \quad \lambda > 0$$

$$= 0, \text{ otherwise}$$

Determine the constant  $k$ , obtain the mean and variance of  $X$ .

[Nagpur Univ. B.Sc. 1990]

13. For the probability density function,

$$\begin{aligned} f(x) &= \frac{2(b+x)}{b(a+b)}, \quad -b \leq x < 0 \\ &= \frac{2(a-x)}{a(a+b)}, \quad 0 \leq x \leq a \end{aligned}$$

Find mean, median and variance.

[Calcutta Univ. B.Sc. 1984]

Ans. Mean =  $(a-b)/3$ , Variance =  $(a^2 + b^2 + ab)/18$ ,

$$\text{Median} = a - \sqrt{a(a+b)/2}$$

(ii) Show that, if terms of order  $(a-b)^2/a^2$  are neglected, then  
 $\text{mean} - \text{median} = (\text{mean} - \text{mode})/4$

14. A variable  $X$  can assume values only between 0 and 5 and the equation of its frequency curve is

$$y = A \sin \frac{1}{5} \pi x, \quad 0 \leq x \leq 5$$

where  $A$  is a constant such that the area under the curve is unity. Determine the value of  $A$  and obtain the median and quartiles of the distribution.

Show also that the variance of the distribution is  $50 \left\{ \frac{1}{8} - \frac{1}{\pi^2} \right\}$ .

Ans. 1/10, 2.5, 4/3, 10/3

15. A continuous variable  $X$  is distributed over the interval [0, 1] with p.d.f.  $ax^2 + bx$ , where  $a, b$  are constants. If the arithmetic mean of  $X$  is 0.5, find the values of  $a$  and  $b$ .

Ans. -6, 6

16. A man leaves his house at the same time every morning and the time taken to journey to work has the following probability density function : less than 30 minutes, zero, between 30 minutes and 60 minutes, uniform with density  $k$ ; between 60 minutes and 70 minutes, uniform with density  $2k$ ; and more than 70 minutes, zero. What is the probability that on one particular day he arrives at work later than on the previous day but not more than 5 minutes later.

17. The density function of sheer strength of spot welds is given by

$$\begin{aligned} f(x) &= x/160,000 \quad \text{for } 0 \leq x \leq 400 \\ &= (800-x)/160,000 \quad \text{for } 400 \leq x \leq 800 \end{aligned}$$

Find the number  $a$  such that

$\text{Prob. } (X < a) = 0.50$  and the number  $b$  such that

$\text{Prob. } (X < b) = 0.90$ . Find the mean, median and variance of  $X$ .

[Delhi Univ. B.E., 1987]

18. A batch of small calibre ammunition is accepted as satisfactory if none of a sample of five shot falls more than 2 feet from the centre of the target at a given range. If  $X$ , the distance from the centre of the target to a given impact point, actually has the density

$$f(x) = k \cdot 2x e^{-x^2}, \quad 0 < x < 3$$

where  $k$  is a number which makes it probability density function, what is the value of  $k$  and what is the probability that the batch will be accepted? \*

[Nagpur Univ. B.E., 1987]

$$\text{Hint. } \int_0^3 f(x) dx = 1 \quad \Rightarrow \quad k = 1/(1 - e^{-9})$$

Reqd. Prob. =  $P$  [ Each of a sample of 5 shots falls within a distance of 2 ft. from the centre ]

$$= [P(0 < X < 2)]^5 = \left[ \int_0^2 f(x) dx \right]^5 = \left[ \frac{1 - e^{-4}}{1 - e^{-9}} \right]^5$$

19. A random variable  $X$  has the p.d.f. :

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

Find (i)  $P\left(X < \frac{1}{2}\right)$ , (ii)  $P\left(\frac{1}{4} < X < \frac{1}{2}\right)$ , (iii)  $P\left(X > \frac{3}{4} \mid X > \frac{1}{2}\right)$ , and  
 (iv)  $P\left(X < \frac{3}{4} \mid X > \frac{1}{2}\right)$ .

(Gorakhpur Univ. B.Sc., 1988)

Ans. (i)  $1/4$ , (ii)  $3/16$ , (iii)  $\frac{P(X > 3/4)}{P(X > 1/2)} = \frac{7/16}{3/4} = \frac{7}{12}$ ; (iv)  $\frac{P(1/2 < X < 3/4)}{P(X > 1/2)}$

**5.4.3. Continuous Distribution Function.** If  $X$  is a continuous random variable with the p.d.f.  $f(x)$ , then the function

$$F_X(x) = P(X \leq x) = \int_{-\infty}^x f(t) dt, \quad -\infty < x < \infty. \quad \dots(5.12)$$

is called the *distribution function* (d.f.) or sometimes the *cumulative distribution function* (c.d.f.) of the random variable  $X$ .

**Remarks 1.**  $0 \leq F(x) \leq 1, \quad -\infty < x < \infty$ .

2. From analysis (Riemann integral), we know that

$$F'(x) = \frac{d}{dx} F(x) = f(x) \geq 0 \quad [\because f(x) \text{ is p.d.f.}]$$

$\Rightarrow F(x)$  is non-decreasing function of  $x$ .

$$3. \quad F(-\infty) = \lim_{x \rightarrow -\infty} F(x) = \lim_{x \rightarrow -\infty} \int_{-\infty}^x f(x) dx = \int_{-\infty}^{-\infty} f(x) dx = 0$$

$$\text{and } F(+\infty) = \lim_{x \rightarrow \infty} F(x) = \lim_{x \rightarrow \infty} \int_{-\infty}^x f(x) dx = \int_{-\infty}^{\infty} f(x) dx = 1$$

4.  $F(x)$  is a continuous function of  $x$  on the right.

5. The discontinuities of  $F(x)$  are at the most countable.

6. It may be noted that

$$\begin{aligned} P(a \leq X \leq b) &= \int_a^b f(x) dx = \int_{-\infty}^b f(x) dx - \int_{-\infty}^a f(x) dx \\ &= P(X \leq b) - P(X \leq a) = F(b) - F(a) \end{aligned}$$

Similarly

$$P(a < X < b) = P(a < X \leq b) = P(a \leq X < b) = \int_a^b f(t) dt$$

7. Since  $F'(x) = f(x)$ , we have

$$\frac{d}{dx} F(x) = f(x) \Rightarrow dF(x) = f(x) dx$$

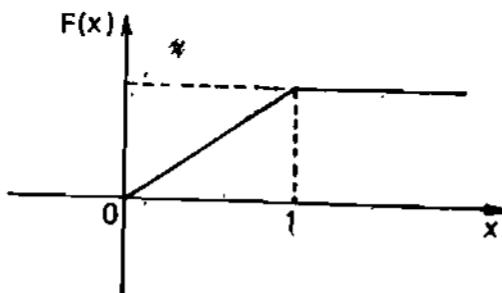
This is known as probability differential of  $X$ .

**Remarks.** 1. It may be pointed out that the properties (2), (3) and (4) above uniquely characterise the distribution functions. This means that any function  $F(x)$  satisfying (2) to (4) is the distribution function of some random variable, and any function  $F(x)$  violating any one or more of these three properties cannot be the distribution function of any random variable.

2. Often, one can obtain a p.d.f. from a distribution function  $F(x)$  by differentiating  $F(x)$ , provided the derivative exists. For example, consider

$$F_x(x) = \begin{cases} 0, & \text{for } x < 0 \\ x, & \text{for } 0 \leq x \leq 1 \\ 1, & \text{for } x > 1 \end{cases}$$

The graph of  $F(x)$  is given by bold lines. Obviously we see that  $F(x)$  is continuous from right as stipulated in (4) and we also see that  $F(x)$  is not continuous at  $x = 0$  and  $x = 1$  and hence is not derivable at  $x = 0$  and  $x = 1$ .



Differentiating  $F(x)$  w.r.t.  $x$ , we get

$$\frac{d}{dx} F(x) = \begin{cases} 1, & 0 < x < 1, \\ 0, & \text{otherwise} \end{cases}$$

[Note the strict inequality in  $0 < x < 1$ , since  $F(x)$  is not derivable at  $x = 0$  and  $x = 1$ ]

Let us define

$$f(x) = \begin{cases} 1, & 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}$$

Then  $f(x)$  is a p.d.f. for  $F$ .

**Example 5.15.** Verify that the following is a distribution function:

$$F(x) = \begin{cases} 0, & x < -a \\ \frac{1}{2} \left( \frac{x}{a} + 1 \right), & -a \leq x \leq a \\ 1, & x > a \end{cases}$$

(Madras Univ. B.Sc., 1992)

**Solution.** Obviously the properties (i), (ii), (iii) and (iv) are satisfied. Also we observe that  $F(x)$  is continuous at  $x = a$  and  $x = -a$ , as well.

Now

$$\begin{aligned} \frac{d}{dx} F(x) &= \begin{cases} \frac{1}{2a}, & -a \leq x \leq a \\ 0, & \text{otherwise} \end{cases} \\ &= f(x), \text{ say} \end{aligned}$$

In order that  $F(x)$  is a distribution function,  $f(x)$  must be a p.d.f. Thus we have to show that

$$\int_{-\infty}^{\infty} f(x) dx = 1$$

$$\text{Now } \int_{-\infty}^{\infty} f(x) dx = \int_{-a}^a f(x) dx = \frac{1}{2a} \int_{-a}^a 1 dx = 1$$

Hence  $F(x)$  is a d.f.

**Example 5.16.** Suppose the life in hours of a certain kind of radio tube has the probability density function :

$$\begin{aligned} f(x) &= \frac{100}{x^2}, \text{ when } x \geq 100 \\ &= 0, \text{ when } x < 100 \end{aligned}$$

Find the distribution function of the distribution. What is the probability that none of three such tubes in a given radio set will have to be replaced during the first 150 hours of operation? What is the probability that all three of the original tubes will have been replaced during the first 150 hours? (Delhi Univ. B.Sc., Oct. 1988)

**Solution.** Probability that a tube will last for first 150 hours is given by

$$\begin{aligned} P(X \leq 150) &= P(0 < X < 100) + P(100 \leq X \leq 150) \\ &= \int_{100}^{150} f(x) dx = \int_{100}^{150} \frac{100}{x^2} dx = \frac{1}{3} \end{aligned}$$

Hence the probability that none of the three tubes will have to be replaced during the first 150 hours is  $(1/3)^3 = 1/27$ .

The probability that a tube will not last for the first 150 hours is  $1 - \frac{1}{3} = \frac{2}{3}$ .

Hence the probability that all three of the original tubes will have to be replaced during the first 150 hours is  $(2/3)^3 = 8/27$ .

**Example 5.17.** Suppose that the time in minutes that a person has to wait at a certain station for a train is found to be a random phenomenon, a probability function specified by the distribution function,

$$\begin{aligned}F(x) &= 0, \text{ for } x \leq 0 \\&= \frac{x}{2}, \text{ for } 0 \leq x < 1 \\&= \frac{1}{2}, \text{ for } 1 \leq x < 2 \\&= \frac{x}{4}, \text{ for } 2 \leq x < 4 \\&= 1, \text{ for } x \geq 4\end{aligned}$$

(a) Is the Distribution Function continuous? If so, give the formula for its probability density function?

(b) What is the probability that a person will have to wait (i) more than 3 minutes, (ii) less than 3 minutes, and (iii) between 1 and 3 minutes?

(c) What is the conditional probability that the person will have to wait for a train for (i) more than 3 minutes, given that it is more than 1 minute, (ii) less than 3 minutes given that it is more than 1 minute? (Calicut Univ. B.Sc., 1985)

**Solution.** (a) Since the value of the distribution function is the same at the points  $x = 0, x = 1, x = 2$ , and  $x = 4$  given by the two forms of  $F(x)$  for  $x < 0$  and  $0 \leq x < 1$ ,  $0 \leq x < 1$  and  $1 \leq x < 2$ ,  $1 \leq x < 2$  and  $2 \leq x < 4$ ,  $2 \leq x < 4$  and  $x \geq 4$ , the distribution function is continuous.

$$\text{Probability density function} = f(x) = \frac{d}{dx} F(x)$$

$$\begin{aligned}\therefore f(x) &= 0, \text{ for } x < 0 \\&= \frac{1}{2}, \text{ for } 0 \leq x < 1 \\&= 0, \text{ for } 1 \leq x < 2 \\&= \frac{1}{4}, \text{ for } 2 \leq x < 4 \\&= 0, \text{ for } x \geq 4\end{aligned}$$

(b) Let the random variable  $X$  represent the waiting time in minutes.

Then

$$(i) \text{ Required probability} = P(X > 3) = 1 - P(X \leq 3) = 1 - F(3)$$

$$= 1 - \frac{1}{4} \cdot 3 = \frac{1}{4}$$

$$(ii) \text{ Required probability} = P(X < 3) = P(X \leq 3) + P(X = 3)$$

$$= F(3) = \frac{3}{4}$$

(Since, the probability that a continuous variable takes a fixed value is zero)

$$(iii) \text{ Required Probability} = P(1 < X < 3) = P(1 < X \leq 3) \\ = F(3) - F(1) = \frac{3}{4} - \frac{1}{2} = \frac{1}{4}$$

(c) Let  $A$  denote the event that a person has to wait for more than 3 minutes and  $B$  the event that he has to wait for more than 1 minute. Then

$$P(A) = P(X > 3) = \frac{1}{4} \quad [\text{c.f. (b), (i)}]$$

$$P(B) = P(X > 1) = 1 - P(X \leq 1) = 1 - F(1) = 1 - \frac{1}{2} = \frac{1}{2}$$

$$P(A \cap B) = P(X > 3 \cap X > 1) = P(X > 3) = \frac{1}{4}$$

(i) Required probability is

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1/4}{1/2} = \frac{1}{2}$$

$$(ii) \text{ Required probability} = P(\bar{A}|B) = \frac{P(\bar{A} \cap B)}{P(B)}$$

$$\text{Now } P(\bar{A} \cap B) = P(X \leq 3 \cap X > 1) = P(1 < X \leq 3) = F(3) - F(1) = \frac{3}{4} - \frac{1}{2} = \frac{1}{4}$$

$$P(\bar{A}|B) = \frac{1/4}{1/2} = \frac{1}{2}$$

**Example 5.18.** A petrol pump is supplied with petrol once a day. If its daily volume  $X$  of sales in thousands of litres is distributed by

$$f(x) = 5(1-x)^4, \quad 0 \leq x \leq 1,$$

what must be the capacity of its tank in order that the probability that its supply will be exhausted in a given day shall be 0.01? (Madras Univ. B.E., 1986)

**Solution.** Let the capacity of the tank (in '000 of litres) be ' $a$ ' such that

$$P(X \geq a) = 0.01 \Rightarrow \int_a^1 f(x) dx = 0.01$$

$$\Rightarrow \int_a^1 5(1-x)^4 dx = 0.01 \quad \text{or} \quad \left[ 5 \cdot \frac{(1-x)^5}{(-5)} \right]_a^1 = 0.01$$

$$\Rightarrow (1-a)^5 = 1/100 \quad \text{or} \quad 1-a = (1/100)^{1/5}$$

$$\therefore a = 1 - (1/100)^{1/5} = 1 - 0.3981 = 0.6019$$

Hence the capacity of the tank =  $0.6019 \times 1000$  litres = 601.9 litres.

**Example 5.19.** Prove that mean deviation is least when measured from the median. [Delhi Univ. B.Sc. (Maths. Hons.), 1989]

**Solution.** If  $f(x)$  is the probability function of a random variable  $X$ ,  $a \leq X \leq b$ , then mean deviation  $M(A)$ , say, about the point  $x = A$  is given by

$$M(A) = \int_a^b |x - A| f(x) dx$$

$$\begin{aligned}
 &= \int_a^A |x - A| f(x) dx + \int_A^b |x - A| f(x) dx \\
 &= \int_a^A (A - x) f(x) dx + \int_A^b (x - A) f(x) dx \quad \dots (1)
 \end{aligned}$$

We want to find the value of 'A' so that  $M(A)$  is minimum. From the principle of maximum and minimum in differential calculus,  $M(A)$  will be minimum for variations in  $A$  if

$$\frac{\partial M(A)}{\partial A} = 0 \text{ and } \frac{\partial^2 M(A)}{\partial A^2} > 0 \quad \dots (2)$$

Differentiating (1) w.r.t. 'A' under the integral sign, since the functions  $(A - x) f(x)$  and  $(x - A) f(x)$  vanish at the point  $x = A$ , we get

$$\frac{\partial M(A)}{\partial A} = \int_a^A f(x) dx - \int_A^b f(x) dx \quad \dots (3)$$

$$\begin{aligned}
 \text{Also } \frac{\partial M(A)}{\partial A} &= \int_a^A f(x) dx - \left[ 1 - \int_a^A f(x) dx \right], \\
 &\quad \left[ \because \int_a^b f(x) dx = 1 \right] \\
 &= 2 \int_a^A f(x) dx - 1 = 2F(A) - 1,
 \end{aligned}$$

where  $F(\cdot)$  is the distribution function of  $X$ . Differentiating again w.r.t.  $A$ , we get

$$\frac{\partial^2}{\partial A^2} M(A) = 2f(A) \quad \dots (4)$$

Now  $\frac{\partial M(A)}{\partial A} = 0$ , on using (3) gives

$$\int_a^A f(x) dx = \int_A^b f(x) dx$$

i.e.,  $A$  is the median value.

Also from (4), we see that

$$\frac{\partial^2 M(A)}{\partial A^2} > 0,$$

assuming that  $f(x)$  does not vanish at the median value. Thus mean deviation is least when taken from median.

\*If  $f(x, \theta)$  is a continuous function of both variables  $x$  and  $\theta$ , possessing continuous partial derivatives  $\frac{\partial^2 f}{\partial x \partial \theta}$ ,  $\frac{\partial^2 f}{\partial \theta \partial x}$  and  $a$  and  $b$  are differentiable functions of  $\theta$ , then

$$\frac{\partial}{\partial \theta} \left[ \int_a^b f(x, \theta) dx \right] = \int_a^b \frac{\partial f}{\partial \theta} dx + f(b, \theta) \frac{db}{d\theta} - f(a, \theta) \frac{da}{d\theta}$$

### EXERCISE 5 (d)

1. (a) Explain the terms (i) probability differential, (ii) probability density function, and (iii) distribution function.

(b) Explain what is meant by a random variable. Distinguish between a discrete and a continuous random variable. Define distribution function of a random variable and show that it is monotonic non-decreasing everywhere and continuous on the right at every point.

[Madras Univ. B.Sc. (Stat Main), 1987]

(c) Show that the distribution function  $F(x)$  of a random variable  $X$  is a non-decreasing function of  $x$ . Determine the jump of  $F(x)$  at a point  $x_0$  of its discontinuity in terms of the probability that the random variable has the value  $x_0$ .

[Calcutta Univ. B.Sc. (Hons.), 1984]

2. The length (in hours)  $X$  of a certain type of light bulb may be supposed to be a continuous random variable with probability density function :

$$f(x) = \begin{cases} \frac{a}{x^3}, & 1500 < x < 2500 \\ 0, & \text{elsewhere.} \end{cases}$$

Determine the constant  $a$ , the distribution function of  $X$ , and compute the probability of the event  $1,700 \leq X \leq 1,900$ .

Ans.  $a = 70,31,250$ ;  $F(x) = \frac{a}{2} \left( \frac{1}{22,50,000} - \frac{1}{x^2} \right)$  and

$$P(1,700 < X < 1,900) = F(1,900) - F(1,700) = \frac{a}{2} \left( \frac{1}{28,90,000} - \frac{1}{36,10,000} \right)$$

3. Define the "distribution function" (or cumulative distribution function) of a random variable and state its essential properties.

Show that, whatever the distribution function  $F(x)$  of a random variable  $X$ ,  $P[a \leq F(x) \leq b] = b - a$ ,  $0 \leq a, b \leq 1$ .

4. (a) The distribution function of a random variable  $X$  is given by

$$F(x) = \begin{cases} 1 - (1+x)e^{-x}, & \text{for } x \geq 0 \\ 0, & \text{for } x < 0 \end{cases}$$

Find the corresponding density function of random variable  $X$ .

(b) Consider the distribution for  $X$  defined by

$$F(x) = \begin{cases} 0, & \text{for } x < 0 \\ 1 - \frac{1}{4} e^{-x}, & \text{for } x \geq 0 \end{cases}$$

Determine  $P(x=0)$  and  $P(x>0)$ .

[Allahabad Univ. B.Sc., 1992]

5. (a) Let  $X$  be a continuous random variable with probability density function given by

$$f(x) = \begin{cases} ax, & 0 \leq x \leq 1 \\ a, & 1 \leq x \leq 2 \\ -ax + 3a, & 2 \leq x \leq 3 \\ 0, & \text{elsewhere} \end{cases}$$

(i) Determine the constant  $a$ .

(ii) Determine  $F(x)$ , and sketch its graph.

(iii) If three independent observations are made, what is the probability that exactly one of these three numbers is larger than 1.5?

[Rajasthan Univ. M.Sc., 1987]

Ans. (i) 1/2, (iii) 3/8.

(b) For the density  $f_X(x) = k e^{-ax^2} (1 - e^{-ax^2}) I_{[0, \infty)}(x)$ , find the normalising constant  $k$ ,  $f_X(x)$  and evaluate  $P(X > 1)$ .

[Delhi Univ. B.Sc. (Maths Hons.), 1989]

Ans.  $k = 2a$ ;  $F(x) = 1 - 2e^{-ax^2} + e^{-2ax^2}$ ;  $P(X > 1) = 2e^{-a} - e^{-2a}$

6. A random variable  $X$  has the density function :

$$f(x) = K \cdot \frac{1}{1+x^2}, \quad \text{if } -\infty < x < \infty \\ = 0, \quad \text{otherwise}$$

Determine  $K$  and the distribution function.

Evaluate the probability  $P(X \geq 0)$ . Find also the mean and variance of  $X$ .

[Karnatak Univ. B.Sc. 1985]

Ans.  $K = 1$ ,  $F(x) = \frac{1}{\pi} \left\{ \tan^{-1} x + \frac{\pi}{2} \right\}$ ,  $P(X \geq 0) = 1/2$ , Mean = 0,

Variance does not exist.

\* \*

7. A continuous random variable  $X$  has the distribution function

$$F(x) = \begin{cases} 0, & \text{if } x \leq 1 \\ k(x-1)^4, & \text{if } 1 < x \leq 3 \\ 1, & \text{if } x > 3 \end{cases}$$

Find (i)  $k$ , (ii) the probability density function  $f(x)$ , and (iii) the mean and the median of  $X$ .

Ans. (i)  $k = \frac{1}{16}$ , (ii)  $f(x) = \frac{1}{4} (x-1)^3$ ,  $1 \leq x \leq 3$

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

Show that

(i)  $k = 1/5$ , (ii)  $F(x) = 0$  for  $x \leq 0$  and  $F(x) = 1 - e^{-2x^5}$ , for  $x > 0$

Using  $F(x)$ , show that

$$(iii) P(3 < X < 5) = 0.1809, (iv) P(X < 4) = 0.5507, (v) P(X > 6) = 0.3012$$

9. A bombing plane carrying three bombs flies directly above a railroad track. If a bomb falls within 40 feet of track, the track will be sufficiently damaged to disrupt the traffic. Within a certain bomb site the points of impact of a bomb have the probability density function :

$$\begin{aligned}f(x) &= (100 + x)/10,000, \text{ when } -100 \leq x \leq 0 \\&= (100 - x)/10,000, \text{ when } 0 \leq x \leq 100 \\&= 0, \text{ elsewhere}\end{aligned}$$

where  $x$  represents the vertical deviation (in feet) from the aiming point, which is the track in this case. Find the distribution function. If all the bombs are used, what is the probability that track will be damaged ?

**Hint.** Probability that track will be damaged by the bomb is given by

$$\begin{aligned}P(|X| < 40) &= P(-40 < X < 40) \\&= \int_{-40}^0 f(x) dx + \int_0^{40} f(x) dx \\&= \int_{-40}^0 \frac{100+x}{10,000} dx + \int_0^{40} \frac{100-x}{10,000} dx = \frac{16}{25}\end{aligned}$$

$$\therefore \text{Probability that a bomb will not damage the track} = 1 - \frac{16}{25} = \frac{9}{25}$$

Probability that none of the three bombs damages the track  
 $= \left(\frac{9}{25}\right)^3 = 0.046656$

Required probability that the track will be damaged =  $1 - 0.046656 = 0.953344$ .

10. The length of time (in minutes) that a certain lady speaks on the telephone is found to be random phenomenon, with a probability function specified by the probability density function  $f(x)$  as

$$\begin{aligned}f(x) &= A e^{-x/5}, \text{ for } x \geq 0 \\&= 0, \text{ otherwise}\end{aligned}$$

(a) Find the value of  $A$  that makes  $f(x)$  a p.d.f.

$$\text{Ans. } A = 1/5$$

(b) What is the probability that the number of minutes that she will talk over the phone is

(i) More than 10 minutes, (ii) less than 5 minutes, and (iii) between 5 and 10 minutes ?

[Shivaji Univ. B.Sc., 1990]

$$\text{Ans. (i) } \frac{1}{e^2}, \text{ (ii) } \frac{e-1}{e}, \text{ (iii) } \frac{e-1}{e^2}.$$

11. The probability that a person will die in the time interval  $(t_1, t_2)$  is given by

$$A \int_{t_1}^{t_2} f(t) dt ;$$

where  $A$  is a constant and the function  $f(t)$  determined from long records, is

$$f(t) = \begin{cases} t^2(100-t)^2, & 0 \leq t \leq 100 \\ 0, & \text{elsewhere} \end{cases}$$

Find the probability that a person will die between the ages 60 and 70 assuming that his age is  $\geq 50$ . [Calcutta Univ. B.A. (Hons.), 1987]

**5.5. Joint Probability Law.** Two random variables  $X$  and  $Y$  are said to be jointly distributed if they are defined on the same probability space. The sample points consist of 2-tuples. If the joint probability function is denoted by  $P_{XY}(x, y)$  then the probability of a certain event  $E$  is given by

$$P_{XY}(x, y) = P[(X, Y) \in E] \quad \dots (5.13)$$

$(X, Y)$  is said to belong to  $E$ , if in the 2 dimensional space the 2-tuples lie in the Borel set  $B$ , representing the event  $E$ .

**5.5.1. Joint Probability Mass Function and Marginal and Conditional Probability Functions.** Let  $X$  and  $Y$  be random variables on a sample space  $S$  with respective image sets  $X(S) = \{x_1, x_2, \dots, x_n\}$  and  $Y(S) = \{y_1, y_2, \dots, y_m\}$ . We make the product set

$$X(S) \times Y(S) = \{x_1, x_2, \dots, x_n\} \times \{y_1, y_2, \dots, y_m\}$$

into a probability space by defining the probability of the ordered pair  $(x_i, y_j)$  to be  $P(X = x_i, Y = y_j)$  which we write  $p(x_i, y_j)$ . The function  $p$  on  $X(S) \times Y(S)$  defined by

$$p_{ij} = P(X = x_i \cap Y = y_j) = p(x_i, y_j) \quad \dots (5.14)$$

is called the *joint probability function* of  $X$  and  $Y$  and is usually represented in the form of the following table :

$X \backslash Y$	$y_1$	$y_2$	$y_3$	...	$y_j$	...	$y_m$	Total
$x_1$	$p_{11}$	$p_{12}$	$p_{13}$	...	$p_{1j}$	...	$p_{1m}$	$p_{1..}$
$x_2$	$p_{21}$	$p_{22}$	$p_{23}$	...	$p_{2j}$	...	$p_{2m}$	$p_{2..}$
$x_3$	$p_{31}$	$p_{32}$	$p_{33}$	...	$p_{3j}$	...	$p_{3m}$	$p_{3..}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
$x_i$	$p_{i1}$	$p_{i2}$	$p_{i3}$	...	$p_{ij}$	...	$p_{im}$	$p_{i..}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
$x_n$	$p_{n1}$	$p_{n2}$	$p_{n3}$	...	$p_{nj}$	...	$p_{nm}$	$p_{n..}$
Total	$p_{.1}$	$p_{.2}$	$p_{.3}$	...	$p_{.j}$	...	$p_{.m}$	1

$$\therefore \sum_{i=1}^n \sum_{j=1}^m p(x_i, y_j) = 1$$

Suppose the joint distribution of two random variables  $X$  and  $Y$  is given, then the probability distribution of  $X$  is determined as follows :

$$\begin{aligned} p_X(x_i) &= P(X=x_i) = P[X=x_i \cap Y=y_1] + P[X=x_i \cap Y=y_2] + \dots \\ &\quad + P[X=x_i \cap Y=y_j] + \dots + P[X=x_i \cap Y=y_m] \\ &= p_{i1} + p_{i2} + \dots + p_{ij} + \dots + p_{im} \\ &= \sum_{j=1}^m p_{ij} = \sum_{j=1}^m p(x_i, y_j) = p_i. \end{aligned} \quad \dots (5.14 a)$$

and is known as *marginal probability function of X*.

$$\text{Also } \sum_{i=1}^n p_{i.} = p_{1.} + p_{2.} + \dots + p_{n.} = \sum_{i=1}^n \sum_{j=1}^m p_{ij} = 1$$

Similarly, we can prove that

$$p_Y(y_j) = P(Y=y_j) = \sum_{i=1}^n p_{ij} = \sum_{i=1}^n p(x_i, y_j) = p_j \quad \dots (5.14 b)$$

which is the *marginal probability function of Y*.

Also

$$P[X=x_i | Y=y_j] = \frac{P[X=x_i \cap Y=y_j]}{P[Y=y_j]} = \frac{p(x_i, y_j)}{p(y_j)} = \frac{p_{ij}}{p_j}$$

This is known as *conditional probability function of X given Y = y<sub>j</sub>*

Similarly

$$P[Y=y_j | X=x_i] = \frac{p(x_i, y_j)}{p(x_i)} = \frac{p_{ij}}{p_i} \quad \dots (5.14 c)$$

is the *conditional probability function of Y given X = x<sub>i</sub>*

$$\text{Also } \sum_{i=1}^n \frac{p_{ij}}{p_i} = \frac{p_{1j} + p_{2j} + \dots + p_{ij} + \dots + p_{nj}}{p_i} = \frac{p_{.j}}{p_i} = 1$$

Similarly

$$\sum_{j=1}^m \frac{p_{ij}}{p_i} = 1$$

Two random variables  $X$  and  $Y$  are said to be *independent* if

$$P(X=x_i, Y=y_j) = P(X=x_i) \cdot P(Y=y_j), \quad \dots (5.14 d)$$

otherwise they are said to be dependent.

**5.5.2. Joint Probability Distribution Function.** Let  $(X, Y)$  be a two-dimensional random variable then their joint distribution function is denoted by  $F_{XY}(x, y)$  and it represents the probability that simultaneously the observation

$(X, Y)$  will have the property ( $X \leq x$  and  $Y \leq y$ ), i.e.,

$$\begin{aligned} F_{XY}(x, y) &= P(-\infty < X \leq x, -\infty < Y \leq y) \\ &= \int_{-\infty}^x \left[ \int_{-\infty}^y f_{XY}(x, y) dx dy \right] \quad \dots (5-15) \end{aligned}$$

(For continuous variables)

where

$$f_{XY}(x, y) \geq 0$$

$$\text{And } \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x, y) dx dy = 1 \text{ or } \sum_x \sum_y f(x, y) = 1$$

### Properties of Joint Distribution Function

1. (i) For the real numbers  $a_1, b_1, a_2$  and  $b_2$

$$\begin{aligned} P(a_1 < X \leq b_1, a_2 < Y \leq b_2) &= F_{XY}(b_1, b_2) + F_{XY}(a_1, a_2) \\ &\quad - F_{XY}(a_1, b_2) - F_{XY}(b_1, a_2) \\ &\quad [\text{For proof, See Example 5-29}] \end{aligned}$$

(ii) Let  $a_1 < a_2, b_1 < b_2$ . We have

$$(X \leq a_1, Y \leq a_2) + (a_1 < X \leq b_1, Y \leq a_2) = (X \leq b_1, Y \leq a_2)$$

and the events on the L.H.S. are mutually exclusive. Taking probabilities on both-sides, we get :

$$\begin{aligned} F(a_1, a_2) + P(a_1 < X \leq b_1, Y \leq a_2) &= F(b_1, a_2) \\ \Rightarrow F(b_1, a_2) - F(a_1, a_2) &= P(a_1 < X \leq b_1, Y \leq a_2) \\ \therefore F(b_1, a_2) &\geq F(a_1, a_2) \quad [\text{Since } P(a_1 < X \leq b_1, Y \leq a_2) \geq 0] \end{aligned}$$

Similarly it follows that

$$\begin{aligned} F(a_1, b_2) - F(a_1, a_2) &= P(X \leq a_1, a_2 < Y \leq b_2) \geq 0 \\ \Rightarrow F(a_1, b_2) &\geq F(a_1, a_2), \end{aligned}$$

which shows that  $F(x, y)$  is monotonic non-decreasing function.

2.  $F(-\infty, y) = 0 = F(x, -\infty), F(+\infty, +\infty) = 1$

3. If the density function  $f(x, y)$  is continuous at  $(x, y)$  then

$$\frac{\partial^2 F}{\partial x \partial y} = f(x, y)$$

**5-5-3. Marginal Distribution Functions.** From the knowledge of joint distribution function  $F_{XY}(x, y)$ , it is possible to obtain the individual distribution functions,  $F_X(x)$  and  $F_Y(y)$  which are termed as marginal distribution function of  $X$  and  $Y$  respectively with respect to the joint distribution function  $F_{XY}(x, y)$ .

$$\begin{aligned} F_X(x) &= P(X \leq x) = P(X \leq x, Y < \infty) = \lim_{y \rightarrow \infty} F_{XY}(x, y). \\ &\approx F_{XY}(x, \infty) \quad \dots (5-16) \end{aligned}$$

Similarly,  $F_Y(y) = P(Y \leq y) = P(X < \infty, Y \leq y)$

$$\approx \lim_{x \rightarrow -\infty} F_{XY}(x, y) = F_{XY}(\infty, y)$$

$F_X(x)$  is termed as the marginal distribution function of  $X$  corresponding to the joint distribution function  $F_{XY}(x, y)$  and similarly  $F_Y(y)$  is called marginal distribution function of the random variable  $Y$  corresponding to the joint distribution function  $F_{XY}(x, y)$ .

In the case of jointly discrete random variables, the marginal distribution functions are given as

$$F_X(x) = \sum_y P(X \leq x, Y = y),$$

$$F_Y(y) = \sum_x P(X = x, Y \leq y)$$

Similarly in the case of jointly continuous random variable, the marginal distribution functions are given as

$$F_X(x) = \int_{-\infty}^x \left\{ \int_{-\infty}^{\infty} f_{XY}(x, y) dy \right\} dx$$

$$F_Y(y) = \int_{-\infty}^y \left\{ \int_{-\infty}^{\infty} f_{XY}(x, y) dx \right\} dy$$

**5-5-4 Joint Density Function, Marginal Density Functions.** From the joint distribution function  $F_{XY}(x, y)$  of two dimensional continuous random variable we get the joint probability density function by differentiation as follows :

$$\begin{aligned} f_{XY}(x, y) &= \frac{\partial^2 F(x, y)}{\partial x \partial y} \\ &= \lim_{\delta x \rightarrow 0, \delta y \rightarrow 0} \frac{P(x \leq X \leq x + \delta x, y \leq Y \leq y + \delta y)}{\delta x \delta y} \end{aligned}$$

Or it may be expressed in the following way also :

"The probability that the point  $(x, y)$  will lie in the infinitesimal rectangular region, of area  $dx dy$  is given by

$$P\left\{x - \frac{1}{2}dx \leq X \leq x + \frac{1}{2}dx, y - \frac{1}{2}dy \leq Y \leq y + \frac{1}{2}dy\right\} = dF_{XY}(x, y)$$

and is denoted by  $f_{XY}(x, y) dx dy$ , where the function  $f_{XY}(x, y)$  is called the joint probability density function of  $X$  and  $Y$ .

The marginal probability function of  $Y$  and  $X$  are given respectively

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x, y) dx \quad (\text{for continuous variables})$$

$$= \sum_x p_{XY}(x, y) \quad (\text{for discrete variables})$$

...(5.17)

$$\text{and } f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y) dy \quad (\text{for continuous variables})$$

$$= \sum_y p_{XY}(x, y) \quad (\text{for discrete variables}) \quad (5-17a)$$

The marginal density functions of  $X$  and  $Y$  can be obtained in the following manner also.

$$\left. \begin{aligned} f_X(x) &= \frac{dF_X(x)}{dx} = \int_{-\infty}^{\infty} f_{XY}(x, y) dy \\ \text{and } f_Y(y) &= \frac{dF_Y(y)}{dy} = \int_{-\infty}^{\infty} f_{XY}(x, y) dx \end{aligned} \right\} \dots (5-17b)$$

**Important Remark.** If we know the joint p.d.f. (p.m.f.)  $f_{XY}(x, y)$  of two random variables  $X$  and  $Y$ , we can obtain the individual distributions of  $X$  and  $Y$  in the form of their marginal p.d.f.'s (p.m.f's)  $f_X(x)$  and  $f_Y(y)$  by using (5-17) and (5-17a). However, the converse is not true i.e., from the marginal distributions of two jointly distributed random variables, we cannot determine the joint distributions of these two random variables.

To verify this, it will suffice to show that two different joint p.m.f's (p.d.f.'s) have the same marginal distribution for  $X$  and the same marginal distribution for  $Y$ . We give below two joint discrete probability distributions which have the same marginal distributions.

### JOINT DISTRIBUTIONS HAVING SAME MARGINALS

*Probability Distribution I*

$\begin{matrix} X \\ Y \end{matrix}$	0	1	$f_Y(y)$
0	0.28	0.37	0.65
1	0.22	0.13	0.35
$f_X(x)$	0.50	0.50	1.00

*Probability Distribution II*

$\begin{matrix} X \\ Y \end{matrix}$	0	1	$f_Y(y)$
0	0.35	0.30	0.65
1	0.15	0.20	0.35
$f_X(x)$	0.50	0.50	1.00

As an illustration for continuous random variables, let  $(X, Y)$  be continuous r.v. with joint p.d.f.

$$f_{XY}(x, y) = x + y ; \quad 0 \leq (x, y) \leq 1 \quad \dots (5-17c)$$

The marginal p.d.f.'s of  $X$  and  $Y$  are given by :

$$f_X(x) = \int_0^1 f(x, y) dy = \int_0^1 (x + y) dy = \left| xy + \frac{y^2}{2} \right|_0^1$$

$$\Rightarrow f_X(x) = x + \frac{1}{2} ; \quad 0 \leq x \leq 1 \\ \text{Similarly } f_Y(y) = \int_0^y f(x, y) dx = y + \frac{1}{2} ; \quad 0 \leq y \leq 1 \quad \left. \right\} \quad \dots (5.17d)$$

Consider another continuous joint p.d.f.

$$g(x, y) = \left( x + \frac{1}{2} \right) \left( y + \frac{1}{2} \right); \quad 0 \leq (x, y) \leq 1 \quad \dots (5.17e)$$

Then marginal p.d.f.'s of  $X$  and  $Y$  are given by :

$$g_1(x) = \int_0^1 g(x, y) dy = \left( x + \frac{1}{2} \right) \int_0^1 \left( y + \frac{1}{2} \right) dy \\ = \left( x + \frac{1}{2} \right) \left| \frac{y^2}{2} + \frac{1}{2} y \right|_0^1 \\ \Rightarrow g_1(x) = x + \frac{1}{2}; \quad 0 \leq x \leq 1 \\ \text{Similarly } g_2(y) = y + \frac{1}{2}; \quad 0 \leq y \leq 1 \quad \left. \right\} \quad \dots (5.17f)$$

(5.17d) and (5.17f) imply that the two joint p.d.f.'s in (5.17c) and (5.17e) have the same marginal p.d.f.'s (5.17d) or (5.17f).

Another illustration of continuous r.v.'s is given in Remark to Bivariate Normal Distribution, § 10-10-2.

**5.5.5. The Conditional Distribution Function and Conditional Probability Density Function.** For two dimensional random variable  $(X, Y)$ , the joint distribution function  $F_{XY}(x, y)$ , for any real numbers  $x$  and  $y$  is given by

$$F_{XY}(x, y) = P(X \leq x, Y \leq y)$$

Now let  $A$  be the event ( $Y \leq y$ ) such that the event  $A$  is said to occur when  $Y$  assumes values up to and inclusive of  $y$ .

Using conditional probabilities we may now write

$$F_{XY}(x, y) = \int_{-\infty}^x P(A | X=x) dF_X(x) \quad \dots (5.18)$$

The *conditional distribution function*  $F_{Y|X}(y|x)$  denotes the distribution function of  $Y$  when  $X$  has already assumed the particular value  $x$ . Hence

$$F_{Y|X}(y|x) = P(Y \leq y | X=x) = P(A | X=x)$$

Using this expression, the joint distribution function  $F_{XY}(x, y)$  may be expressed in terms of the conditional distribution function as follows :

$$F_{XY}(x, y) = \int_{-\infty}^x F_{Y|X}(y|x) dF_X(x) \quad \dots (5.18a)$$

Similarly

$$F_{XY}(x, y) = \int_{-\infty}^y F_{X|Y}(x|y) dF_Y(y) \quad \dots (5.18b)$$

The *conditional probability density function* of  $Y$  given  $X$  for two random variables  $X$  and  $Y$  which are jointly continuously distributed is defined as follows, for two real numbers  $x$  and  $y$ :

$$f_{r|x}(y|x) = \frac{\partial}{\partial y} F_{r|x}(y|x) \quad \dots (5.19)$$

**Remarks :** 1.  $f_X(x) > 0$ , then

$$f_{r|x}(y|x) = \frac{f_{x,y}(x,y)}{f_X(x)}$$

**Proof.** We have

$$\begin{aligned} F_{x,y}(x,y) &= \int_{-\infty}^x F_{r|x}(y|x) dF_X(x) \\ &= \int_{-\infty}^x F_{r|x}(y|x) f_X(x) dx \end{aligned}$$

Differentiating w.r.t.  $x$ , we get

$$\frac{\partial}{\partial x} F_{x,y}(x,y) = F_{r|x}(y|x) f_X(x)$$

Differentiating w.r.t.  $y$ , we get

$$\frac{\partial}{\partial y} \left[ \frac{\partial}{\partial x} F_{x,y}(x,y) \right] = f_{r|x}(y|x) f_X(x)$$

$$\Rightarrow f_{x,y}(x,y) = f_{r|x}(y|x) f_X(x)$$

$$\Rightarrow f_{r|x}(y|x) = \frac{f_{x,y}(x,y)}{f_X(x)}$$

2. If  $f_Y(y) > 0$ , then

$$f_{x|r}(x|y) = \frac{f_{x,y}(x,y)}{f_Y(y)}$$

3. In terms of the differentials, we have

$$\begin{aligned} P(x < X \leq x+dx | y < Y \leq y+dy) \\ &= \frac{P(x < X \leq x+dx, y < Y \leq y+dy)}{P(y < Y \leq y+dy)} \\ &= \frac{f_{x,y}(x,y) dx dy}{f_Y(y) dy} = f_{x|r}(x|y) dx \end{aligned}$$

Whence  $f_{x|r}(x|y)$  may be interpreted as the conditional density function of  $X$  on the assumption  $Y = y$ .

**5.5.6. Stochastic Independence.** Let us consider two random variables  $X$  and  $Y$  (of discrete or continuous type) with joint p.d.f.  $f_{x,y}(x,y)$  and marginal p.d.f.'s  $f_X(x)$  and  $f_Y(y)$  respectively. Then by the compound probability theorem

$$f_{x,y}(x,y) = f_X(x) f_Y(y|x)$$

where  $g_Y(y|x)$  is the conditional p.d.f. of  $Y$  for given value of  $X = x$ .

If we assume that  $g(y|x)$  does not depend on  $x$ , then by the definition of marginal p.d.f.'s, we get for continuous r.v.'s

$$\begin{aligned} g(y) &= \int_{-\infty}^{\infty} f(x, y) dx \\ &= \int_{-\infty}^{\infty} f_X(x) g(y|x) dx \\ &= g(y|x) \int_{-\infty}^{\infty} f_X(x) dx \\ &\quad [\text{since } g(y|x) \text{ does not depend on } x] \\ &= g(y|x) \quad [\because f(\cdot) \text{ is p.d.f. of } X] \end{aligned}$$

Hence

$$g(y) = g(y|x)$$

and  $f_{XY}(x, y) = f_X(x) g_Y(y)$  ... (\*)  
provided  $g(y|x)$  does not depend on  $x$ . This motivates the following definition of independent random variables.

**Independent Random variables.** Two r.v.'s  $X$  and  $Y$  with joint p.d.f.  $f_{XY}(x, y)$  and marginal p.d.f.'s  $f_X(x)$  and  $g_Y(y)$  respectively are said to be stochastically independent if and only if

$$f_{XY}(x, y) = f_X(x) g_Y(y) \quad \dots (5.20)$$

**Remarks.** 1. In terms of the distribution function, we have the following definition :

Two jointly distributed random variables  $X$  and  $Y$  are stochastically independent if and only if their joint distribution function  $F_{X,Y}(\dots)$  is the product of their marginal distribution functions  $F_X(\cdot)$  and  $G_Y(\cdot)$ , i.e., if for real  $(x, y)$

$$F_{X,Y}(x, y) = F_X(x) G_Y(y) \quad \dots (5.20a)$$

2. The variables which are not stochastically independent are said to be stochastically dependent.

**Theorem 5.8.** Two random variables  $X$  and  $Y$  with joint p.d.f.  $f(x, y)$  are stochastically independent if and only if  $f_{XY}(x, y)$  can be expressed as the product of a non-negative function of  $x$  alone and a non-negative function of  $y$  alone, i.e., if

$$f_{XY}(x, y) = h_X(x) k_Y(y) \quad \dots (5.21)$$

where  $h(\cdot) \geq 0$  and  $k(\cdot) \geq 0$ .

**Proof.** If  $X$  and  $Y$  are independent then by definition, we have

$$f_{XY}(x, y) = f_X(x) \cdot g_Y(y)$$

where  $f(x)$  and  $g(y)$  are marginal p.d.f. of  $X$  and  $Y$  respectively. Thus condition (5.21) is satisfied.

Conversely if (5.21) holds, then we have to prove that  $X$  and  $Y$  are independent. For continuous random variables  $X$  and  $Y$ , the marginal p.d.f.'s are given by

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f(x, y) dy = \int_{-\infty}^{\infty} h(x) k(y) dy \\ &= h(x) \int_{-\infty}^{\infty} k(y) dy = c_1 h(x), \text{ say} \end{aligned} \quad \dots (*)$$

$$\begin{aligned} \text{and } g_Y(y) &= \int_{-\infty}^{\infty} f(x, y) dx = \int_{-\infty}^{\infty} h(x) k(y) dx \\ &= k(y) \int_{-\infty}^{\infty} h(x) dx = c_2 k(y), \text{ say.} \end{aligned} \quad \dots (**)$$

where  $c_1$  and  $c_2$  are constants independent of  $x$  and  $y$ . Moreover

$$\begin{aligned} &\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = 1 \\ \Rightarrow &\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x) k(y) dx dy = 1 \\ \Rightarrow &\left( \int_{-\infty}^{\infty} h(x) dx \right) \left( \int_{-\infty}^{\infty} k(y) dy \right) = 1 \\ \Rightarrow &c_2 c_1 = 1 \end{aligned} \quad \dots (***)$$

Finally, we get

$$\begin{aligned} f_{X,Y}(x, y) &= h_X(x) k_Y(y) = c_1 c_2 h_X(x) k_Y(y) \quad [\text{using } (***)] \\ &= (c_1 h_X(x)) (c_2 k_Y(y)) \\ &= f_X(x) g_Y(y) \quad [\text{from } (*) \text{ and } (**)] \end{aligned}$$

$\Rightarrow X$  and  $Y$  are stochastically independent.

**Theorem 5.9.** If the random variables  $X$  and  $Y$  are stochastically independent, then for all possible selections of the corresponding pairs of real numbers  $(a_1, b_1)$ ,  $(a_2, b_2)$  where  $a_i \leq b_i$  for all  $i = 1, 2$  and where the values  $\pm \infty$  are allowed, the events  $(a_1 < X \leq b_1)$  and  $(a_2 < Y \leq b_2)$  are independent, i.e.,

$$P[(a_1 < X \leq b_1) \cap (a_2 < Y \leq b_2)] = P(a_1 < X \leq b_1) P(a_2 < Y \leq b_2)$$

**Proof.** Since  $X$  and  $Y$  are stochastically independent, we have in the usual notations

$$f_{X,Y}(x, y) = f_X(x) g_Y(y) \quad \dots (*)$$

In case of continuous r.v.'s, we have

$$P[(a_1 < X \leq b_1) \cap (a_2 < Y \leq b_2)] = \int_{a_1}^{b_1} \int_{a_2}^{b_2} f(x, y) dx dy$$

$$\begin{aligned}
 &= \left( \int_{a_1}^{b_1} f_X(x) dx \right) \left( \int_{a_2}^{b_2} g_Y(y) dy \right) \\
 &= P(a_1 < X \leq b_1) P(a_2 < Y \leq b_2)
 \end{aligned}
 \quad [\text{from } (*)]$$

as desired.

**Remark.** In case of discrete r.v.'s theorems 5.8 and 5.9 can be proved on replacing integration by summation over the given range of the variables.

**Example 5.20.** For the following bivariate probability distribution of  $X$  and  $Y$ , find

- (i)  $P(X \leq 1, Y = 2)$ , (ii)  $P(X \leq 1)$ , (iii)  $P(Y = 3)$ , (iv)  $P(Y \leq 3)$  and  
 (v)  $P(X < 3, Y \leq 4)$

$X \backslash Y$	1	2	3	4	5	6
0	0	0	$\frac{1}{32}$	$\frac{2}{32}$	$\frac{2}{32}$	$\frac{3}{32}$
1	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$
2	$\frac{1}{32}$	$\frac{1}{32}$	$\frac{1}{64}$	$\frac{1}{64}$	0	$\frac{2}{64}$

**Solution.** The marginal distributions are given below :

$X \backslash Y$	1	2	3	4	5	6	$p_X(x)$
0	0	0	$\frac{1}{32}$	$\frac{2}{32}$	$\frac{2}{32}$	$\frac{3}{32}$	$\frac{8}{32}$
1	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{10}{16}$
2	$\frac{1}{32}$	$\frac{1}{32}$	$\frac{1}{64}$	$\frac{1}{64}$	0	$\frac{2}{64}$	$\frac{8}{64}$
$p_Y(y)$	$\frac{3}{32}$	$\frac{3}{32}$	$\frac{11}{64}$	$\frac{13}{64}$	$\frac{6}{32}$	$\frac{16}{64}$	$\Sigma p(x) = 1$
							$\Sigma p(y) = 1$

$$\begin{aligned}
 (i) \quad P(X \leq 1, Y = 2) &= P(X = 0, Y = 2) + P(X = 1, Y = 2) \\
 &= 0 + \frac{1}{16} = \frac{1}{16}
 \end{aligned}$$

$$\begin{aligned}
 (ii) \quad P(X \leq 1) &= P(X = 0) + P(X = 1) \\
 &= \frac{8}{32} + \frac{10}{16} = \frac{7}{8} \quad (\text{From above table})
 \end{aligned}$$

$$(iii) \quad P(Y = 3) = \frac{11}{64} \quad (\text{From above table})$$

$$\begin{aligned}
 (iv) \quad P(Y \leq 3) &= P(Y = 1) + P(Y = 2) + P(Y = 3) \\
 &= \frac{3}{32} + \frac{3}{32} + \frac{11}{64} = \frac{23}{64}
 \end{aligned}$$

$$\begin{aligned}
 (v) \quad P(X < 3, Y \leq 4) &= P(X = 0, Y \leq 4) + P(X = 1, Y \leq 4) \\
 &\quad + P(X = 2, Y \leq 4) \\
 &= \left( \frac{1}{32} + \frac{2}{32} \right) + \left( \frac{1}{16} + \frac{1}{16} + \frac{1}{8} + \frac{1}{8} \right) \\
 &\quad + \left( \frac{1}{32} + \frac{1}{32} + \frac{1}{64} + \frac{1}{64} \right) = \frac{9}{16}
 \end{aligned}$$

**Example 5.21.** The joint probability distribution of two random variables  $X$  and  $Y$  is given by :

$$p(x, y) = \frac{2}{n(n+1)}, \quad x = 1, 2, \dots, n \\ y = 1, 2, \dots, x$$

Examine whether  $X$  and  $Y$  are independent. (Calicut Univ. B.Sc., 1991)

**Solution.** The joint probability distribution table along with the marginal distributions of  $X$  and  $Y$  is given below.

$\backslash$	$X$	1	2	3	.....	$n$	$p_Y(y)$
$Y$		$\frac{2}{n(n+1)}$	$\frac{2}{n(n+1)}$	$\frac{2}{n(n+1)}$	.....	$\frac{2}{n(n+1)}$	$\frac{2n}{n(n+1)}$
1		$\frac{2}{n(n+1)}$	$\frac{2}{n(n+1)}$	$\frac{2}{n(n+1)}$	.....	$\frac{2}{n(n+1)}$	$\frac{2n}{n(n+1)}$
2	-		$\frac{2}{n(n+1)}$	$\frac{2}{n(n+1)}$	.....	$\frac{2}{n(n+1)}$	$\frac{2(n-1)}{n(n+1)}$
3	-	-		$\frac{2}{n(n+1)}$	.....	$\frac{2}{n(n+1)}$	$\frac{2(n-2)}{n(n+1)}$
:	:				.....		
$n-1$	-	-	-		$\frac{2}{n(n+1)}$	$\frac{2}{n(n+1)}$	$\frac{2 \times 2}{n(n+1)}$
$n$	-	-	-	-		$\frac{2}{n(n+1)}$	$\frac{2}{n(n+1)}$
$p_X(x)$		$\frac{2}{n(n+1)}$	$\frac{2 \times 2}{n(n+1)}$	$\frac{2 \times 3}{n(n+1)}$	.....	$\frac{2 \times n}{n(n+1)}$	

Note that  $y = 1, 2, \dots, x$ .

When  $x = 1$ ,  $y = 1$ ; when  $x = 2$ ,  $y = 1, 2$ ; when  $x = 3$ ,  $y = 1, 2, 3$  and so on.

From the above table, we see that

$$\begin{aligned}
 p_{XY}(x, y) &\neq p_X(x)p_Y(y); \quad \forall x, y \\
 \Rightarrow X \text{ and } Y \text{ are not independent.}
 \end{aligned}$$

**Example 5.22.** Given the following bivariate probability distribution, obtain (i) marginal distributions of  $X$  and  $Y$ , (ii) the conditional distribution of  $X$  given  $Y = 2$ .

$\backslash$	$X$	-1	0	1
$Y$				
0		$\frac{1}{15}$	$\frac{2}{15}$	$\frac{1}{15}$
1		$\frac{3}{15}$	$\frac{2}{15}$	$\frac{1}{15}$
2		$\frac{2}{15}$	$\frac{1}{15}$	$\frac{2}{15}$

(Mysore Univ. B.Sc., Oct. 1987)

**Solution.**

$\backslash$	$X$	-1	0	1	$\sum_x p(x,y)$
$Y$					
0		$\frac{1}{15}$	$\frac{2}{15}$	$\frac{1}{15}$	$\frac{4}{15}$
1		$\frac{3}{15}$	$\frac{2}{15}$	$\frac{1}{15}$	$\frac{6}{15}$
2		$\frac{2}{15}$	$\frac{1}{15}$	$\frac{2}{15}$	$\frac{5}{15}$
$\sum_y p(x,y)$		$\frac{6}{15}$	$\frac{5}{15}$	$\frac{4}{15}$	1

(i) Marginal distribution of  $X$ . From the above table, we get

$$P(X = -1) = \frac{6}{15} = \frac{2}{5}; P(X = 0) = \frac{5}{15} = \frac{1}{3}; P(X = 1) = \frac{4}{15}$$

Marginal distribution of  $Y$ :

$$P(Y = 0) = \frac{4}{15}; P(Y = 1) = \frac{6}{15} = \frac{2}{5}; P(Y = 2) = \frac{5}{15} = \frac{1}{3}$$

(ii) Conditional distribution of  $X$  given  $Y = 2$ . We have

$$\begin{aligned} P(X = x \cap Y = 2) &= P(Y = 2), P(X = x | Y = 2) \\ \Rightarrow P(X = x | Y = 2) &= \frac{P(X = x \cap Y = 2)}{P(Y = 2)} \end{aligned}$$

$$\therefore P(X = -1 | Y = 2) = \frac{P(X = -1 \cap Y = 2)}{P(Y = 2)} = \frac{2/15}{1/3} = \frac{2}{5}$$

**Example 5.23.**  $X$  and  $Y$  are two random variables having the joint density function,  $f(x, y) = \frac{1}{27}(2x + y)$ , where  $x$  and  $y$  can assume only the integer values 0, 1 and 2. Find the conditional distribution of  $Y$  for  $X = x$ .

(South Gujarat Univ. B.Sc., 1988)

**Solution.** The joint probability function

$$f(x, y) = \frac{1}{27}(2x + y); \quad x = 0, 1, 2; \quad y = 0, 1, 2$$

gives the following table of joint probability distribution of  $X$  and  $Y$ .

JOINT PROBABILITY DISTRIBUTION  $f(x, y)$  OF  $X$  AND  $Y$

$X \downarrow \backslash Y \rightarrow$	0	1	2	$f_X(x)$
0	0	1/27	2/27	3/27
1	2/27	3/27	4/27	9/27
2	4/27	5/27	6/27	15/27

For example  $f(0, 0) = \frac{1}{27}(0 + 2 \times 0) = 0$

$f(1, 0) = \frac{1}{27}(0 + 2 \times 1) = \frac{2}{27}; \quad f(2, 0) = \frac{1}{27}(0 + 2 \times 2) = \frac{4}{27}$   
and so on.

The marginal probability distribution of  $X$  is given by

$$f_X(x) = \sum_y f(x, y),$$

and is tabulated in last column of above table.

The conditional distribution of  $Y$  for  $X = x$  is given by

$$f_{Y|X}(Y = y | X = x) = \frac{f(x, y)}{f_X(x)}$$

and is obtained in the following table.

CONDITIONAL DISTRIBUTION OF  $Y$  FOR  $X = x$

$X \backslash Y$	0	1	2
0	0	1/3	2/3
1	2/9	3/9	4/9
2	4/15	5/15	6/15

**Example 5.24.** Two discrete random variables  $X$  and  $Y$  have the joint probability density function :

$$p(x, y) = \frac{\lambda^x e^{-\lambda} p^y (1-p)^{x-y}}{y! (x-y)!}, \quad y = 0, 1, 2, \dots, x; \quad x = 0, 1, 2, \dots$$

where  $\lambda, p$  are constants with  $\lambda > 0$  and  $0 < p < 1$ .

*Find (i) The marginal probability density functions of  $X$  and  $Y$ .*

*(ii) The conditional distribution of  $Y$  for a given  $X$  and of  $X$  for a given  $Y$ .*

(Poona Univ. B.Sc., 1986 ; Nagpur Univ. M.Sc., 1989)

**Solution.** (i)

$$\begin{aligned} p_X(x) &= \sum_{y=0}^{\infty} p(x, y) = \sum_{y=0}^{\infty} \frac{\lambda^x e^{-\lambda} p^y (1-p)^{x-y}}{y! (x-y)!} \\ &= \frac{\lambda^x e^{-\lambda}}{x!} \sum_{y=0}^{\infty} \frac{x! p^y (1-p)^{x-y}}{y! (x-y)!} = \frac{\lambda^x e^{-\lambda}}{x!} \sum_{y=0}^{\infty} {}^x C_y p^y (1-p)^{x-y} \\ &= \frac{\lambda^x e^{-\lambda}}{x!}, \quad x = 0, 1, 2, \dots \end{aligned}$$

which is the probability function of a Poisson distribution with parameter  $\lambda$ .

$$\begin{aligned} p_Y(y) &= \sum_{x=0}^{\infty} p(x, y) = \sum_{x=y}^{\infty} \frac{\lambda^x e^{-\lambda} p^y (1-p)^{x-y}}{y! (x-y)!} \\ &= \frac{(\lambda p)^y e^{-\lambda}}{y!} \sum_{x=y}^{\infty} \frac{[\lambda(1-p)]^{x-y}}{(x-y)!} = \frac{(\lambda p)^y e^{-\lambda}}{y!} e^{\lambda(1-p)} \\ &= \frac{e^{-\lambda p} (\lambda p)^y}{y!}, \quad y = 0, 1, 2, \dots \end{aligned}$$

which is the probability function of a Poisson distribution with parameter  $\lambda p$ .

(ii) The conditional distribution of  $Y$  for given  $X$  is

$$\begin{aligned} p_{Y|X}(y|x) &= \frac{p_{XY}(x,y)}{p_X(x)} = \frac{\lambda^x e^{-\lambda} p^y (1-p)^{x-y} x!}{y! (x-y)! \lambda^x e^{-\lambda}} \\ &= \frac{x!}{y! (x-y)!} p^y (1-p)^{x-y} = {}^x C_y p^y (1-p)^{x-y}, \quad x > y \end{aligned}$$

The conditional probability distribution of  $X$  for given  $Y$  is

$$\begin{aligned} p_{X|Y}(x|y) &= \frac{p_{XY}(x,y)}{p_Y(y)} \\ &= \frac{\lambda^x e^{-\lambda} p^y (1-p)^{x-y}}{y! (x-y)!} \cdot \frac{y!}{e^{-\lambda p} (\lambda p)^y} \quad [\text{cf. Part (i)}] \\ &= \frac{e^{-\lambda q} (\lambda q)^{x-y}}{(x-y)!}; \quad q = 1-p, \quad x > y \end{aligned}$$

**Example 5-25.** The joint p.d.f. of two random variables  $X$  and  $Y$  is given by :

$$f(x, y) = \frac{9(1+x+y)}{2(1+x)^4(1+y)^4}; \quad \begin{cases} 0 \leq x < \infty \\ 0 < y < \infty \end{cases}$$

Find the marginal distributions of  $X$  and  $Y$ , and the conditional distribution of  $Y$  for  $X = x$ .

**Solution.** Marginal p.d.f. of  $X$  is given by

$$\begin{aligned} f_X(x) &= \int_0^\infty f(x, y) dy \\ &= \frac{9}{2(1+x)^4} \int_0^\infty \frac{(1+y)+x}{(1+y)^4} dy \\ &= \frac{9}{2(1+x)^4} \cdot \int_0^\infty [(1+y)^{-3} + x(1+y)^{-4}] dy \\ &= \frac{9}{2(1+x)^4} \left[ \left| \frac{-1}{2(1+y)^2} \right|_0^\infty + x \left| \frac{-1}{3(1+y)^3} \right|_0^\infty \right] \\ &= \frac{9}{2(1+x)^4} \cdot \left[ \frac{1}{2} + \frac{x}{3} \right] \\ &= \frac{3}{4} \cdot \frac{3+2x}{(1+x)^4}; \quad 0 < x < \infty \end{aligned}$$

Since  $f(x, y)$  is symmetric in  $x$  and  $y$ , the marginal p.d.f. of  $Y$  is given by

$$\begin{aligned} f_Y(y) &= \int_0^\infty f(x, y) dx \\ &= \frac{3}{4} \cdot \frac{3+2y}{(1+y)^4}; \quad 0 < y < \infty \end{aligned}$$

The conditional distribution of  $Y$  for  $X = x$  is given by

$$\begin{aligned} f_{XY}(Y=y | X=x) &= \frac{f_{XY}(x, y)}{f_X(x)} \\ &= \frac{9(1+x+y)}{2(1+x)^4(1+y)^4} \cdot \frac{4(1+x)^4}{3(3+2x)} \\ &= \frac{6(1+x+y)}{(1+y)^4(3+2x)}; \quad 0 < y < \infty \end{aligned}$$

**Example 5-26.** The joint probability density function of a two-dimensional random variable  $(X, Y)$  is given by

$$\begin{aligned} f(x, y) &= 2; \quad 0 < x < 1, \quad 0 < y < x \\ &= 0, \quad \text{elsewhere} \end{aligned}$$

(i) Find the marginal density functions of  $X$  and  $Y$ ,

(ii) find the conditional density function of  $Y$  given  $X=x$  and conditional density function of  $X$  given  $Y=y$ , and

(iii) check for independence of  $X$  and  $Y$ .

[M.S.Baroda Univ. B.Sc., 1987; Karnataka Univ. B.Sc., Oct. 1988]

**Solution.** Evidently  $f(x, y) \geq 0$  and

$$\int_0^1 \int_0^x 2 dx dy = 2 \int_0^1 x dx = 1$$

(i) The marginal p.d.f.'s of  $X$  and  $Y$  are given by

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y) dy = \int_0^x 2 dy = 2x, \quad 0 < x < 1 \\ = 0, \text{ elsewhere}$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x, y) dx = \int_y^1 2 dx = 2(1-y), \quad 0 < y < 1 \\ = 0, \text{ elsewhere}$$

(ii) The conditional density function of  $Y$  given  $X$  is

$$f_{Y|X}(y|x) = \frac{f_{XY}(x, y)}{f_X(x)} = \frac{2}{2x} = \frac{1}{x}, \quad 0 < x < 1$$

The conditional density function of  $X$  given  $Y$  is

$$f_{X|Y}(x|y) = \frac{f_{XY}(x, y)}{f_Y(y)} = \frac{2}{2(1-y)} = \frac{1}{(1-y)}, \quad 0 < y < 1$$

(iii) Since  $f_X(x)f_Y(y) = 2(2x)(1-y) \neq f_{XY}(x, y)$ ,  $X$  and  $Y$  are not independent.

**Example 5.27.** A gun is aimed at a certain point (origin of the coordinate system). Because of the random factors, the actual hit point can be any point  $(X, Y)$  in a circle of radius  $R$  about the origin. Assume that the joint density of  $X$  and  $Y$  is constant in this circle given by :

$$f_{XY}(x, y) = k, \text{ for } x^2 + y^2 \leq R^2 \\ = 0, \text{ otherwise}$$

(i) Compute  $k$ , (ii) show that

$$f_X(x) = \frac{2}{\pi R} \left\{ 1 - \left( \frac{x}{R} \right)^2 \right\}^{1/2}, \quad \text{for } -R \leq x \leq R \\ = 0, \text{ otherwise}$$

[Calcutta Univ. B.Sc.(Stat. Hons.), 1987]

**Solution.** (i) The constant  $k$  is computed from the consideration that the total probability is 1, i.e.,

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = 1 \Rightarrow \iint_{x^2 + y^2 \leq R^2} k dx dy = 1$$

$$\Rightarrow 4 \int_I k dx dy = 1$$

where region  $I$  is the first quadrant of the circle  
 $x^2 + y^2 = R^2$ .

$$\begin{aligned} \Rightarrow 4k \int_0^R \left( \int_0^{\sqrt{R^2 - x^2}} 1 \cdot dy \right) dx &= 1 \\ \Rightarrow 4k \int_0^R \sqrt{R^2 - x^2} dx &= 1 \\ \Rightarrow 4k \left| x \sqrt{R^2 - x^2} + \frac{R^2}{2} \sin^{-1} \left( \frac{x}{R} \right) \right|_0^R &= 1 \\ \Rightarrow 4k \cdot \left( \frac{R^2}{2} \cdot \frac{\pi}{2} \right) &= 1 \quad \Rightarrow \quad k = \frac{1}{\pi R^2} \\ \therefore f_{XY}(x, y) &= \frac{1}{\pi R^2} ; \quad x^2 + y^2 \leq R^2 \\ &= 0, \quad \text{otherwise} \end{aligned}$$

$$\begin{aligned} (ii) \quad f_X(x) &= \int_{-\infty}^{\infty} f(x, y) dy = \frac{1}{\pi R^2} \int_{-\sqrt{R^2 - x^2}}^{\sqrt{R^2 - x^2}} 1 \cdot dy \\ &\quad [ \text{because } x^2 + y^2 \leq R^2 \Rightarrow -\sqrt{R^2 - x^2} \leq y \leq \sqrt{R^2 - x^2}] \\ &= \frac{2}{\pi R^2} \int_0^{\sqrt{R^2 - x^2}} 1 \cdot dy = \frac{2}{\pi R^2} (R^2 - x^2)^{1/2} \\ &= \frac{2}{\pi R} \left[ 1 - \left( \frac{x}{R} \right)^2 \right]^{1/2} \end{aligned}$$

**Example 5.28.** Given:

$$f(x, y) = e^{-(x+y)} I_{(0, \infty)}(x) \cdot I_{(0, \infty)}(y),$$

find (i)  $P(X > 1)$ , (ii)  $P(X < Y | X < 2Y)$ , (iii)  $P(1 < X + Y < 2)$

[Delhi Univ. B.Sc. (Maths Hons.), 1987]

**Solution.** We are given:

$$\begin{aligned} f(x, y) &= e^{-(x+y)} ; \quad 0 \leq x < \infty, 0 \leq y < \infty \quad \dots (1) \\ &= (e^{-x})(e^{-y}) \\ &= f_X(x) \cdot f_Y(y) ; \quad 0 \leq x < \infty, 0 \leq y < \infty \end{aligned}$$

$\Rightarrow X$  and  $Y$  are independent and

$$f_X(x) = e^{-x} ; \quad x \geq 0 \quad \text{and} \quad f_Y(y) = e^{-y} ; \quad y \geq 0 \quad \dots (2)$$

$$(i) P(X > 1) = \int_1^{\infty} f_X(x) dx = \int_1^{\infty} e^{-x} dx$$

$$= \left| \frac{e^{-x}}{-1} \right|_1^{\infty} = \frac{1}{e}$$

$$(ii) P(X < Y | X < 2Y) = \frac{P(X < Y \cap X < 2Y)}{P(X < 2Y)}$$

$$= \frac{P(X < Y)}{P(X < 2Y)} \quad \dots (3)$$



$$P(X < Y) = \int_0^{\infty} \left[ \int_0^y f(x, y) dx \right] dy$$

$$= \int_0^{\infty} \left[ e^{-x}, \left| \frac{e^{-x}}{-1} \right|_0^y \right] dy = - \int_0^{\infty} e^{-x} (e^{-y} - 1) dy$$

$$= - \left| \frac{e^{-2y}}{-2} + e^{-y} \right|_0^{\infty} = 1 - \frac{1}{2} = \frac{1}{2}$$

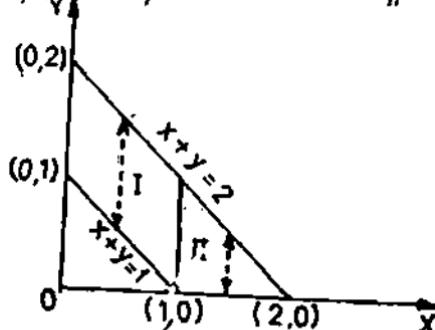
$$P(X < 2Y) = \int_0^{\infty} \left[ \int_0^{2y} f(x, y) dx \right] dy = - \int_0^{\infty} e^{-x} (e^{-2y} - 1) dy$$

$$= - \left| \frac{e^{-3y}}{-3} + e^{-y} \right|_0^{\infty} = 1 - \frac{1}{3} = \frac{2}{3}$$

Substituting in (3),

$$P(X < Y | X < 2Y) = \frac{1/2}{2/3} = \frac{3}{4}$$

$$(iii) P(1 < X + Y < 2) = \iint_{I'} f(x, y) dx dy = \iint_{II} f(x, y) dx dy$$



$$\begin{aligned}
 &= \int_0^1 \left( \int_{1-x}^{2-x} f(x, y) dy \right) dx + \int_1^2 \left( \int_0^{2-x} f(x, y) dy \right) dx \\
 &= \int_0^1 \left( e^{-x} \int_{1-x}^{2-x} e^{-y} dy \right) dx + \int_1^2 \left( e^{-x} \int_0^{2-x} e^{-y} dy \right) dx \\
 &= \int_0^1 \frac{e^{-x}}{-1} (e^{x-2} - e^{x-1}) dx + \int_1^2 \frac{e^{-x}}{-1} (e^{x-2} - 1) dx \\
 &= - (e^{-2} - e^{-1}) \int_0^1 1 \cdot dx - \int_1^2 (e^{-2} - e^{-x}) dx \\
 &= - (e^{-2} - e^{-1}) \left| x \right|_0^1 - \left| e^{-2} \cdot x + e^{-x} \right|_1^2 \\
 &= 2/e - 3/e^2
 \end{aligned}$$

**Example 5-29.** (i) Let  $F(x, y)$  be the d.f. of  $X$  and  $Y$ . Show that  $P(a < X \leq b, c < Y \leq d) = F(b, d) - F(b, c) - F(a, d) + F(a, c)$  where  $a, b, c, d$  are real constants  $a < b ; c < d$ .

Deduce that if:  $F(x, y) = 1$ , for  $x + 2y \geq 1$

$$F(x, y) = 0, \text{ for } x + 2y < 1,$$

then  $F(x, y)$  cannot be joint distribution function of variables  $X$  and  $Y$ .

(ii) Show that, with usual notation : for all  $x, y$ ,

$$F_X(x) + F_Y(y) - 1 \leq F_{XY}(x, y) \leq \sqrt{F_X(x) F_Y(y)}$$

[Delhi Univ. B.Sc. (Maths Hons.), 1985]

**Solution.** (i) Let us define the events :

$$A : \{X \leq a\}; B : \{X \leq b\}; C : \{Y \leq c\}; D : \{Y \leq d\};$$

for  $a < b; c < d$ .

$$P(a < X \leq b \cap c < Y \leq d)$$

$$= P[(B - A) \cap (D - C)]$$

$$= P[B \cap (D - C) - A \cap (D - C)] \quad \dots (*)$$

(By distributive property of sets)

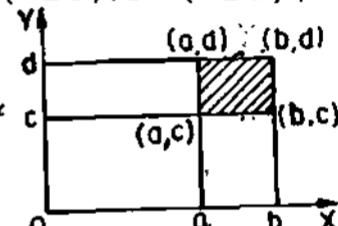
We know that if  $E \subset F \Rightarrow E \cap F = E$ , then

$$P(F - E) = P(\bar{E} \cap F) = P(F) - P(E \cap F) = P(F) - P(E) \quad \dots (**)$$

Obviously  $A \subset B \Rightarrow [A \cap (D - C)] \subset [B \cap (D - C)]$

Hence using (\*\*), we get from (\*)

$$\begin{aligned}
 P(a < X \leq b \cap c < Y \leq d) &= P[B \cap (D - C)] - P[A \cap (D - C)] \\
 &= P[(B \cap D) - (B \cap C)] - P[(A \cap D) - (A \cap C)]
 \end{aligned}$$



$$= P(B \cap D) - P(B \cap C) - P(A \cap D) + P(A \cap C) \dots (***)$$

[On using (\*\*), since  $C \subset D \Rightarrow (B \cap C) \subset (B \cap D)$  and  $(A \cap C) \subset (A \cap D)$ ]

We have:

$$P(B \cap D) = P[X \leq b \cap Y \leq d] = F(b, d).$$

Similarly

$$P(B \cap C) = F(b, c); P(A \cap D) = F(a, d) \text{ and } P(A \cap C) = F(a, c)$$

Substituting in (\*\*\*) we get:

$$P(a < X \leq b \cap c < Y \leq d) = F(b, d) - F(b, c) - F(a, d) + F(a, c) \dots (1)$$

$$\text{We are given } F(x, y) = \begin{cases} 1, & \text{for } x+2y \geq 1 \\ 0, & \text{for } x+2y < 1 \end{cases} \dots (2)$$

In (1) let us take:  $a = 0, b = 1/2, c = 1/4, d = 3/4$  s.t.  $a < b$  and  $c < d$ . Then using (2) we get:

$$F(b, d) = 1; F(b, c) = 1; F(a, d) = 1; F(a, c) = 0.$$

Substituting in (1) we get:

$$P(a < X \leq b \cap c < Y \leq d) = 1 - 1 - 1 + 0 = -1;$$

which is not possible since  $P(\cdot) \geq 0$ .

Hence  $F(x, y)$  defined in (2) cannot be the distribution function of variates  $X$  and  $Y$ .

(ii) Let us define the events:  $A = \{X \leq x\}; B = \{Y \leq y\}$

$$\text{Then } P(A) = P(X \leq x) = F_X(x); P(B) = P(Y \leq y) = F_Y(y) \quad \text{and, } P(A \cap B) = P(X \leq x \cap Y \leq y) = F_{XY}(x, y) \dots (3)$$

$$(A \cap B) \subset A \Rightarrow P(A \cap B) \leq P(A) \Rightarrow F_{XY}(x, y) \leq F_X(x)$$

$$(A \cap B) \subset B \Rightarrow P(A \cap B) \leq P(B) \Rightarrow F_{XY}(x, y) \leq F_Y(y)$$

Multiplying these inequalities we get:

$$F_{XY}(x, y) \leq F_X(x) F_Y(y) \Rightarrow F_{XY}(x, y) \leq \sqrt{F_X(x) F_Y(y)} \dots (4)$$

$$\text{Also } P(A \cup B) \leq 1 \Rightarrow P(A) + P(B) - P(A \cap B) \leq 1$$

$$\Rightarrow P(A) + P(B) - 1 \leq P(A \cap B)$$

$$\Rightarrow F_X(x) + F_Y(y) - 1 \leq F_{XY}(x, y) \dots (5)$$

From (4) and (5) we get:

$$F_X(x) + F_Y(y) - 1 \leq F_{XY}(x, y) \leq \sqrt{F_X(x) F_Y(y)}, \text{ as required.}$$

**Example 5-30.** If  $X$  and  $Y$  are two random variables having joint density function

$$f(x, y) = \frac{1}{8} (6 - x - y); 0 < x < 2, 2 < y < 4 \\ = 0, \text{ otherwise}$$

Find (i)  $P(X < 1 \cap Y < 3)$ , (ii)  $P(X + Y < 3)$  and (iii)  $P(X < 1 | Y < 3)$

(Madras Univ. B.Sc., Nov. 1986)

**Solution.** We have

$$(i) \quad P(X < 1 \cap Y < 3) = \int_{-\infty}^1 \int_{-\infty}^3 f(x, y) dx dy \\ = \int_0^1 \int_2^3 \frac{1}{8}(6-x-y) dx dy = \frac{3}{8}$$

(ii) The probability that  $X + Y$  will be less than 3 is

$$P(X + Y < 3) = \int_0^1 \int_2^3 \frac{1}{8}(6-x-y) dx dy = \frac{5}{24}$$

(iii) The probability that  $X < 1$  when it is known that  $Y < 3$  is

$$P(X < 1 | Y < 3) = \frac{P(X < 1 \cap Y < 3)}{P(Y < 3)} = \frac{3/8}{5/8} = \frac{3}{5} \\ [P(Y < 3) = \int_0^2 \int_2^3 \frac{1}{8}(6-x-y) dx dy = \frac{5}{8}]$$

**Example 5-31.** If the joint distribution function of  $X$  and  $Y$  is given by :

$$F(x, y) = 1 - e^{-x} - e^{-y} + e^{-(x+y)} ; \quad x > 0, y > 0 \\ = 0 ; \quad \text{elsewhere}$$

(a) Find the marginal densities of  $X$  and  $Y$ .

(b) Are  $X$  and  $Y$  independent?

(c) Find  $P(X \leq 1 \cap Y \leq 1)$  and  $P(X + Y \leq 1)$ . (I.C.S., 1989)

**Solution.** (a) & (b) The joint p.d.f. of the r.v.'s  $(X, Y)$  is given by:

$$f_{XY}(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y} = \frac{\partial}{\partial x} [e^{-y} - e^{-(x+y)}] \\ = e^{-(x+y)} ; \quad x \geq 0, y \geq 0 \\ = 0 ; \quad \text{otherwise} \quad \dots (i)$$

We have

$$f_{XY}(x, y) = e^{-x} \cdot e^{-y} = f_X(x) f_Y(y) \quad \dots (ii)$$

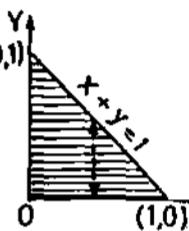
$$\text{where } f_X(x) = e^{-x} ; \quad x \geq 0 ; \quad f_Y(y) = e^{-y} ; \quad y \geq 0 \quad \dots (iii)$$

(ii)  $\Rightarrow X$  and  $Y$  are independent,

and (iii) gives the marginal p.d.f.'s of  $X$  and  $Y$ .

$$(c) \quad P(X \leq 1 \cap Y \leq 1) = \int_0^1 \int_0^1 f(x, y) dx dy \\ = \left( \int_0^1 e^{-x} dx \right) \left( \int_0^1 e^{-y} dy \right) \\ = (1 - e^{-1})^2$$

$$\begin{aligned}
 P(X+Y \leq 1) &= \int \int_{x+y \leq 1} f(x,y) = \int_0^1 \left( \int_0^{1-x} f(x,y) dy \right) dx \\
 &= \int_0^1 \left[ e^{-x} \int_0^{1-x} e^{-y} dy \right] dx \\
 &\approx \int_0^1 e^{-x} \left( 1 - e^{-(1-x)} \right) dx = 1 - 2e^{-1}
 \end{aligned}$$



**Example 5.32.** Joint distribution of  $X$  and  $Y$  is given by

$$f(x,y) = 4xy e^{-(x^2+y^2)}; \quad x \geq 0, y \geq 0.$$

Test whether  $X$  and  $Y$  are independent.

For the above joint distribution, find the conditional density of  $X$  given  $Y = y$ . (Calicut Univ. B.Sc., 1986)

**Solution.** Joint p.d.f. of  $X$  and  $Y$  is

$$f(x,y) = 4xy e^{-(x^2+y^2)}; \quad x \geq 0, y \geq 0.$$

Marginal density of  $X$  is given by

$$\begin{aligned}
 f_1(x) &= \int_0^\infty f(x,y) dy = \int_0^\infty 4xy e^{-(x^2+y^2)} dy \\
 &= 4x e^{-x^2} \int_0^\infty y e^{-y^2} dy \\
 &= 4x e^{-x^2} \cdot \int_0^\infty e^{-t} \cdot \frac{dt}{2} \quad (\text{Put } y^2 = t) \\
 &= 2x \cdot e^{-x^2} \Big| - e^{-t} \Big|_0^\infty \\
 \Rightarrow f_1(x) &= 2x e^{-x^2}; \quad x \geq 0
 \end{aligned}$$

Similarly, the marginal p.d.f. of  $Y$  is given by

$$f_2(y) = \int_0^\infty f(x,y) dx = 2y e^{-y^2}; \quad y \geq 0$$

Since  $f(x,y) = f_1(x) \cdot f_2(y)$ ,  $X$  and  $Y$  are independently distributed. The conditional distribution of  $X$  for given  $Y$  is given by :

$$f(X=x | Y=y) = \frac{f(x,y)}{f_2(y)}$$

$$= 2x e^{-x^2}; \quad x \geq 0.$$

**EXERCISE 5(e)**

1. (a) Two fair dice are tossed simultaneously. Let  $X$  denote the number on the first die and  $Y$  denote the number on the second die.

(i) Write down the sample space of this experiment.

(ii) Find the following probabilities :

- (1)  $P(X+Y=8)$ , (2)  $P(X+Y \geq 8)$ , (3)  $P(X=Y)$ ,  
 (4)  $P(X+Y=6 | Y=4)$ , (5)  $P(X-Y=2)$ .

(Sardar Patel Univ. B.Sc., 1991)

2. (a) Explain the concepts (i) conditional probability, (ii) random variable, (iii) independence of random variables, and (iv) marginal and conditional probability distributions.

(b) Explain the notion of the joint distribution of two random variables. If  $F(x, y)$  be the joint distribution function of  $X$  and  $Y$ , what will be the distribution functions for the marginal distribution of  $X$  and  $Y$ ?

What is meant by the conditional distribution of  $Y$  under the condition that  $X=x$ ? Consider separately the cases where (i)  $X$  and  $Y$  are both discrete and (ii)  $X$  and  $Y$  are both continuous.

3. The joint probability distribution of a pair of random variables is given by the following table :-

$\backslash$	$X$	1	2	3
$Y$		0.1	0.1	0.2
1		0.2	0.3	0.1
2				

Find :

- (i) The marginal distributions.  
 (ii) The conditional distribution of  $X$  given  $Y=1$ .  
 (iii)  $P\{(X+Y) < 4\}$ .

4. (a) What do you mean by marginal and conditional distributions? The following table represents the joint probability distribution of the discrete random variable  $(X, Y)$

$\backslash$	$X$	1	2	3
$Y$		$\frac{1}{12}$	$\frac{1}{6}$	0
1		0	$\frac{1}{6}$	$\frac{1}{5}$
2		$\frac{1}{18}$	$\frac{1}{4}$	$\frac{2}{15}$
3				

- (i) Evaluate marginal distribution of  $X$ .

(ii) Evaluate the conditional distribution of  $Y$  given  $X = 2$ ,

(Aligarh Univ. B.Sc., 1992)

(b) Two discrete random variables  $X$  and  $Y$  have

$$P(X=0, Y=0) = \frac{2}{9}; P(X=0, Y=1) = \frac{1}{9}$$

$$P(X=1, Y=0) = \frac{1}{9}; P(X=1, Y=1) = \frac{5}{9}$$

Examine whether  $X$  and  $Y$  are independent.

(Kerala Univ. B.Sc., Oct. 1987)

5. (a) Let the joint p.m.f. of  $X_1$  and  $X_2$  be

$$\begin{aligned} p(x_1, x_2) &= \frac{x_1 + x_2}{21}; x_1 = 1, 2, 3; x_2 = 1, 2 \\ &= 0, \text{ otherwise} \end{aligned}$$

Show that marginal p.m.f.'s of  $X_1$  and  $X_2$  are

$$p_1(x_1) = \frac{2x_1 + 3}{21}; x_1 = 1, 2, 3; \quad p_2(x_2) = \frac{6 + 3x_2}{21}; x_2 = 1, 2$$

(b) Let

$$\begin{aligned} f(x_1, x_2) &= C(x_1 x_2 + e^{x_1}); 0 < (x_1, x_2) < 1 \\ &= 0, \text{ elsewhere} \end{aligned}$$

(i) Determine  $C$ .

(ii) Examine whether  $X_1$  and  $X_2$  are stochastically independent.

$$\text{Ans. (i)} C = \frac{4}{4e - 3}, \quad \text{(ii)} \quad g(x_1) = C\left(\frac{1}{2}x_1 + e^{x_1}\right),$$

$$g(x_2) = C\left(\frac{1}{2}x_2 + e^{-1}\right)$$

Since  $g(x_1) \cdot g(x_2) \neq f(x_1, x_2)$ ,  $X_1$  and  $X_2$  are not stochastically independent.

6. Find  $k$  so that  $f(x, y) = kxy$ ,  $1 \leq x \leq y \leq 2$  will be a probability density function. (Mysore Univ. B.Sc., 1986)

$$\text{Hint. } \int \int f(x, y) dx dy = 1 \Rightarrow k \int_1^2 x \left( \int_x^2 y dy \right) dx = 1 \Rightarrow k = 8/9$$

$$7. (a) \text{ If } f(x, y) = e^{-(x+y)}; x \geq 0, y \geq 0 \\ = 0, \text{ elsewhere}$$

is the joint probability density function of random variables  $X$  and  $Y$ , find

(i)  $P(X < 1)$ , (ii)  $P(X > Y)$ , and (iii)  $P(X + Y < 1)$ .

$$\text{Ans. (i)} 1 - \frac{1}{e}, \quad \text{(ii)} \frac{1}{2} \quad \text{and (iii)} 1 - \frac{2}{e}$$

(b) The joint frequency function of  $(X, Y)$  is given to be

$$f(x, y) = A e^{-x-y}; \quad 0 \leq x \leq y, \quad 0 \leq y < +\infty \\ = 0; \quad \text{otherwise}$$

(i) Determine  $A$ .

(ii) Find the marginal density function of  $X$ .

(iii) Find the marginal density function of  $Y$ .

(iv) Examine if  $X$  and  $Y$  are independent.

(v) Find the conditional density function of  $Y$  given  $X = 2$ .

[Madras Univ. B.Sc. (Main Stat.), 1992]

(c) Suppose that the random variables  $X$  and  $Y$  have the joint p.d.f.

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

(i) Evaluate the constant  $k$ .

(ii) Find the marginal probability density functions of the random variables.  
[South Gujarat Univ. B.Sc., 1988]

8. (a) Two-dimensional random variable  $(X, Y)$  have the joint density

$$f(x, y) = 8xy, \quad 0 < x < y < 1 \\ = 0, \quad \text{otherwise}$$

(i) Find,  $P(X < 1/2 \cap Y < 1/4)$ .

(ii) Find the marginal and conditional distributions.

(iii) Are  $X$  and  $Y$  independent? Give reasons for your answer.

[South Gujarat Univ. B.Sc., 1992]

$$f_1(x) = 4x(1-x^2), \quad 0 < x < 1 \quad f_1(x|y) = 2x/y^2; \quad 0 < x < y, \quad 0 < y < 1$$

Ans.  $= 0, \text{ otherwise}$

$$f_2(y) = 4y^3, \quad 0 < y < 1 \quad f_2(y|x) = 2y/(1-x^2); \quad x < y < 1, \quad 0 < x < 1$$

9. (a) The random variables  $X$  and  $Y$  have the joint density function :

$$f(x, y) = 2, \quad \text{if } x+y \leq 1, \quad x \geq 0 \text{ and } y \geq 0 \\ = 0, \quad \text{otherwise}$$

Find the conditional distribution of  $Y$ , given  $X = x$ .

[Calcutta Univ. B.Sc. (Hons.), 1984]

(b) The random variables  $X$  and  $Y$  have the joint distribution given by the probability density function :

$$f(x, y) = \begin{cases} 6(1-x-y), & \text{for } x > 0, \quad y > 0, \quad x+y < 1 \\ 0, & \text{elsewhere} \end{cases}$$

Find the marginal distributions of  $X$  and  $Y$ . Hence examine if  $X$  and  $Y$  are independent. [Calcutta Univ. B.Sc. (Hons.), 1986]

10. If the joint distribution function of  $X$  and  $Y$  is given by

$$F(x, y) = (1 - e^{-x})(1 - e^{-y}) \quad \text{for } x > 0, \quad y > 0 \\ = 0, \quad \text{elsewhere}$$

Find  $P(1 < X < 3, 1 < Y < 2)$ . [Delhi Univ. M.A.(Econ.), 1988]

$$\text{Hint. Reqd. Prob.} = \left( \int_1^3 e^{-x} dx \right) \left( \int_1^2 e^{-y} dy \right) = (1 - e^{-3})(1 - e^{-2})$$

11. Let  $X$  and  $Y$  be two random variables with the joint probability density function

$$f(x, y) = \begin{cases} 8xy, & 0 < x \leq y < 1 \\ 0, & \text{otherwise} \end{cases}$$

Obtain :

- (i) the joint distribution function of  $X$  and  $Y$ .
- (ii) the marginal probability density function of  $Y$ ; and
- (iii)  $P(X \leq \frac{1}{4} | \frac{1}{2} < Y \leq 1)$ .

12. Let  $X$  and  $Y$  be jointly distributed with p.d.f.

$$f(x, y) = \begin{cases} \frac{1}{4}(1+xy), & |x| < 1, |y| < 1 \\ 0, & \text{otherwise} \end{cases}$$

Show that  $X$  and  $Y$  are not independent but  $X^2$  and  $Y^2$  are independent.

$$\text{Hint. } f_1(x) = \int_{-1}^1 f(x, y) dy = \frac{1}{2}, \quad -1 < x < 1;$$

$$f_2(y) = \int_{-1}^1 f(x, y) dx = \frac{1}{2}, \quad -1 < y < 1$$

Since  $f(x, y) \neq f_1(x)f_2(y)$ ,  $X$  and  $Y$  are not independent. However,

$$P(X^2 \leq x) = P(|X| \leq \sqrt{x}) = \int_{-\sqrt{x}}^{\sqrt{x}} f_1(x) dx = \sqrt{x}$$

$$\begin{aligned} P(X^2 \leq x \cap Y^2 \leq y) &= P(|X| \leq \sqrt{x} \cap |Y| \leq \sqrt{y}) \\ &= \int_{-\sqrt{x}}^{\sqrt{x}} \left[ \int_{-\sqrt{y}}^{\sqrt{y}} f(u, v) dv \right] du \\ &= \sqrt{x} \cdot \sqrt{y} \\ &= P(X^2 \leq x) \cdot P(Y^2 \leq y) \end{aligned}$$

$\Rightarrow X^2$  and  $Y^2$  are independent.

13. (a) The joint probability density function of the two dimensional random variable  $(X, Y)$  is given by :

$$f(x, y) = \begin{cases} x^3 y^3 / 16 & , 0 \leq x \leq 2, 0 \leq y \leq 2 \\ 0 & , elsewhere \end{cases}$$

Find the marginal densities of  $X$  and  $Y$ . Also find the cumulative distribution functions for  $X$  and  $Y$ .  
 (Annamalai Univ. B.E., 1986)

**Ans.**  $f_x(x) = \frac{x^3}{4} ; 0 \leq x \leq 2 ; f_y(y) = \frac{y^3}{4} ; 0 \leq y \leq 2$

$$F_x(x) = \begin{cases} 0 & ; x < 0 \\ x^4 / 16 & ; 0 \leq x \leq 2 \\ 1 & ; x > 2 \end{cases} \quad F_y(y) = \begin{cases} 0 & ; y < 0 \\ y^4 / 16 & ; 0 \leq y \leq 2 \\ 1 & ; y > 2 \end{cases}$$

(b) The joint probability density function of the two dimensional random variable  $(X, Y)$  is given by :

$$f(x, y) = \begin{cases} \frac{8}{9} xy & , 1 \leq x \leq y \leq 2 \\ 0 & , elsewhere \end{cases}$$

- (i) Find the marginal density functions of  $X$  and  $Y$ ,
- (ii) Find the conditional density function of  $Y$  given  $X = x$ , and conditional density function of  $X$  given  $Y = y$ .

[Madras Univ. B.Sc. (Stat. Main), 1987]

**Ans.** (i)  $f_x(x) = \int_x^2 f(x, y) dy = \frac{4}{9} x (4 - x^2) ; 1 \leq x \leq 2$   
 $= 0 ; \text{ otherwise}$

$$f_y(y) = \int_1^y f(x, y) dx = \frac{4}{9} y (y^2 - 1) ; 1 \leq y \leq 2$$

$$f_{x|y}(x|y) = \frac{2x}{y^2 - 1} ; 1 \leq x \leq y$$

$$f_{y|x}(y|x) = \frac{f(x, y)}{f_x(y)} = \frac{2y}{4 - x^2} ; x \leq y \leq 2$$

14. The two random variables  $X$  and  $Y$  have, for  $X = x$  and  $Y = y$ , the joint probability density function :

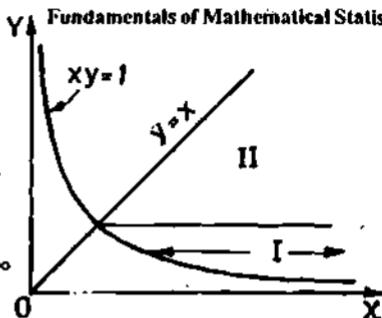
$$f(x, y) = \frac{1}{2x^2 y}, \text{ for } 1 \leq x < \infty \text{ and } \frac{1}{x} < y < x$$

Derive the marginal distributions of  $X$  and  $Y$ . Further obtain the conditional distribution of  $Y$  for  $X = x$  and also that of  $X$  given  $Y = y$ .

(Civil Services Main, 1986)

**Hint.**  $f_x(x) = \int_y f(x, y) dy = \int_y^x f(x, y) dy$

$$\begin{aligned}
 f_Y(y) &= \int_{-\infty}^{\infty} f(x, y) dx \\
 &= \int_0^{\infty} f(x, y) dx; \quad 0 \leq y \leq 1 \\
 &= \int_y^{\infty} f(x, y) dx; \quad 1 \leq y < \infty
 \end{aligned}$$



15. Show that the conditions for the function

$$f(x, y) = k \exp [Ax^2 + 2Hxy + By^2], \quad -\infty < (x, y) < \infty$$

to be a bivariate p.d.f. are

$$(i) A \leq 0, \quad (ii) B \leq 0 \quad (iii) AB - H^2 \geq 0.$$

Further show that under these conditions

$$k = \frac{1}{\pi} (AB - H^2)^{1/2}$$

**Hint.**  $f(x, y)$  will represent the p.d.f. of a bivariate distribution if and only if

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = 1$$

$$\Rightarrow k \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp [Ax^2 + 2Hxy + By^2] dx dy = 1 \quad \dots (*)$$

We have

$$\begin{aligned}
 Ax^2 + 2Hxy + By^2 &= A \left[ x^2 + \frac{2H}{A} xy + \frac{B}{A} y^2 \right] \\
 &= A \left[ \left( x + \frac{H}{A} y \right)^2 + \frac{AB - H^2}{A^2} y^2 \right] \quad \dots (**)
 \end{aligned}$$

Similarly, we can write

$$Ax^2 + 2Hxy + By^2 = B \left[ \left( y + \frac{H}{B} x \right)^2 + \frac{AB - H^2}{B^2} x^2 \right] \quad \dots (***)$$

Substituting from (\*\*) and (\*\*\*)) in (\*) we observe that the double integral on the left hand side will converge if and only if

$$A \leq 0, \quad B \leq 0 \quad \text{and} \quad AB - H^2 \geq 0,$$

as desired.

Let us take  $A = -a$ ;  $B = -b$ ;  $H = h$  so that  $AB - H^2 = ab - h^2$ , where  $a > 0, b > 0$ .

Substituting in (\*), we get

$$\begin{aligned} & k \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp \left[ -\frac{ab-h^2}{a} y^2 - \frac{1}{a} (-ax+hy)^2 \right] dx dy = 1 \\ \Rightarrow & k \int_{-\infty}^{\infty} \left[ \exp \left( -\frac{ab-h^2}{a} y^2 \right) \cdot \int_{-\infty}^{\infty} \exp \left( -\frac{1}{a} (ax-hy)^2 \right) dx \right] dy \\ & \qquad \qquad \qquad = 1 \quad \dots (***) \end{aligned}$$

(By Fubini's theorem)

$$\begin{aligned} \text{Now } \int_{-\infty}^{\infty} \exp \left\{ -\frac{1}{a} (ax-hy)^2 \right\} dx &= \int_{-\infty}^{\infty} \exp \left( -\frac{u^2}{a} \right) \frac{du}{a} \\ &= \frac{1}{a} \sqrt{\pi} \sqrt{a} = \sqrt{\frac{\pi}{a}} \\ &\quad \left( \because \int_{-\infty}^{\infty} e^{-c^2 u^2} du = \frac{\sqrt{\pi}}{c} \right) \end{aligned}$$

Hence from (\*\*\*) , we get

$$\begin{aligned} & k \sqrt{\frac{\pi}{a}} \int_{-\infty}^{\infty} \exp \left\{ -\frac{ab-h^2}{a} y^2 \right\} dy = 1 \\ \Rightarrow & k \sqrt{\frac{\pi}{a}} \cdot \sqrt{\frac{\pi a}{ab-h^2}} = 1 \\ \Rightarrow & k = \frac{1}{\pi} \sqrt{ab-h^2} = \frac{1}{\pi} \sqrt{AB-H^2}. \end{aligned}$$

### OBJECTIVE TYPE QUESTIONS

I. Which of the following statements are TRUE or FALSE.

(i) Given a continuous random variable  $X$  with probability density function  $f(x)$ , then  $f(x)$  cannot exceed unity.

(ii) A random variable  $X$  has the following probability density function :

$$\begin{aligned} f(x) &= x, \quad 0 < x < 1 \\ &= 0, \text{ elsewhere} \end{aligned}$$

(iii) The function defined as

$$\begin{aligned} f(x) &= |x|, \quad -1 < x < 1 \\ &= 0, \text{ elsewhere} \end{aligned}$$

is a possible probability density function.

(iv) The following represents joint probability distribution.

		$X$		
		1	2	3
$Y$	-1	1/9	1/18	1/18
	0	1/18	2/9	3/9
	1	1/8	1/18	1/18

**II. Fill in the blanks :**

(i) If  $p_1(x)$  and  $p_2(y)$  be the marginal probability functions of two independent discrete random variables  $X$  and  $Y$ , then their joint probability function

$$p(x, y) = \dots$$

(ii) The function  $f(x)$  defined as

$$\begin{aligned} f(x) &= |x|, -1 < x < 1 \\ &= 0, \text{ elsewhere} \end{aligned}$$

is a possible .....

**5.6. Transformation of One-dimensional Random Variable.** Let  $X$  be a random variable defined on the event space  $S$  and let  $g(\cdot)$  be a function such that  $Y = g(X)$  is also a r.v. defined on  $S$ . In this section we shall deal with the following problem :

"Given the probability density of a r.v.  $X$ , to determine the density of a new r.v.  $Y = g(X)$ ."

It can be proved in general that, if  $g(\cdot)$  is any continuous function, then the distribution of  $Y = g(X)$  is uniquely determined by that of  $X$ . The proof of this result is rather difficult and beyond the scope of this book. Here we shall consider the following, relatively simple theorem.

**Theorem 5.9.** Let  $X$  be a continuous r.v. with p.d.f.  $f_X(x)$ . Let  $y = g(x)$  be strictly monotonic (increasing or decreasing) function of  $x$ . Assume that  $g(x)$  is differentiable (and hence continuous) for all  $x$ . Then the p.d.f. of the r.v.  $Y$  is given by

$$h_Y(y) = f_X(x) \left| \frac{dx}{dy} \right|,$$

where  $x$  is expressed in terms of  $y$ .

**Proof.** Case (i).  $y = g(x)$  is strictly increasing function of  $x$  (i.e.,  $dy/dx > 0$ ). The d.f. of  $Y$  is given by

$$H_Y(y) = P(Y \leq y) = P[g(X) \leq y] = P(X \leq g^{-1}(y)).$$

the inverse exists and is unique, since  $g(\cdot)$  is strictly increasing.

$$\therefore H_Y(y) = F_X[g^{-1}(y)], \text{ where } F \text{ is the d.f. of } X$$

$$= F_X(x) \quad [\because y = g(x) \Rightarrow g^{-1}(y) = x]$$

Differentiating w.r.t  $y$ , we get

$$\begin{aligned} h_Y(y) &= \frac{d}{dy} [F_X(x)] = \frac{d}{dx} (F_X(x)) \frac{dx}{dy} \\ &= f_X(x) \frac{dx}{dy} \end{aligned} \quad \dots (*)$$

**Case (ii).**  $y = g(x)$  is strictly monotonic decreasing.

$$\begin{aligned}H_Y(y) &= P(Y \leq y) = P[g(X) \leq y] = P[X \geq g^{-1}(y)] \\&= 1 - P[X \leq g^{-1}(y)] = 1 - F_X[g^{-1}(y)] = 1 - F_X(x),\end{aligned}$$

where  $x = g^{-1}(y)$ , the inverse exists and is unique. Differentiating w.r.t.  $y$ , we get

$$\begin{aligned}h_Y(y) &= \frac{d}{dx} [1 - F_X(x)] \frac{dx}{dy} = -f_X(x) \cdot \frac{dx}{dy} \\&= f_X(x) \cdot \frac{-dx}{dy} \quad \dots (**)\end{aligned}$$

Note that the algebraic sign (-ive) obtained in (\*\*) is correct, since  $y$  is a decreasing function of  $x \Rightarrow x$  is a decreasing function of  $y \Rightarrow dx/dy < 0$ .

The results in (\*) and (\*\*) can be combined to give

$$h_Y(y) = f_X(x) \left| \frac{dx}{dy} \right|$$

**Example 5.33.** If the cumulative distribution function of  $X$  is  $F(x)$ , find the cumulative distribution function of

- (i)  $Y = X + a$ ,      (ii)  $Y = X - b$ ,      (iii)  $Y = aX$ ,  
 (iv)  $Y = X^3$ , and      (v)  $Y = X^2$

What are the corresponding probability density functions?

**Solution.** Let  $G(\cdot)$  be the c.d.f. of  $Y$ . Then

- (i)  $G(x) = P(Y \leq x) = P[X + a \leq x] = P[X \leq x - a] = F(x - a)$   
 (ii)  $G(x) = P(Y \leq x) = P[X - b \leq x] = P[X \leq x + b] = F(x + b)$

$$\begin{aligned}\text{(iii)} \quad G(x) &= P[aX \leq x] = P\left[X \leq \frac{x}{a}\right], \quad a > 0 \\&= F\left(\frac{x}{a}\right), \quad \text{if } a > 0\end{aligned}$$

$$\begin{aligned}\text{and} \quad G(x) &= P\left[X \geq \frac{x}{a}\right] = 1 - P\left[X < \frac{x}{a}\right] \\&= 1 - F\left(\frac{x}{a}\right), \quad \text{if } a < 0\end{aligned}$$

- (iv)  $G(x) = P[Y \leq x] = P[X^3 \leq x] = P[X \leq x^{1/3}] = F(x^{1/3})$   
 (v)  $G(x) = P[X^2 \leq x] = [-x^{1/2} \leq X \leq x^{1/2}]$   
 $= P[X \leq x^{1/2}] - P[X \leq -x^{1/2}]$

$$= 0, \quad \text{if } x < 0 \\ = F(\sqrt{x}) - F(-\sqrt{x} - 0), \quad \text{if } x > 0$$

Variable	df.	p.d.f.
$X$	$F(x)$	$f(x)$
$X - a$	$F(x+a)$	$f(x+a)$
$aX$	$\begin{cases} F(x/a) & a > 0 \\ 1 - F(x/a), & a < 0 \end{cases}$	$\begin{cases} (1/a) f(x/a), & a > 0 \\ (-1/a) f(x/a), & a < 0 \end{cases}$
$X^2$	$\begin{cases} F(\sqrt{x}) - F(-\sqrt{x} - 0) & \text{for } x > 0 \\ 0, & \text{otherwise} \end{cases}$	$\begin{cases} \frac{1}{2\sqrt{x}} [f(\sqrt{x}) + f(-\sqrt{x})] & \text{for } x > 0 \\ = 0 \text{ for } x \leq 0 \end{cases}$
$X^3$	$F(x^{1/3})$	$\frac{1}{3} f(x^{1/3}) \cdot \frac{1}{x^{2/3}}$

### EXERCISE 5(f)

1. (a) A random variable  $X$  has  $F(x)$  as its distribution function [ $f(x)$  is the density function]. Find the distribution and the density functions of the random variable :

- (i)  $Y = a + bX$ ,  $a$  and  $b$  are real numbers, (ii)  $Y = X^{-1}$ ,  $[P(X=0)=0]$ ,
- (iii)  $Y = \tan X$ , and (iv)  $Y = \cos X$ .

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

be the p.d.f. of the r.v.  $X$ . Find the distribution function and the p.d.f. of  $Y = X^2$ .  
[ Delhi Univ. B.Sc. (Maths Hons.), 1988 ]

Hint.  $F(x) = P(X \leq x) = \int_{-1}^x f(x) dx = \frac{1}{2}(x+1)$  ... (\*)

Distribution function  $G(\cdot)$  of  $Y = X^2$  is given by :

$$G_Y(x) = F(\sqrt{x}) - F(-\sqrt{x}) \quad ; \quad x > 0 \quad [\text{c.f. Example 5-33 (v)}]$$

$$\begin{aligned} &\approx \frac{1}{2}(\sqrt{x} + 1) - \frac{1}{2}(-\sqrt{x} + 1) \\ &= \sqrt{x} \quad ; \quad 0 < x < 1 \end{aligned} \quad [\text{From (*)}]$$

(As  $-1 < x < 1$ ,  $Y = X^2$  lies between 0 and 1)

$$\text{p.d.f. of } Y = X^2 \text{ is } g(x) = G'(x) = \frac{1}{2\sqrt{x}} \quad ; \quad 0 < x < 1$$

2. Let  $X$  be a continuous random variable with p.d.f.  $f(x)$ . Let  $Y = X^2$ . Show that the random variable  $Y$  has p.d.f. given by

$$g(y) = \begin{cases} \frac{1}{2\sqrt{y}} [f(\sqrt{y}) + f(-\sqrt{y})], & y > 0 \\ 0, & y \leq 0 \end{cases}$$

3. Find the distribution and density functions for (i)  $Y = aX + b$ ,  $a \neq 0$ ,  $b$  real, (ii)  $Y = e^X$ , assuming that  $F(x)$  and  $f(x)$ , the distribution and the density of  $X$  are known.

$$\text{Ans. (i)} \quad G(y) = F[(y-b)/a], \quad \text{if } a > 0 \\ G(y) = 1 - F[(y-b)/a], \quad \text{if } a < 0 \\ g(y) = \frac{1}{|a|} f\left(\frac{y-b}{a}\right)$$

$$\text{(ii)} \quad G(y) = F(\log y), \quad y > 0 \\ = 0, \quad y \leq 0 \quad g(y) = \frac{1}{y} f(\log y), \quad y > 0 \\ = 0, \quad y \leq 0$$

4. (a) The random variable  $X$  has an exponential distribution

$$f(x) = e^{-x}, \quad 0 < x \leq \infty$$

Find the density function of the variable (i)  $Y = 3X + 5$ , (ii)  $Y = X^3$ .

(b) Suppose that  $X$  has p.d.f.,

$$f(x) = 2x, \quad 0 < x < 1 \\ = 0, \quad \text{elsewhere}$$

Find the p.d.f. of  $Y = 3X + 1$ .

$$\text{Ans. } g(y) = \frac{2}{9}(y-1), \quad 1 < y < 4$$

5. Let  $X$  be a random variable with p.d.f.

$$f(x) = \frac{2}{9}(x+1), \quad -1 < x < 2 \\ = 0, \quad \text{elsewhere}$$

Find the p.d.f. of  $U = X^2$ .

[ Poona Univ. B.E., 1992 ]

6. Let the p.d.f. of  $X$  be

$$f(x) = \frac{1}{6}, \quad -3 \leq x \leq 3 \\ = 0, \quad \text{elsewhere}$$

Find the p.d.f. of  $Y = 2X^2 - 3$ .

7. Let  $X$  be a random variable with the distribution function :

$$F_X(x) = \begin{cases} 0, & x < 0 \\ x, & 0 \leq x \leq 1 \\ 1, & x > 1 \end{cases}$$

Determine the distribution function  $F_Y(y)$  of the random variable  $Y = \sqrt{X}$  and hence compute mean of  $Y$ . [ Calcutta Univ. B.A.(Hons.), 1986 ]

**5.7. Transformation of Two-dimensional Random Variable.** In this section we shall consider the problem of change of variables in the two-dimensional

case. Let the r.v.'s  $U$  and  $V$  by the transformation  $u = u(x, y)$ ,  $v = v(x, y)$ , where  $u$  and  $v$  are continuously differentiable functions for which Jacobian of transformation

$$J = \frac{\partial(x, y)}{\partial(u, v)} = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix}$$

is either  $> 0$  or  $< 0$  throughout the  $(x, y)$  plane so that the inverse transformation is uniquely given by  $x = x(u, v)$ ,  $y = y(u, v)$ .

**Theorem 5.10.** The joint p.d.f.  $g_{uv}(u, v)$  of the transformed variables  $U$  and  $V$  is given by

$$g_{uv}(u, v) = f_{xy}(x, y) |J|$$

where  $|J|$  is the modulus value of the Jacobian of transformation and  $f_{xy}(x, y)$  is expressed in terms of  $u$  and  $v$ .

**Proof.**  $P(x < X \leq x + dx, y < Y \leq y + dy)$

$$= P(u < U \leq u + du, v < V \leq v + dv)$$

$$\Rightarrow f_{xy}(x, y) dx dy = g_{uv}(u, v) du dv$$

$$\Rightarrow g_{uv}(u, v) du dv = f_{xy}(x, y) \left| \frac{\partial(x, y)}{\partial(u, v)} \right| du dv$$

$$\Rightarrow g_{uv}(u, v) = f_{xy}(x, y) \left| \frac{\partial(x, y)}{\partial(u, v)} \right| = f_{xy}(x, y) |J|$$

**Theorem 5.11.** If  $X$  and  $Y$  are independent continuous r.v.'s, then the p.d.f. of  $U = X + Y$  is given by

$$h(u) = \int_{-\infty}^{\infty} f_X(v) f_Y(u-v) dv$$

**Proof.** Let  $f_{xy}(x, y)$  be the joint p.d.f. of independent continuous r.v.'s  $X$  and  $Y$  and let us make the transformation :

$$u = x + y, v = x \Rightarrow x = v, y = u - v$$

$$J = \frac{\partial(x, y)}{\partial(u, v)} = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} 0 & 1 \\ 1 & -1 \end{vmatrix} = -1$$

Thus the joint p.d.f. of r.v.'s  $U$  and  $V$  is given by

$$g_{uv}(u, v) = f_{xy}(x, y) |J|$$

$$= f_X(x) \cdot f_Y(y) |J|$$

(Since  $X$  and  $Y$  are independent)

$$= f_X(v) \cdot f_Y(u-v)$$

The marginal density of  $U$  is given by

$$\begin{aligned} h(u) &= \int_{-\infty}^{\infty} g_{uv}(u, v) dv \\ &= \int_{-\infty}^{\infty} f_X(v) f_Y(u-v) dv \end{aligned}$$

**Remark.** The function  $h(\cdot)$  is given a special name and is said to be the *convolution* of  $f_X(\cdot)$  and  $f_Y(\cdot)$  and we write

$$h(\cdot) = f_X(\cdot) * f_Y(\cdot)$$

**Example 5.34.** Let  $(X, Y)$  be a two-dimensional non-negative continuous r.v. having the joint density :

$$f(x, y) = \begin{cases} 4xy e^{-(x^2+y^2)} & ; x \geq 0, y \geq 0 \\ 0 & , elsewhere \end{cases}$$

Prove that the density function of  $U = \sqrt{x^2 + y^2}$  is .

$$h(u) = \begin{cases} 2u^3 e^{-u^2} & , 0 \leq u < \infty \\ 0 & , elsewhere \end{cases}$$

[ Meerut Univ. M.Sc., 1986 ]

**Solution.** Let us make the transformation :

$$u = \sqrt{x^2 + y^2} \text{ and } v = x$$

$$\Rightarrow v \geq 0, u \geq 0 \text{ and } u \geq v \Rightarrow u \geq 0 \text{ and } 0 \leq v \leq u$$

The Jacobian of transformation  $J$  is given by

$$\frac{1}{J} = \frac{\partial(u, v)}{\partial(x, y)} = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial y}{\partial u} \\ \frac{\partial x}{\partial v} & \frac{\partial y}{\partial v} \end{vmatrix} = -\frac{y}{\sqrt{x^2 + y^2}}$$

The joint p.d.f. of  $U$  and  $V$  is given by

$$\begin{aligned} g(u, v) &= f(x, y) |J| \\ &= 4xy e^{-(x^2+y^2)} \left| \begin{matrix} -\frac{\sqrt{x^2+y^2}}{y} \\ 1 \end{matrix} \right| \\ &= 4x \sqrt{x^2+y^2} e^{-(x^2+y^2)} \\ &= \begin{cases} 4vu \cdot e^{-u^2} & ; u \geq 0, 0 \leq v \leq u \\ 0 & , otherwise \end{cases} \end{aligned}$$

Hence the density function of  $U = \sqrt{X^2 + Y^2}$  is

$$h(u) = \int_0^u g(u, v) dv = 4u e^{-u^2} \int_0^u v dv \\ = \begin{cases} 2u^3 e^{-u^2}, & u \geq 0 \\ 0, & \text{elsewhere} \end{cases}$$

**Example 5.35.** Let the probability density function of the random variable  $(X, Y)$  be

$$f(x, y) = \begin{cases} \alpha^{-2} e^{-(x+y)/\alpha} & ; x, y > 0, \alpha > 0 \\ 0 & , \text{elsewhere} \end{cases}$$

Find the distribution of  $\frac{1}{2}(X - Y)$ . [ Nagpur Univ. B.E., 1988 ]

**Solution.** Let us make the transformation :

$$u = \frac{1}{2}(x - y) \text{ and } v = y$$

$$\Rightarrow x = 2u + v \text{ and } y = v$$

The Jacobian of the transformation is :

$$J = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} 2 & 1 \\ 0 & 1 \end{vmatrix} = 2$$

Thus, the joint p.d.f. of the random variables  $(U, V)$  is given by :

$$g(u, v) = \begin{cases} \frac{2}{\alpha^2} e^{-(2/\alpha)(u+v)}, & -\infty < u < \infty, v > -2u, \text{ if } u < 0 \\ v > 0 \text{ if } u \geq 0 \text{ and } \alpha > 0 \\ 0, & \text{elsewhere} \end{cases}$$

The marginal p.d.f. of  $U$  is given by

$$g_U(u) = \begin{cases} \int_{-2u}^{\infty} \frac{2}{\alpha^2} e^{-(2/\alpha)(u+v)} dv \\ = \frac{1}{\alpha} e^{-2u/\alpha}, & u < 0 \\ \int_0^{\infty} \frac{2}{\alpha} e^{-(2/\alpha)(u+v)} dv \\ = \frac{1}{\alpha} e^{-2u/\alpha}, & u \geq 0 \end{cases}$$

Hence

$$g_U(u) = \frac{1}{\alpha} e^{-(2/\alpha)|u|} ; -\infty < u < \infty$$

**Example 5.36.** Given the joint density function of  $X$  and  $Y$  as

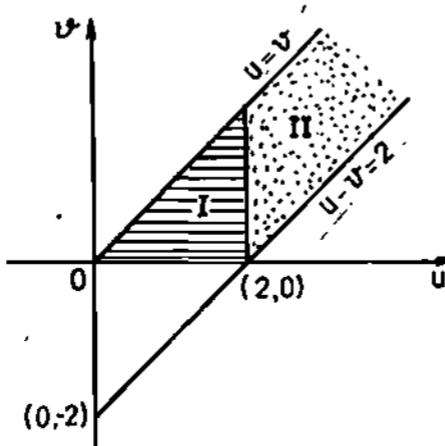
$$f(x, y) = \frac{1}{2}x e^{-y}; 0 < x < 2, y > 0 \\ = 0, \text{ elsewhere}$$

Find the distribution of  $X + Y$ .

**Solution.** Let us make the transformation :

$$u = x + y \text{ and } v = y \Rightarrow y = v, x = u - v$$

The Jacobian of transformation  $J = \frac{\partial(x, y)}{\partial(u, v)} = 1$  and the region  $0 < x < 2$  and  $y > 0$  transforms to  $0 < u - v < 2$  and  $v > 0$  as shown in the following figure.



The joint density function of  $U$  and  $V$  is given by

$$g(u, v) = \frac{1}{2}(u - v)e^{-v}; 0 < v < u, u > 0$$

To find the density of  $U = X + Y$ , we split the range of  $U$  into two parts  
(i)  $0 < u \leq 2$  (region I) (ii)  $u > 2$  (region II) (which is suggested by the diagram).

For  $0 < u \leq 2$ , (Region I) :

$$h(u) = \int_0^u g(u, v) dv = \frac{1}{2} \int_0^u (u - v)e^{-v} dv \\ = \frac{1}{2} \left[ -e^{-v}(u - v) + e^{-v} \right]_{v=0}^{v=u} \quad (\text{Integration by parts}) \\ = \frac{1}{2} (e^{-u} + u - 1)$$

For  $2 < u < \infty$ , (Region II) :

$$\begin{aligned}
 h(u) &= \frac{1}{2} \int_{u-2}^u (u-v) e^{-v} dv \\
 &= \frac{1}{2} \left[ e^{-v} (1+v-u) \right]_{v=u-2}^{v=u} \\
 &= \frac{1}{2} e^{-u} (1+e^2)
 \end{aligned}
 \quad (\text{on simplification})$$

Hence

$$g(u) = \begin{cases} \frac{1}{2}(e^{-u} + u - 1), & 0 < u \leq 2 \\ \frac{1}{2} e^{-u} (1+e^2), & 2 < u < \infty \\ 0, & \text{elsewhere} \end{cases}$$

### 1 MISCELLANEOUS EXERCISE ON CHAPTER FIVE

1. 4 coins are tossed. Let  $X$  be the number of heads and  $Y$  be the number of heads minus the number of tails. Find the probability function of  $X$ , the probability function of  $Y$  and  $P(-2 \leq Y < 4)$ .

**Ans.** Probability function of  $X$  is

Values of $X$ , $x$	0	1	2	3	4
$p_1(x)$	$\frac{1}{16}$	$\frac{4}{16}$	$\frac{6}{16}$	$\frac{4}{16}$	$\frac{1}{16}$

Probability function of  $Y$  is

Values of $Y$ , $y$	4	2	0	-2	-4
$p_2(y)$	$\frac{1}{16}$	$\frac{4}{16}$	$\frac{6}{16}$	$\frac{4}{16}$	$\frac{1}{16}$

$$P(-2 \leq Y < 4) = \frac{4+6+4}{16} = \frac{7}{8}.$$

2. A random process gives measurements  $X$  between 0 and 1 with a probability density function

$$\begin{aligned}
 f(x) &= 12x^2 - 21x^2 + 10x, \quad 0 \leq x \leq 1 \\
 &= 0, \quad \text{elsewhere}
 \end{aligned}$$

(i) Find  $P(X \leq \frac{1}{2})$  and  $P(X > \frac{1}{2})$

(ii) Find a number  $k$  such that  $P(X \leq k) = \frac{1}{2}$ .

**Ans.** (i)  $\frac{1}{16}, \frac{7}{16}$ , (ii)  $k = 0.452$ .

3. Show that for the distribution

$$dF = y_o \left[ 1 - \frac{|x-b|}{a} \right] dx, \quad b-a < x < b+a \\ = 0, \text{ otherwise,} \\ y_o = \frac{1}{a}, \text{ mean } = b \text{ and variance } = a^2/6$$

4. A ray of light is sent in a random direction towards the  $x$ -axis from a station  $Q(0, 1)$  on the  $y$ -axis and the ray meets the  $x$ -axis at a point  $P$ . Find the probability density function of the abscissa of  $P$ .

[Calcutta Univ. B.Sc.(Hons.), 1982]

5. Let  $X$  be a continuous variate with p.d.f.

$$f(x) = k(x-x^2); \quad a < x < b, \quad k > 0$$

What are the possible values of  $a$  and  $b$  and what is  $k$ ?

[Delhi Univ. B.Sc.(Maths Hons.), 1989]

6. Pareto distribution with parameters  $r$  and  $A$  is given by the probability density function

$$f(x) = rA^r \frac{1}{x^{r+1}}, \quad \text{for } x \geq A \\ = 0, \quad x < A, \quad r > 0$$

Show that it has a finite  $n$ th moment if and only if  $n < r$ . Find the mean and variance of the distribution.

7. For a continuous random variable  $X$ , defined in the range  $(0 \leq x < \infty)$ , the probability distribution is such that

$$P(X \leq x) = 1 - e^{-\beta x^2}, \quad \text{where } \beta > 0$$

Find the median of the distribution. Also if  $m$ ,  $m_o$  and  $\sigma$  denote the mean, mode and standard deviation respectively of the distribution, prove that

$$2m_o^2 - m^2 = \sigma^2 \quad \text{and} \quad m_o = m \sqrt{2/\pi}$$

What is the sign of skewness of the distribution?

8. (a) Two dice are rolled,  $S = \{(a, b) \mid a, b = 1, 2, \dots, 6\}$ . Let  $X$  denote the sum of the two faces and  $Y$  the absolute value of their difference, i.e.,  $X$  is distributed over the integers 2, 3, ..., 12 and  $Y$  over 0, 1, 2, ..., 5. Assuming the dice are fair, find the probabilities that (i)  $X = 5 \cap Y = 1$ , (ii)  $X = 7 \cap Y \geq 3$ , (iii)  $X = Y$ , and (iv)  $X + Y = 4 \cap X - Y = 2$ .

Ans. (i)  $1/8$ , (ii)  $1/9$ , (iii) 0 and (iv)  $1/18$ .

9. The joint probability density function of the two-dimensional variable  $(X, Y)$  is of the form

$$f(x, y) = k e^{-(x+y)}, \quad 0 \leq y < x < \infty \\ = 0, \quad \text{elsewhere}$$

(i) Determine the constant  $k$ . (ii) Find the conditional probability density function  $f_1(x|y)$  and (iii) Compute  $P(Y \geq 3)$ .

[Sardar Patel Univ. B.Sc., 1986]

, (iv) Find the marginal frequency function  $f_1(x)$  of  $X$ .

(v) Find the marginal frequency function  $f_2(y)$  of  $Y$ .

(vi) Examine if  $X, Y$  are independent.

(vii) Find the conditional frequency function of  $Y$  given  $X = 2$ .

**Ans.** (i)  $k = 1$ , (ii)  $f_1(x|y) = e^{-x}$ , (iii)  $e^{-3}$ .

**10.** Let

$$f(x, y) = \begin{cases} \binom{y}{x} p^x (1-p)^{y-x} \frac{e^{-\lambda} \lambda^y}{y!} \\ ; x = 0, 1, 2, \dots; y = 0, 1, 2, \dots; \text{ with } y \geq x \\ 0, \text{ elsewhere} \end{cases}$$

Find the marginal density function of  $X$  and the marginal density function of  $Y$ . Also determine whether the random variables  $X$  and  $Y$  are independent.

[I.S.I., 1987]

**11.** Consider the following function :

$$f(x|y) = \begin{cases} \frac{y^x e^{-y}}{x!}, & x = 0, 1, 2, \dots \\ 0, & \text{otherwise} \end{cases}$$

(i) Show that  $f(x|y)$  is the conditional probability function of  $X$  given  $Y$ ;  $y \geq 0$ .

(ii) If the marginal p.d.f. of  $Y$  is

$$f_Y(y) = \begin{cases} \lambda e^{-\lambda y}, & y > 0 \\ 0 & y \leq 0, \lambda > 0 \end{cases}$$

what is the joint p.d.f. of  $X$  and  $Y$  ?

(iii) Obtain the marginal probability function of  $X$ .

[Delhi Univ. M.A.(Econ.), 1989]

**12.** The probability density function of  $(x_1, x_2)$  is given as

$$f(x_1, x_2) = \begin{cases} \theta_1 \theta_2 e^{-\theta_1 x_1 - \theta_2 x_2} & \text{if } x_1, x_2 > 0 \\ 0 & \text{otherwise.} \end{cases}$$

Find the density function of  $(y_1, y_2)$  where

$$y_1 = \frac{2x_1}{x_2} + 1, \quad y_2 = 3x_1 + x_2 \quad \text{almost everywhere.}$$

[Punjab Univ. M.A.(Econ.), 1992]

**13. (a)** Let  $X_1, X_2$  be a random sample of size 2 from a distribution with probability density function,

$$f(x) = e^{-x}, \quad 0 < x < \infty$$

$$= 0, \text{ elsewhere}$$

Show

$$Y_1 = X_1 + X_2, \text{ and } Y_2 = \frac{X_1}{X_1 + X_2}$$

are independent.

[Sardar Patel Univ. B.Sc., Sept. 1986]

(b)  $X_1, X_2, X_3$  denote random sample of size 3 drawn from the distribution:

$$f(x) = e^{-x}, 0 < x < \infty$$

$$= 0, \text{ elsewhere}$$

Show that

$$Y_1 = \frac{X_1}{X_1 + X_2}, Y_2 = \frac{X_1 + X_2}{X_1 + X_2 + X_3} \text{ and } Y_3 = X_1 + X_2 + X_3$$

are mutually independent.

14. If the probability density function of the random variables  $X$  and  $Y|X$  is given by

$$f(x) = \begin{cases} e^{-x}, & x \geq 0 \\ 0, & \text{elsewhere} \end{cases}$$

$$\text{and } f_{Y|X}(y|x) = \begin{cases} \frac{e^{-x} x^y}{y!}, & y \geq 0 \\ 0, & \text{elsewhere} \end{cases}$$

respectively, find the probability density function of the random variable  $Y$ .

[Jiwaji Univ. M.Sc., 1987]

15. (a) The random variable  $X$  and  $Y$  have a joint p.d.f.  $f(x, y)$  given by

$$f(x, y) = g(x+y), \quad x > 0, y > 0$$

$$= 0, \quad \text{otherwise.}$$

Obtain the distribution function  $H(z)$  of  $Z = X + Y$  and hence show that its p.d.f. is

$$h(z) = z g(z), \quad z > 0$$

$$= 0 \quad z \leq 0.$$

(b) The joint density function of two random variables is given by

$$f(x, y) = e^{-(x+y)} ; x > 0, y > 0. \text{ Show that the p.d.f. of}$$

$$U = \frac{X+Y}{2} \text{ is } g(u) = 4ue^{-2u}$$

[Calicut Univ. B.Sc., 1986]

16. The time  $X$  taken by a garage to repair a car is a continuous random variable with probability density function

$$f_1(x) = \begin{cases} \frac{3}{4}x(2-x), & 0 \leq x \leq 2 \\ 0, & \text{elsewhere} \end{cases}$$

If, on leaving his car, a motorist goes to keep an engagement lasting for a time  $Y$ , where  $Y$  is a continuous random variable, independent of  $X$ , with probability function

$$f_2(y) = \begin{cases} \frac{1}{2}y, & 0 \leq y \leq 2 \\ 0, & \text{elsewhere}; \end{cases}$$

determine the probability that the car will not be ready on his return.

[Calcutta Univ. B.A.(Hons.), 1988]

17. If  $X$  and  $Y$  are two independent random variables such that

$$f(x) = e^{-x}, x \geq 0 \text{ and } g(y) = 3e^{-3y}, y \geq 0;$$

find the probability distribution of  $Z = X/Y$ .

[Madurai Univ. B.Sc., Oct. 1987]

18. The random variables  $X$  and  $Y$  are independent and their probability density functions are, respectively given by

$$f(x) = \frac{1}{\pi} \cdot \frac{1}{\sqrt{1+x^2}}, |x| < 1 \text{ and } g(y) = y e^{-y^2/2}, y > 0.$$

Find the joint probability density of  $Z$  and  $W$ , where  $Z = XY$  and  $W = X$ . Deduce the probability density of  $Z$ . [Calcutta Univ. B.Sc.(Hons.), 1985]

## CHAPTER SIX

# *Mathematical Expectation, Generating Functions and Law of Large Numbers*

**6.1. Mathematical Expectation.** Let  $X$  be a random variable (r.v.) with p.d.f. (p.m.f.)  $f(x)$ . Then its mathematical expectation, denoted by  $E(X)$  is given by :

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx, \quad (\text{for continuous r.v.}) \quad \dots(6.1)$$

$$= \sum_{x=-\infty}^{\infty} x f(x), \quad (\text{for discrete r.v.}) \quad \dots(6.1a)$$

provided the righthand integral or series is absolutely convergent, i.e., provided

$$\int_{-\infty}^{\infty} |x f(x)| dx = \int_{-\infty}^{\infty} |x| f(x) dx < \infty \quad \dots(6.2)$$

$$\text{or} \quad \sum_{x=-\infty}^{\infty} |x f(x)| = \sum_{x=-\infty}^{\infty} |x| f(x) < \infty \quad \dots(6.2a)$$

**Remarks.** 1. Since absolute convergence implies ordinary convergence, if (6.2) or (6.2a) holds then the integral or series in (6.1) and (6.1a) also exists, i.e., has a finite value and in that case we define  $E(X)$  by (6.1) or (6.1a). It should be clearly understood that although  $X$  has an expectation only if L.H.S. in (6.2) or (6.2a) exists, i.e., converges to a finite limit, its value is given by (6.1) or (6.1a).

2.  $E(X)$  exists iff  $E|X|$  exists.

3. The expectation of a random variable is thought of as a long-term average. [See Remark to Example (6.2a), page 6.19.]

4. **Expected value and variance of an Indicator Variable.** Consider the indicator variable :  $X = I_A$  so that

$$X = 1 \quad \text{if } A \text{ happens}$$

$$= 0 \quad \text{if } \bar{A} \text{ happens}$$

$$\therefore E(X) = 1 \cdot P(X = 1) + 0 \cdot P(X = 0)$$

$$\Rightarrow E(I_A) = 1 \cdot P[I_A = 1] + 0 \cdot P[I_A = 0]$$

$$\Rightarrow E(I_A) = P(A)$$

This gives us a very useful tool to find  $P(A)$ , rather than to evaluate  $E(X)$ .

Thus  $P(A) = E(I_A) \quad \dots(6.2b)$

For illustration of this result, see Example 6.14, page 6.27.

$$E(X^2) = 1^2 \cdot P(X = 1) + 0^2 \cdot P(X = 0) = P(I_A = 1) = P(A)$$

$$\therefore \text{Var } X = E(X^2) - [E(X)]^2 = P(A) - [P(A)]$$

$$= P(A)[1 - P(A)]$$

$$= P(A) P(\bar{A}) \quad \dots(6.2c)$$

**Illustrations.** If the r.v.  $X$  takes the values  $0!, 1!, 2!, \dots$  with probability

law

$$P(X = x!) = \frac{e^{-1}}{x!}; \quad x = 0, 1, 2, \dots$$

$$\text{then } \sum_{x=0}^{\infty} x! P(X = x!) = e^{-1} \sum_{x=0}^{\infty} 1$$

which is a divergent series. In this case  $E(X)$  does not exist.

More rigorously, let us consider a random variable  $X$  which takes the values

$$x_i = (-1)^{i+1} (i+1); \quad i = 1, 2, 3, \dots$$

with the probability law

$$p_i = P(X = x_i) = \frac{1}{i(i+1)}; \quad i = 1, 2, 3, \dots$$

$$\text{Here } \sum_{i=1}^{\infty} x_i P(X = x_i) = \sum_{i=1}^{\infty} (-1)^{i+1} \left( \frac{1}{i} \right) = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \dots$$

Using Leibnitz test for alternating series the series on right hand side is conditionally convergent since the terms alternate in sign, are monotonically decreasing and converge to zero. By conditional convergence we mean that although  $\sum p_i x_i$  converges,  $\sum |p_i x_i|$  does not converge. So, rigorously speaking, in the above example  $E(X)$  does not exist, although  $\sum p_i x_i$  is finite, viz.,  $\log_2 2$ .

As another example, let us consider the r.v.  $X$  which takes the values

$$x_k = \frac{(-1)^k \cdot 2^k}{k}; \quad k = 1, 2, 3, \dots$$

with probabilities  $p_k = 2^{-k}$ .

Here also we get

$$\begin{aligned} \sum_{k=1}^{\infty} x_k p_k &= \sum_{k=1}^{\infty} \frac{(-1)^k}{k} \\ &= - \left[ 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \dots \right] = -\log_e 2 \end{aligned}$$

$$\text{and } \sum_{k=1}^{\infty} |x_k| p_k = \sum_{k=1}^{\infty} \frac{1}{k},$$

which is a divergent series. Hence in this case also expectation does not exist.

As an illustration of a continuous r.v. let us consider the r.v.  $X$  with p.d.f.

$$f(x) = \frac{1}{\pi} \cdot \frac{1}{1+x^2} ; \quad -\infty < x < \infty$$

which is p.d.f. of Standard Cauchy distribution. [c.f.s. 8.9].

$$\begin{aligned} \int_{-\infty}^{\infty} |x| f(x) dx &= \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{|x|}{1+x^2} dx = \frac{2}{\pi} \int_0^{\infty} \frac{x}{1+x^2} dx \\ &\quad (\because \text{Integrand is an even function of } x) \\ &= \frac{1}{\pi} \left| \log(1+x^2) \right|_0^{\infty} \rightarrow \infty \end{aligned}$$

Since this integral does not converge to a finite limit,  $E(X)$  does not exist.

**6.2. Expectation of a Function of a Random Variable.** Consider a r.v.  $X$  with p.d.f. (p.m.f.)  $f(x)$  and distribution function  $F(x)$ . If  $g(\cdot)$  is a function such that  $g(X)$  is a r.v. and  $E[g(X)]$  exists (i.e., is defined), then

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) dF(x) = \int_{-\infty}^{\infty} g(x) f(x) dx \quad \dots(6.3)$$

$$= \sum_x g(x) f(x) \quad \dots(6.3a)$$

(For discrete r.v.)

By definition, the expectation of  $Y = g(X)$  is

$$E[g(X)] = E(Y) = \int y \cdot dH_Y(y) = \int y h(y) dy \quad \dots(6.4)$$

$$\text{or} \quad E(Y) = \sum_y y h(y) \quad \dots(6.4a)$$

where  $H_Y(y)$  is the distribution function of  $Y$  and  $h(y)$  is p.d.f. of  $Y$ .

(The proof of equivalence of (6.3) and (6.4) is beyond the scope of the book.)

This result extends into higher dimensions. If  $X$  and  $Y$  have a joint p.d.f.  $f(x, y)$  and  $Z = h(x, y)$  is a random variable for some function  $h$ , and if  $E(Z)$  exists, then

$$E(Z) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x, y) f(x, y) dx dy \quad \dots(6.5)$$

$$\text{or} \quad E(Z) = \sum_x \sum_y h(x, y) f(x, y) \quad \dots(6.5a)$$

**Particular Cases.** 1. If we take  $g(X) = X^r$ ,  $r$  being a positive integer, in (6.3) we get :

$$E(X^r) = \int_{-\infty}^{\infty} x^r \cdot f(x) dx \quad \dots(6.5b)$$

which is defined as  $\mu_r'$ , the  $r$ th moment (about origin) of the probability distribution.

Thus  $\mu_r'$  (about origin) =  $E(X^r)$ . In particular

$$\mu_1'$$
 (about origin) =  $E(X)$  and  $\mu_2'$  (about origin) =  $E(X^2)$

Hence  $\text{Mean} = \bar{x} = \mu_1' \text{ (about origin)} = E(X) \quad \dots(6-6)$

and  $\mu_2 = \mu_2' - \mu_1'^2 = E(X^2) - [E(X)]^2 \quad \dots(6-6a)$

2. If  $g(X) = [X - E(X)]^r = (X - \bar{x})^r$ , then from (6-3) we get :

$$E[X - E(X)]^r = \int_{-\infty}^{\infty} [x - E(X)]^r f(x) dx = \int_{-\infty}^{\infty} (x - \bar{x})^r f(x) dx \quad \dots(6-7)$$

which is  $\mu_r$ , the  $r$ th moment about mean.

In particular, if  $r = 2$ , we get

$$\mu_2 = E[X - E(X)]^2 = \int_{-\infty}^{\infty} (x - \bar{x})^2 f(x) dx \quad \dots(6-8)$$

Formulae (6-6a) and (6-8) give the variance of the probability distribution of a r.v.  $X$  in terms of expectation.

3. Taking  $g(x) = \text{constant} = c$ , say in (6-3) we get

$$E(c) = \int_{-\infty}^{\infty} c \cdot f(x) dx = c \int_{-\infty}^{\infty} f(x) dx = c \quad \dots(6-9)$$

$$E(c) = c \quad \dots(6-9a)$$

**Remark.** The corresponding results for a discrete r.v.  $X$  can be obtained on replacing integration by summation ( $\Sigma$ ) over the given range of the variable  $X$  in the formulae (6-5) to (6-9).

In the following sections, we shall establish some more results on Expectation in the form of theorems, for continuous r.v.'s only. The corresponding results for discrete r.v.'s can be obtained similarly on replacing integration by summation ( $\Sigma$ ) over the given range of the variable  $X$  and are left as an exercise to the reader.

### 6-3. Addition Theorem of Expectation

**Theorem 6-1.** If  $X$  and  $Y$  are random variables then

$$E(X + Y) = E(X) + E(Y), \quad \dots(6-10)$$

provided all the expectations exist.

**Proof.** Let  $X$  and  $Y$  be continuous r.v.'s with joint p.d.f.  $f_{X,Y}(x, y)$  and marginal p.d.f.'s  $f_X(x)$  and  $f_Y(y)$  respectively. Then by definition :

$$E(X) = \int_{-\infty}^{\infty} x f_X(x) dx \quad \dots(6-11)$$

$$E(Y) = \int_{-\infty}^{\infty} y f_Y(y) dy \quad \dots(6-12)$$

$$E(X + Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x + y) f_{X,Y}(x, y) dx dy$$

$$\begin{aligned}
 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f_{XY}(x, y) dx dy \\
 &\quad + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y f_{XY}(x, y) dx dy \\
 &= \int_{-\infty}^{\infty} x \left[ \int_{-\infty}^{\infty} f_{XY}(x, y) dy \right] dx \\
 &\quad + \int_{-\infty}^{\infty} y \left[ \int_{-\infty}^{\infty} f_{XY}(x, y) dx \right] dy \\
 &= \int_{-\infty}^{\infty} x f_X(x) dx + \int_{-\infty}^{\infty} y f_Y(y) dy \\
 &= E(X) + E(Y) \quad [\text{On using (6.11) and (6.12)}]
 \end{aligned}$$

The result in (6.10) can be extended to  $n$  variables as given below.

**Theorem 6.1(a).** *The mathematical expectation of the sum of  $n$  random variables is equal to the sum of their expectations, provided all the expectations exist.*

Symbolically, if  $X_1, X_2, \dots, X_n$  are random variables then

$$E(X_1 + X_2 + \dots + X_n) = E(X_1) + E(X_2) + \dots + E(X_n) \quad \dots(6.13)$$

$$\text{or} \quad E\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n E(X_i), \quad \dots(6.13a)$$

if all the expectations exist.

**Proof.** Using (6.10), for two r.v.'s  $X_1$  and  $X_2$  we get :

$$\begin{aligned}
 E(X_1 + X_2) &= E(X_1) + E(X_2) \\
 \Rightarrow (6.13) \text{ is true for } n = 2. \quad \dots(*) 
 \end{aligned}$$

Let us now suppose that (6.13) is true for  $n = r$  (say), so that

$$E\left(\sum_{i=1}^r X_i\right) = \sum_{i=1}^r E(X_i) \quad \dots(6.14)$$

$$\begin{aligned}
 E\left(\sum_{i=1}^{r+1} X_i\right) &= E\left[\sum_{i=1}^r X_i + X_{r+1}\right] \\
 &= E\left(\sum_{i=1}^r X_i\right) + E(X_{r+1}) \quad [\text{Using (6.10)}] \\
 &= \sum_{i=1}^r E(X_i) + E(X_{r+1}) \quad [\text{Using (6.14)}] \\
 &= \sum_{i=1}^{r+1} E(X_i)
 \end{aligned}$$

Hence if (6.13)' is true for  $n = r$ , it is also true for  $n = r + 1$ . But we have proved in (\*) above that (6.13) is true for  $n = 2$ . Hence it is true for  $n = 2 + 1 = 3$ ;  $n = 3 + 1 = 4$ ; ... and so on. Hence by the principle of mathemati-

cal Introduction (6.13) is true for all positive integral values of  $n$ .

#### 6.4. Multiplication Theorem of Expectation

**Theorem 6.2.** If  $X$  and  $Y$  are independent random variables, then

$$E(XY) = E(X) \cdot E(Y) \quad \dots(6.15)$$

**Proof.** Proceeding as in Theorem 6.1, we have :

$$\begin{aligned} E(XY) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f_{XY}(x, y) dx dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f_X(x) f_Y(y) dx dy \\ &\quad [\text{Since } X \text{ and } Y \text{ are independent}] \\ &= \int_{-\infty}^{\infty} x f_X(x) dx \int_{-\infty}^{\infty} y f_Y(y) dy \\ &= E(X) \cdot E(Y), \quad [\text{Using (6.11) and (6.12)}] \end{aligned}$$

provided  $X$  and  $Y$  are independent.

**Generalisation to  $n$ -variables.**

**Theorem 6.2(a).** The mathematical expectation of the product of a number of independent random variables is equal to the product of their expectations. Symbolically, if  $X_1, X_2, \dots, X_n$  are  $n$  independent random variables, then

$$E(X_1 X_2 \dots X_n) = E(X_1) E(X_2) \dots E(X_n) \quad \left. \begin{array}{l} \\ i.e., \quad E\left(\prod_{i=1}^n X_i\right) = \prod_{i=1}^n E(X_i) \end{array} \right\} \quad \dots(6.16)$$

provided all the expectations exist.

**Proof.** Using (6.15), for two independent random variables  $X_1$  and  $X_2$ , we get:

$$E(X_1 X_2) = E(X_1) E(X_2)$$

$$\Rightarrow \quad (6.16) \text{ is true for } n = 2. \quad \dots(*)$$

Let us now suppose that (6.16) is true for  $n = r$ , (say) so that :

$$\begin{aligned} E\left(\prod_{i=1}^r X_i\right) &= \prod_{i=1}^r E(X_i) \quad \dots(6.17) \\ E\left(\prod_{i=1}^{r+1} X_i\right) &= E\left(\prod_{i=1}^r X_i \times X_{r+1}\right) \\ &= E\left(\prod_{i=1}^r X_i\right) E(X_{r+1}) \quad [\text{Using (6.15)}] \\ &= \prod_{i=1}^r (E X_i) E(X_{r+1}) \quad [\text{Using (6.17)}] \\ &= \prod_{i=1}^{r+1} (E X_i) \end{aligned}$$

Hence if (6.16) is true for  $n = r$ , it is also true for  $n = r + 1$ . Hence using (\*), by the principle of mathematical induction we conclude that (6.16) is true for all positive integral values of  $n$ .

**Theorem 6.3.** If  $X$  is a random variable and 'a' is constant, then

$$(i) \quad E[a\Psi(X)] = aE[\Psi(X)] \quad \dots(6.18)$$

$$(ii) \quad E[\Psi(X) + a] = E[\Psi(X)] + a, \quad \dots(6.19)$$

where  $\Psi(X)$ , a function of  $X$ , is a r.v. and all the expectations exist.

**Proof.**

$$(i) \quad E[a\Psi(X)] = \int_{-\infty}^{\infty} a\Psi(x) f(x) dx = a \int_{-\infty}^{\infty} \Psi(x) f(x) dx = aE[\Psi(X)]$$

$$\begin{aligned} (ii) \quad E[\Psi(X) + a] &= \int_{-\infty}^{\infty} [\Psi(x) + a] f(x) dx \\ &= \int_{-\infty}^{\infty} \Psi(x) f(x) dx + a \int_{-\infty}^{\infty} f(x) dx \\ &= E[\Psi(X)] + a \quad \left( \because \int_{-\infty}^{\infty} f(x) dx = 1 \right) \end{aligned}$$

**Cor. (i)** If  $\Psi(X) = X$ , then

$$E(aX) = aE(X) \text{ and } E(X + a) = E(X) + a \quad \dots(6.20)$$

$$(ii) \quad \text{If } \Psi(X) = 1, \text{ then } E(a) = a. \quad \dots(6.21)$$

**Theorem 6.4.** If  $X$  is a random variable and  $a$  and  $b$  are constants, then

$$E(aX + b) = aE(X) + b \quad \dots(6.22)$$

provided all the expectations exist.

**Proof.** By definition, we have

$$\begin{aligned} E(aX + b) &= \int_{-\infty}^{\infty} (ax + b) f(x) dx \\ &= a \int_{-\infty}^{\infty} x f(x) dx + b \int_{-\infty}^{\infty} f(x) dx \\ &= aE(X) + b \end{aligned}$$

**Cor. 1.** If  $b = 0$ , then we get

$$E(aX) = a \cdot E(X) \quad \dots(6.22a)$$

**Cor. 2.** Taking  $a = 1$ ,  $b = -\bar{X} = -E(X)$ , we get

$$E(X - \bar{X}) = 0$$

**Remark.** If we write,

$$g(X) = aX + b \quad \dots(6.23)$$

$$\text{then } g[E(X)] = aE(X) + b \quad \dots(6.23a)$$

Hence from (6.22) and (6.23a) we get

$$E[g(X)] = g[E(X)] \quad \dots(6.24)$$

Now (6.23) and (6.24) imply that expectation of a linear function is the same linear function of the expectation. The result, however, is not true if  $g(\cdot)$  is not linear. For instance

$$E(1/X) = (1/E(X)) ; \quad E(X^{\frac{1}{2}}) = [E(X)]^{\frac{1}{2}},$$

$$E[\log(X)] = \log[E(X)] ; \quad E(X^2) = [E(X)]^2,$$

since all the functions stated above are non-linear. As an illustration, let us consider a random variable  $X$  which assumes only two values +1 and -1, each with equal probability  $\frac{1}{2}$ . Then

$$E(X) = 1 \times \frac{1}{2} + (-1) \times \frac{1}{2} = 0.$$

$$\text{and} \quad E(X^2) = 1^2 \times \frac{1}{2} + (-1)^2 \times \frac{1}{2} = 1.$$

$$\text{Thus} \quad E(X^2) \neq [E(X)]^2$$

For a non-linear function  $g(X)$ , it is difficult to obtain expressions for  $E[g(X)]$  in terms of  $g[E(X)]$ , say, for  $E[\log(X)]$  or  $E(X^2)$  in terms of  $\log[E(X)]$  or  $[E(X)]^2$ . However, some results in the form of inequalities between  $E[g(X)]$  and  $g[E(X)]$  are available, as discussed in Theorem 6.12 (Jensen's Inequality) page 6.15.

### 6.5. Expectation of a Linear Combination of Random Variables

Let  $X_1, X_2, \dots, X_n$  be any  $n$  random variables and if  $a_1, a_2, \dots, a_n$  are any  $n$  constants, then

$$E\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i E(X_i) \quad \dots(6.25)$$

provided all the expectations exist.

**Proof.** The result is obvious from (6.13) and (6.20).

**Theorem 6.5 (a).** If  $X \geq 0$  then  $E(X) \geq 0$ .

**Proof.** If  $X$  is a continuous random variable s.t.  $X \geq 0$  then

$$E(X) = \int_{-\infty}^{\infty} x \cdot p(x) dx = \int_0^{\infty} x \cdot p(x) dx > 0,$$

[ $\because$  If  $X \geq 0$ ,  $p(x) = 0$  for  $x < 0$ ]

provided the expectation exists.

**Theorem 6.5 (b).** Let  $X$  and  $Y$  be two random variables such that  $Y \leq X$  then

$$E(Y) \leq E(X),$$

provided the expectations exist.

**Proof.** Since  $Y \leq X$ , we have the r.v.

$$\begin{aligned} Y - X &\leq 0 &\Rightarrow X - Y &\geq 0 \\ \text{Hence} \quad E(X - Y) &\geq 0 &\Rightarrow E(X) - E(Y) &\geq 0 \\ \Rightarrow \quad E(X) &\geq E(Y) &\Rightarrow E(Y) &\leq E(X), \end{aligned}$$

as desired.

**Theorem 6.6.**  $|E(X)| \leq E|X|$ , ... (6.26)  
provided the expectations exist.

**Proof.** Since  $X \leq |X|$ , we have by Theorem 6.5(b)

$$E(X) \leq E|X| \quad \dots(*)$$

Again since  $-X \leq |X|$ , we have by Theorem 6.5(b)

$$E(-X) \leq E|X| \quad \dots(**)$$

$$\Rightarrow -E(X) \leq E|X|$$

From (\*) and (\*\*), we get the desired result  $|E(X)| \leq E|X|$ .

**Theorem 6.7.** If  $\mu_s'$  exists, then  $\mu_s'$  exists for all  $1 \leq s \leq r$ .

Mathematically, if  $E(X^r)$  exists, then  $E(X^s)$  exists for all  $1 \leq s \leq r$ , i.e.,

$$E(X^r) < \infty \Rightarrow E(X^s) < \infty \quad \forall \quad 1 \leq s \leq r \quad \dots(6.27)$$

**Proof.**  $\int_{-\infty}^{\infty} |x|^s dF(x) = \int_{-1}^1 |x|^s dF(x)$

$$+ \int_{|x| > 1} |x|^s dF(x)$$

If  $s < r$ , then  $|x|^s < |x|^r$  for  $|x| > 1$ .

$$\begin{aligned} \therefore \int_{-\infty}^{\infty} |x|^s dF(x) &\leq \int_{-1}^1 |x|^s dF(x) + \int_{|x| > 1} |x|^r dF(x) \\ &\leq \int_{-1}^1 dF(x) + \int_{|x| > 1} |x|^r dF(x), \end{aligned}$$

since for  $-1 < x < 1$ ,  $|x|^s < 1$ .

$$\therefore \int_{-\infty}^{\infty} |x|^s dF(x) \leq 1 + E|X|^r < \infty$$

$$\Rightarrow E(x^s) \text{ exists } \forall \quad 1 \leq s \leq r \quad [\because E(X^r) \text{ exists}]$$

**Remark.** The above theorem states that if the moments of a specified order exist, then all the lower order moments automatically exist. However, the converse is not true, i.e., we may have distributions for which all the moments of a specified order exist but no higher order moments exist. For example, for the r.v. with p.d.f.

$$\begin{aligned} p(x) &= 2/x^3 \quad ; \quad x \geq 1 \\ &= 0 \quad ; \quad x < 1 \end{aligned}$$

we have :

$$E(X) = \int_1^{\infty} x p(x) dx = 2 \int_1^{\infty} x^{-2} dx = \left[ \left( \frac{-2}{x} \right) \right]_1^{\infty} = 2$$

$$E(X^2) = \int_1^\infty x^2 p(x) dx = 2 \int_1^\infty \frac{1}{x} dx = \infty$$

Thus for the above distribution, 1st order moment (mean) exists but 2nd order moment (variance) does not exist.

As another illustration, consider a r.v.  $X$  with p.d.f.

$$p(x) = \frac{(r+1) a^{r+1}}{(x+a)^{r+2}} ; x \geq 0, ; a > 0$$

$$\mu_r' = E(X^r) = (r+1) a^{r+1} \int_0^\infty \frac{x^r}{(x+a)^{r+2}} dx$$

Put  $x = ay$  and using Beta integral :

$$\int_0^\infty \frac{x^{m-1}}{(1+x)^{m+n}} = \beta(m, n),$$

we shall get on simplification :

$$\mu_r' = (r+1) a^r \cdot \beta(r+1, 1) = a^r$$

However,

$$\mu_{r+1}' = E(X^{r+1}) = (r+1) a^{r+1} \int_0^\infty \frac{x^{r+1}}{(x+a)^{r+2}} dx \rightarrow \infty,$$

as the integral is not convergent. Hence in this case only the moments up to  $r$ th order exist and higher order moments do not exist.

**Theorem 6.8.** If  $X$  is a random variable, then

$$V(aX + b) = a^2 V(X), \quad \dots(6.28)$$

where  $a$  and  $b$  are constants.

**Proof.** Let  $Y = aX + b$

$$\text{Then } E(Y) = aE(X) + b$$

$$\therefore Y - E(Y) = a[X - E(X)]$$

Squaring and taking expectation of both sides, we get

$$E(Y - E(Y))^2 = a^2 E(X - E(X))^2$$

$$\Rightarrow V(Y) = a^2 V(X) \Rightarrow V(aX + b) = a^2 V(X),$$

where  $V(X)$  is written for variance of  $X$ .

$$\text{Cor. (i) If } b = 0, \text{ then } V(aX) = a^2 V(X)$$

$\Rightarrow$  Variance is not independent of change of scale.

$$(ii) \text{ If } a = 0, \text{ then } V(b) = 0$$

$\Rightarrow$  Variance of a constant is zero.

$$(iii) \text{ If } a = 1, \text{ then } V(X+b) = V(X)$$

$\Rightarrow$  Variance is independent of change of origin.

**6-6. Covariance:** If  $X$  and  $Y$  are two random variables, then covariance between them is defined as

$$\text{Cov}(X, Y) = E[(X - E(X))(Y - E(Y))] \quad \dots(6-29)$$

$$\begin{aligned} &= E[XY - XE(Y) - YE(X) + E(X)E(Y)] \\ &= E(XY) - E(Y)E(X) - E(X)E(Y) + E(X)E(Y) \\ &= E(XY) - E(X)E(Y) \end{aligned} \quad \dots(6-29a)$$

If  $X$  and  $Y$  are independent then  $E(XY) = E(X)E(Y)$  and hence in this case

$$\text{Cov}(X, Y) = E(X)E(Y) - E(X)E(Y) = 0 \quad \dots(6-29b)$$

$$\text{Remarks. 1. } \text{Cov}(aX, bY) = E[(aX - E(aX))(bY - E(bY))]$$

$$\begin{aligned} &= E[a(X - E(X))b(Y - E(Y))] \\ &= abE[(X - E(X))(Y - E(Y))] \\ &= ab\text{Cov}(X, Y) \end{aligned} \quad \dots(6-30)$$

$$2. \quad \text{Cov}(X+a, Y+b) = \text{Cov}(X, Y) \quad \dots(6-30a)$$

$$3. \quad \text{Cov}\left(\frac{X-\bar{X}}{\sigma_X}, \frac{Y-\bar{Y}}{\sigma_Y}\right) = \frac{1}{\sigma_X \sigma_Y} \text{Cov}(X, Y) \quad \dots(6-30b)$$

4. Similarly, we shall get :

$$\text{Cov}(aX+b, cY+d) = ac\text{Cov}(X, Y) \quad \dots(6-30c)$$

$$\text{Cov}(X+Y, Z) = \text{Cov}(X, Z) + \text{Cov}(Y, Z) \quad \dots(6-30d)$$

$$\text{Cov}(aX+bY, cX+dY) = ac\sigma_X^2 + bd\sigma_Y^2 + (ad+bc)\text{Cov}(X, Y) \quad \dots(6-30e)$$

If  $X$  and  $Y$  are independent,  $\text{Cov}(X, Y) = 0$  : [c.f. (6-29b)]

However, the converse is not true.

(For details see Theorem 10-2)

**6-6-1. Correlation Coefficient.** The correlation coefficient ( $\rho_{XY}$ ), between the variables  $X$  and  $Y$  is defined as :

$$\rho_{XY} = \text{Correlation Coefficient}(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} \quad \dots(6-30f)$$

For detailed discussion on correlation coefficient, see Chapter 10.

#### 6-7. Variance of a Linear Combination of Random Variables

**Theorem 6-9.** Let  $X_1, X_2, \dots, X_n$  be  $n$  random variables then

$$V\left[\sum_{i=1}^n a_i X_i\right] = \sum_{i=1}^n a_i^2 V(X_i) + 2 \sum_{i=1}^n \sum_{j=1, i < j}^n a_i a_j \text{Cov}(X_i, X_j) \quad \dots(6-31)$$

**Proof.** Let  $U = a_1 X_1 + a_2 X_2 + \dots + a_n X_n$

$$\therefore E(U) = a_1 E(X_1) + a_2 E(X_2) + \dots + a_n E(X_n)$$

$$\therefore U - E(U) = a_1 [X_1 - E(X_1)] + a_2 [X_2 - E(X_2)] + \dots + a_n [X_n - E(X_n)]$$

Squaring and taking expectation of both sides, we get

$$\begin{aligned} E[U - E(U)]^2 &= a_1^2 E[X_1 - E(X_1)]^2 + a_2^2 E[X_2 - E(X_2)]^2 + \dots \\ &\quad + a_n^2 E[X_n - E(X_n)]^2 \\ &\quad + 2 \sum_{i=1}^n \sum_{\substack{j=1 \\ i < j}}^n a_i a_j E[(X_i - E(X_i))(X_j - E(X_j))] \\ \Rightarrow V(U) &= a_1^2 V(X_1) + a_2^2 V(X_2) + \dots + a_n^2 V(X_n) \\ &\quad + 2 \sum_{i=1}^n \sum_{\substack{j=1 \\ i < j}}^n a_i a_j Cov(X_i, X_j) \\ \Rightarrow V\left[\sum_{i=1}^n a_i X_i\right] &= \sum_{i=1}^n a_i^2 V(X_i) + 2 \sum_{i=1}^n \sum_{\substack{j=1 \\ i < j}}^n a_i a_j Cov(X_i, X_j) \end{aligned}$$

**Remarks.** 1. If  $a_i = 1$ ;  $i = 1, 2, \dots, n$  then

$$V(X_1 + X_2 + \dots + X_n) = V(X_1) + V(X_2) + \dots + V(X_n)$$

$$+ 2 \sum_{i=1}^n \sum_{\substack{j=1 \\ i < j}}^n Cov(X_i, X_j) \quad \dots(6.31a)$$

2. If  $X_1, X_2, \dots, X_n$  are independent (pairwise) then  $Cov(X_i, X_j) = 0$ , ( $i \neq j$ ).

Thus from (6.31) and (6.31a), we get

$$\begin{aligned} \{V(a_1 X_1 + a_2 X_2 + \dots + a_n X_n) = a_1^2 V(X_1) + a_2^2 V(X_2) + \dots + a_n^2 V(X_n)\} \\ \text{and } V(X_1 + X_2 + \dots + X_n) = V(X_1) + V(X_2) + \dots + V(X_n) \quad \dots(6.31b) \end{aligned}$$

3. If  $a_1 = 1 = a_2$  and  $a_3 = a_4 = \dots = a_n = 0$ , then from (6.31), we get

$$V(X_1 + X_2) = V(X_1) + V(X_2) + 2 Cov(X_1, X_2)$$

Again if  $a_1 = 1$ ,  $a_2 = -1$  and  $a_3 = a_4 = \dots = a_n = 0$ , then

$$V(X_1 - X_2) = V(X_1) + V(X_2) - 2 Cov(X_1, X_2)$$

Thus we have

$$V(X_1 \pm X_2) = V(X_1) + V(X_2) \pm 2 Cov(X_1, X_2) \quad \dots(6.31c)$$

If  $X_1$  and  $X_2$  are independent, then  $Cov(X_1, X_2) = 0$  and we get

$$V(X_1 \pm X_2) = V(X_1) + V(X_2). \quad \dots(6.31d)$$

**Theorem 6.10.** If  $X$  and  $Y$  are independent random variables then

$$E[h(X) \cdot k(Y)] = E[h(X)] E[k(Y)] \quad \dots(6.32)$$

where  $h(\cdot)$  is a function of  $X$  alone and  $k(\cdot)$  is a function of  $Y$  alone, provided expectations on both sides exist.

**Proof.** Let  $f_X(x)$  and  $g_Y(y)$  be the marginal p.d.f.'s of  $X$  and  $Y$  respectively. Since  $X$  and  $Y$  are independent, their joint p.d.f.  $f_{XY}(x, y)$  is given by

$$f_{XY}(x, y) = f_X(x) f_Y(y) \quad \dots(*)$$

By definition, for continuous r.v.'s

$$\begin{aligned} E[h(X) \cdot k(Y)] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x) k(y) f(x, y) dx dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x) k(y) f(x) g(y) dx dy \end{aligned}$$

[From (\*)]

Since  $E[h(X) \cdot k(Y)]$  exists, the integral on the right hand side is absolutely convergent and hence by Fubini's theorem for integrable functions we can change the order of integration to get

$$\begin{aligned} E[h(X) \cdot k(Y)] &= \left[ \int_{-\infty}^{\infty} h(x) f(x) dx \right] \left[ \int_{-\infty}^{\infty} k(y) g(y) dy \right] \\ &= E[h(X)] \cdot E[k(Y)], \end{aligned}$$

as desired.

**Remark.** The result can be proved for discrete random variables  $X$  and  $Y$  on replacing integration by summation over the given range of  $X$  and  $Y$ .

**Theorem 6-11. Cauchy-Schwartz Inequality.** If  $X$  and  $Y$  are random variables taking real values, then

$$[E(XY)]^2 \leq E(X^2) \cdot E(Y^2) \quad \dots(6.33)$$

**Proof.** Let us consider a real valued function of the real variable  $t$ , defined by

$$Z(t) = E(X + tY)^2$$

which is always non-negative, since  $(X + tY)^2 \geq 0$ , for all real  $X$ ,  $Y$  and  $t$ .

$$\text{Thus } Z(t) = E(X + tY)^2 \geq 0 \quad \forall t.$$

$$\Rightarrow Z(t) = E[X^2 + 2tXY + t^2Y^2] \\ = E(X^2) + 2t \cdot E(XY) + t^2 E(Y^2) \geq 0, \text{ for all } t. \quad (*)$$

Obviously,  $Z(t)$  is a quadratic expression in ' $t$ '.

We know that the quadratic expression of the form :



$\Psi(t) = A t^2 + B t + C \geq 0$  for all  $t$ , implies that the graph of the function  $\Psi(t)$  either touches the  $t$ -axis at only one point or not at all, as exhibited in the diagrams.

This is equivalent to saying that the discriminant of the function  $\Psi(t)$ , viz.,  $B^2 - 4AC \leq 0$ , since the condition  $B^2 - 4AC > 0$  implies that the function  $\Psi(t)$  has two distinct real roots which is a contradiction to the fact that  $\Psi(t)$  meets the  $t$ -axis either at only one point or not at all. Using this result, we get from (\*),

$$\begin{aligned} 4 \cdot [E(XY)]^2 - 4 \cdot E(X^2) \cdot E(Y^2) &\leq 0 \\ \Rightarrow [E(XY)]^2 &\leq E(X^2) \cdot E(Y^2) \end{aligned}$$

**Remarks.** 1. The sign of equality holds in (6.33) or (\*) iff

$$\begin{aligned} E(X+tY)^2 = 0 \quad \forall t &\Rightarrow P[(X+tY)^2 = 0] = 1, \\ \Rightarrow P[X+tY = 0] = 1 &\Rightarrow P\left[Y = -\frac{X}{t}\right] = 1 \\ \Rightarrow P[Y = AX] = 1 &; \quad (A = -1/t) \end{aligned} \quad \dots(6.33b)$$

2. If the r.v.  $X$  takes the real values  $x_1, x_2, \dots, x_n$  and r.v.  $Y$  takes the real values  $y_1, y_2, \dots, y_n$  then Cauchy-Schwartz inequality implies :

$$\begin{aligned} \left(\frac{1}{n} \sum_{i=1}^n x_i y_i\right)^2 &\leq \left(\frac{1}{n} \sum_{i=1}^n x_i^2\right) \cdot \left(\frac{1}{n} \sum_{i=1}^n y_i^2\right) \\ \Rightarrow \left(\sum_{i=1}^n x_i y_i\right)^2 &\leq \left(\sum_{i=1}^n x_i^2\right) \cdot \left(\sum_{i=1}^n y_i^2\right) \end{aligned}$$

the sign of equality holding if and only if:

$$\frac{x_i}{y_i} = \text{constant} = k, \text{ (say)} \quad \text{for all } i = 1, 2, \dots, n$$

i.e. iff  $\frac{x_1}{y_1} = \frac{x_2}{y_2} = \dots = \frac{x_n}{y_n} = k$ , (say).

3. Replacing  $X$  by  $|X - E(X)| = |X - \mu_x|$  and taking  $Y = 1$  in (6.33) we get

$$\begin{aligned} [E|X - \mu_x|]^2 &= E|X - \mu_x|^2, E(1) \\ \Rightarrow [\text{Mean Deviation about mean}]^2 &\leq \text{Variance}(X) \\ \Rightarrow M.D. &\leq S.D. \Rightarrow S.D. \geq M.D. \end{aligned} \quad \dots(6.33a)$$

#### 6.7. Jensen's Inequality

**Continuous Convex Function. (Definition).** A continuous function  $g(x)$  on the interval  $I$  is convex if for every  $x_1$  and  $x_2$ ,  $(x_1 + x_2)/2 \in I$ , we have

$$g\left(\frac{x_1 + x_2}{2}\right) \leq \frac{1}{2} g(x_1) + \frac{1}{2} g(x_2) \quad \dots(6.34)$$

**Remarks.** 1. If  $x_1, x_2 \in I$ , then  $(x_1 + x_2)/2 \in I$ .

2. Sometimes (6.34) is replaced by the stronger condition :

For  $x_1, x_2 \in I$ ,

$$g[\lambda x_1 + (1-\lambda)x_2] \leq \lambda g(x_1) + (1-\lambda)g(x_2); 0 \leq \lambda \leq 1 \quad \dots(6.35)$$

(6-34) and (6-35) agree at  $\lambda = \frac{1}{2}$ .

3. If we do not assume the continuity of  $g(x)$ , then (6-35) is required to define convexity. There are certain 'non-measurable' non-constant functions  $g(\cdot)$  satisfying (6-34) but not (6-35). If  $g(\cdot)$  is measurable, then (6-34) and (6-35) are equivalent.

4. A function satisfying (6-35) is continuous except possibly at the end points of the interval  $I$  (if it has end points).

5. If  $g$  is twice differentiable, i.e.,  $g''(x)$  exists for  $X \in [\text{interior of } I]$ , and  $g''(x) \geq 0$  for such  $x$ , then  $g$  is convex on the interior points.

6. For any point  $x_0$  interior to  $I$ ,  $\exists$  a straight line  $y = ax + b$ , which passes through  $(x_0, g(x_0))$  and satisfies  $g(x) \geq ax + b$ , for all  $x \in I$ .

**Theorem 6-12. (Jensen's Inequality).** If  $g$  is continuous and convex function on the interval  $I$ , and  $X$  is a random variable whose values are in  $I$  with probability 1, then

$$E[g(X)] \geq g[E(X)], \quad \dots(6-36)$$

provided the expectations exist.

**Proof.** First of all we shall show that  $E(X) \in I$ .

The various possible cases for  $I$  are :

$$I = (-\infty, \infty); I = (a, \infty); I = [a, \infty); I = (-\infty, b);$$

$$I = (-\infty, b], I = (a, b) \text{ and variations of this.}$$

If  $E(X)$  exists, then  $-\infty < E(X) < \infty$ .

If  $X \geq a$  almost surely (a.s.), i.e., with probability 1, then  $E(X) \geq a$ .

If  $X \leq b$ , a.s. then  $E(X) \leq b$ .

Thus  $E(X) \in I$ . Now  $E(X)$  can be either a left end or a right end point (if end points exist) of  $I$  or an interior point of  $I$ .

Suppose  $I$  has a left end point 'a', i.e.,  $X \geq a$  and  $E(X) = a$ . Then  $X - a \geq 0$  a.s. and  $E(X - a) = 0$ .

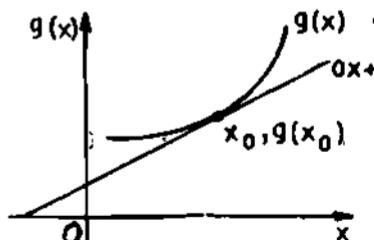
Thus  $P(X = a) = 1$  or  $P[(X - a) = 0] = 1$ .

$$\therefore E[g(X)] = E[g(a)] \quad [\because g(x) = g(a) \text{ a.s.}] \\ = g(a) \quad (\because g(a) \text{ is a constant}) \\ = g(E(X)).$$

The result can be established similarly if  $I$  has a right end point 'b' and  $E(X) = b$ .

Thus we are now required to establish (6-36) when  $E(X) = x_0$ , is an interior point of  $I$ .

Let  $ax + b$  pass through the point



$(x_0, g(x_0))$  and let it be below  $g$

[c.f. Remark 6 above].

$$\begin{aligned} E[g(X)] &\geq E(ax + b) = aE(X) + b \\ &= ax_0 + b \\ &= g(x_0) - g[E(X)] \end{aligned}$$

$$\Rightarrow E[g(X)] \leq g^*(X).$$

**Continuous Concave Function.** (Definition). A continuous function  $g$  is concave on an interval  $I$  if  $(-g)$  is convex.

**Corollary to Theorem 6-12.** If  $g$  is a continuous and concave function on the interval  $I$  and  $X$  is a r.v. whose values are in  $I$  with probability 1, then

$$E[g(X)] \leq g[E(X)] \quad \dots(6-37)$$

provided the expectations exist.

**Remarks.** 1. Equality holds in Theorem (6-12) and corollary (6-37), if and only if

$$P[g(X) = aX + b] = 1,$$

for some  $a$  and  $b$ .

2. Jenson's inequality extends to random vectors. If  $I$  is a convex set in  $n$ -dimensional Euclidean space, i.e., the interval  $I$  in theorem 6-12 is transferred to convex set,  $g$  is continuous on  $I$ , (6-34) holds whenever  $X_1$  and  $X_2$  are any arbitrary vectors in  $I$ . The condition  $g''(x) \geq 0$  for  $x$  interior to  $I$  implies

$$\left( \frac{\partial^2 g(x)}{\partial x_i \partial x_j} \right) = M(x), \text{ (say),}$$

is non-negative definite for all  $x$  interior to  $I$ .

### SOME ILLUSTRATIONS OF JENSON'S INEQUALITY

1. If  $E(X^2)$  exists, then

$$E(X^2) \geq [E(X)]^2, \quad \dots(6-38)$$

since  $g(X) = X^2$  is convex function of  $X$  as  $g''(X) = 2 > 0$ .

2. If  $X > 0$  a.s. i.e.,  $X$  assumes only positive values and  $E(X)$  and  $E(1/X)$  exist then

$$E\left(\frac{1}{X}\right) \geq \frac{1}{E(X)}, \quad \dots(6-38a)$$

because  $g(X) = \frac{1}{X}$  is a convex function of  $X$  since

$$g''(X) = \frac{2}{X^3} > 0, \text{ for } X > 0.$$

3. If  $X > 0$ , a.s. then

$$E(X^{1/2}) \leq [E(X)]^{1/2}, \quad \dots(6-38b)$$

since  $g(X) = X^{1/2}$ ,  $X > 0$  is a concave function

as  $g''(X) = -\frac{1}{4}X^{-3/2} < 0$ , for  $X > 0$ .

4. If  $X > 0$ , a.s. then

$$E[\log(X)] \leq \log[E(X)], \quad \dots(6.38c)$$

provided the expectations exist, because  $\log X$  is a concave function of  $X$ .

5. Since  $g(X) = E(e^{tX})$  is a convex function of  $X$  for all  $t$  and all  $X$ , if  $E(e^{tX})$  and  $E(X)$  exist then

$$E(e^{tX}) \geq e^{tE(X)} \quad \dots(6.38d)$$

If  $E(X) = 0$ , then

$$M_X(t) = E(e^{tX}) \geq 1, \text{ for all } t.$$

Thus if  $M_X(t)$  exists, then it has a lower bound 1, provided  $E(X) = 0$ . Further, this bound is attained at  $t = 0$ . Thus  $M_X(t)$  has a minimum at  $t = 0$ .

**6.7.1 ANOTHER USEFUL INEQUALITY.** Let  $f$  and  $g$  be monotone functions on some subset of the real line and  $X$  be a r.v. whose range is in the subset almost surely (a.s.) If the expectations exist, then

$$E[f(X)g(X)] \geq E[f(X)] \cdot E[g(X)] \quad \dots(6.39)$$

$$E[f(X)g(X)] \leq E[f(X)] \cdot E[g(X)] \quad \dots(6.39a)$$

or

according as  $f$  and  $g$  are monotone in the same or in the opposite directions.

**Proof.** Let us consider the case when both the functions  $f$  and  $g$  are monotone in the same direction. Let  $x$  and  $y$  lie in the domain of  $f$  and  $g$  respectively.

If  $f$  and  $g$  are both monotonically increasing, then

$$\begin{aligned} y \geq x &\Rightarrow f(y) \geq f(x) \quad \text{and} \quad g(y) \geq g(x) \\ \Rightarrow f(y) - f(x) &\geq 0 \quad \text{and} \quad g(y) - g(x) \geq 0 \\ \Rightarrow [f(y) - f(x)] \cdot [g(y) - g(x)] &\geq 0 \end{aligned} \quad \dots(*)$$

If  $f$  and  $g$  are both monotonically decreasing then for  $y \geq x$ , we have

$$\begin{aligned} f(y) &\leq f(x) \quad \text{and} \quad g(y) \leq g(x) \\ \Rightarrow f(y) - f(x) &\leq 0 \quad \text{and} \quad g(y) - g(x) \leq 0 \\ \Rightarrow [f(y) - f(x)] \cdot [g(y) - g(x)] &\geq 0 \end{aligned} \quad \dots(**)$$

Hence if  $f$  and  $g$  are both monotonic in the same direction, then from (\*) and (\*\*), we get the same result, viz.,

$$[f(y) - f(x)] \cdot [g(y) - g(x)] \geq 0.$$

Let us now consider independently and identically distributed (i.i.d.) random variables  $X$  and  $Y$ . Then from above, we get

$$\begin{aligned} E[(f(Y) - f(X))(g(Y) - g(X))] &\geq 0. \\ \Rightarrow E[f(Y) \cdot g(Y)] - E[f(Y) \cdot g(X)] - E[f(X) \cdot g(Y)] \\ &\quad + E[f(X) \cdot g(X)] \geq 0 \end{aligned} \quad \dots(6.40)$$

Since  $X$  and  $Y$  are i.i.d. r.v.'s, we have

$$E[f(Y)g(Y)] = E[f(X)g(X)];$$

$$E[f(Y)g(X)] = E[f(Y)]E[g(X)] = E[f(X)]E[g(X)]$$

( $\because X$  and  $Y$  are independent) ( $\because X$  and  $Y$  are identical)

and  $E[f(X)g(Y)] = E[f(X)] \cdot E[g(Y)] = E[f(X)] \cdot E[g(X)]$

Substituting in (6.40) we get

$$2 E[f(X) \cdot g(X)] - 2 E[f(X)] \cdot E[g(X)] \geq 0 ,$$

$$\Rightarrow E[f(X) \cdot g(X)] \geq E[f(X)] \cdot E[g(X)]$$

which establishes the result in (6.39).

Similarly, (6.39a) can be established, if  $f$  and  $g$  are monotonic in opposite directions, i.e., if  $f$  is monotonically increasing (decreasing) and  $g$  is monotonically decreasing (increasing). The proof is left as an exercise to the reader.

### SOME ILLUSTRATIONS OF INEQUALITY (6.39).

1. If  $X$  is a r.v. which takes only non-negative values, i.e., if  $X \geq 0$  a.s. then for  $\alpha > 0, \beta > 0, f(X) = X^\alpha$  and  $g(X) = X^\beta$  are monotonic in the same direction. Hence if the expectations exist,

$$\begin{aligned} E(X^\alpha \cdot X^\beta) &\geq E(X^\alpha) \cdot E(X^\beta) \\ \Rightarrow E(X^{\alpha+\beta}) &\geq E(X^\alpha) E(X^\beta); \quad \alpha > 0, \beta > 0 \end{aligned} \quad \dots(6.41)$$

In particular, taking  $\alpha = \beta \approx 1$ , we get

$$E(X^2) \geq [E(X)]^2 ,$$

a result already obtained in (6.38).

2. If  $X \geq 0$ , a.s. and  $E(X^\alpha)$  and  $E(X^{-1})$  exist, then for  $\alpha > 0$ , we get from (6.39a)

$$\begin{aligned} E(X^\alpha \cdot X^{-\alpha}) &\leq E(X^\alpha) \cdot E(X^{-1}) \\ \Rightarrow E(X^\alpha) E\left(\frac{1}{X}\right) &\geq E(X^{\alpha-1}); \quad \alpha > 0. \end{aligned} \quad \dots(6.42)$$

In particular with  $\alpha = 1$ , we get

$$E(X) E\left(\frac{1}{X}\right) \geq 1 ,$$

a result already obtained in (6.38a)

Taking  $\alpha = 2$  in (6.42), we get

$$\begin{aligned} E(X^2) E\left(\frac{1}{X}\right) &\geq E(X) \\ \Rightarrow E(X^2) &\geq \frac{E(X)}{E\left(\frac{1}{X}\right)} \geq \frac{1}{\left[E\left(\frac{1}{X}\right)\right]^2}, \end{aligned} \quad \dots(6.43)$$

on using (6.38a).

Taking  $\alpha = 2$  and  $\beta = -2$  in  $f(X) = X^\alpha, g(X) = X^\beta$  and using (6.39a), we get

$$E(X^2) \cdot E(X^{-2}) \geq E(X^2 \cdot X^{-2}) = 1.$$

$$\Rightarrow E(X^2) \geq \frac{1}{E(X^{-2})} \quad \dots(6.43a)$$

which is a weaker inequality than (6.43).

3. If  $M_X(t) = E(e^{tX})$  exists for all  $t$  and for some r.v.  $X$ , then

$$\begin{aligned} M_{X(u+v)} &= E[e^{(u+v)X}] = E(e^{uX}, e^{vX}) \\ &\geq E(e^{uX}) \cdot E(e^{vX}) \\ &= M_X(u) \cdot M_X(v) \end{aligned}$$

$\therefore M_X(u+v) \geq M_X(u) \cdot M_X(v)$ , for  $u, v \geq 0$ .

**Example 6-1.** Let  $X$  be a random variable with the following probability distribution :

$x$	:	-3	6	9
$P_r(X=x)$	:	$1/6$	$1/2$	$1/3$

Find  $E(X)$  and  $E(X^2)$  and using the laws of expectation, evaluate  $E(2X+1)^2$ .

(Gauhati Univ. B.Sc., 1992)

**Solution.**  $E(X) = \Sigma x \cdot p(x)$

$$= (-3) \times \frac{1}{6} + 6 \times \frac{1}{2} + 9 \times \frac{1}{3} = \frac{11}{2}$$

$$\begin{aligned} E(X^2) &= \Sigma x^2 p(x) \\ &= 9 \times \frac{1}{2} + 36 \times \frac{1}{2} + 81 \times \frac{1}{3} = \frac{93}{2} \end{aligned}$$

$$\therefore E(2X+1)^2 = E[4X^2 + 4X + 1] = 4E(X^2) + 4E(X) + 1 \\ = 4 \times \frac{93}{2} + 4 \times \frac{11}{2} + 1 = 209$$

**Example 6-2. (a)** Find the expectation of the number on a die when thrown.

**(b)** Two unbiased dice are thrown. Find the expected values of the sum of numbers of points on them.

**Solution.** (a) Let  $X$  be the random variable representing the number on a die when thrown. Then  $X$  can take any one of the values 1, 2, 3, ..., 6 each with equal probability  $1/6$ . Hence

$$\begin{aligned} E(X) &= \frac{1}{6} \times 1 + \frac{1}{6} \times 2 + \frac{1}{6} \times 3 + \dots + \frac{1}{6} \times 6 \\ &= \frac{1}{6}(1+2+3+\dots+6) = \frac{1}{6} \times \frac{6 \times 7}{2} = \frac{7}{2} \quad \dots(*) \end{aligned}$$

**Remark.** This does not mean that in a random throw of a dice, the player will get the number  $(7/2) = 3.5$ . In fact, one can never get this (fractional) number in a throw of a dice. Rather, this implies that if the player tosses the dice for a "long" period, then on the average toss he will get  $(7/2) = 3.5$ .

(b) The probability function of  $X$  (the sum of numbers obtained on two dice), is

Value of $X : x$	2	3	4	5	6	7	.....	11	12
Probability	$1/36$	$2/36$	$3/36$	$4/36$	$5/36$	$6/36$	.....	$2/36$	$1/36$

$$\begin{aligned}
 E(X) &= \sum_i p_i x_i \\
 &= 2 \times \frac{1}{36} + 3 \times \frac{2}{36} + 4 \times \frac{3}{36} + 5 \times \frac{4}{36} + 6 \times \frac{5}{36} + 7 \times \frac{6}{36} \\
 &\quad + 8 \times \frac{5}{36} + 9 \times \frac{4}{36} + 10 \times \frac{3}{36} + 11 \times \frac{2}{36} + 12 \times \frac{1}{36} \\
 &= \frac{1}{36} (2 + 6 + 12 + 20 + 30 + 42 + 40 + 36 + 30 + 22 + 12) \\
 &= \frac{1}{36} \times 252 = 7
 \end{aligned}$$

**Aliter.** Let  $X_i$  be the number obtained on the  $i$ th dice ( $i = 1, 2$ ) when thrown. Then the sum of the number of points on two dice is given by

$$\begin{aligned}
 S &= X_1 + X_2 \\
 \Rightarrow E(S) &= E(X_1) + E(X_2) = \frac{7}{2} + \frac{7}{2} = 7 \quad [\text{On using (*)}]
 \end{aligned}$$

**Remark.** This result can be generalised to the sum of points obtained in a random throw of  $n$  dice. Then

$$E(S) = \sum_{i=1}^n E(X_i) = \sum_{i=1}^n (7/2) = \frac{7n}{2}$$

**Example 6-3.** A box contains  $2^n$  tickets among which  $"C_i$  tickets bear the number  $i$ ;  $i = 0, 1, 2, \dots, n$ . A group of  $m$  tickets is drawn. What is the expectation of the sum of their numbers?

**Solution.** Let  $X_i$ ;  $i = 1, 2, \dots, m$  be the variable representing the number on the  $i$ th ticket drawn. Then the sum 'S' of the numbers on the tickets drawn is given by

$$\begin{aligned}
 S &= X_1 + X_2 + \dots + X_m = \sum_{i=1}^m X_i \\
 \therefore E(S) &= \sum_{i=1}^m E(X_i)
 \end{aligned}$$

Now  $X_i$  is a random variable which can take any one of the possible values 0, 1, 2, ...,  $n$  with respective probabilities.

$"C_0/2^n, "C_1/2^n, "C_2/2^n, \dots, "C_n/2^n,$

$$\begin{aligned}
 \therefore E(X_i) &= \frac{1}{2^n} [1."C_1 + 2."C_2 + 3."C_3 + \dots + n."C_n] \\
 &= \frac{1}{2^n} \left[ 1.n + 2.\frac{n(n-1)}{2!} + 3.\frac{n(n-1)(n-2)}{3!} + \dots + n.1 \right] \\
 &= \frac{n}{2^n} \left[ 1 + (n-1) + \frac{(n-1)(n-2)}{2!} + \dots + 1 \right] \\
 &= \frac{n}{2^n} [".^{-1}C_0 + ".^{-1}C_1 + ".^{-1}C_2 + \dots + ".^{-1}C_{n-1}] \\
 &= \frac{n}{2^n} \cdot (1+1)^{n-1} = \frac{n}{2}
 \end{aligned}$$

$$\text{Hence } E(S) = \sum_{i=1}^m (n/2) = \frac{m \cdot n}{2}$$

**Example 6·4.** In four tosses of a coin, let  $X$  be the number of heads. Tabulate the 16 possible outcomes with the corresponding values of  $X$ . By simple counting, derive the distribution of  $X$  and hence calculate the expected value of  $X$ .

**Solution.** Let  $H$  represent a head,  $T$  a tail and  $X$ , the random variable denoting the number of heads.

S. No.	Outcomes	No. of Heads ( $X$ )	S. No.	Outcomes	No. of Heads ( $X$ )
1	$H H H H$	4	9	$H T H T$	2
2	$H H H T$	3	10	$T H T H$	2
3	$H H T H$	3	11	$T H H T$	2
4	$H T H H$	3	12	$H T T T$	1
5	$T H H H$	3	13	$T H T T$	1
6	$H H T T$	2	14	$T T H T$	1
7	$H T T H$	2	15	$T T T H$	1
8	$T T H H$	2	16	$T T T T$	0

The random variable  $X$  takes the values 0, 1, 2, 3 and 4. Since, from the above table, we find that the number of cases favourable to the coming of 0, 1, 2, 3 and 4 heads are 1, 4, 6, 4 and 1 respectively, we have

$$P(X=0) = \frac{1}{16}, \quad P(X=1) = \frac{4}{16} = \frac{1}{4}, \quad P(X=2) = \frac{6}{16} = \frac{3}{8},$$

$$P(X=3) = \frac{4}{16} = \frac{1}{4} \text{ and } P(X=4) = \frac{1}{16}.$$

Thus the probability distribution of  $X$  can be summarised as follows :

$x :$	0	1	2	3	4
$p(x) :$	$\frac{1}{16}$	$\frac{1}{4}$	$\frac{3}{8}$	$\frac{1}{4}$	$\frac{1}{16}$
$E(X) = \sum_{x=0}^4 x p(x) = 1 \cdot \frac{1}{4} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{4} + 4 \cdot \frac{1}{16}$					
$= \frac{1}{4} + \frac{3}{4} + \frac{3}{4} + \frac{1}{4} = 2.$					

**Example 6·5.** A coin is tossed until a head appears. What is the expectation of the number of tosses required ?

[Delhi Univ. B.Sc., Oct. 1989]

**Solution.** Let  $X$  denote the number of tosses required to get the first head. Then  $X$  can materialise in the following ways :

$$\therefore E(X) = \sum_{x=1}^{\infty} x p(x)$$

Event	$x$	Probability $p(x)$
$H$	1	$1/2$
$TH$	2	$1/2 \times 1/2 = 1/4$
$TTH$	3	$1/2 \times 1/2 \times 1/2 = 1/8$
$\vdots$	$\vdots$	$\vdots$
		$= 1 \times \frac{1}{2} + 2 \times \frac{1}{4} + 3 \times \frac{1}{8} + 4 \times \frac{1}{16} + \dots$ ...(*)

This is an arithmetic-geometric series with ratio of GP being  $r = 1/2$ .

$$\text{Let } S = 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} + 3 \cdot \frac{1}{8} + 4 \cdot \frac{1}{16} + \dots$$

$$\text{Then } \frac{1}{2}S = \frac{1}{4} + 2 \cdot \frac{1}{8} + 3 \cdot \frac{1}{16} + \dots$$

$$\therefore (1 - \frac{1}{2})S = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \dots$$

$$\Rightarrow \frac{1}{2}S = \frac{1/2}{1 - (1/2)} = 1$$

[Since the sum of an infinite G.P. with first term  $a$  and common ratio  $r (< 1)$  is  $a/(1-r)$ ]

$$\Rightarrow S = 2$$

Hence, substituting in (\*), we get

$$E(X) = 2$$

**Example 6-6.** What is the expectation of the number of failures preceding the first success in an infinite series of independent trials with constant probability  $p$  of success in each trial? [Delhi Univ. B.Sc., Oct. 1991]

**Solution.** Let the random variable  $X$  denote the number of failures preceding the first success. Then  $X$  can take the values  $0, 1, 2, \dots, \infty$ . We have

$$p(x) = P(X=x) = P[x \text{ failures precede the first success}] = q^x p$$

where  $q = 1 - p$  is the probability of failure in a trial. Then by def.

$$\begin{aligned} E(X) &= \sum_{x=0}^{\infty} x p(x) = \sum_{x=0}^{\infty} x \cdot q^x p = pq \sum_{x=1}^{\infty} x q^{x-1} \\ &= pq [1 + 2q + 3q^2 + 4q^3 + \dots] \end{aligned}$$

Now  $1 + 2q + 3q^2 + 4q^3 + \dots$  is an infinite arithmetic-geometric series.

$$\text{Let } S = 1 + 2q + 3q^2 + 4q^3 + \dots$$

$$qS = q + 2q^2 + 3q^3 + \dots$$

$$\therefore (1 - q)S = 1 + q + q^2 + q^3 + \dots = \frac{1}{1 - q}$$

$$\Rightarrow S = \frac{1}{(1 - q)^2}$$

$$\therefore 1 + 2q + 3q^2 + 4q^3 + \dots = \frac{1}{(1-q)^2}$$

$$\text{Hence } E(X) = \frac{pq}{(1-q)^2} = \frac{pq}{p^2} = \frac{q}{p}$$

**Example 6-7.** A box contains 'a' white and 'b' black balls. 'c' balls are drawn. Find the expected value of the number of white balls drawn.

[Allahabad Univ. B.Sc., 1989; Indian Forest Service 1987]

**Solution.** Let a variable  $X_i$ , associated with  $i$ th draw, be defined as follows:

$$X_i = 1, \text{ if } i\text{th ball drawn is white}$$

$$\text{and } X_i = 0, \text{ if } i\text{th ball drawn is black}$$

Then the number 'S' of the white balls among 'c' balls drawn is given by

$$S = X_1 + X_2 + \dots + X_c = \sum_{i=1}^c X_i \Rightarrow E(S) = \sum_{i=1}^c E(X_i)$$

$$\text{Now } P(X_i = 1) = P(\text{of drawing a white ball}) = \frac{a}{a+b}$$

$$\text{and } P(X_i = 0) = P(\text{of drawing a black ball}) = \frac{b}{a+b}$$

$$\therefore E(X_i) = 1 \cdot P(X_i = 1) + 0 \cdot P(X_i = 0) = \frac{a}{a+b}$$

$$\text{Hence } E(S) = \sum_{i=1}^c \left( \frac{a}{a+b} \right) = \frac{ca}{a+b}$$

**Example 6-8.** Let variate  $X$  have the distribution

$$P(X=0) = P(X=2) = p; P(X=1) = 1-2p, \text{ for } 0 \leq p \leq \frac{1}{2}.$$

For what  $p$  is the  $\text{Var}(X)$  a maximum?

[Delhi Univ. B.Sc. (Maths Hons.) 1987, 85]

**Solution.** Here, the r.v.  $X$  takes the values 0, 1 and 2 with respective probabilities  $p$ ,  $1-2p$  and  $p$ ,  $0 \leq p \leq \frac{1}{2}$ .

$$\therefore E(X) = 0 \times p + 1 \times (1-2p) + 2 \times p = 1$$

$$E(X^2) = 0 \times p + 1^2 \times (1-2p) + 2^2 \times p = 1 + 2p$$

$$\therefore \text{Var}(X) = E(X^2) - [E(X)]^2 = 2p; 0 \leq p \leq \frac{1}{2}$$

Obviously  $\text{Var}(X)$  is maximum when  $p = \frac{1}{2}$ , and

$$[\text{Var}(X)]_{\max} = 2 \times \frac{1}{2} = 1$$

**Example 6-9.**  $\text{Var}(X) = 0 \Rightarrow P[X = E(X)] = 1$ . Comment.

**Solution.**  $\text{Var}(X) = E[X - E(X)]^2 = 0$

$$\Rightarrow [X - E(X)]^2 = 0, \text{ with probability 1}$$

$$\Rightarrow [X - E(X)] = 0, \text{ with probability 1}$$

$$\Rightarrow P[X = E(X)] = 1$$

**Example 6.10.** Explain by means of an example that a probability distribution is not uniquely determined by its moments.

**Solution.** Consider a r.v.  $X$  with p.d.f. [c.f. Log-Normal distribution § 8.2.15]

$$f(x) = \frac{1}{\sqrt{2\pi}x} \cdot \exp\left[-\frac{1}{2}(\log x)^2\right]; x > 0 \\ = 0; \text{ otherwise}$$
 ...(\*)

Consider, another r.v.  $Y$  with p.d.f.

$$g(y) = [1 + a \sin(2\pi \log y)] f(y) = g_a(y), \text{ (say)}, \quad y > 0 \quad ...(**)$$

which, for  $-1 \leq a \leq 1$ , represents a family of probability distributions.

$$\begin{aligned} E(Y') &= \int_0^\infty y' \{1 + a \sin(2\pi \log y)\} f(y) dy \\ &= \int_0^\infty y' f(y) dy + a \cdot \int_0^\infty y' \cdot \sin(2\pi \log y) f(y) dy \\ &= EX' + a \cdot \frac{1}{\sqrt{2\pi}} \int_0^\infty y' \cdot \sin(2\pi \log y) \cdot \frac{1}{y} \exp\left[-\frac{1}{2}(\log y)^2\right] dy \\ &= EX' + \frac{a}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{z^2/2} \cdot \sin(2\pi z) dz \\ &\quad [\log y = z \Rightarrow y = e^z] \\ &= EX' + \frac{a \cdot e^{r^2/2}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(z-r)^2} \cdot \sin(2\pi z) dz \\ &= EX' + \frac{a \cdot e^{r^2/2}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-y^2/2} \cdot \sin(2\pi y) dy \\ &\quad [z - r = y \Rightarrow \sin(2\pi z) = \sin(2\pi r + 2\pi y) = \sin 2\pi y, \\ &\quad r \text{ being a positive integer}.] \\ &= EX', \end{aligned}$$

the value of the integral being zero, since the integrand is an odd function of  $y$ .

$\Rightarrow E(Y')$  is independent of 'a' in (\*\*).

Hence,  $\{g(y) = g_a(y); -1 \leq a \leq 1\}$ , represents a family of distributions, each different from the other, but having the same moments. This explains that the moments may not determine a distribution uniquely.

**Example 6.11.** Starting from the origin, unit steps are taken to the right with probability  $p$  and to the left with probability  $q$  ( $\approx 1-p$ ). Assuming independent movements, find the mean and variance of the distance moved from origin after  $n$  steps (Random Walk Problem).

**Solution.** Let us associate a variable  $X_i$  with the  $i$ th step defined as follows :

$X_i = +1$ , if the  $i$ th step is towards the right,

= -1, if the  $i$ th step is towards the left.

Then  $S = X_1 + X_2 + \dots + X_n = \sum X_i$ , represents the random distance moved from origin after  $n$  steps.

$$\begin{aligned} E(X_i) &= 1 \times p + (-1) \times q = p - q \\ E(X_i^2) &= 1^2 \times p + (-1)^2 \times q = p + q = 1 \\ \therefore \text{Var}(X_i) &= E(X_i^2) - [E(X_i)]^2 = (p+q)^2 - (p-q)^2 = 4pq \\ \therefore E(S_n) &= \sum_{i=1}^n E(X_i) = n(p-q) \\ V(S_n) &= \sum_{i=1}^n V(X_i) = 4npq \end{aligned}$$

[ $\because$  Movements of steps are independent].

**Example 6-12.** Let r.v.  $X$  have a density function  $f(\cdot)$ , cumulative distribution function  $F(\cdot)$ , mean  $\mu$  and variance  $\sigma^2$ . Define  $Y = \alpha + \beta X$ , where  $\alpha$  and  $\beta$  are constants satisfying  $-\infty < \alpha < \infty$  and  $\beta > 0$ .

- (a) Select  $\alpha$  and  $\beta$  so that  $Y$  has mean 0 and variance 1.
- (b) What is the correlation coefficient  $\rho_{XY}$  between  $X$  and  $Y$ ?
- (c) Find the cumulative distribution function of  $Y$  in terms of  $\alpha$ ,  $\beta$  and  $F(\cdot)$ .
- (d) If  $X$  is symmetrically distributed about  $\mu$ , is  $Y$  necessarily symmetrically distributed about its mean?

**Solution.** (a)  $E(X) = \mu$ ,  $\text{Var}(X) = \sigma^2$ . We want  $\alpha$  and  $\beta$  s.t.

$$E(Y) = E(\alpha + \beta X) = \alpha + \beta\mu = 0 \quad \dots(1)$$

$$\text{Var}(Y) = \text{Var}(\alpha + \beta X) = \beta^2 \cdot \sigma^2 = 1 \quad \dots(2)$$

Solving (1) and (2) we get :

$$\beta = 1/\sigma, (\beta > 0) \text{ and } \alpha = -\mu/\sigma \quad \dots(3)$$

$$(b) \text{Cov}(X, Y) = E(XY) - E(X)E(Y) = E[X(\alpha + \beta X)]$$

$$[\because E(Y) = 0]$$

$$= \alpha \cdot E(X) + \beta \cdot E(X^2) = \alpha\mu + \beta[\sigma^2 + \mu^2]$$

$$\therefore \rho_{XY} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{\alpha\mu + \beta[\sigma^2 + \mu^2]}{\sigma \cdot 1} \quad (\because \sigma_Y = 1)$$

$$= \frac{1}{\sigma^2} [-\mu^2 + \sigma^2 + \mu^2] = 1$$

[On using (3)]

(c) Distribution function  $G_Y(\cdot)$  of  $Y$  is given by :

$$\begin{aligned} G_Y(y) &= P(Y \leq y) = P[\alpha + \beta X \leq y] \\ &= P(X \leq (y - \alpha)/\beta) \end{aligned}$$

$$\Rightarrow G_Y(y) = F_X\left(\frac{y - \alpha}{\beta}\right);$$

$$(d) \text{We have : } Y = \alpha + \beta X = \frac{1}{\sigma}(X - \mu) = \beta(X - \mu) \quad [\text{On using (3)}]$$

Since  $X$  is given to be symmetrically distributed about mean  $\mu$ ,  $(X - \mu)$  and  $-(X - \mu)$  have the same distribution.

Hence  $Y = \beta(X - \mu)$  and  $-Y = -\beta(X - \mu)$  have the same distribution. Since  $E(Y) = 0$ , we conclude that  $Y$  is symmetrically distributed about its mean.

**Example 6-13.** Let  $X$  be a r.v. with mean  $\mu$  and variance  $\sigma^2$ . Show that  $E(X - b)^2$ , as a function of  $b$ , is minimised when  $b = \mu$ .

$$\begin{aligned}\text{Solution. } E(X - b)^2 &= E[(X - \mu) + (\mu - b)]^2 \\ &= E(X - \mu)^2 + (\mu - b)^2 + 2(\mu - b)E(X - \mu) \\ &= \text{Var}(X) + (\mu - b)^2 \quad [\because E(X - \mu) = 0] \\ \Rightarrow E(X - b)^2 &\geq \text{Var}(X), \quad \dots(*)\end{aligned}$$

since  $(\mu - b)^2$ , being the square of a real quantity is always non-negative.

The sign of equality holds in (\*) iff

$$(\mu - b)^2 = 0 \Rightarrow \mu = b.$$

Hence  $E(X - b)^2$  is minimised when  $\mu = b$  and its minimum value is  $E(X - \mu)^2 = \sigma_X^2$ .

**Remark.** This result states that the sum of squares of deviations is minimum when taken about mean.

[Also see § 2-4, Property 3 of Arithmetic Mean]

**Example 6-14.** Let  $a_1, a_2, \dots, a_n$  be arbitrary real numbers and  $A_1, A_2, \dots, A_n$  be events. Prove that

$$\sum_{i=1}^n \sum_{j=1}^n a_i a_j P(A_i A_j) \geq 0$$

[Delhi Univ. B.A. (Spl. Course - Stat. Hons.), 1986]

**Solution.** Let us define the indicator variable :

$$\begin{aligned}X_i &= I_{A_i} = 1 \quad \text{if } A_i \text{ occurs} \\ &= 0 \quad \text{if } \bar{A}_i \text{ occurs}.\end{aligned}$$

Then using (6-2b) :

$$E(X_i) = P(A_i); \quad (i = 1, 2, \dots, n) \quad \dots(i)$$

$$\begin{aligned}\text{Also } X_i X_j &= I_{A_i \cap A_j} \\ \Rightarrow E(X_i X_j) &= P(A_i A_j) \quad \dots(ii)\end{aligned}$$

Consider, for real numbers  $a_1, a_2, \dots, a_n$ , the expression  $\left( \sum_{i=1}^n a_i X_i \right)^2$ , which is always non-negative.

$$\begin{aligned}\Rightarrow \left( \sum_{i=1}^n a_i X_i \right)^2 &\geq 0 \\ \Rightarrow \left( \sum_{i=1}^n a_i X_i \right) \left( \sum_{j=1}^n a_j X_j \right) &\geq 0 \\ \Rightarrow \sum_{i=1}^n \sum_{j=1}^n a_i a_j X_i X_j &\geq 0, \quad \dots(iii)\end{aligned}$$

for all  $a_i$ 's and  $a_j$ 's.

Since expected value of a non-negative quantity is always non-negative, on taking expectations of both sides in (iii) and using (i) and (ii) we get :

$$\sum_{i=1}^n \sum_{j=1}^n a_i a_j E(X_i X_j) \geq 0 \Rightarrow \sum_{i=1}^n \sum_{j=1}^n a_i a_j P(A_i A_j) \geq 0.$$

**Example 6.15.** In a sequence of Bernoulli trials, let  $X$  be the length of the run of either successes or failures starting with the first trial. Find  $E(X)$  and  $V(X)$ .

**Solution.** Let ' $p$ ' denote the probability of success. Then  $q = 1 - p$  is the probability of failure.  $X = 1$  means that we can have any one of the possibilities  $SF$  and  $FS$  with respective probabilities  $pq$  and  $qp$ .

$$\therefore P(X=1) = P(SF) + P(FS) = pq + qp = 2pq$$

Similarly

$$P(X=2) = P(SSF) + P(FFS) = p^2q + q^2p$$

In general

$$P(X=r) = P[SSS...SF] + P[FFF...FS] = p^r \cdot q + q^r \cdot p$$

$$\begin{aligned}\therefore E(X) &= \sum_{r=1}^{\infty} r P(X=r) = \sum_{r=1}^{\infty} r (p^r \cdot q + q^r \cdot p) \\&= pq \left[ \sum_{r=1}^{\infty} r \cdot p^{r-1} + \sum_{r=1}^{\infty} r \cdot q^{r-1} \right] \\&= pq [(1+2p+3p^2+\dots)+(1+2q+3q^2+\dots)] \\&= pq [(1-p)^{-2} + (1-q)^{-2}] = pq [q^{-2} + p^{-2}]\end{aligned}$$

(See Remark to Example 6.17)

$$= pq \left[ \frac{1}{q^2} + \frac{1}{p^2} \right] = \frac{p}{q} + \frac{q}{p}$$

$$V(X) = E(X^2) - [E(X)]^2 = E[X(X-1)] + E(X) - [E(X)]^2$$

Now

$$\begin{aligned}E[X(X-1)] &= \sum_{r=2}^{\infty} r(r-1) P(X=r) = \sum_{r=2}^{\infty} r(r-1) (p^r q + q^r p) \\&= \sum_{r=2}^{\infty} r(r-1)p^r q + \sum_{r=2}^{\infty} r(r-1)q^r p \\&= p^2 q \sum_{r=2}^{\infty} r(r-1)p^{r-2} + q^2 p \sum_{r=2}^{\infty} r(r-1)q^{r-2} \\&= 2p^2 q \sum_{r=2}^{\infty} \frac{r(r-1)}{2} p^{r-2} + 2q^2 p \sum_{r=2}^{\infty} \frac{r(r-1)}{2} q^{r-2} \\&= 2p^2 q (1-p)^{-3} + 2q^2 p (1-q)^{-3} \\&= 2 \left( \frac{p^2}{q^2} + \frac{q^2}{p^2} \right)\end{aligned}$$

$$V(X) = 2 \left( \frac{p^2}{q^2} + \frac{q^2}{p^2} \right) + \left( \frac{p}{q} + \frac{q}{p} \right) - \left( \frac{p}{q} + \frac{q}{p} \right)^2 \\ = \left( \frac{p}{q} - \frac{q}{p} \right)^2 + \left( \frac{p}{q} + \frac{q}{p} \right)$$

**Aliter.** Proceed as in Example 6.17.

**Example 6.16.** A deck of  $n$  numbered cards is thoroughly shuffled and the cards are inserted into  $n$  numbered cells one by one. If the card number ' $i$ ' falls in the cell ' $i$ ', we count it as a match, otherwise not. Find the mean and variance of total number of such matches. [Delhi Univ. B.Sc., (Stat. Hons.), 1988]

**Solution.** Let us associate a random variable,  $X_i$  with the  $i$ th draw defined as follows :

$$X_i = \begin{cases} 1, & \text{if the } i\text{th card dealt has the number 'i' on it} \\ 0, & \text{otherwise} \end{cases}$$

Then the total number of matches 'S' is given by

$$S = X_1 + X_2 + \dots + X_n = \sum_{i=1}^n X_i \\ \therefore E(S) = \sum_{i=1}^n E(X_i)$$

$$\text{Now } E(X_i) = 1.P(X_i = 1) + 0.P(X_i = 0) = P(X_i = 1) = \frac{1}{n}$$

$$\text{Hence } E(S) = \sum_{i=1}^n \left( \frac{1}{n} \right) = n \cdot \frac{1}{n} = 1$$

$$V(S) = V(X_1 + X_2 + \dots + X_n) \\ = \sum_{i=1}^n V(X_i) + 2 \sum_{i,j=1}^n \text{Cov}(X_i, X_j) \quad \dots(1)$$

$$\text{Now } V(X_i) = E(X_i^2) - [E(X_i)]^2 \\ = 1^2.P(X_i = 1) + 0^2.P(X_i = 0) - \left( \frac{1}{n} \right)^2 \\ = \frac{1}{n} - \frac{1}{n^2} = \frac{n-1}{n^2} \quad \dots(2)$$

$$\text{Cov}(X_i, X_j) = E(X_i X_j) - E(X_i) E(X_j) \quad \dots(3)$$

$$E(X_i X_j) = 1.P(X_i X_j = 1) + 0.P(X_i X_j = 0) \\ = \frac{(n-2)!}{n!} = \frac{1}{n(n-1)},$$

since  $X_i X_j = 1$  if and only if both card numbers  $i$  and  $j$  are in their respective matching places and there are  $(n-2)!$  arrangements of the remaining cards that correspond to this event.

Substituting in (3), we get

$$\text{Cov}(X_i, X_j) = \frac{1}{n(n-1)} - \frac{1}{n} \cdot \frac{1}{n} = \frac{1}{n^2(n-1)} \quad \dots(4)$$

Substituting from (2) and (4) in (1), we have

$$\begin{aligned} V(S) &= \sum_{i=1}^n \left( \frac{n-1}{n^2} \right) + 2 \sum_{\substack{i=1 \\ i \neq j}}^n \sum_{j=1}^n \left[ \frac{1}{n^2(n-1)} \right] \\ &= n \left( \frac{n-1}{n^2} \right) + 2 \cdot C_2 \frac{1}{n^2(n-1)} = \frac{n-1}{n} + \frac{1}{n} = 1 \end{aligned}$$

**Example 6.17.** If  $t$  is any positive real number, show that the function defined by

$$p(x) = e^{-t} (1 - e^{-t})^{x-1} \quad \dots(*)$$

can represent a probability function of a random variable  $X$  assuming the values 1, 2, 3, ... Find the  $E(X)$  and  $\text{Var}(X)$  of the distribution.

[Nagpur Univ. B.Sc., 1988]

**Solution.** We have

$$e' > 1, \forall t > 0 \Rightarrow e^{-t} < 1 \Rightarrow 1 - e^{-t} > 0$$

$$\text{Also } e^{-t} = \frac{1}{e'} > 0, \forall t > 0$$

$$\text{Hence } p(x) = e^{-t} (1 - e^{-t})^{x-1} \geq 0 \quad \forall t > 0, x = 1, 2, 3, \dots$$

$$\begin{aligned} \text{Also } \sum_{x=1}^{\infty} p(x) &= e^{-t} \sum_{x=1}^{\infty} (1 - e^{-t})^{x-1} = e^{-t} \sum_{x=1}^{\infty} a^{x-1}; \quad [a = 1 - e^{-t}] \\ &= e^{-t} (1 + a + a^2 + a^3 + \dots) = e^{-t} \times \frac{1}{(1-a)} \\ &= e^{-t} [1 - (1 - e^{-t})]^{-1} = e^{-t} (e^{-t})^{-1} = 1 \end{aligned}$$

Hence  $p(x)$  defined in (\*) represents the probability function of a r.v.  $X$ .

$$\begin{aligned} E(X) &= \sum x \cdot p(x) = e^{-t} \sum_{x=1}^{\infty} x (1 - e^{-t})^{x-1} \\ &= e^{-t} \sum_{x=1}^{\infty} x \cdot a^{x-1}; \quad [a = 1 - e^{-t}] \\ &= e^{-t} (1 + 2a + 3a^2 + 4a^3 + \dots) = e^{-t} (1 - a)^{-2} \quad \dots(*) \\ &= e^{-t} (e^{-t})^{-2} = e^t \end{aligned}$$

$$E(X^2) = \sum x^2 p(x) = e^{-t} \sum_{x=1}^{\infty} x^2 \cdot a^{x-1}$$

$$= e^{-t} [1 + 4a + 9a^2 + 16a^3 + \dots]$$

$$= e^{-t} (1 + a) (1 - a)^{-3} = e^{-t} (2 - e^{-t}) e^{3t}$$

$$\begin{aligned} \text{Hence } \text{Var}(X) &= E(X^2) - [E(X)]^2 = e^{-t} (2 - e^{-t}) e^{3t} - e^{2t} \\ &= e^{2t} [(2 - e^{-t}) - 1] = e^{2t} (1 - e^{-t}) \\ &= e^t (e^t - 1) \end{aligned}$$

**Remark.**

(i) Consider  $S = 1 + 2a + 3a^2 + 4a^3 + \dots$  (Arithmetico-geometric series)

$$\begin{aligned} \Rightarrow aS &= a + 2a^2 + 3a^3 + \dots \\ \Rightarrow (1-a)S &= 1 + a + a^2 + a^3 + \dots = \frac{1}{a}(1-a) \Rightarrow S = (1-a)^{-2} \\ \sum_{x=1}^{\infty} x a^{x-1} &= 1 + 2a + 3a^2 + 4a^3 + \dots = (1-a)^{-2} \quad \dots(*) \end{aligned}$$

(ii) Consider

$$\begin{aligned} S &= 1 + 2^2 \cdot a + 3^2 \cdot a^2 + 4^2 \cdot a^3 + 5^2 \cdot a^4 + \dots \\ \Rightarrow S &= 1 + 4a + 9a^2 + 16a^3 + 25a^4 + \dots \\ - 3aS &= -3a - 12a^2 - 27a^3 - 48a^4 - \dots \\ + 3a^2S &= \quad + 3a^2 + 12a^3 + 27a^4 + \dots \\ - a^3S &= \quad - a^3 - 4a^4 - \dots \end{aligned}$$

Adding the above equations we get :

$$\begin{aligned} (1-a)^3S &= 1 + a \Rightarrow S = (1+a)(1-a)^{-3} \quad \dots(**) \\ \sum_{x=1}^{\infty} x^2 a^{x-1} &= 1 + 4a + 9a^2 + 16a^3 + \dots = (1+a)(1-a)^{-3} \end{aligned}$$

The results in (\*) and (\*\*) are quite useful for numerical problems and should be committed to memory.

**Example 6.18.** A man with  $n$  keys wants to open his door and tries the keys independently and at random. Find the mean and variance of the number of trials required to open the door (i) if unsuccessful keys are not eliminated from further selection, and (ii) if they are. [Rajasthan Univ. B.Sc.(Hons.), 1992]

**Solution.** (i) Suppose the man gets the first success at the  $x$ th trial, i.e., he is unable to open the door in the first  $(x-1)$  trials. If unsuccessful keys are not eliminated then  $X$  is a random variable which can take the values  $1, 2, 3, \dots$  ad infinity.

Probability of success at the first trial =  $\frac{1}{n}$

$\therefore$  Probability of failure at the first trial =  $1 - (\frac{1}{n})$

If unsuccessful keys are not eliminated then the probability of success and consequently of failure is constant for each trial.

Hence  $p(x) = \text{Probability of 1st success at the } x\text{th trial}$

$$= \left(1 - \frac{1}{n}\right)^{x-1} \cdot \frac{1}{n}$$

$$\begin{aligned} \therefore E(X) &= \sum_{x=1}^{\infty} x p(x) = \sum_{x=1}^{\infty} x \left(1 - \frac{1}{n}\right)^{x-1} \cdot \frac{1}{n} \\ &= \frac{1}{n} \sum_{x=1}^{\infty} x A^{x-1}, \text{ where } A = 1 - \frac{1}{n} \end{aligned}$$

$$E(X) = \frac{1}{n} [1 + 2A + 3A^2 + 4A^3 + \dots] = \frac{1}{n} (1-A)^{-2}$$

[See (\*), Example (6.17)]

$$\begin{aligned}
 \frac{1}{n} \left[ 1 - \left( 1 - \frac{1}{n} \right) \right]^{-2} &= n \\
 E(X^2) &= \sum_{x=1}^n x^2 p(x) = \sum_{x=1}^n x^2 \left( 1 - \frac{1}{n} \right)^{x-1} \cdot \frac{1}{n} \\
 &= \frac{1}{n} \sum_{x=1}^n x^2 A^{x-1} \\
 &= \frac{1}{n} [1 + 2^2 \cdot A + 3^2 \cdot A^2 + 4^2 \cdot A^3 + \dots] \\
 &= \frac{1}{n} (1+A)(1-A)^{-3} \quad [\text{See } (**), \text{ Example (6-17)}] \\
 &= \frac{1}{n} \left[ 1 + \left( 1 - \frac{1}{n} \right) \right] \left[ 1 - \left( 1 - \frac{1}{n} \right) \right]^{-3} \\
 &= (2n-1)n
 \end{aligned}$$

Hence  $V(X) = E(X^2) - [E(X)]^2 = (2n-1)n - n^2 = n^2 - n = n(n-1)$

(ii) If unsuccessful keys are eliminated from further selection, then the random variable  $X$  will take the values from 1 to  $n$ . In this case, we have

Probability of success at the first trial =  $\frac{1}{n}$

Probability of success at the 2nd trial =  $\frac{1}{n-1}$

Probability of success at the 3rd trial =  $\frac{1}{n-2}$

and so on.

Hence probability of 1st success at 2nd trial =  $\left( 1 - \frac{1}{n} \right) \frac{1}{n-1} = \frac{1}{n}$

Probability of first success at the third trial

$$= \left( 1 - \frac{1}{n} \right) \left( 1 - \frac{1}{n-1} \right) \cdot \frac{1}{n-2} = \frac{1}{n}$$

and so on. In general, we have

$$p(x) = \text{Probability of first success at the } x\text{th trial} = \frac{1}{n}$$

$$\therefore E(X) = \sum_{x=1}^n x p(x) = \frac{1}{n} \sum_{x=1}^n x = \frac{n+1}{2}$$

$$E(X^2) = \sum_{x=1}^n x^2 p(x) = \frac{1}{n} \sum_{x=1}^n x^2 = \frac{(n+1)(2n+1)}{6}$$

Hence  $V(X) = E(X^2) - [E(X)]^2 = \frac{(n+1)(2n+1)}{6} - \left( \frac{n+1}{2} \right)^2$

$$= \frac{n+1}{12} \left[ 2(2n+1) - 3(n+1) \right] = \frac{n^2-1}{12}$$

**Example 6.19.** In a lottery  $m$  tickets are drawn at a time out of  $n$  tickets numbered 1 to  $n$ . Find the expectation and the variance of the sum  $S$  of the numbers on the tickets drawn. [Delhi Univ. B.Sc. (Maths Hons.), 1987]

**Solution.** Let  $X_i$  denote the score on the  $i$ th ticket drawn.

$$\text{Then } S = X_1 + X_2 + \dots + X_m = \sum_{i=1}^m X_i,$$

is the total score on the  $m$  tickets drawn.

$$\therefore E(S) = \sum_{i=1}^m E(X_i)$$

Now each  $X_i$  is a random variable which assumes the values 1, 2, 3, ...,  $n$  each with equal probability  $1/n$ .

$$\therefore E(X_i) = \frac{1}{n}(1+2+3+\dots+n) = \frac{(n+1)}{2}$$

$$\text{Hence } E(S) = \sum_{i=1}^m \left( \frac{n+1}{2} \right) = \frac{m(n+1)}{2}$$

$$\begin{aligned} V(S) &= V(X_1 + X_2 + \dots + X_m) \\ &= \sum_{i=1}^m V(X_i) + 2 \sum_{\substack{i,j \\ i \neq j}} \text{Cov}(X_i, X_j) \end{aligned}$$

$$\begin{aligned} E(X_i^2) &= \frac{1}{n}(1^2 + 2^2 + 3^2 + \dots + n^2) \\ &= \frac{1}{n} \cdot \frac{n(n+1)(2n+1)}{6} = \frac{(n+1)(2n+1)}{6} \end{aligned}$$

$$\begin{aligned} V(X_i) &= E(X_i^2) - [E(X_i)]^2 \\ &= \frac{(n+1)(2n+1)}{6} - \left( \frac{n+1}{2} \right)^2 = \frac{n^2-1}{12} \end{aligned}$$

$$\text{Also } \text{Cov}(X_i, X_j) = E(X_i X_j) - E(X_i) E(X_j)$$

To find  $E(X_i X_j)$  we note that the variables  $X_i$  and  $X_j$  can take the values as shown below :

$X_i$	$X_j$
1	2, 3, ..., $n$
2	1, 3, ..., $n$
$\vdots$	$\vdots$
$n$	1, 2, ..., ( $n-1$ )

Thus the variable  $X_i X_j$  can take  $n(n-1)$  possible values and  $P(X_i = l \cap X_j = k) = \frac{1}{n(n-1)}$ ,  $k \neq l$ . Hence

$$\begin{aligned}
 E(X_i X_j) &= \frac{1}{n(n-1)} \left[ \begin{array}{l} 1.2 + 1.3 + \dots + 1.n \\ + 2.1 + 2.3 + \dots + 2.n \\ + \dots \dots \dots \dots \dots \\ + n.1 + n.2 + \dots + n.(n-1) \end{array} \right] \\
 &\stackrel{(1)}{=} \frac{1}{n(n-1)} \left[ \begin{array}{l} 1(1+2+3+\dots+n)-1^2 \\ + 2(1+2+3+\dots+n)-2^2 \\ + \dots \dots \dots \dots \dots \\ + n(1+2+\dots+n-1+n)-n^2 \end{array} \right] \\
 &= \frac{1}{n(n-1)} [(1+2+3+\dots+n)^2 - (1^2 + 2^2 + \dots + n^2)] \\
 &= \frac{1}{n(n-1)} \left[ \left\{ \frac{n(n+1)}{2} \right\}^2 - \frac{n(n+1)(2n+1)}{6} \right] \\
 &= \frac{(n+1)(3n^2-n-2)}{12(n-1)}
 \end{aligned}$$

$$\begin{aligned}
 \therefore \text{Cov}(X_i, X_j) &= \frac{(n+1)(3n^2-n-2)}{12(n-1)} - \left( \frac{n+1}{2} \right)^2 \\
 &= \frac{(n+1)}{12(n-1)} [3n^2 - n - 2 - 3(n^2 - 1)] \\
 &= -\frac{(n+1)}{12}
 \end{aligned}$$

$$\begin{aligned}
 \text{Hence } V(S) &= \sum_{i=1}^n \left( \frac{n^2-1}{12} \right) + 2 \sum_{i < j=1}^n \left\{ -\frac{(n+1)}{12} \right\} \\
 &= \frac{m(n^2-1)}{12} + 2 \cdot \frac{m(m-1)}{2!} \left\{ -\frac{(n+1)}{12} \right\},
 \end{aligned}$$

[since there are  $C_2$  covariance terms in  $\text{Cov}(Y_i, X_j)$ ]

$$\therefore V(S) = \frac{m(n+1)}{12} [(n-1) - (m-1)] = \frac{m(n+1)(n-m)}{12}$$

**Example 6.20.** A die is thrown  $(n+2)$  times. After each throw a '+' is recorded for 4, 5 or 6 and '-' for 1, 2 or 3, the signs forming an ordered sequence. To each, except the first and the last sign, is attached a characteristic random variable which takes the value 1 if both the neighbouring signs differ from the one between them and 0 otherwise. If  $X_1, X_2, \dots, X_n$  are characteristic random variables, find the mean and variance of  $X = \sum_{i=1}^n X_i$ .

$$\text{Solution. } X = \sum_{i=1}^n X_i \Rightarrow E(X) = \sum_{i=1}^n E(X_i)$$

Now  $E(X_i) = 1 \cdot P(X_i = 1) + 0 \cdot P(X_i = 0) = P(X_i = 1)$

For  $X_i = 1$ , there are the following two mutually exclusive possibilities :

(i) - + -, (ii) + - +

and since the probability of each sign is  $\frac{1}{2}$ , we have by addition probability theorem:

$$P(X_i = 1) = P(i) + P(ii) = \left(\frac{1}{2}\right)^3 + \left(\frac{1}{2}\right)^3 = \frac{1}{4}$$

$$\therefore E(X_i) = \frac{1}{4}$$

$$\text{Hence } E(X) = \sum_{i=1}^n \left(\frac{1}{4}\right) = \frac{n}{4},$$

$$\begin{aligned} V(X) &= V(X_1 + X_2 + \dots + X_n) \\ &= \sum_{i=1}^n V(X_i) + 2 \sum_{i < j} \text{Cov}(X_i, X_j) \end{aligned} \quad \dots(*)$$

$$\text{Now } E(X_i^2) = 1^2 \cdot P(X_i = 1) + 0^2 \cdot P(X_i = 0) = P(X_i = 1) = \frac{1}{4}$$

$$\therefore V(X) = E(X^2) - [E(X)]^2 = \frac{1}{4} - \frac{1}{16} = \frac{3}{16}$$

$$\text{Now } E(X_i X_j) = 1 \cdot P(X_i = 1 \cap X_j = 1) + 0 \cdot P(X_i = 0 \cap X_j = 0)$$

$$4 + 0 \cdot P(X_i = 1 \cap X_j = 0) + 0 \cdot P(X_i = 0 \cap X_j = 1)$$

$$= P(X_i = 1 \cap X_j = 1)$$

Since there are the following two mutually exclusive possibilities for the event :  $(X_i = 1 \cap X_j = 1)$ ,

(i) - + - +

(ii) + - + - , we have

$$P(X_i = 1 \cap X_j = 1) = P(i) + P(ii) = \left(\frac{1}{2}\right)^4 + \left(\frac{1}{2}\right)^4 = \frac{1}{8}$$

$$\therefore \text{Cov}(X_i, X_j) = E(X_i X_j) - E(X_i) E(X_j)$$

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

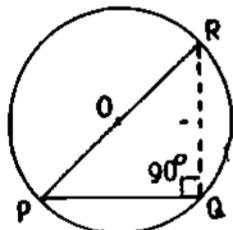
$$\begin{aligned} \text{Hence } V(X) &= \sum_{i=1}^n (3/16) + 2 \sum_{i < j} \text{Cov}(X_i, X_j) \quad [\text{From } (*)] \\ &= \frac{3n}{16} + 2 [\text{Cov}(X_1, X_2) + \text{Cov}(X_2, X_3) + \\ &\quad \dots + \text{Cov}(X_{n-1}, X_n)] \\ &= \frac{3n}{16} + 2(n-1) \cdot \frac{1}{16} = \frac{5n-2}{16} \end{aligned}$$

**Example 6.21.** From a point on the circumference of a circle of radius 'a', a chord is drawn in a random direction, (all directions are equally likely). Show that the expected value of the length of the chord is  $4a/\pi$  and that the variance of the

length is  $2a^2(1 - 8/\pi^2)$ . Also show that the chance is  $1/3$  that the length of the chord will exceed the length of the side of an equilateral triangle inscribed in the circle.

**Solution.** Let  $P$  be any point on the circumference of a circle of radius ' $a$ ' and centre ' $O$ '. Let  $PQ$  be any chord drawn at random and let  $\angle OPQ = \theta$ . Obviously,

$\theta$  ranges from  $-\pi/2$  to  $\pi/2$ . Since all the directions are equally likely, the probability differential of  $\theta$  is given by the rectangular distribution (c.f. Chapter 8) :



$$dF(\theta) = f(\theta) d(\theta) = \frac{d\theta}{\pi/2 - (-\pi/2)} = \frac{d\theta}{\pi}, \quad -\frac{\pi}{2} \leq \theta \leq \frac{\pi}{2}$$

Now, since  $\angle PQR$  is a right angle, (angle in a semi-circle), we have

$$\frac{PQ}{PR} = \cos \theta \Rightarrow PQ = PR \cos \theta = 2a \cos \theta$$

$$E(PQ) = \int_{-\pi/2}^{\pi/2} (PQ) f(\theta) d(\theta) = \frac{2a}{\pi} \int_{-\pi/2}^{\pi/2} \cos \theta d\theta \\ = \frac{2a}{\pi} \left| \sin \theta \right|_{-\pi/2}^{\pi/2} = \frac{4a}{\pi}$$

$$E[(PQ)^2] = \int_{-\pi/2}^{\pi/2} [PQ]^2 f(\theta) d\theta = \frac{4a^2}{\pi} \int_{-\pi/2}^{\pi/2} \cos^2 \theta d\theta \\ = \frac{4a^2}{\pi} \int_0^{\pi/2} 2 \cos^2 \theta d\theta = \frac{4a^2}{\pi} \int_0^{\pi/2} (1 + \cos 2\theta) d\theta \\ \text{(Since } \cos^2 \theta \text{ is an even function of } \theta\text{)} \\ = \frac{4a^2}{\pi} \left| \theta + \frac{\sin 2\theta}{2} \right|_0^{\pi/2} = \frac{4a^2}{\pi} \cdot \frac{\pi}{2} = 2a^2$$

$$\therefore V(PQ) = E[(PQ)^2] - [E(PQ)]^2 = 2a^2 - \frac{16a^2}{\pi^2} = 2a^2 \left(1 - \frac{8}{\pi^2}\right)$$

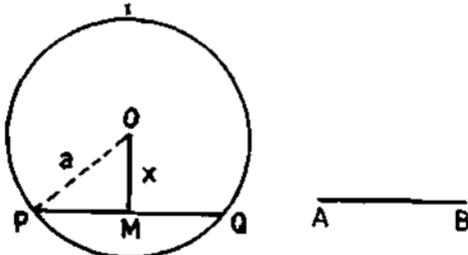
We know that the length of the side of an equilateral triangle inscribed in a circle of radius ' $a$ ' is  $a\sqrt{3}$ . Hence

$$P(PQ > a\sqrt{3}) = P(2a \cos \theta > a\sqrt{3}) = P\left(\cos \theta > \frac{\sqrt{3}}{2}\right) \\ = P\left(|\theta| < \frac{\pi}{6}\right) = P\left(-\frac{\pi}{6} < \theta < \frac{\pi}{6}\right) \\ = \int_{-\pi/6}^{\pi/6} f(\theta) d\theta = \frac{1}{\pi} \int_{-\pi/6}^{\pi/6} 1 \cdot d\theta = \frac{1}{\pi} \cdot \frac{\pi}{3} = \frac{1}{3}$$

**Example 6-22.** A chord of a circle of radius ' $a$ ' is drawn parallel to a given straight line, all distances from the centre of the circle being equally likely. Show that the expected value of the length of the chord is  $\pi a/2$  and that the variance of the length is  $a^2(32 - 3\pi^2)/12$ . Also show that the chance is  $1/2$  that the length of

*the chord will exceed the length of the side of an equilateral triangle inscribed in the circle.*

**Solution.** Let  $PQ$  be the chord of a circle with centre  $O$  and radius ' $a$ ' drawn at random parallel to the given straight line  $AB$ . Draw  $OM \perp PQ$ . Let  $OM = x$ . Obviously  $x$  ranges from  $-a$  to  $a$ . Since all distances from the centre are equally



likely, the probability that a random value of  $x$  will lie in the small interval ' $dx$ ' is given by the rectangular distribution [c.f. Chapter 8]:

$$dF(x) = f(x) dx = \frac{dx}{a - (-a)} = \frac{dx}{2a}, \quad -a \leq x \leq a$$

Length of the chord is

$$PQ = 2PM = 2\sqrt{a^2 - x^2}$$

$$\begin{aligned} \text{Hence } E(PQ) &= \int_{-a}^a PQ dF(x) = \frac{2}{2a} \int_{-a}^a \sqrt{a^2 - x^2} dx \\ &= \frac{2}{a} \int_0^a \sqrt{a^2 - x^2} dx, \end{aligned}$$

(since integrand is an even function of  $x$ ).

$$\begin{aligned} &= \frac{2}{a} \left[ \frac{1}{2} x \sqrt{a^2 - x^2} + \frac{1}{2} a^2 \sin^{-1} \left( \frac{x}{a} \right) \right]_0^a \\ &= \frac{2}{a} \cdot \frac{a^2}{2} \cdot \frac{\pi}{2} = \frac{\pi a}{2} \end{aligned}$$

$$\begin{aligned} E[(PQ)^2] &= \int_{-a}^a (PQ)^2 dF(x) = \frac{4}{2a} \int_{-a}^a (a^2 - x^2) dx \\ &= \frac{4}{a} \int_0^a (a^2 - x^2) dx = \frac{4}{a} \left[ a^2 x - \frac{x^3}{3} \right]_0^a \\ &= \frac{4}{a} \cdot \frac{2a^3}{3} = \frac{8a^2}{3} \end{aligned}$$

Hence

$$\begin{aligned} \text{Var (length of chord)} &= E[(PQ)^2] - [E(PQ)]^2 = \frac{8a^2}{3} - \frac{\pi^2 a^2}{4} \\ &= \frac{a^2}{12} (32 - 3\pi^2) \end{aligned}$$

The length of the chord is greater than the side of the equilateral triangle inscribed in the circle if

$$2\sqrt{a^2 - x^2} > a\sqrt{3} \Rightarrow 4(a^2 - x^2) > 3a^2 \\ i.e., \quad x^2 < a^2/4 \Rightarrow |x| < a/2$$

Hence the required probability is

$$P(|x| < a/2) = P\left(-\frac{a}{2} < X < \frac{a}{2}\right) = \int_{-a/2}^{a/2} dF(x) \\ = \frac{1}{2a} \int_{-a/2}^{a/2} 1 \cdot dx = \frac{1}{2}$$

**Example 6-23.** Let  $X_1, X_2, \dots, X_n$  be a sequence of mutually independent random variables with common distribution. Suppose  $X_k$  assumes only positive integral values and  $E(X_k) = a$ , exists;  $k = 1, 2, \dots, n$ . Let  $S_n = X_1 + X_2 + \dots + X_n$ .

(i) Show that  $E\left(\frac{S_m}{S_n}\right) = \frac{m}{n}$ , for  $1 \leq m \leq n$

(ii) Show that  $E(S_n^{-1})$  exists and

$$E\left(\frac{S_m}{S_n}\right) = 1 + (m-n)a E(S_n^{-1}), \text{ for } 1 \leq n \leq m$$

(iii) Verify and use the inequality  $x + x^{-1} \geq 2$ , ( $x > 0$ ) to show that

$$E\left(\frac{S_m}{S_n}\right) \geq \frac{m}{n} \text{ for } m, n \geq 1 \quad [\text{Delhi Univ. M.Sc. (Stat.), 1988}]$$

**Solution.** (i) We have

$$E\left[\frac{X_1 + X_2 + \dots + X_n}{X_1 + X_2 + \dots + X_n}\right] = E(1) = 1$$

$$\Rightarrow E\left[\frac{X_1 + X_2 + \dots + X_n}{S_n}\right] = 1$$

$$\Rightarrow E\left(\frac{X_1}{S_n}\right) + E\left(\frac{X_2}{S_n}\right) + \dots + E\left(\frac{X_n}{S_n}\right) = 1$$

Since  $X_i$ 's, ( $i = 1, 2, \dots, n$ ) are identically distributed random variables,  $(X_i/S_n)$ , ( $i = 1, 2, \dots, n$ ) are also identically distributed random variables.

$$\therefore n E\left(\frac{X_i}{S_n}\right) = 1$$

$$\Rightarrow E\left(\frac{X_i}{S_n}\right) = \frac{1}{n}; \quad i = 1, 2, \dots, n \quad \dots(*)$$

Now

$$E\left(\frac{S_m}{S_n}\right) = E\left(\frac{X_1 + X_2 + \dots + X_n}{S_n}\right) = E\left[\frac{X_1}{S_n} + \frac{X_2}{S_n} + \dots + \frac{X_n}{S_n}\right]$$

$$\begin{aligned}
 &= E\left(\frac{X_1}{S_n}\right) + E\left(\frac{X_2}{S_n}\right) + \dots + E\left(\frac{X_m}{S_n}\right) \\
 &= \frac{1}{n} + \frac{1}{n} + \dots + \frac{1}{n} \quad [(m \text{ times})] \quad [\text{Using } (*)] \\
 &= \frac{m}{n}, \quad (m < n)
 \end{aligned}$$

(ii) Since  $X_i$ 's assume only positive integral values, we have  
 $n \leq X_1 + X_2 + \dots + X_n < \infty$

$$\Rightarrow \frac{1}{n} \geq \frac{1}{S_n} > 0 \Rightarrow 0 < S_n^{-1} \leq \frac{1}{n}$$

Since  $S_n^{-1}$  lies between two finite quantities 0 and  $\frac{1}{n}$ , we get

$$0 < E(S_n^{-1}) \leq \frac{1}{n}$$

Hence  $E(S_n^{-1})$  exists.

$$\begin{aligned}
 E\left(\frac{S_m}{S_n}\right) &= E\left[\frac{X_1 + X_2 + \dots + X_n + X_{n+1} + \dots + X_m}{S_n}\right], \quad m \geq n \\
 &= E\left[1 + \frac{X_{n+1}}{S_n} + \dots + \frac{X_m}{S_n}\right] \\
 &= 1 + E\left(\frac{X_{n+1}}{S_n}\right) + \dots + E\left(\frac{X_m}{S_n}\right)
 \end{aligned}$$

Since  $X_{n+1}, X_{n+2}, \dots, X_m$  are independent of  $S_n = X_1 + X_2 + \dots + X_n$ , they are independent of  $S_n^{-1}$  also.

$$\begin{aligned}
 \therefore E\left(\frac{S_m}{S_n}\right) &= 1 + E(X_{n+1})E(S_n^{-1}) + \dots + E(X_m)E(S_n^{-1}) \\
 &= 1 + [E(X_{n+1}) + \dots + E(X_m)]E(S_n^{-1}) \\
 &= 1 + (m-n)aE(S_n^{-1}), \quad 1 \leq n \leq m
 \end{aligned}$$

$[\because E(X_i) = a \quad \forall i]$

(iii) Verification of  $x + \frac{1}{2} \geq 2, \quad (x > 0)$ .

$$x + \frac{1}{2} \geq 2$$

$$\Rightarrow x^2 + 1 \geq 2x \quad (\text{multiplication valid only if } x > 0)$$

$$\Rightarrow (x - 1)^2 \geq 0$$

which is always true for  $x > 0$ .

If  $1 \leq m \leq n$ , result follows from (\*).

If  $1 \leq n \leq m$ , then using (\*\*), we have to prove that

$$\begin{aligned}
 1 + (m-n) a E(S_n^{-1}) &\geq \frac{m}{n} \\
 \Rightarrow (m-n)a E(S_n^{-1}) &\geq \frac{m-n}{n} \\
 \Rightarrow E(S_n^{-1}) &\geq \frac{1}{an} \quad \dots (***) \\
 \end{aligned}$$

In (\*\*), taking  $x = \frac{S_n}{an} > 0$ , we get

$$\begin{aligned}
 E(x) + E(x^{-1}) &\geq 2 \\
 \Rightarrow E\left(\frac{S_n}{an}\right) + E\left(\frac{S_n}{an}\right)^{-1} &\geq 2 \\
 \Rightarrow \frac{1}{an} \cdot E(S_n) + an E(S_n^{-1}) &\geq 2 \\
 \Rightarrow \frac{1}{an} \cdot an + an E(S_n^{-1}) &\geq 2 \\
 \Rightarrow an E(S_n^{-1}) &\geq 1 \\
 \Rightarrow E(S_n^{-1}) &\geq \frac{1}{an},
 \end{aligned}$$

which was to be proved in (\*\*\*)�.

**Example 6-24.** Let  $X$  be a r.v. for which  $\beta_1$  and  $\beta_2$  exist. Then for any real  $k$ , prove that :

$$\beta_2 \geq \beta_1 - (2k + k^2) \quad \dots (*)$$

Deduce that (i)  $\beta_2 \geq \beta_1$  (ii)  $\beta_2 \geq 1$ . When is  $\beta_2 = 1$  ?

**Solution.** Without any loss of generality we can take  $E(X) = 0$ . [If  $E(X) \neq 0$ , then we may start with the random variable  $Y = X - E(X)$  so that  $E(Y) = 0$ .]

Consider the real valued function of the real variable  $t$  defined by :

$$Z(t) = E[X^2 + tX + k\mu_2]^2 \geq 0 \quad \forall t,$$

$$\text{where } \mu_r = EX^r, \quad \dots (i)$$

is the  $r$ th moment of  $X$  about mean.

$$\begin{aligned}
 \therefore Z(t) &= E[X^4 + t^2 X^2 + k^2 \mu_2^2 + 2tX^3 + 2k\mu_2 X^2 + 2k\mu_2 tX] \\
 &= \mu_4 + t^2 \mu_2 + k^2 \mu_2^2 + 2t\mu_3 + 2k\mu_2^2 \quad [\text{Using (i) and } E(X) = 0] \\
 &= t^2 \mu_2 + 2t\mu_3 + \mu_4 + k^2 \mu_2^2 + 2k\mu_2^2 \geq 0 \quad \text{for all } t. \quad \dots (ii)
 \end{aligned}$$

Since  $Z(t)$  is a quadratic form in  $t$ ,  $Z(t) \geq 0$  for all  $t$  iff its discriminant is  $\leq 0$ , i.e.,

$$\begin{aligned}
 \text{iff } (2\mu_3)^2 - 4\mu_2 [ \mu_4 + k^2 \mu_2^2 + 2k\mu_2^2 ] &\leq 0 \\
 \Rightarrow \frac{\mu_3^2}{\mu_2^2} - \left[ \frac{\mu_4}{\mu_2^2} + k^2 + 2k \right] &\leq 0 \quad [\text{Dividing by } 4\mu_2^2 > 0]
 \end{aligned}$$

$$\Rightarrow \beta_1 - \beta_2 - (2k + k^2) \leq 0 \\ \Rightarrow \beta_2 \geq \beta_1 - (2k + k^2)$$

**Deductions.** (i) Taking  $k = 0$  in (\*) we get  $\beta_2 \geq \beta_1$  ...(\*\*)

(ii) Taking  $k = -1$  in (\*) we get:  $\beta_2 \geq \beta_1 + 1$  ...(\*\*\*)

a result, which is established differently in Example 6.26.

(iii) Since  $\beta_1 = \mu_3^2/\mu_2^3$  is always non-negative, we get from (\*\*\*) :

$$\beta_2 \geq 1 \quad \dots(\text{****})$$

**Remark.** The sign of equality holds in (\*\*\*\*), i.e.,  $\beta_2 = 1$  iff

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = 1 \quad \Rightarrow \quad \mu_4 = \mu_2^2$$

$$\begin{aligned} \Rightarrow E[X - E(X)]^4 &= [E(X - E(X))]^2 \\ \Rightarrow E(Y^2) - [E(Y)]^2 &= 0, \quad (Y = [X - E(X)]^2) \\ \Rightarrow \text{Var}(Y) &= 0 \\ \Rightarrow P[Y = E(Y)] &= 1 \quad [\text{See Example 6.9}] \\ \Rightarrow P[(X - \mu)^2 = E(X - \mu)^2] &= 1 \\ \Rightarrow P[(X - \mu)^2 = \sigma^2] &= 1 \\ \Rightarrow P[(X - \mu) = \pm \sigma] &= 1 \\ \Rightarrow P[X = \mu \pm \sigma] &= 1 \end{aligned}$$

Thus  $X$  takes only two values  $\mu + \sigma$  and  $\mu - \sigma$  with respective probabilities  $p$  and  $q$ , (say).

$$\therefore E(X) = p(\mu + \sigma) + q(\mu - \sigma) = \mu \\ \Rightarrow p + q = 1 \quad \text{and} \quad (p - q)\sigma = 0$$

But since  $\sigma \neq 0$ , (since in this case  $\beta_2$  is defined) we have:

$$p + q = 1 \quad \text{and} \quad p - q = 0 \Rightarrow p = q = \frac{1}{2}.$$

Hence  $\beta_2 = 1$  iff the r.v.  $X$  assumes only two values, each with equal probability  $\frac{1}{2}$ .

**Example 6.25.** Let  $X$  and  $Y$  be two variates having finite means.

Prove or disprove :

- (a)  $E[\min(X, Y)] \leq \min[E(X), E(Y)]$
- (b)  $E[\max(X, Y)] \geq \max[E(X), E(Y)]$
- (c)  $E[\min(X, Y) + \max(X, Y)] = E(X) + E(Y)$

[Delhi Univ. B.A. (Stat. Hons.), Spl. Course, 1989]

**Solution.** We know that

$$\min(X, Y) = \frac{1}{2}(X + Y) - |X - Y| \quad \dots(i)$$

$$\text{and} \quad \max(X, Y) = \frac{1}{2}(X + Y) + |X - Y| \quad \dots(ii)$$

$$(a) \therefore E[\min(X, Y)] = \frac{1}{2}E(X+Y) - E|X-Y| \quad \dots(iii)$$

We have :  $|E(X-Y)| \leq E|X-Y|$   
 $\Rightarrow -|E(X-Y)| \geq -E|X-Y|$   
 $\Rightarrow E|X-Y| \leq -|E(X-Y)| = -|E(X)-E(Y)| \quad \dots(*)$

Substituting in (iii) we get :

$$\begin{aligned} E[\min(X, Y)] &\leq \frac{1}{2}E(X+Y) - E|X-Y| \\ &\leq \frac{1}{2}[E(X)+E(Y)] - |E(X)-E(Y)| \quad \dots[\text{From } (*)] \end{aligned}$$

$$\Rightarrow E[\min(X, Y)] \leq \min[E(X), E(Y)]$$

(b) Similarly from (ii) we get :

$$\begin{aligned} E[\max(X, Y)] &= \frac{1}{2}E(X+Y) + E|X-Y| \\ &\geq \frac{1}{2}E(X+Y) + |E(X)-E(Y)| \\ &\quad (\because |E(X-Y)| \leq E|X-Y|) \\ &= \frac{1}{2}[E(X)+E(Y)] + |E(X)-E(Y)| \\ &= \max[E(X), E(Y)] \end{aligned}$$

$$\text{i.e., } E[\max(X, Y)] \geq \max[E(X), E(Y)]$$

$$\begin{aligned} (c) \quad [ \min(X, Y) + \max(X, Y) ] &= [X+Y] \\ \Rightarrow E[\min(X, Y) + \max(X, Y)] &= E(X+Y) \\ &= E(X) + E(Y), \end{aligned}$$

as required.

Hence all the results in (a), (b) and (c) are true.

**Example 6.26.** Use the relation  $E(AX^a + BX^b + CX^c)^2 \geq 0$ ,  $X$  being a random variable with  $E(X) = 0$ ,  $E$  denoting the mathematical expectation, to show that

$$\begin{vmatrix} \mu_{2a} & \mu_{a+b} & \mu_{a+c} \\ \mu_{a+b} & \mu_{2b} & \mu_{b+c} \\ \mu_{a+c} & \mu_{b+c} & \mu_{2c} \end{vmatrix} \geq 0, \quad \dots(*)$$

$\mu_n$  denoting the  $n$ th moment about mean.

Hence or otherwise show that Pearson Beta-coefficients satisfy the inequality

$$\beta_2 - \beta_1 - 1 \geq 0.$$

Also deduce that  $\beta_2 \geq 1$ .

**Solution.** Since  $E(X) = 0$ , we get

$$E(X') = \mu_r \quad \dots(**)$$

We are given that

$$E(AX^a + BX^b + CX^c)^2 \geq 0$$

$$\begin{aligned} \Rightarrow E[A^2X^{2a} + B^2X^{2b} + C^2X^{2c} + 2ABX^{a+b} + 2ACX^{a+c} + 2BCX^{b+c}] &\geq 0 \\ \Rightarrow A^2\mu_{2a} + B^2\mu_{2b} + C^2\mu_{2c} + 2AB\mu_{a+b} + 2AC\mu_{a+c} + 2BC\mu_{b+c} &\geq 0 \\ &\quad [\text{From } (**)] \quad \dots(***) \end{aligned}$$

for all values of  $A, B, C$ .

We know from matrix theory that the conditions for the quadratic form

$$a'x^2 + b'y^2 + c'z^2 + 2f'yz + 2g'zx + 2h'xy,$$

to be non-negative for all values of  $x, y$  and  $z$  are

$$a' \geq 0, \quad (ii) \quad \begin{vmatrix} a' & h' \\ h' & b' \end{vmatrix} \geq 0, \text{ and} \quad (iii) \quad \begin{vmatrix} a' & h' & g' \\ h' & b' & f' \\ g' & f' & c' \end{vmatrix} \geq 0$$

Comparing with (\*\*), we have

$$a' = \mu_{2a}, b' = \mu_{2b}, c' = \mu_{2c}, f' = \mu_{b+c}, g' = \mu_{a+c}, h' = \mu_{a+b}$$

Substituting these values in condition (iii) above, we get the required result.

Taking  $a = 0, b = 1$  and  $c = 2$  in (\*) and noting that  $\mu_0 = 1$  and  $\mu_1 = 0$ , we get

$$\begin{aligned} & \begin{vmatrix} 1 & 0 & \mu_2 \\ 0 & \mu_2 & \mu_3 \\ \mu_2 & \mu_3 & \mu_4 \end{vmatrix} \geq 0 \\ \Rightarrow \quad & \mu_2 \mu_4 - \mu_3^2 + \mu_2 (-\mu_2^2) \geq 0 \end{aligned}$$

Dividing throughout by  $\mu_2^3$  (assuming that  $\mu_2$  is finite, for otherwise  $\beta_2$  will become infinite), we get

$$\begin{aligned} & \frac{\mu_4}{\mu_2^2} - \frac{\mu_3^2}{\mu_2^3} - 1 \geq 0 \\ \Rightarrow \quad & \beta_2 - \beta_1 - 1 \geq 0 \\ \Rightarrow \quad & \beta_2 \geq \beta_1 + 1 \end{aligned}$$

Further since  $\beta_1 \geq 0$ , we get  $\beta_2 \geq 1$ .

**Example 6.27.** Let  $X$  be a non-negative random variable with distribution function  $F$ . Show that

$$E(X) = \int_0^\infty [1 - F(x)] dx. \quad \dots(i)$$

Conjecture a corresponding expression for  $E(X^2)$ .

[Delhi Univ. M.Sc.(Stat). 1988]

**Solution.** Since  $X \geq 0$ , we have :

$$R.H.S. = \int_0^\infty [1 - P(X \leq x)] dx = \int_0^\infty \left[ 1 - \int_0^x f(u) du \right] dx,$$

where  $f(\cdot)$  is the p.d.f. of r.v.  $X$ .

$$\therefore R.H.S. = \int_0^\infty \left[ \int_x^\infty f(u) du \right] dx, \quad \dots(ii)$$

From the integral in bracket (ii), we have,  $u \geq x$  and since  $x$  ranges from 0 to  $\infty$ ,  $u$  also range from 0 to  $\infty$ . Further  $u \geq x \Rightarrow x \leq u$  and since  $x$  is non-negative,

we have  $0 \leq x \leq u$ . [See Region  $R_1$  in Remark 2 below]. Hence changing the order of integration in (ii), [by Fubini's theorem for non-negative functions], we get

$$\begin{aligned} \text{R.H.S.} &= \int_0^{\infty} \left[ \int_0^x 1 \cdot dx \right] f(u) du = \int_0^{\infty} u \cdot f(u) du \\ &= E(X) \quad [\text{Since } f(\cdot) \text{ is p.d.f. of } X] \end{aligned}$$

**Conjecture for  $E(X^2)$ .** Consider the integral:

$$\int_0^{\infty} 2x [1 - F(x)] dx = \int_0^{\infty} 2x \left( \int_x^{\infty} f(u) du \right) dx$$

$$= \int_0^{\infty} \left( \int_0^u 2x dx \right) f(u) du,$$

(By Fubini's theorem for non-negative functions).

$$= \int_0^{\infty} u^2 \cdot f(u) du = E(X^2)$$

**Remarks.** 1. If  $X$  is a non-negative r.v. then

$$\text{Var}X = EX^2 - [E(X)]^2 = \int_0^{\infty} 2x [1 - F(x)] dx - \mu_x^2 \quad \dots(iii)$$

2. If we do not restrict ourselves to non-negative random variables only, we have the following more generalised result.

If  $F$  denotes the distribution function of the random variable  $X$  then :

$$E(X) = \int_0^{\infty} [1 - F(x)] dx = \int_{-\infty}^{\infty} F(x) dx, \quad \dots(iv)$$

provided the integrals exist finitely.

*Proof of (iv).* The first integral has already been evaluated in the above example, i.e.,

$$\int_0^{\infty} [1 - F(x)] dx = \int_0^{\infty} u \cdot f(u) du \quad \dots(v)$$

Consider :

$$\int_{-\infty}^0 F(x) dx = \int_{-\infty}^0 P(X \leq x) dx = \int_{-\infty}^0 \left( \int_{-\infty}^x f(u) du \right) dx$$

$$= \int_{-\infty}^0 \left( \int_u^0 1 \cdot dx \right) f(u) du$$

[Changing the order of integration in the Region  $R_2$  where  $u \leq x$ ].

$$= \int_{-\infty}^0 u \cdot f(u) du \quad \dots(vi)$$

Subtracting (vi) from (v), we get :

$$\begin{aligned} \int_0^\infty [1 - F(x)] dx - \int_{-\infty}^0 F(x) dx &= \int_0^\infty u f(u) du + \int_{-\infty}^0 u f(u) du \\ &= \int_{-\infty}^\infty u f(u) du \\ &= E(X), \quad (\text{Since } f(\cdot) \text{ is p.d.f. of } X) \end{aligned}$$

as desired.

In this generalised case,

$$\text{Var}(X) = \int_0^\infty 2x [1 - F_X(x) + F_X(-x)] dx = [E(X)]^2$$

3. The corresponding analogue of the above result for discrete random variable is given in the next Example 6-28.

**Example 6-28.** If the possible values of a variate  $X$  are  $0, 1, 2, 3, \dots$ , then

$$E(X) = \sum_{n=0}^{\infty} P(X > n)$$

[Delhi Univ. B.Sc. (Maths Hons.), 1987]

**Solution.** Let  $P(X = n) = p_n$ ,  $n = 0, 1, 2, 3, \dots$  ...(i)

If  $E(X)$  exists, then by definition :

$$E(X) = \sum_{n=0}^{\infty} n \cdot P(X = n) = \sum_{n=1}^{\infty} n \cdot p_n \quad \dots(ii)$$

Consider :

$$\begin{aligned} \sum_{n=0}^{\infty} P(X > n) &= P(X > 0) + P(X > 1) + P(X > 2) + \dots \\ &= P(X \geq 1) + P(X \geq 2) + P(X \geq 3) + \dots \\ &= (p_1 + p_2 + p_3 + p_4 + \dots) \\ &\quad + (p_2 + p_3 + p_4 + \dots) \\ &\quad + (p_3 + p_4 + p_5 + \dots) \\ &\quad + \dots \\ &= p_1 + 2p_2 + 3p_3 + \dots \\ &= \sum_{n=1}^{\infty} n p_n \\ &= E(X) \quad \text{[From (ii)]} \end{aligned}$$

As an illustration of this result, see Problem 24 in Exercise 6(a).

$$\text{Aliter. R.H.S.} = \sum_{n=0}^{\infty} P(X > n) = \sum_{n=1}^{\infty} P(X \geq n)$$

$$= \sum_{x=n}^{\infty} \left( \sum_{x=n}^{\infty} p(x) \right)$$

Since the series is convergent and  $p(x) \geq 0 \forall x$ , by Fubini's theorem, changing the order of summation we get:

$$\text{R.H.S.} = \sum_{x=1}^{\infty} \left( \sum_{n=1}^x p(x) \right) = \sum_{x=1}^{\infty} \left\{ p(x) \sum_{n=1}^x 1 \right\}$$

since  $x \geq n \Rightarrow n \leq x$  and  $x$  assumes only positive integral values:

$$\therefore \text{R.H.S.} = \sum_{x=1}^{\infty} x p(x) = \sum_{x=0}^{\infty} x p(x) = E(X)$$

**Example 6-29.** For any variates  $X$  and  $Y$ , show that

$$[E(X+Y)^2]^{1/2} \leq [E(X^2)]^{1/2} + [E(Y^2)]^{1/2} \quad \dots(*)$$

**Solution.** Squaring both sides in (\*), we have to prove

$$\begin{aligned} E(X+Y)^2 &\leq [(E(X^2))^{1/2} + (E(Y^2))^{1/2}]^2 \\ \Rightarrow E(X^2) + E(Y^2) + 2E(XY) &\leq E(X^2) + E(Y^2) + 2\sqrt{E(X^2)E(Y^2)} \\ \Rightarrow E(XY) &\leq \sqrt{E(X^2)E(Y^2)} \\ \Rightarrow [E(XY)]^2 &\leq E(X^2) \cdot E(Y^2). \end{aligned}$$

This is nothing but Cauchy-Schwartz inequality. [For proof see Theorem 6-11 page 6-13.]

**Example 6-30.** Let  $X$  and  $Y$  be independent non-degenerate variates. Prove that

$$\text{Var}(XY) = \text{Var}(X) \cdot \text{Var}(Y)$$

$$\text{iff } E(X) = 0, E(Y) = 0$$

[Delhi Univ. B.Sc. (Maths Hons.), 1989]

**Solution.** We have

$$\begin{aligned} \text{Var}(XY) &= E(XY)^2 - [E(XY)]^2 = E(X^2Y^2) - [E(XY)]^2 \\ &= E(X^2)E(Y^2) - [E(X)]^2[E(Y)]^2, \end{aligned} \quad \dots(*)$$

since  $X$  and  $Y$  are independent.

$$\text{If } E(X) = 0 = E(Y)$$

$$\text{then } \text{Var}(X) = E(X^2) \text{ and } \text{Var}(Y) = E(Y^2) \quad \dots(**)$$

Substituting from (\*\*) in (\*), we get

$$\text{Var}(XY) = \text{Var}(X) \cdot \text{Var}(Y),$$

as desired..

**Only If.** We have to prove that if

$$\text{Var}(XY) = \text{Var}(X) \cdot \text{Var}(Y) \quad \dots(***)$$

$$\text{then } E(X) = 0 \text{ and } E(Y) = 0.$$

Now (\*\*\* ) gives, [on using (\*)]

$$E(X^2)E(Y^2) - [E(X)]^2[E(Y)]^2 = [E(X^2) - [E(X)]^2] \times [E(Y^2) - [E(Y)]^2]$$

$$\begin{aligned}
 &= E(X^2)E(Y^2) - E(X^2)[E(Y)]^2 - [E(X)]^2E(Y^2) + [E(X)]^2[E(Y)]^2 \\
 \Rightarrow & E(X^2)[E(Y)]^2 - [E(X)]^2[E(Y)]^2 + [E(X)]^2E(Y^2) - [E(X)]^2[E(Y)]^2 = 0 \\
 \Rightarrow & [E(Y)]^2(E(X^2) - [E(X)]^2) + [E(X)]^2(E(Y^2) - [E(Y)]^2) = 0 \\
 \Rightarrow & [E(Y)]^2 \text{Var}(X) + [E(X)]^2 \text{Var}(Y) = 0 \quad \dots (***) \\
 \end{aligned}$$

Since each of the quantities  $[E(X)]^2$ ,  $[E(Y)]^2$ ,  $\text{Var}(X)$  and  $\text{Var}(Y)$  is non-negative and since  $X$  and  $Y$  are given to be non-degenerate random variables such that  $\text{Var}(X) > 0$  and  $\text{Var}(Y) > 0$ ,  $(***)$  holds only if we have  $E(X) = 0 = E(Y)$ , as required.

### EXERCISE 6(a)

1. (a) Define a random variable and its mathematical expectation.

(b) Show that the mathematical expectation of the sum of two random variables is the sum of their individual expectations and if two variables are independent, the mathematical expectation of their product is the product of their expectations.

Is the condition of independence necessary for the latter? If not, what is the necessary condition?

(c) If  $X$  is a random variable, prove that  $|E(X)| \leq E(|X|)$ .

(d) If  $X$  and  $Y$  are two random variables such that  $X \leq Y$ , prove that

$$E(X) \leq E(Y).$$

(e) Prove that  $E[(X - c)^2] = [\text{Var}(X)] + [E(X) - c]^2$ , where  $c$  is a constant.

2. Prove that

(a)  $E(aX + bY) = aE(X) + bE(Y)$ .

where  $a$  and  $b$  are any constants.

(b)  $E(a) = a$ ,  $a$  being a constant.

(c)  $E[ag(X)] = aE[g(X)]$

(d)  $E[g_1(X) + g_2(X) + \dots + g_n(X)] = E[g_1(X)] + E[g_2(X)] + \dots + E[g_n(X)]$

(e)  $|E[g(X)]| \leq E[|g(X)|]$ .

(f) If  $g(X) \geq 0$ , everywhere then  $E[g(X)] \geq 0$ .

(g) If  $g(X) \geq 0$  everywhere and  $E[g(X)] = 0$ , then  $g(X) = 0$ , i.e., the random variable  $g(X)$  has a one point distribution at  $X = 0$ .

3. Show that if  $X$  is non-negative random variable such that both  $E(X)$  and  $E(1/X)$  exist, then

$$E(1/X) \geq 1/E(X).$$

4. If  $X, Y$  are independent random variables with  $E(X) = \alpha$ ,  $E(X^2) = \beta$ , and  $E(Y^k) = a_k$ ;  $k = 1, 2, 3, 4$ , find  $E(XY + Y^2)^2$ .

5. (a) If  $X$  and  $Y$  are two independent random variables, show that

$$\text{Var}(aX + bY) = a^2 \text{Var}(X) + b^2 \text{Var}(Y).$$

(b) With usual notations, show that

$$\text{Cov}(aX + bY, cX + dY) = ac \text{Var}(X) + bd \text{Var}(Y) + (ad + bc) \text{Cov}(X, Y)$$

(c) Show that

$$\text{Cov} \left( \sum_{i=1}^n a_i X_i, \sum_{j=1}^m b_j X_j \right) = \sum_{j=1}^m \sum_{i=1}^n a_i b_j \text{Cov}(X_i, X_j)$$

6. (a) Define the indicator function  $I_A(x)$  and show that  $E(I_A(X)) = P(A)$ .

(b) Prove that the probability function  $P(X \in A)$  for set  $A$  and the distribution function  $F_X(x)$ , ( $-\infty < x < \infty$ ), can be regarded as expectations of some random variable.

**Hint.** Define the indicator functions :

$$\begin{array}{ll} I_A(x) = 1 & \text{if } x \in A \\ & = 0 \text{ if } x \notin A \end{array} \quad \begin{array}{ll} I_y(x) = 1 & \text{if } x \leq y \\ & = 0 \text{ if } x > y \end{array}$$

Then we shall get :

$$E(I_A(X)) = P(X \in A) \text{ and } E(I_y(X)) = P(X \leq y) = F_X(y)$$

7. (a) Let  $X$  be a continuous random variable with median  $m$ . Minimise  $E|X - b|$ , as an function of  $b$ .

**Ans.**  $E|X - b|$  is minimum when  $b = m$  = Median. This states that absolute sum of deviations of a given set of observations is minimum when taken about median. [See Example 5.19.]

(b) Let  $X$  be a random variable such that  $E|X| < \infty$ . Show that  $E|X - C|$  is minimised if we choose  $C$  equal to the median of the distribution.

[Delhi Univ. B.Sc. (Maths Hons.), 1988]

8. If  $X$  and  $Y$  are symmetric, show that

$$E\left(\frac{X}{X+Y}\right) = \frac{1}{2}$$

**Hint.**  $1 = E\left[\frac{X+Y}{X+Y}\right] = E\left[\frac{X}{X+Y}\right] + E\left[\frac{Y}{X+Y}\right]$

$$\Rightarrow 1 = 2E\left[\frac{X}{X+Y}\right] \quad (\because X \text{ and } Y \text{ are symmetric.})$$

9. (a) If a r.v.  $X$  has a symmetric density about the point ' $a$ ' and if  $E(X)$  exists, then

$$\text{Mean}(X) = \text{Median}(X) = a$$

**Hint.** Given  $f(a-x) = f(a+x)$ ;  $f(x)$  p.d.f. of  $X$ . prove that

$$E(X-a) = \int_{-\infty}^{\infty} (x-a)f(x)dx = \int_{-\infty}^{\infty} (x-a)f(x)dx + \int_a^{\infty} (x-a)f(x)dx = 0$$

(b) If  $X$  and  $Y$  are two random variables with finite variances, then show that

$$E^2(XY) \leq E(X^2) \cdot E(Y^2) \quad \dots (*)$$

When does the equality sign hold in (\*)? [Indian Civil Service, 1987]

10. Let  $X$  be a non-negative arbitrary r.v. with distribution function  $F$ , show that

$$E(X) = \int_0^{\infty} [1 - F_X(x)] dx = \int_{-\infty}^0 F_X(x) dx,$$

in the sense that, if either side exists, so does the other and the two are equal.

[Delhi Univ. B.Sc. (Maths Hons.), 1992]

11. Show that if  $Y$  and  $Z$  are independent random values of a variable  $X$ , the expected value of  $(Y - Z)^2$  is twice the variance of the distribution of  $X$ .

[Allahabad Univ. B.Sc., 1989]

**Hint.**  $E(Y) = E(Z) = E(X) = \mu$ , (say);  $\sigma_y^2 = \sigma_z^2 = \sigma_x^2 = \sigma^2$ , (say). ... (\*)

$$E(Y - Z)^2 = E(Y^2) + E(Z^2) - 2E(Y)E(Z)$$

( $\because Y, Z$  are independent)

$$\equiv (\sigma_y^2 + \mu_y^2) + (\sigma_z^2 + \mu_z^2) - 2\mu^2$$

$$= 2\sigma^2 = 2\sigma_x^2$$

[On using (\*)]

12. Given the following table :

$x$	-3	-2	-1	0	1	2	3
$p(x)$	0.05	0.10	0.30	0	0.30	0.15	0.10

Compute (i)  $E(X)$ , (ii)  $E(2X \pm 3)$ , (iii)  $E(4X + 5)$ , (iv)  $E(X^2)$

(v)  $V(X)$ , and (vi)  $V(2X \pm 3)$ .

13. (a)  $A$  and  $B$  throw with one die for a stake of Rs. 44 which is to be won by the player who first throws a 6. If  $A$  has the first throw, what are their respective expectations?

**Ans.** Rs. 24, Rs. 20.

- (b) A contractor has to choose between two jobs. The first promises a profit of Rs. 1,20,000 with a probability of  $3/4$  or a loss of Rs. 30,000 due to delays with a probability of  $1/4$ ; the second promises a profit of Rs. 1,80,000 with a probability of  $1/2$  or a loss of Rs. 45,000 with a probability of  $1/2$ . Which job should the contractor choose so as to maximise his expected profit?

- (c) A random variable  $X$  can assume any positive integral value  $n$  with a probability proportional to  $1/3^n$ . Find the expectation of  $X$ .

[Delhi Univ. B.Sc., Oct. 1987]

14. Three tickets are chosen at random without replacement from 100 tickets numbered 1, 2, ..., 100. Find the mathematical expectation of the sum of the numbers on the tickets drawn.

15. (a) Three urns contain respectively 3 green and 2 white balls, 5 green and 6 white balls and 2 green and 4 white balls. One ball is drawn from each urn. Find the expected number of white balls drawn out.

**Hint.** Let us define the r.v.

$X_i = 1$ , if the ball drawn from  $i$ th urn is white  
 $= 0$ , otherwise

Then the number of white balls drawn is  $S = X_1 + X_2 + X_3$ .

$$E(S) = E(X_1) + E(X_2) + E(X_3) = 1 \times \frac{2}{5} + 1 \times \frac{6}{11} + 1 \times \frac{4}{6} = \frac{266}{165}.$$

(b) Urn A contains 5 cards numbered from 1 to 5 and urn B contains 4 cards numbered from 1 to 4. One card is drawn from each of these urns. Find the probability function of the number which appears on the cards drawn and its mathematical expectation.

Ans. 11/4.

16. (a) Thirteen cards are drawn from its pack simultaneously. If the values of aces are 1, face cards 10 and others according to denomination, find the expectation of the total score in all the 13 cards.

[Madurai Univ. B.Sc., Oct. 1990]

(b) Let  $X$  be a random variable with p.d.f. as given below :

$x$ :	0	1	2	3
$p(x)$ :	1/3	1/2	1/24	.1/8

Find the expected value of  $Y = (X - 1)^2$ . [Aligarh Univ. B.Sc. (Hons.), 1992]

17. A player losses 3 fair coins. He wins Rs. 8, if three heads occur; Rs. 3, if 2 heads occur and Re. 1, if one head occurs. If the game is to be fair, how much should he lose, if no heads occur? [Punjab Univ. M.A. (Econ.), 1987]

Hint.  $X$  : Player's prize in Rs.

$x$	8	3	1	$a$
No. of heads	3	2	1	0
$p(x)$	1/8	3/8	3/8	1/8

$$E(X) = \sum x p(x) = \frac{1}{8}(8 + 9 + 3 + a)$$

For the game to be fair, we have :

$$E(X) = 0 \Rightarrow 20 + a = 0 \Rightarrow a = -20$$

Hence the player loses Rs. 20, if

no heads come up.

18. (a) A coin is tossed until a tail appears. What is the expectation of the number of tosses ?

Ans. 2.

(b) Find the expectation of (i) the sum, and (ii) the product, of number of points on  $n$  dice when thrown.

Ans. (i)  $7n/2$ , (ii)  $(7/2)^n$

19. (a) Two cards are drawn at random from ten cards numbered 1 to 10. Find the expectation of the sum of points on two cards.

(b) An urn contains  $n$  cards marked from 1 to  $n$ . Two cards are drawn at a time. Find the mathematical expectation of the product of the numbers on the cards.

[Mysore Univ. B.Sc., 1991]

(c) In a lottery  $m$  tickets are drawn out of  $n$  tickets numbered from 1 to  $n$ . What is the expectation of the sum of the squares of numbers drawn?

(d) A bag contains  $n$  white and 2 black balls. Balls are drawn one by one without replacement until a black is drawn. If 0, 1, 2, 3, ... white balls are drawn before the first black, a man is to receive  $0^2, 1^2, 2^2, 3^2, \dots$  rupees respectively. Find his expectation. [Rajasthan Univ. B.Sc., 1992]

(e) Find the expectation and variance of the number of successes in a series of independent trials, the probability of success in the  $i$ th trial being  $p_i$ ; ( $i = 1, 2, \dots, n$ ). [Nagarjuna Univ. B.Sc., 1991]

20. Balls are taken one by one out of an urn containing  $w$  white and  $b$  black balls until the first white ball is drawn. Prove that the expectation of the number of black balls preceding the first white ball is  $b/(w+1)$ .

[Allahabad Univ. B.Sc. (Hons.), 1992]

21. (a)  $X$  and  $Y$  are independent variables with means 10 and 20, and variances 2 and 3 respectively. Find the variance of  $3X + 4Y$ .

Ans. 66.

(b) Obtain the variance of  $Y = 2X_1 + 3X_2 + 4X_3$ , where  $X_1, X_2$  and  $X_3$  are three random variables with means given by 3, 4, 5 respectively, variances by 10, 20, 30 respectively, and co-variances by  $\sigma_{X_1 X_2} = 0, \sigma_{X_2 X_3} = 0, \sigma_{X_1 X_3} = 5$ , where  $\sigma_{X_i X_j}$  stands for the co-variance of  $X_i$  and  $X_j$ .

22. (a) Suppose that  $X$  is a random variable for which  $E(X) = 10$  and  $\text{Var}(X) = 25$ . Find the positive values of  $a$  and  $b$  such that  $Y = aX - b$ , has expectation 0 and variance 1.

Ans.  $a = 1/5, b = 2$

(b) Let  $X_1$  and  $X_2$  be two stochastic random variables having variances  $k$  and 2 respectively. If the variance of  $Y = 3X_2 - X_1$  is 25, find  $k$ .

[Poona Univ. B.Sc., 1990]

Ans.  $k = 7$ .

23. A bag contains  $2n$  counters, of which half are marked with odd numbers and half with even numbers, the sum of all the numbers being  $S$ . A man is to draw two counters. If the sum of the numbers drawn is odd, he is to receive that number of rupees, if even he is to pay that number of rupees. Show that his expectation is  $S/[n(2n-1)]$  rupees. [I.F.S., 1989]

24. A jar has  $n$  chips numbered 1, 2, ...,  $n$ . A person draws a chip, returns it, draws another, returns it, and so on, until a chip is drawn that has been drawn before. Let  $X$  be the number of drawings. Find  $E(X)$ .

[Delhi Univ. B.A. (Stat. Hons.), Spl. Course, 1986]

Hint. Obviously  $P(X > 1) = 1$ , because we must have at least two draws to get the chip which has been drawn before.

$$P(X > r) = P[\text{Distinct number on } i\text{th draw}; i = 1, 2, \dots, r]$$

$$P(X > r) = \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{r-1}{n}\right); r = 1, 2, 3, \dots \quad \dots (*)$$

Hence, using the result in Example 6.28

$$\begin{aligned} E(X) &= \sum_{r=0}^{\infty} P(X > r) \\ &= P(X > 0) + P(X > 1) + P(X > 2) + \dots \\ &= 1 + 1 + \left(1 - \frac{1}{n}\right) + \left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right) + \\ &\dots + \left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{r-1}{n}\right) + \dots \quad [\text{Using } (*)] \end{aligned}$$

25. A coin is tossed four times. Let  $X$  denote the number of times a head is followed immediately by a tail. Find the distribution, mean and variance of  $X$ .

**Hint.**  $S = \{H, T\} \times \{H, T\} \times \{H, T\} \times \{H, T\}$   
 $= \{HHHH, HHHT, HHTH, HTHH, HTHT, \dots, TTTT\}$

$X:$	0,	1,	1,	1,	2,	...,	0
$x$	0	1	2				
$p(x)$	$\frac{5}{16}$	$\frac{10}{16}$	$\frac{1}{16}$				

$$E(X) = \frac{3}{4}, E(X^2) = \frac{7}{8}$$

$$\text{Var } X = \frac{7}{8} - \frac{9}{16} = \frac{5}{16}.$$

26. An urn contains balls numbered 1, 2, 3. First a ball is drawn from the urn and then a fair coin is tossed the number of times as the number shown on the drawn ball. Find the expected number of heads.

[Delhi Univ. B.Sc. (Maths Hons.), 1984]

**Hint.**  $B_j$  : Event of drawing the ball numbered  $j$ .

$$P(B_j) = \frac{1}{3}; j = 1, 2, 3.$$

$X$  : No. of heads shown.  $X$  is a r.v. taking the values 0, 1, 2, and 3.

$$P(X = x) = \sum_{j=1}^3 P(B_j) \cdot P(X = x | B_j) = \frac{1}{3} \sum_{j=1}^3 P(X = x | B_j)$$

$$\therefore P(X = 0) = \frac{1}{3} [P(X = 0 | B_1) + P(X = 0 | B_2) + P(X = 0 | B_3)] \\ \leq \frac{1}{3} \left[ \frac{1}{2} + \frac{1}{4} + \frac{1}{8} \right] = \frac{7}{24}$$

$$P(X = 1) = \frac{1}{3} [P(X = 1 | B_1) + P(X = 1 | B_2) + P(X = 1 | B_3)] \\ = \frac{1}{3} \left[ \frac{1}{2} + \frac{2}{4} + \frac{3}{8} \right]$$

$$\text{e.g., } P(X = 0 | B_2) = P[\text{No head when two coins are tossed}] = \frac{1}{4}$$

$$P(X = 1 | B_3) = P[\text{1 head when three coins are tossed}] = \frac{3}{8}$$

$$\text{Similarly } P(X = 2) = \frac{1}{3} \left(0 + \frac{1}{4} + \frac{3}{8}\right) = \frac{5}{24}$$

$$P(X=3) = \frac{1}{3} \left( 0 + 0 + \frac{1}{8} \right) = \frac{1}{24}$$

$$\therefore E(X) = \sum_{x=0}^3 x P(X=x) = \frac{11}{24} + \frac{10}{24} + \frac{3}{24} = 1$$

27. An urn contains  $pN$  white and  $qN$  black balls, the total number of balls being  $N$ ,  $p+q=1$ . Balls are drawn one by one (without being returned to the urn) until a certain number  $n$  of balls is reached.

Let  $X_i = 1$ , if the  $i$ th ball drawn is white.

$= 0$ , if the  $i$ th ball drawn is black.

(i) Show that  $E(X_i) = p$ ,  $\text{Var}(X_i) = pq$ .

(ii) Show that the co-variance between  $X_j$  and  $X_k$  is  $-\frac{pq}{n-1}$ , ( $j \neq k$ )

(iii) From (i) and (ii), obtain the variance of  $S_n = X_1 + X_2 + \dots + X_n$ .

28. Two similar decks of  $n$  distinct cards each are put into random order and are matched against each other. Prove that the probability of having exactly  $r$  matches is given by

$$\frac{1}{r!} \sum_{k=0}^{n-r} \frac{(-1)^k}{k!}, r = 0, 1, 2, \dots, n$$

Prove further that the expected number of matches and its variance are equal and are independent of  $n$ .

29. (a) If  $X$  and  $Y$  are two independent random variables, such that  $E(X) = \lambda_1$ ,  $V(X) = \sigma_1^2$  and  $E(Y) = \lambda_2$ ,  $V(Y) = \sigma_2^2$ , then prove that

$$V(X+Y) = \sigma_1^2 \sigma_2^2 + \lambda_1^2 \sigma_2^2 + \lambda_2^2 \sigma_1^2 \quad [\text{Gorakhpur Univ. B.Sc., 1992}]$$

(b) If  $X$  and  $Y$  are two independent random variables, show that

$$\frac{V(XY)}{[E(X)]^2 [E(Y)]^2} = C_X^2 C_Y^2 + C_X^2 + C_Y^2$$

where

$$C_X = \frac{\sqrt{V(X)}}{E(X)}, C_Y = \frac{\sqrt{V(Y)}}{E(Y)}$$

are the so-called coefficients of variation of  $X$  and  $Y$ ? [Patna Univ. B.Sc., 1991]

30. A point  $P$  is taken at random in a line  $AB$  of length  $2a$ , all positions of the point being equally likely. Show that the expected value of the area of the rectangle  $AP \cdot PB$  is  $2a^2/3$  and the probability of the area exceeding  $1/2a^2$  is  $1/\sqrt{2}$ . [Delhi Univ. B.Sc. (Maths Hons.), 1986]

31. If  $X$  is a random variable with  $E(X) = \mu$  satisfying  $P(X \leq 0) = 0$ , show that  $P(X > 2\mu) \leq 1/2$ . [Delhi Univ. B.Sc. (Maths Hons.), 1992]

### OBJECTIVE TYPE QUESTIONS

1. Fill in the blanks :

(i) Expected value of a random variable  $X$  exists if .....

(ii) If  $E(X')$  exists then  $E(X')$  also exists for .....

(iii) When  $X$  is a random variable, expectation of  $(X - \text{constant})^2$  is mini-

num when the constant is ....

- (iv)  $E|X - A|$  is minimum when  $A = \dots$
- (v)  $\text{Var}(c) = \dots$ , where  $c$  is a constant
- (vi)  $\text{Var}(X + c) = \dots$ , where  $c$  is a constant
- (vii)  $\text{Var}(aX + b) = \dots$ , where  $a$  and  $b$  are constants.
- (viii) If  $X$  is a r.v. with mean  $\mu$  and variance  $\sigma^2$  then

$$E\left(\frac{X - \mu}{\sigma}\right) = \dots, \text{Var}\left(\frac{X - \mu}{\sigma}\right) = \dots$$

- (ix)  $[E(XY)]^2 \dots E(X^2) \cdot E(Y^2)$ .

- (x)  $V(aX \pm bY) = \dots$

where  $a$  and  $b$  are constants.

## II. Mark the correct answer in the following :

- (i) For two random variables  $X$  and  $Y$ , the relation

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

holds good

- (a) if  $X$  and  $Y$  are statistically independent,
- (b) for all  $X$  and  $Y$ ,
- (c) if  $X$  and  $Y$  are identical.
- (ii)  $\text{Var}(2X \pm 3)$  is
  - (a) 5
  - (b) 13
  - (c) 4, if  $\text{Var } X = 1$ .
- (iii)  $E(X - k)^2$  is minimum when
  - (a)  $k < E(X)$ ,
  - (b)  $k > E(X)$ ,
  - (c)  $k = E(X)$ .

## III. Comment on the following :

If  $X$  and  $Y$  are mutually independent variables, then

- (i)  $E(XY + Y + 1) = E(X + 1)E(Y) = 0$
- (ii)  $X$  and  $Y$  are independent if and only if

$$\text{Cov}(X, Y) = 0$$

- (iii) For every univariable distribution :
  - (a)  $V(cX) = c^2V(X)$
  - (b)  $E(c/X) = c/E(X)$
- (iv) Expected value of a r.v. always exists.

## IV. Mark true or false with reasons for your answers :

- (a)  $\text{Cov}(X, Y) = 0 \Rightarrow X$  and  $Y$  are independent.
- (b) If  $\text{Var}(X) > \text{Var}(Y)$ , then  $X + Y$  and  $X - Y$  are dependent.

- (c) If  $\text{Var}(X) = \text{Var}(Y)$  and if  $2X + Y$  and  $X - Y$  are independent, then  $X$  and  $Y$  are dependent.

- (d) If  $\text{Cov}(aX + bY, bX + aY) = ab \text{Var}(X + Y)$ , then  $X$  and  $Y$  are dependent.

**6.8. Moments of Bivariate Probability Distributions.** The mathematical expectation of a function  $g(x, y)$  of two-dimensional random variable  $(X, Y)$  with

p.d.f.  $f(x, y)$  is given by

$$E[g(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f(x, y) dx dy \quad \dots(6.43)$$

(If  $X$  and  $Y$  are continuous variables)

$$= \sum_i \sum_j x_i y_j P(X = x_i \cap Y = y_j), \quad \dots(6.43a)$$

(If  $X$  and  $Y$  are discrete variables)

provided the expectation exists.

In particular, the  $r$ th and  $s$ th product moment about origin of the random variables  $X$  and  $Y$  respectively is defined as

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

$$\text{or } \mu_{rs}' = \sum_i \sum_j x_i' y_j' P(X = x_i \cap Y = y_j) \quad \dots(6.44)$$

The joint  $r$ th central moment of  $X$  and  $s$ th central moment of  $Y$  is given by

$$\begin{aligned} \mu_{rs} &= E[(X - E(X))' (Y - E(Y))] \\ &= E[(X - \mu_X)' (Y - \mu_Y)], \quad [E(X) = \mu_X, E(Y) = \mu_Y] \end{aligned} \quad \dots(6.45)$$

In particular

$$\mu_{00}' = 1 = \mu_{00}, \quad \mu_{10}' = 0 = \mu_{01}$$

$$\mu_{10}' = E(X), \quad \mu_{01}' = E(Y)$$

$$\mu_{20} = \sigma_X^2, \quad \mu_{02} = \sigma_Y^2 \text{ and } \mu_{11} = \text{Cov}(X, Y).$$

### 6.9. Conditional Expectation and Conditional Variance.

**Discrete Case.** The conditional expectation or mean value of a continuous function  $g(X, Y)$  given that  $Y = y_j$ , is defined by

$$\begin{aligned} E[g(X, Y) | Y = y_j] &= \frac{\sum_{i=1}^n g(x_i, y_j) P(X = x_i | Y = y_j)}{\sum_{i=1}^n P(Y = y_j)} \\ &= \frac{\sum_{i=1}^n g(x_i, y_j) P(X = x_i \cap Y = y_j)}{P(Y = y_j)} \end{aligned} \quad \dots(6.46)$$

i.e.,  $E[g(X, Y) | Y = y_j]$  is nothing but the expectation of the function  $g(X, y_j)$  of  $X$  w.r.t. the conditional distribution of  $X$  when  $Y = y_j$ .

In particular, the conditional expectation of a discrete random variable  $X$  given  $Y = y_j$  is

$$E(X | Y = y_j) = \sum_{i=1}^n x_i P(X = x_i | Y = y_j) \quad \dots(6.47)$$

The conditional variance of  $X$  given  $Y = y_j$  is likewise given by

$$V(X | Y = y_j) = E[(X - E(X | Y = y_j))^2 | Y = y_j] \quad \dots(6.47a)$$

The conditional expectation of  $g(X, Y)$  and the conditional variance of  $Y$  given  $X = x_i$  may also be defined in an exactly similar manner.

**Continuous Case.** The conditional expectation of  $g(X, Y)$  on the hypothesis  $Y = y$  is defined by

$$\begin{aligned} E\{g(X, Y) | Y = y\} &= \int_{-\infty}^{\infty} g(x, y) f_{X|Y}(x | y) dx \\ &= \frac{\int_{-\infty}^{\infty} g(x, y) f(x, y) dx}{f_Y(y)} \end{aligned} \quad \dots(6-48)$$

In particular, the conditional mean of  $X$  given  $Y = y$  is defined by

$$E(X | Y = y) = \frac{\int_{-\infty}^{\infty} x f(x, y) dx}{f_Y(y)}$$

Similarly, we define

$$E(Y | X = x) = \frac{\int_{-\infty}^{\infty} y f(x, y) dy}{f_X(x)} \quad \dots(6-48 a)$$

The conditional variance of  $X$  may be defined as

$$V(X | Y = y) = E[(X - E(X | Y = y))^2 | Y = y]$$

Similarly, we define

$$V(Y | X = x) = E[(Y - E(Y | X = x))^2 | X = x] \quad \dots(6-49)$$

**Theorem 6-13.** The expected value of  $X$  is equal to the expectation of the conditional expectation of  $X$  given  $Y$ . Symbolically,

$$E(X) = E[E(X | Y)] \quad \dots(6-50)$$

[Calicut Univ. B.Sc. (Main Stat.), 1980]

**Proof.**  $E[E(X | Y)] = E[\sum_i x_i P(X = x_i | Y = y_i)]$

$$= E\left[\sum_i x_i \frac{P(X = x_i \cap Y = y_i)}{P(Y = y_i)}\right]$$

$$= \sum_j \left[ \sum_i \left\{ x_i \frac{P(X = x_i \cap Y = y_i)}{P(Y = y_i)} \right\} \right] P(Y = y_j)$$

$$= \sum_j \sum_i x_i P(X = x_i \cap Y = y_i)$$

$$= \sum_i \left[ x_i \left\{ \sum_j P(X = x_i \cap Y = y_i) \right\} \right]$$

$$= \sum_i x_i P(X = x_i) = E(X).$$

**Theorem 6.14.** The variance of  $X$  can be regarded as consisting of two parts, the expectation of the conditional variance and the variance of the conditional expectation. Symbolically,

$$V(X) = E[V(X|Y)] + V[E(X|Y)] \quad \dots(6.51)$$

**Proof.**  $E[V(X|Y)] + V[E(X|Y)]$

$$\begin{aligned} &= E\left[E(X^2|Y) - [E(X|Y)]^2\right] \\ &\quad + E[(E(X|Y))^2] - [E(E(X|Y))]^2 \\ &= E[E(X^2|Y)] - E[(E(X|Y))^2] \\ &\quad + E[(E(X|Y))^2] - [E(E(X|Y))]^2 \\ &= E[E(X^2|Y)] - [E(X)]^2 \quad (\text{c.f. Theorem 6.13}) \\ &= E[\sum_i x_i^2 P(X=x_i|Y=y_i)] - [E(X)]^2 \\ &= E\left[\sum_i x_i^2 \frac{P(X=x_i \cap Y=y_i)}{P(Y=y_i)}\right] - [E(X)]^2 \\ &= \sum_j \left\{ \left[ \sum_i x_i^2 \frac{P(X=x_i \cap Y=y_j)}{P(Y=y_j)} \right] P(Y=y_j) \right\} - [E(X)]^2 \\ &= \sum_i x_i^2 \sum_j P(X=x_i \cap Y=y_j) - [E(X)]^2 \\ &= \sum_i x_i^2 P(X=x_i) - [E(X)]^2 \\ &= E(X^2) - [E(X)]^2 = \text{Var}(X) \end{aligned}$$

Hence the theorem.

**Remarks** The proofs of Theorems 6.13 and 6.14 for continuous r.v.'s  $X$  and  $Y$  are left as an exercise to the reader.

**Theorem 6.15.** Let  $A$  and  $B$  be two mutually exclusive events, then

$$E(X|A \cup B) = \frac{P(A)E(X|A) + P(B)E(X|B)}{P(A \cup B)} \quad \dots(6.52)$$

where by def.,

$$E(X|A) = \frac{1}{P(A)} \sum_{x_i \in A} x_i P(X=x_i)$$

$$\text{Proof. } E(X|A \cup B) = \frac{1}{P(A \cup B)} \sum_{x_i \in A \cup B} x_i P(X=x_i)$$

Since  $A$  and  $B$  are mutually exclusive events,

$$\sum_{x_i \in A \cup B} x_i P(X=x_i) = \sum_{x_i \in A} x_i P(X=x_i) + \sum_{x_i \in B} x_i P(X=x_i)$$

$$\therefore E(X|A \cup B) = \frac{1}{P(A \cup B)} [P(A)E(X|A) + P(B)E(X|B)]$$

$$\text{Cor. } E(X) = P(A)E(X|A) + P(\bar{A})E(X|\bar{A}) \quad \dots(6.53)$$

The corollary follows by putting  $B = \bar{A}$  in the above Theorem.

**Example 6.31.** Two ideal dice are thrown. Let  $X_1$  be the score on the first die and  $X_2$  the score on the second die. Let  $Y$  denote the maximum of  $X_1$  and  $X_2$ , i.e.,  $Y = \max(X_1, X_2)$ .

(i) Write down the joint distribution of  $Y$  and  $X_1$ ,

(ii) Find the mean and variance of  $Y$  and co-variance ( $Y, X_1$ ).

**Solution.** Each of the random variables  $X_1$  and  $X_2$  can take six values 1, 2, 3, 4, 5, 6 each with probability  $1/6$ , i.e.,

$$P(X_1 = i) = P(X_2 = i) = 1/6 ; \quad i = 1, 2, 3, 4, 5, 6 \quad \dots(i)$$

$$Y = \text{Max}(X_1, X_2).$$

Obviously

$$P(X_1 = i, Y = j) = 0, \text{ if } j < i = 1, 2, \dots, 6$$

$$\begin{aligned} P(X_1 = i, Y = i) &= P(X_1 = i, X_2 \leq i) = \sum_{j=1}^i P(X_1 = i, X_2 = j) \\ &= \sum_{j=1}^i P(X_1 = i) P(X_2 = j) \quad (\because X_1, X_2 \text{ are independent.}) \\ &= \sum_{j=1}^i \left(\frac{1}{36}\right) = \frac{i}{36}; \quad i = 1, 2, \dots, 6. \end{aligned}$$

$$P(X_1 = i, Y = j) = P(X_1 = i, X_2 = j) ; \quad j > i$$

$$= P(X_1 = i) P(X_2 = j) = \frac{1}{36}; \quad j > i = 1, 2, \dots, 6.$$

The joint probability table of  $X_1$  and  $Y$  is given as follows:

$X_1 \backslash Y$	1	2	3	4	5	6	Marginal Totals
1	1/36	1/36	1/36	1/36	1/36	1/36	6/36
2	0	2/36	1/36	1/36	1/36	1/36	6/36
3	0	0	3/36	1/36	1/36	1/36	6/36
4	0	0	0	4/36	1/36	1/36	6/36
5	0	0	0	0	5/36	1/36	6/36
6	0	0	0	0	0	6/36	6/36
Marginal Totals	1/36	3/36	5/36	7/36	9/36	11/36	1

$$E(Y) = 1 \cdot \frac{1}{36} + 2 \cdot \frac{3}{36} + 3 \cdot \frac{5}{36} + 4 \cdot \frac{7}{36} + 5 \cdot \frac{9}{36} + 6 \cdot \frac{11}{36}$$

$$= \frac{1}{36} [1 + 6 + 15 + 28 + 45 + 66] = \frac{161}{36}$$

$$E(Y^2) = 1^2 \cdot \frac{1}{36} + 2^2 \cdot \frac{3}{36} + 3^2 \cdot \frac{5}{36} + 4^2 \cdot \frac{7}{36} + 5^2 \cdot \frac{9}{36} + 6^2 \cdot \frac{11}{36} = \frac{791}{36}$$

$$V(Y) = E(Y^2) - [E(Y)]^2 = \frac{791}{36} - \left(\frac{161}{36}\right)^2 = \frac{2555}{1296}$$

$$E(X_1) = \frac{6}{36} [1 + 2 + 3 + 4 + 5 + 6] = \frac{126}{36} = \frac{21}{6}$$

$$E(X_1 Y) = 1 \cdot \frac{1}{36} + 2 \cdot \frac{1}{36} + 3 \cdot \frac{1}{36} + 4 \cdot \frac{1}{36} + 5 \cdot \frac{1}{36} + 6 \cdot \frac{1}{36}$$

$$+ 4 \cdot \frac{2}{36} + 6 \cdot \frac{1}{36} + 8 \cdot \frac{1}{36} + 10 \cdot \frac{1}{36} + 12 \cdot \frac{1}{36}$$

$$+ 9 \cdot \frac{3}{36} + 12 \cdot \frac{1}{36} + 15 \cdot \frac{1}{36} + 18 \cdot \frac{1}{36}$$

$$+ 16 \cdot \frac{4}{36} + 20 \cdot \frac{1}{36} + 24 \cdot \frac{1}{36}$$

$$+ 25 \cdot \frac{5}{36} + 30 \cdot \frac{1}{36} + 36 \cdot \frac{6}{36}$$

$$= \frac{1}{36} [21 + 44 + 72 + 108 + 155 + 216] = \frac{1}{36} \times 616$$

$$\text{Cov}(X_1, Y) = E(X_1 Y) - E(X_1)E(Y)$$

$$= \frac{616}{36} - \frac{21}{6} \cdot \frac{161}{36} = \frac{3696 - 3381}{216} = \frac{315}{216}.$$

**Example 6-32.** Let  $X$  and  $Y$  be two random variables each taking three values  $-1, 0$  and  $1$ , and having the joint probability distribution :

(i) Show that  $X$  and  $Y$  have different expectations.

$X \backslash Y$	-1	0	1	Total
-1	0	1	1	2
0	2	2	2	6
1	0	1	1	2
Total	2	4	4	10

(ii) Prove that  $X$  and  $Y$  are uncorrelated.

(iii) Find  $\text{Var } X$  and  $\text{Var } Y$ .

(iv) Given that  $Y = 0$ , what is the conditional probability distribution of  $X$ ?

(v) Find  $V(Y|X = -1)$ .

**Solution.** (i)  $E(Y) = \sum p_i y_i = -1(2) + 0(6) + 1(2) = 0$

$$E(X) = \sum p_i x_i = -1(2) + 0(4) + 1(4) = 2$$

$$\therefore E(X) \neq E(Y)$$

$$(ii) E(XY) = \sum p_{ij} y_i x_j$$

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

$$+ 0(-1)(2) + 0(0)(2) + 0(1)(2)$$

$$+ 1(-1)(0) + 1(0)(1) + 1(1)(1)$$

$$= -0 \cdot 1 + 0 \cdot 1 = 0$$

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

$\Rightarrow X$  and  $Y$  are uncorrelated (c.f. § 10.5)

$$\begin{aligned}
 (iii) \quad E(Y^2) &= (-1)^2(2) + 0(4) + 1^2(4) = -4 \\
 \therefore V(Y) &= E(Y^2) - [E(Y)]^2 = -4 \\
 E(X^2) &= (-1)^2(2) + 0(4) + 1^2(4) = -2 + -4 = -6 \\
 V(X) &= -6 - -4 = -56 \\
 (iv) \quad P(X = -1 | Y = 0) &= \frac{P(X = -1 \cap Y = 0)}{P(Y = 0)} = \frac{-2}{-6} = \frac{1}{3} \\
 P(X = 0 | Y = 0) &= \frac{P(X = 0 \cap Y = 0)}{P(Y = 0)} = \frac{-2}{-6} = \frac{1}{3} \\
 P(X = 1 | Y = 0) &= \frac{P(X = 1 \cap Y = 0)}{P(Y = 0)} = \frac{-2}{-6} = \frac{1}{3} \\
 (v) \quad V(Y | X = -1) &= E(Y | X = -1)^2 - [E(Y | X = -1)]^2 \\
 E(Y | X = -1) &= \sum y P(Y = y | X = -1) = (-1)0 + 0(2) + 1(0) = 0 \\
 E(Y | X = -1)^2 &= \sum y^2 P(Y = y | X = -1) = 1(0) + 0(2) + 1(0) = 0 \\
 \therefore V(Y | X = -1) &= 0.
 \end{aligned}$$

**Example 6-33.** Two tetrahedra with sides numbered 1 to 4 are tossed. Let  $X$  denote the number on the downturned face of the first tetrahedron and  $Y$  denote the larger of the downturned numbers. Investigate the following :

- (a) Joint density function of  $X$ ,  $Y$  and marginals  $f_X$  and  $f_Y$ ,
- (b)  $P(X \leq 2, Y \leq 3)$ ,
- (c)  $p(X, Y)$ ,
- (d)  $E(Y | X = 2)$ ,
- (e) Construct joint density different from that in part (a) but possessing same marginals  $f_X$  and  $f_Y$ . [Delhi Univ. B.A. (Stat. Hons.), Spl. Course, 1985]

**Hint.** The sample space is  $S = \{1, 2, 3, 4\} \times \{1, 2, 3, 4\}$  and each of the 16 sample points (outcomes) has probability  $p = \frac{1}{16}$  of occurrence.

Let  $X$ : Number on the first dice and  $Y$ : Larger of the numbers on the two dice. Then the above 16 sample points, in that order, give the following distribution of  $X$  and  $Y$ .

Sample Point	:	(1, 1)	(1, 2)	(1, 3)	(1, 4)	(2, 1)	(2, 2)	(2, 3)	(2, 4)
$X$	:	1	1	1	1	2	2	2	2
$Y$	:	1	2	3	4	2	2	3	4
Sample Point	:	(3, 1)	(3, 2)	(3, 3)	(3, 4)	(4, 1)	(4, 2)	(4, 3)	(4, 4)
$X$	:	3	3	3	3	4	4	4	4
$Y$	:	3	3	3	4	4	4	4	4

Since each sample point has probability  $p = \frac{1}{16}$ , the joint density functions of  $X$  and  $Y$  and the marginal densities  $f_X$  and  $f_Y$  are given on page 6-61.

Here  $p = \frac{1}{16}$ .

$$(b) \quad P(X \leq 2, Y \leq 3) = p + p + 2p + p + p = 6p = \frac{3}{8}.$$

$$(c) \quad \text{Var}(X) = EX^2 - [E(X)]^2 = \frac{15}{2} - \frac{25}{4} = \frac{5}{4} \quad (\text{Try it})$$

	(a) $x$				Total ( $f_x$ )		(e) $x$				Total ( $f_y$ )
	1	2	3	4			1	2	3	4	
1	p	0	0	0	p	I	p	0	0	0	p
2	p	2p	0	0	3p	2	p	2p	0	0	3p
y	3	p	p	3p	0	3	p	p+ε	3p-ε	0	5p
4	p	p	p	4p	7p	4	p	p-ε	p+ε	4p	7p
Total ( $f_y$ )	4p	4p	4p	4p	16p=1	Total ( $f_x$ )	4p	4p	4p	4p	1

$$\text{Var}(Y) = E(Y^2) - [E(Y)]^2 = \frac{85}{8} - \left(\frac{25}{8}\right)^2 = \frac{55}{64} \quad (\text{Try it})$$

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y) = \frac{135}{16} - \frac{5}{2} \times \frac{25}{8} = \frac{5}{8} \quad (\text{Try it})$$

$$\therefore \rho(X, Y) = \frac{\frac{5}{8}}{\sqrt{5/4 \times 55/64}} = \frac{2}{\sqrt{11}}$$

$$(d) E(Y|X=2) = \sum y \cdot f(y|x=2) = \sum y \cdot \frac{f(x=2 \cap y)}{f(x=2)}$$

$$= 4 \cdot \sum y f(2, y) = 4 [0 + 4p + 3p + 4p] = 44p = \frac{11}{4}$$

(e) Let  $0 < \epsilon < p$ . The joint density of  $X$  and  $Y$  given in (e) above is different from that in (a) but has the same marginals as in (a).

**Example 6.34.** (a) Given two variates  $X_1$  and  $X_2$ , with joint density function  $f(x_1, x_2)$ , prove that conditional mean of  $X_2$  (given  $X_1$ ) coincides with (unconditional) mean only if  $X_1$  and  $X_2$  are independent (stochastically).

(b) Let  $f(x_1, x_2) = 21x_1^2 x_2^3$ ,  $0 < x_1 < x_2 < 1$ , and zero elsewhere be the joint p.d.f. of  $X_1$  and  $X_2$ . Find the conditional mean and variance of  $X_1$  given  $X_2 = x_2$ ,  $0 < x_2 < 1$ . [Delhi Univ. M.A. (Eco.), 1986]

**Solution.** (a) Conditional mean of  $X_2$  given  $X_1$  is given by :

$$E(X_2 | X_1 = x_1) = \int_{x_2} x_2 f(x_2 | x_1) dx_2 \quad ...(*)$$

where  $f(x_2 | x_1)$  is conditional p.d.f. of  $X_2$  given  $X_1 = x_1$ .

But the joint p.d.f. of  $X_1$  and  $X_2$  is given by

$$f(x_1, x_2) = f_1(x_1) \cdot f(x_2 | x_1)$$

$$\Rightarrow f(x_2 | x_1) = \frac{f(x_1, x_2)}{f_1(x_1)}$$

where  $f_1(\cdot)$  is marginal p.d.f. of  $X_1$ .

Substituting in (\*), we get

$$E(X_2 | X_1 = x_1) = \int_{x_2} \left[ \frac{x_2 f(x_1, x_2)}{f_1(x_1)} \right] dx_2, \quad \dots (**)$$

Unconditional mean of  $X_2$  is given by

$$E(X_2) = \int_{x_2} x_2 f_2(x_2) dx_2, \quad \dots (***)$$

where  $f_2(\cdot)$  is marginal p.d.f. of  $X_2$ .

From (\*\*) and (\*\*\*) we conclude that the conditional mean of  $X_2$  (given  $X_1$ ) will coincide with unconditional mean of  $X_2$  only if

$$\frac{f(x_1, x_2)}{f_1(x_1)} = f_2(x_2)$$

$$\Rightarrow f(x_1, x_2) = f_1(x_1) \cdot f_2(x_2)$$

i.e., if  $X_1$  and  $X_2$  are (stochastically) independent.

$$(b) \quad f(x_1, x_2) = 21 x_1^2 x_2^3; \quad 0 < x_1 < x_2 < 1 \\ = 0, \quad \text{otherwise}$$

Marginal p.d.f. of  $X_2$  is given by

$$f_2(x_2) = \int_0^{x_2} f(x_1, x_2) dx_1 = 21 x_2^3 \int_0^{x_2} x_1^2 dx_1 \\ = 21 x_2^3 \cdot \left| \frac{x_1^3}{3} \right|_0^{x_2} = 7 x_2^6; \quad 0 < x_2 < 1$$

$\therefore$  Conditional p.d.f. of  $X_1$  (given  $X_2$ ) is given by

$$f_1(x_1 | x_2) = \frac{f(x_1, x_2)}{f_2(x_2)} = \frac{3}{x_2^3} \cdot \frac{x_1^2}{x_2^3}; \quad 0 < x_1 < x_2; \quad 0 < x_2 < 1$$

Conditional mean of  $X_1$  is

$$E(X_1 | X_2 = x_2) = \int_0^{x_2} x_1 f_1(x_1 | x_2) dx_1 = \frac{3}{x_2^3} \int_0^{x_2} x_1^4 dx_1 \\ = \frac{3}{x_2^3} \cdot \left| \frac{x_1^5}{5} \right|_0^{x_2} = \frac{3x_2^2}{4}; \quad 0 < x_2 < 1$$

Now

$$E(X_1^2 | X_2 = x_2) = \int_0^{x_2} x_1^2 f_1(x_1 | x_2) dx_1 = \frac{3}{x_2^3} \int_0^{x_2} x_1^4 dx_1 \\ = \frac{3}{x_2^3} \cdot \frac{x_2^5}{5} = \frac{3}{5} x_2^2$$

$$\therefore \text{Var}(X_1 | X_2 = x_2) = E(X_1^2 | X_2 = x_2) - [E(X_1 | X_2 = x_2)]^2 \\ = \frac{3}{5} x_2^2 - \frac{9}{16} x_2^2 = \frac{3}{80} x_2^2; \quad 0 < x_2 < 1.$$

**Example 6.35.** Two random variables  $X$  and  $Y$  have the following joint probability density function :

$$f(x, y) = 2 - x - y; \quad 0 \leq x \leq 1, \quad 0 \leq y \leq 1 \\ = 0, \text{ otherwise}$$

- Find*
- Marginal probability density functions of X and Y.*
  - Conditional density functions.*
  - Var (X) and Var (Y).*
  - Co-variance between X and Y.*

[Dibrugarh Univ. B.Sc. (Hons.), 1991]

**Solution.** (i)  $f_X(x) = \int_{-\infty}^{+\infty} f(x, y) dy$

$$= \int_0^1 (2 - x - y) dy = \frac{3}{2} - x$$

$$\therefore f_X(x) = \begin{cases} \frac{3}{2} - x, & 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}$$

Similarly  $f_Y(y) = \begin{cases} \frac{3}{2} - y, & 0 < y < 1 \\ 0, & \text{otherwise} \end{cases}$

(ii)  $f_{X|Y}(x | y) = \frac{f_{XY}(x, y)}{f_Y(y)} = \frac{(2 - x - y)}{\left(\frac{3}{2} - y\right)}, \quad 0 < (x, y) < 1$

$$f_{Y|X}(y | x) = \frac{f_{XY}(x, y)}{f_X(x)} = \frac{(2 - x - y)}{\left(\frac{3}{2} - x\right)}, \quad 0 < (x, y) < 1$$

$$E(X) = \int_0^1 x f_X(x) dx = \int_0^1 x \left( \frac{3}{2} - x \right) dx = \frac{5}{12}$$

$$E(Y) = \int_0^1 y f_Y(y) dy = \int_0^1 y \left( \frac{3}{2} - y \right) dy = \frac{5}{12}$$

(iii)  $E(X^2) = \int_0^1 x^2 \left( \frac{3}{2} - x \right) dx = \left[ \frac{3}{6} x^3 - \frac{x^4}{4} \right]_0^1 = \frac{1}{4}$

$$V(X) = E(X^2) - [E(X)]^2 = \frac{1}{4} - \frac{25}{144} = \frac{11}{144}$$

Similarly  $V(Y) = \frac{11}{144}$

(iv)  $E(XY) = \int_0^1 \int_0^1 xy (2 - x - y) dx dy$

$$= \int_0^1 \left| 2 \frac{x^2 y}{2} - \frac{x^3 y}{3} - \frac{x^2 y^2}{2} \right|_0^1 dy$$

$$= \int_0^1 \left( \frac{2}{3} y - \frac{1}{2} y^2 \right) dy$$

$$= \left| \frac{y^2}{3} - \frac{y^3}{6} \right|_0^1 = \frac{1}{6}$$

$$= \left| \begin{array}{cc} \frac{y^2}{3} - \frac{y^3}{6} & 1 \\ 0 & 0 \end{array} \right| = \frac{1}{6}$$

$$\therefore \text{Cov}(X, Y) = E(XY) - E(X)E(Y) = \frac{1}{6} - \frac{5}{12} \cdot \frac{5}{12} = -\frac{1}{144}.$$

**Example 6.36.** Let  $f(x, y) = 8xy$ ,  $0 < x < y < 1$ ;  $f(x, y) = 0$ , elsewhere. Find  
(a)  $E(Y|X=x)$ , (b)  $E(XY|X=x)$ , (c)  $\text{Var}(Y|X=x)$ . [Calcutta Univ.  
B.Sc. (Maths Hons.), 1988; Delhi Univ. B.Sc. (Maths Hons.), 1990]

**Solution.**  $f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy$

$$= 8x \int_x^1 y dy$$

$$= 4x(1-x^2), \quad 0 < x < 1$$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx$$

$$= 8y \int_0^y x dx$$

$$= 4y^3, \quad 0 < y < 1$$

$$f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)} = \frac{2x}{y^2}, \quad f_{Y|X}(y|x) = \frac{2y}{1-x^2}, \quad 0 < x < y < 1.$$

$$(a) \quad E(Y|X=x) = \int_x^1 y \left( \frac{2y}{1-x^2} \right) dy = \frac{2}{3} \left( \frac{1-x^4}{1-x^2} \right) = \frac{2}{3} \left( \frac{1+x+x^2}{1+x} \right)$$

$$(b) \quad E(XY|X=x) = x E(Y|X=x) = \frac{2}{3} \cdot \frac{x(1+x+x^2)}{(1+x)}$$

$$(c) \quad E(Y^2|X=x) = \int_x^1 y^2 \left( \frac{2y}{1-x^2} \right) dy = \frac{1}{2} \left( \frac{1-x^4}{1-x^2} \right) = \frac{1+x^2}{2}$$

$$\text{Var}(Y|X=x) = E(Y^2|X=x) - [E(Y|X=x)]^2$$

$$= \frac{1+x^2}{2} - \frac{4}{9} \cdot \frac{(1+x+x^2)^2}{(1+x)^2}$$

### EXERCISE 6(b)

1. The joint probability distribution of  $X$  and  $Y$  is given by the following table:

		1	3	9
X	Y			
	2	1/8	1/24	1/12
	4	1/4	1/4	0
	6	1/8	1/24	1/12

- (i) Find the marginal probability distribution of  $Y$ .

- (ii) Find the conditional distribution of  $Y$  given that  $X = 2$ ,  
 (iii) Find the covariance of  $X$  and  $Y$ , and  
 (iv) Are  $X$  and  $Y$  independent?

2. A fair coin is tossed four times. Let  $X$  denote the number of heads occurring and let  $Y$  denote the longest string of heads occurring.

(i) Determine the joint distribution of  $X$  and  $Y$ , and (ii) Find  $\text{Cov}(X, Y)$ .

**Hint.**

$X \backslash Y$	0	1	2	3	4	Total
$X$	1/16	0	0	0	0	1/16
0	0	4/16	0	0	0	4/16
1	0	3/16	3/16	0	0	6/16
2	0	0	2/16	2/16	0	4/16
3	0	0	0	0	1/16	1/16
4	0	0	0	0	1/16	1/16
Total	1/16	7/16	5/16	2/16	1/16	1

$$(ii) \text{Cov}(X, Y) = 0.85.$$

3.  $X$  and  $Y$  are jointly discrete random variables with probability function

$$p(x, y) = 1/4 \text{ at } (x, y) = (-3, -5), (-1, -1), (1, 1), (3, 5) \\ = 0, \text{ otherwise}$$

Compute  $E(X)$ ,  $E(Y)$ ,  $E(XY)$  and  $E(X|Y)$ . Are  $X$  and  $Y$  independent?

4.  $X_1$  and  $X_2$  have a bivariate distribution given by

$$P(X_1 = x_1 \cap X_2 = x_2) = \frac{x_1 + 3x_2}{24}, \text{ where } (x_1, x_2) = (1, 1), (1, 2), (2, 1), (2, 2)$$

Find the conditional mean and variance of  $X_1$ , given  $X_2 = 2$ .

5. Two random variables  $X$  and  $Y$  have the following joint probability density function :

$$f(x, y) = k(4 - x - y); 0 \leq x \leq 2; 0 \leq y \leq 2 \\ = 0, \text{ otherwise}$$

Find (i) the constant  $k$ ,

(ii) marginal density functions of  $X$  and  $Y$ ,

(iii) conditional density functions, and

(iv)  $\text{Var}(X)$ ,  $\text{Var}(Y)$  and  $\text{Cov}(X, Y)$ . (Poona Univ. B.Sc., Oct. 1991)

6. Let the joint probability density function of the random variables  $X$  and  $Y$  be

$$f(x, y) = 2(x + y - 3xy^2); 0 < x < 1, 0 < y < 1 \\ = 0, \text{ otherwise}$$

(i) Find the marginal distributions of  $X$  and  $Y$ .

(ii) Is  $E(XY) = E(X)E(Y)$ ?

(iii) Find  $E(X+Y)$  and  $E(X-Y)$ . [Calicut Univ. B.Sc., Oct. 1990]

7. (a) Let  $X$  and  $Y$  have the joint probability density function

$$f(x, y) = 2, \quad 0 < x < y < 1 \\ = 0, \text{ otherwise}$$

Show that the conditional mean and variance of  $X$  given  $Y = y$  are  $y/2$  and  $y^2/12$  respectively.

(b) If  $f(x, y) = 2$ ;  $0 < x < y, 0 < y < 1$

Find (i)  $E(Y|X)$ , (ii)  $E(X|Y)$ .

8. Give an example to show that  $E(Y)$  may not exist though  $E(XY)$  and  $E(Y|X)$  may both exist? [Delhi Univ. B.A. (Stat. Hons.) Spl. Course, 1985]

**Hint.** Consider the joint p.d.f. :

$$f(x, y) = x \cdot e^{-x(1+y)}; \quad x \geq 0, y \geq 0 \\ = 0, \text{ otherwise.}$$

Then we shall get :

$$f_X(x) = \int_0^\infty f(x, y) dy = e^{-x}; \quad x \geq 0$$

$$f_Y(y) = \int_0^\infty f(x, y) dx = \frac{1}{(1+y)^2}; \quad y \geq 0$$

$$f(Y|X) = \frac{f(x, y)}{f_X(x)} = x e^{-xy}; \quad y \geq 0$$

$$\therefore E(Y) = \int_0^\infty y f(y) dy = \infty \Rightarrow E(Y) \text{ does not exist.}$$

$$E(XY) = \int_0^\infty \int_0^\infty xy \cdot f(x, y) dx dy$$

$$E(Y|X=x) = \int_0^\infty y \cdot f(y|x) dy = \frac{1}{x}$$

$\Rightarrow$  Both  $E(XY)$  and  $(E(Y|X=x))$  exist, though  $E(Y)$  does not exist.

9. Three coins are tossed. Let  $X$  denote the number of heads on the first two coins,  $Y$  denote the number of tails on the last two and  $Z$  denote the number of heads on the last two.

(a) Find the joint distribution of (i)  $X$  and  $Y$ , (ii)  $X$  and  $Z$ .

(b) Find the conditional distribution of  $Y$  given  $X = 1$ .

(c) Find  $E(Z|X=1)$ .

(d) Find  $P_{X,Y}$  and  $P_{X,Z}$ .

(e) Give a joint distribution that is not the joint distribution of  $X$  and  $Z$  in (a) and yet has the same marginals as  $f(x, z)$  has in part (a).

[Delhi Univ. B.Sc. (Maths Hons.), 1989].

**Hint.** The sample space is  $S' = \{H, T\} \times \{H, T\} \times \{H, T\}$

$$= \{H, T\} \times \{HH, HT, TH, TT\}$$

and each of the 8 sample points (outcomes) has the probability  $p = 1/8$  of occurrence.

$X$  : Number of heads on the 1st two coins.

$Y$  : Number of tails on the last two coins.

$Z$  : Number of heads on the last two coins.

Then the distribution of  $X$ ,  $Y$  and  $Z$  is given below :

Sample Point :	HHH	HHT	HTH	HTT	THH	THT	TTH	TTT
Probability	$p$							
$X$	2	2	1	1	1	1	0	0
$Y$	0	1	1	2	0	1	1	2
$Z$	2	1	1	0	2	1	1	0

Joint Distribution of  $X$  and  $Y$

	$y$			Total
	0	1	2	$(f_y)$
$x$	0	0	1/8	1/8
	1	1/8	2/8	1/8
	2	1/8	1/8	0
Total ( $f_y$ )	1/4	1/2	1/4	1

Joint Distribution of  $X$  and  $Z$

	$z$			Total
	0	1	2	$(f_z)$
$x$	0	1/8	1/8	0
	1	1/8	2/8	1/8
	2	0	1/8	1/8
Total ( $f_z$ )	1/4	1/2	1/4	1

$$(b) P(Y=0|X=1) = \frac{P(Y=0, X=1)}{P(X=1)} = \frac{1/8}{1/2} = \frac{1}{4}$$

$$\text{Similarly, } P(Y=1|X=1) = \frac{2/8}{1/2} = \frac{1}{2}; P(Y=2|X=1) = \frac{1/8}{1/2} = \frac{1}{4}$$

$$(c) E(Z|X=1) = \sum z \cdot P(Z|X=1) = 0 \times \frac{1/8}{1/2} + 1 \times \frac{2/8}{1/2} + 2 \times \frac{1/8}{1/2} = 1$$

$$(d) \rho_{XY} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{-1/4}{\sqrt{1/2 \times 1/2}} = -\frac{1}{2},$$

$$\rho_{XZ} = \frac{\text{Cov}(X, Z)}{\sigma_X \sigma_Z} = \frac{-1/4}{\sqrt{1/2 \times 1/2}} = -\frac{1}{2}$$

(e) Let  $0 \leq \epsilon \leq 1/8$ . The joint probability distribution of  $(X, Z)$  given below has the same marginals as in part (a).

	0	$Z$	2	Total ( $f_z$ )
$X$	0	1/8	1/8	1/4
	1	1/8	2/8 + $\epsilon$	1/2
	2	0	1/8 - $\epsilon$	1/4
Total ( $f_z$ )	1/4	1/2	1/4	1

10. Let  $f_{XY}(x, y) = e^{-(x+y)}$ ;  $0 < x < \infty, 0 < y < \infty$

Find :

- |                         |                                      |
|-------------------------|--------------------------------------|
| (a) $P(X > 1)$          | (d) $m$ so that $P(X + Y < m) = 1/2$ |
| (b) $P(1 < X + Y < 2)$  | (e) $P(0 < X < 1   Y = 2)$           |
| (c) $P(X < Y   X < 2Y)$ | (f) $\rho_{XY}$                      |

Ans.  $f_X(x) = e^{-x}; x \geq 0$ ;  $f_Y(y) = e^{-y}; y \geq 0$

(a)  $1/e$  (b) Hint.  $X + Y$  is a Gamma variable with parameter  $n = 2$ .

[See Chapter 8]  $(2/e - 3/e^2)$ .

$$(c) P(X < Y | X < 2Y) = \frac{P(X < Y \cap X < 2Y)}{P(X < 2Y)} = \frac{P(X < Y)}{P(X < 2Y)} = \frac{1/2}{2/3} = \frac{3}{4}$$

(d) Use hint in (b).  $e^{-m}(1+m) = 1/2$ ; (e)  $(e-1)/e$

(f)  $f_{XY}(x, y) = f_X(x)f_Y(y) \Rightarrow X$  and  $Y$  are independent  $\Rightarrow \rho_{XY} = 0$ .

11. The joint p.d.f. of  $X$  and  $Y$  is given by :

$$f(x, y) = 3(x+y); 0 \leq x \leq 1, 0 \leq y \leq 1; 0 \leq x+y \leq 1$$

Find : (a) Marginal density of  $X$ . (b)  $P(X + Y < 1/2)$

$$(c) E(Y | X = x) (d) \text{Cov}(X, Y)$$

Ans. (a)  $f_X(x) = \frac{3}{2}(1-x^2); 0 \leq x \leq 1$ .

$$(b) P(X + Y < 1/2) = \int_0^{0.5} \left[ \int_0^{1-x} 3(x+y) dy \right] dx = \frac{1}{8}$$

$$(c) \frac{(1-x)(x+2)}{3(1+x)} (d) E(XY) = \int_0^1 \left[ \int_0^{1-x} xy f(x, y) dy dx \right] = \frac{1}{10}$$

$$\text{Cov}(x, y) = E(XY) - E(X)E(Y) = \frac{1}{10} - \frac{3}{8} \times \frac{3}{8} = -\frac{13}{320}.$$

6-10. Moment Generating Function. The moment generating function (m.g.f.) of a random variable  $X$  (about origin) having the probability function  $f(x)$  is given by

$$(6-54) \dots \begin{cases} M_X(t) = E(e^{tX}) = \int e^{tx} f(x) dx, \\ \quad \quad \quad \text{(for continuous probability distribution)} \\ \quad \quad \quad \sum e^{tx} f(x), \\ \quad \quad \quad \text{(for discrete probability distribution)} \end{cases}$$

the integration or summation being extended to the entire range of  $x$ ,  $t$  being the real parameter and it is being assumed that the right-hand side of (6-54) is absolutely convergent for some positive number  $h$  such that  $-h < t < h$ . Thus

$$\begin{aligned} M_X(t) = E(e^{tX}) &= E \left[ 1 + tX + \frac{t^2 X^2}{2!} + \dots + \frac{t^r X^r}{r!} + \dots \right] \\ &= 1 + tE(X) + \frac{t^2}{2!} E(X^2) + \dots + \frac{t^r}{r!} E(X^r) + \dots \end{aligned}$$

$$= 1 + t \mu_1' + \frac{t^2}{2!} \mu_2' + \dots + \frac{t^r}{r!} \mu_r' + \dots \quad \dots(6.55)$$

where  $\mu_r' = E(X^r) = \int x^r f(x) dx$ , for continuous distribution  
 $= \sum x^r p(x)$ , for discrete distribution.

is the  $r$ th moment of  $X$  about origin. Thus we see that the coefficient of  $\frac{t^r}{r!}$  in  $M_X(t)$  gives  $\mu_r'$  (above origin). Since  $M_X(t)$  generates moments, it is known as moment generating function.

Differentiating (6.55) w.r.t.  $t$  and then putting  $t = 0$ , we get

$$\begin{aligned} \left[ \frac{d^r}{dt^r} \{ M_X(t) \} \right]_{t=0} &= \left[ \frac{\mu_1'}{r!} \cdot r! + \mu_{r+1}' t + \mu_{r+2}' \cdot \frac{t^2}{2!} + \dots \right]_{t=0} \\ \Rightarrow \mu_r' &= \left[ \frac{d^r}{dt^r} \{ M_X(t) \} \right]_{t=0} \end{aligned} \quad \dots(6.56)$$

In general, the moment generating function of  $X$  about the point  $X = a$  is defined as

$$\begin{aligned} M_X(t) \text{ (about } X = a) &= E[e^{(X-a)t}] \\ &= E \left[ 1 + t(X-a) + \frac{t^2}{2!}(X-a)^2 + \dots + \frac{t^r}{r!}(X-a)^r + \dots \right] \\ &= 1 + t \mu_1' + \frac{t^2}{2!} \mu_2' + \dots + \frac{t^r}{r!} \mu_r' + \dots \end{aligned} \quad \dots(6.57)$$

where  $\mu_r' = E[(X-a)^r]$ , is the  $r$ th moment about the point  $X = a$ .

**6.10.1. Some Limitations of Moment Generating Functions.** Moment generating function suffers from some drawbacks which have restricted its use in Statistics. We give below some of the deficiencies of m.g.f.'s with illustrative examples.

1. A random variable  $X$  may have no moments although its m.g.f. exists. For example, let us consider a discrete r.v. with probability function

$$f(x) = \begin{cases} \frac{1}{x(x+1)} & ; x=1, 2, \dots \\ 0, & \text{otherwise} \end{cases}$$

$$\begin{aligned} \text{Here } E(X) &= \sum_{x=1}^{\infty} x f(x) = \sum_{x=1}^{\infty} \frac{1}{(x+1)} \\ &= \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \dots \\ &= \left[ \sum_{x=1}^{\infty} \frac{1}{x} \right] - 1 \end{aligned}$$

Since  $\sum_{x=1}^{\infty} \frac{1}{x}$  is a divergent series,  $E(X)$  does not exist and consequently no moment of  $X$  exists. However, the m.g.f. of  $X$  is given by

$$\begin{aligned}
 M_X(t) &= \sum_{x=1}^{\infty} e^{tx} \cdot f(x) = \sum_{x=1}^{\infty} \frac{e^{tx}}{x(x+1)} \\
 &= \sum_{x=1}^{\infty} \frac{z^x}{x(x+1)}, \quad (z = e^t) \quad \dots(*) \\
 &= \frac{z}{1.2} + \frac{z^2}{2.3} + \frac{z^3}{3.4} + \frac{z^4}{4.5} + \dots \\
 &= z \left[ 1 - \frac{1}{2} \right] + z^2 \left[ \frac{1}{2} - \frac{1}{3} \right] + z^3 \left[ \frac{1}{3} - \frac{1}{4} \right] + z^4 \left[ \frac{1}{4} - \frac{1}{5} \right] + \dots \\
 &= \left[ z + \frac{z^2}{2} + \frac{z^3}{3} + \frac{z^4}{4} + \dots \right] - \frac{z}{2} - \frac{z^2}{3} - \frac{z^3}{4} - \frac{z^4}{5} - \dots \\
 &= -\log(1-z) - \frac{1}{z} \left[ \frac{z^2}{2} + \frac{z^3}{3} + \frac{z^4}{4} + \dots \right] \\
 &= -\log(1-z) + 1 + \frac{1}{z} \log(1-z), |z| < 1 \\
 &= 1 + \left[ \frac{1}{z} - 1 \right] \log(1-z), |z| < 1 \\
 &= 1 + (e^{-t} - 1) \log(1 - e^t), t < 0
 \end{aligned}$$

\$\because |z| < 1 \Rightarrow |e^t| < 1 \Rightarrow t < 0\$]

And \$M\_X(t) = 1\$, for \$t = 0\$,

[From \$(\*)\$]

while for \$t > 0\$, \$M\_X(t)\$ does not exist.

2. A random variable \$X\$ can have m.g.f. and some (or all) moments, yet the m.g.f. does not generate the moments. For example, consider a discrete r.v. with probability function

$$P(X = 2^x) = \frac{e^{-1}}{x!}; x = 0, 1, 2, \dots$$

$$\begin{aligned}
 \text{Here } E(X') &= \sum_{x=0}^{\infty} (2^x)' P(X = 2^x) = e^{-1} \sum_{x=0}^{\infty} \frac{(2^x)^x}{x!} \\
 &= e^{-1} \cdot \exp(2^x) = \exp(2^x - 1)
 \end{aligned}$$

Hence all the moments of \$X\$ exist.

The m.g.f. of \$X\$, if it exists is given by

$$M_X(t) = \sum_{x=0}^{\infty} \exp(t \cdot 2^x) \left( \frac{e^{-1}}{x!} \right) = e^{-1} \sum_{x=0}^{\infty} \exp(t \cdot 2^x) \frac{1}{x!}$$

By D'Alembert's ratio test, the series on the R.H.S. converges for \$t \leq 0\$ and diverges for \$t > 0\$. Hence \$M\_X(t)\$ cannot be differentiated at \$t = 0\$ and has no Maclaurin's expansion and consequently it does not generate moments.

3. A r.v. \$X\$ can have all or some moments; but m.g.f. does not exist except perhaps at one point.

For example, let \$X\$ be a r.v. with probability function

$$P(X = \pm 2^x) = \frac{e^{-1}}{2x!}; x = 0, 1, 2, \dots$$

= 0, otherwise.

Since the distribution is symmetric about the line  $X = 0$ , all moments of odd order about origin vanish, i.e.,

$$\begin{aligned} E(X^{2r+1}) &= 0 \Rightarrow \mu_{2r+1} = 0 \\ E(X^{2r}) &= \sum_{x=0}^{\infty} (\pm 2^x)^{2r} \left( \frac{1}{2ex!} \right) = \frac{1}{e} \sum_{x=0}^{\infty} \frac{(2^x)^{2r}}{x!} \\ &= \frac{1}{e} \cdot \exp(2^2) = \exp(2^2 - 1) \end{aligned}$$

Thus all the moments of  $X$  exist. The m.g.f. of  $X$ , if it exists, is given by

$$\begin{aligned} M_X(t) &= \sum_{x=0}^{\infty} \left[ (e^{t \cdot 2^x} + e^{-t \cdot 2^x}) \frac{1}{2ex!} \right] \\ &= e^{-1} \sum_{x=0}^{\infty} \left[ \frac{\cos(t \cdot 2^x)}{x!} \right] \end{aligned}$$

which converges only for  $t = 0$ .

As an illustration of a continuous probability distribution, consider Pareto distribution with o.d.f.

$$p(x) = \frac{\theta \cdot a^\theta}{x^{\theta+1}}; x \geq a; \theta > 1$$

$$E(X^r) = \theta \cdot a^\theta \int_a^{\infty} x^{r-\theta-1} dx = \theta \cdot a^\theta \cdot \left[ \frac{x^{r-\theta}}{r-\theta} \right]_a^{\infty},$$

which is finite iff  $r - \theta < 0 \Rightarrow \theta > r$  and then

$$E(X^r) = \theta a^\theta \left[ 0 - \frac{a^{r-\theta}}{r-\theta} \right] = \frac{\theta \cdot a^r}{\theta - r}; \theta > r$$

However, the m.g.f. is given by:

$$M_X(t) = \theta \cdot a^\theta \int_a^{\infty} \frac{e^{tx}}{x^{\theta+1}} dx,$$

which does not exist, since  $e^{tx}$  dominates  $x^{\theta+1}$  and  $(e^{tx}/x^{\theta+1}) \rightarrow \infty$  as  $x \rightarrow \infty$  and hence the integral is not convergent.

For more illustrations see Student's t-distribution and Snedecor's F-distributions, for which m.g.f.'s do not exist, though the moments of all orders exist. [c.f. Chapter 14, § 14·2·4 and 14·5·2.]

**Remark.** The reason that m.g.f. is a poor tool in comparison with characteristic function (c.f. § 6·12) is that the domain of the dummy parameter 't' of the m.g.f. depends on the distribution of the r.v. under consideration, while characteristic function exists for all real  $t$ ,  $(-\infty < t < \infty)$ . If m.g.f. is valid for  $t$  lying in an interval containing zero, then m.g.f. can be expanded with perhaps some additional restrictions.

### 6-10-2. Theorems on Moment Generating Functions.

**Theorem 6-17.**  $M_{cX}(t) = M_X(ct)$ ,  $c$  being a constant. ... (6-58)

**Proof.** By def.,

$$\text{L.H.S.} = M_{cX}(t) = E(e^{tCX})$$

$$\text{R.H.S.} = M_X(ct) = E(e^{ctX}) = \text{L.H.S.}$$

**Theorem 6-18.** The moment generating function of the sum of a number of independent random variables is equal to the product of their respective moment generating functions.

Symbolically, if  $X_1, X_2, \dots, X_n$  are independent random variables, then the moment generating function of their sum  $X_1 + X_2 + \dots + X_n$  is given by

$$M_{X_1 + X_2 + \dots + X_n}(t) = M_{X_1}(t) M_{X_2}(t) \dots M_{X_n}(t) \quad \dots (6-59)$$

**Proof.** By definition,

$$\begin{aligned} M_{X_1 + X_2 + \dots + X_n}(t) &= E\left[e^{t(X_1 + X_2 + \dots + X_n)}\right] \\ &= E\left[e^{tX_1} \cdot e^{tX_2} \dots e^{tX_n}\right] \\ &= E(e^{tX_1}) E(e^{tX_2}) \dots E(e^{tX_n}) \\ &\quad (\because X_1, X_2, \dots, X_n \text{ are independent}) \\ &= M_{X_1}(t) M_{X_2}(t) \dots M_{X_n}(t) \end{aligned}$$

Hence the theorem.

**Theorem 6-19.** Effect of change of origin and scale on M.G.F. Let us transform  $X$  to the new variable  $U$  by changing both the origin and scale in  $X$  as follows :

$$U = \frac{X - a}{h}, \text{ where } a \text{ and } h \text{ are constants}$$

M.G.F. of  $U$  (about origin) is given by

$$\begin{aligned} M_U(t) &= E(e^{tU}) = E[\exp\{t(x-a)/h\}] \\ &= E[e^{tX/h} \cdot e^{-at/h}] = e^{-at/h} E(e^{tX/h}) \\ &= e^{-at/h} E(e^{tX/h}) = e^{-at/h} M_X(t/h) \quad \dots (6-60) \end{aligned}$$

where  $M_X(t)$  is the m.g.f. of  $X$  about origin.

In particular, if we take  $a = E(X) = \mu$  (say) and  $h = \sigma_X = \sigma$  (say), then

$$U = \frac{X - E(X)}{\sigma_X} = \frac{X - \mu}{\sigma} = Z \text{ (say),}$$

is known as a standard variate. Thus the m.g.f. of a standard variate  $Z$  is given by

$$M_Z(t) = e^{-\mu t/\sigma} M_X(t/\sigma) \quad \dots (6-61)$$

**Remark.**  $E(Z) = E\left(\frac{X - \mu}{\sigma}\right) = \frac{1}{\sigma}E(X - \mu)$   
 $= \frac{1}{\sigma}[E(X) - \mu] = \frac{1}{\sigma}(\mu - \mu) = 0$

and  $V(Z) = V\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma^2} V(X-\mu)$  [c.f. Cor. (i) Theorem 6-8]  
 $= \frac{1}{\sigma^2} V(X) = \frac{1}{\sigma^2} \sigma^2 = 1$  [c.f. Cor. (iii) Theorem 6-8]

$\therefore E(Z) = 0$  and  $V(Z) = 1$ , i.e., the mean and variance of a standard variate are 0 and 1 respectively.

**6-10-3. Uniqueness Theorem of Moment Generating Function.** *The moment generating function of a distribution, if it exists, uniquely determines the distribution.* This implies that corresponding to a given probability distribution, there is only one m.g.f. (provided it exists) and corresponding to a given m.g.f., there is only one probability distribution. Hence  $M_X(t) = M_Y(t) \Rightarrow X$  and  $Y$  are identically distributed. [For detailed discussion, see Uniqueness Theorem of Characteristic Functions - Theorem 6-27, page 6-90]

**6-11. Cumulants.** Cumulants generating function  $K(t)$  is defined as

$$K_X(t) = \log_e M_X(t), \quad \dots(6-62)$$

provided the right-hand side can be expanded as a convergent series in powers of  $t$ . Thus

$$\begin{aligned} K_X(t) &= \kappa_1 t + \kappa_2 \frac{t^2}{2!} + \dots + \kappa_r \frac{t^r}{r!} + \dots = \log M_X(t) \\ &= \log \left[ 1 + \mu_1' t + \mu_2' \frac{t^2}{2!} + \mu_3' \frac{t^3}{3!} + \dots + \mu_r' \frac{t^r}{r!} + \dots \right] \end{aligned} \quad \dots(6-62a)$$

where  $\kappa_r$  = coefficient of  $\frac{t^r}{r!}$  in  $K_X(t)$  is called the  $r$ th cumulant. Hence

$$\begin{aligned} \kappa_1 t + \kappa_2 \frac{t^2}{2!} + \kappa_3 \frac{t^3}{3!} + \kappa_4 \frac{t^4}{4!} + \dots \\ = \left[ \left( \mu_1' t + \mu_2' \frac{t^2}{2!} + \mu_3' \frac{t^3}{3!} + \mu_4' \frac{t^4}{4!} + \dots \right) \right. \\ \left. - \frac{1}{2} \left( \mu_1' t + \mu_2' \frac{t^2}{2!} + \mu_3' \frac{t^3}{3!} + \dots \right)^2 \right. \\ \left. + \frac{1}{3} \left( \mu_1' t + \mu_2' \frac{t^2}{2!} + \dots \right)^3 - \frac{1}{4} \left( \mu_1' t + \mu_2' \frac{t^2}{2!} + \dots \right)^4 + \dots \right] \end{aligned}$$

Comparing the coefficients of like powers of ' $t$ ' on both sides, we get the relationship between the moments and cumulants. Hence, we have

$$\kappa_1 = \mu_1' = \text{Mean}, \quad \frac{\kappa_2}{2!} = \frac{\mu_2'}{2!} - \frac{\mu_1'^2}{2!} \Rightarrow \kappa_2 = \mu_2' - \mu_1'^2 = \mu_2$$

$$\frac{\kappa_3}{3!} = \frac{\mu_3'}{3!} - \frac{1}{2} \frac{2\mu_1'\mu_2'}{2!} + \frac{\mu_1'^3}{3!} \Rightarrow \kappa_3 = \mu_3' - 3\mu_2'\mu_1' + 2\mu_1'^3 = \mu_3$$

Also

$$\begin{aligned}\kappa_4 &= \frac{\mu'_4}{4!} - \frac{1}{2} \left( \frac{\mu_2'^2}{4} + \frac{2\mu_1'\mu_3'}{3!} \right) + \frac{1}{3} \frac{3\mu_1'^2\mu_2'}{2} - \frac{\mu_1'^4}{4} \\ \Rightarrow \kappa_4 &= \mu'_4 - 3\mu_2'^2 - 4\mu_1'\mu_3' + 12\mu_1'^2\mu_2' - 6\mu_1'^4 \\ &= (\mu'_4 - 4\mu_3'\mu_1' + 6\mu_2'\mu_1'^2 - 3\mu_1'^4) - 3(\mu_2'^2 - 2\mu_2'\mu_1'^2 + \mu_1'^4) \\ &= \mu_4 - 3(\mu_2' - \mu_1'^2)^2 = \mu_4 - 3\kappa_2^2 \quad (\because \mu_2 = \kappa_2) \\ \Rightarrow \mu_4 &= \kappa_4 + 3\kappa_2^2\end{aligned}$$

Hence we have obtained :

$$\left. \begin{array}{l} \text{Mean} = \kappa_1 \\ \mu_2 = \kappa_2 \pm \text{Variance} \\ \mu_3 = \kappa_3 \\ \mu_4 = \kappa_4 + 3\kappa_2^2 \end{array} \right\} \quad \dots(6.62b)$$

**Remarks.** 1. These results are of fundamental importance and should be committed to memory.

2. If we differentiate both sides of (6.62a) w.r.t.  $t$  'r' times and then put  $t=0$ , we get

$$\kappa_r = \left[ \frac{d^r}{dt^r} K_X(t) \right]_{t=0} \quad \dots(6.62c)$$

**6.11.1. Additive Property of Cumulants.** *The rth cumulant of the sum of independent random variables is equal to the sum of the rth cumulants of the individual variables. Symbolically,*

$$\kappa_r(X_1 + X_2 + \dots + X_n) = \kappa_r(X_1) + \kappa_r(X_2) + \dots + \kappa_r(X_n), \quad \dots(6.63)$$

where  $X_i$ ;  $i = 1, 2, \dots, n$  are independent random variables.

**Proof.** We have, since  $X_i$ 's are independent.

$$M_{X_1 + X_2 + \dots + X_n}(t) = M_{X_1}(t) M_{X_2}(t) \dots M_{X_n}(t)$$

Taking logarithm of both sides, we get

$$K_{X_1 + X_2 + \dots + X_n}(t) = K_{X_1}(t) + K_{X_2}(t) + \dots + K_{X_n}(t)$$

Differentiating both sides w.r.t.  $t$  'r' times and putting  $t=0$ , we get

$$\begin{aligned}\left[ \frac{d^r}{dt^r} K_{X_1 + X_2 + \dots + X_n}(t) \right]_{t=0} &= \left[ \frac{d^r}{dt^r} K_{X_1}(t) \right]_{t=0} \\ &\quad + \left[ \frac{d^r}{dt^r} K_{X_2}(t) \right]_{t=0} + \dots + \left[ \frac{d^r}{dt^r} K_{X_n}(t) \right]_{t=0}\end{aligned}$$

$\Rightarrow \kappa_r(X_1 + X_2 + \dots + X_n) = \kappa_r(X_1) + \kappa_r(X_2) + \dots + \kappa_r(X_n)$ , which establishes the result.

**6.11.2. Effect of Change of Origin and Scale on Cumulants.** If we take

$$U = \frac{X - a}{h}, \text{ then } M_U(t) = \exp(-at/h) M_X(t/h)$$

$$\therefore K_U(t) = -\frac{at}{h} + K_X(t/h)$$

$$K_1' + K_2' \frac{t^2}{2!} + \dots + K_r' \frac{t^r}{r!} + \dots = -\frac{at}{h} + K_1(t/h)$$

$$+ K_2 \frac{(t/h)^2}{2!} + \dots + K_r \frac{(t/h)^r}{r!} + \dots$$

where  $K'_r$  and  $K_r$  are the  $r$ th cumulants of  $U$  and  $X$  respectively.

Comparing coefficients, we get

$$K_1' = \frac{K_1 - a}{h} \quad \text{and} \quad K_r' = \frac{K_r}{h^r}; \quad r = 2, 3, \dots \quad \dots(6.63a)$$

Thus we see that except the first cumulant, all cumulants are independent of change of origin. But the cumulants are not invariant of change of scale as the  $r$ th cumulant of  $U$  is  $(1/h^r)$  times the  $r$ th cumulant of the distribution of  $X$ .

**Example 6.37.** Let the random variable  $X$  assume the value ' $r$ ' with the probability law :

$$P(X=r) = q^{r-1} p; \quad r = 1, 2, 3, \dots$$

Find the m.g.f. of  $X$  and hence its mean and variance.

[Calicut Univ. B.Sc., Oct. 1992]

**Solution.**  $M_X(t) = E(e^{tX})$

$$\begin{aligned} &= \sum_{r=1}^{\infty} e^{tr} q^{r-1} p = \frac{p}{q} \sum_{r=1}^{\infty} (qe^t)^r \\ &= \frac{p}{q} qe^t \sum_{r=1}^{\infty} (qe^t)^{r-1} = pe^t [1 + qe^t + (qe^t)^2 + \dots] \\ &= \left( \frac{pe^t}{1 - qe^t} \right) \end{aligned}$$

If dash (') denotes the differentiation w.r.t.  $t$ , then we have

$$M_X'(t) = \frac{pe^t}{(1 - qe^t)^2}, \quad M_X''(t) = pe^t \frac{(1 + qe^t)}{(1 - qe^t)^3}$$

$$\therefore \mu_1'(\text{about origin}) = M_X'(0) = \frac{p}{(1 - q)^2} = \frac{1}{p}$$

$$\mu_2'(\text{about origin}) = M_X''(0) = \frac{p(1 + q)}{(1 - q)^3} = \frac{1 + q}{p^2}$$

$$\text{Hence} \quad \text{mean} = \mu_1'(\text{about origin}) = \frac{1}{p}$$

$$\text{and} \quad \text{variance} = \mu_2' = \mu_2' - \mu_1'^2 = \frac{1 + q}{p^2} - \frac{1}{p^2} = \frac{q}{p^2}$$

**Example 6.38.** The probability density function of the random variable  $X$  follows the following probability law :

$$p(x) = \frac{1}{2\theta} \exp\left(-\frac{|x-\theta|}{\theta}\right), -\infty < x < \infty$$

Find M.G.F. of  $X$ . Hence or otherwise find  $E(X)$  and  $V(X)$ .

[Punjab Univ. M.A.(Eco.), 1991]

**Solution.** The moment generating function of  $X$  is

$$\begin{aligned} M_X(t) &= E(e^t) = \int_{-\infty}^{\infty} \frac{1}{2\theta} \exp\left(-\frac{|x-\theta|}{\theta}\right) e^x dx \\ &= \int_{-\infty}^{\theta} \frac{1}{2\theta} \exp\left(-\frac{|\theta-x|}{\theta}\right) e^x dx \\ &\quad + \int_{\theta}^{\infty} \frac{1}{2\theta} \exp\left(-\frac{|x-\theta|}{\theta}\right) e^x dx. \end{aligned}$$

For  $x \in (-\infty, \theta)$ ,  $x-\theta < 0 \Rightarrow \theta-x > 0$

$$\therefore |x-\theta| = \theta-x \quad \forall x \in (-\infty, \infty)$$

Similarly,  $|x-\theta| = x-\theta \quad \forall x \in (\theta, \infty)$

$$\begin{aligned} \therefore M_X(t) &= \frac{e^{-1}}{2\theta} \int_{-\infty}^{\theta} \exp\left[x\left(t+\frac{1}{\theta}\right)\right] dx + \frac{e}{2\theta} \int_{\theta}^{\infty} \exp\left[-x\left(\frac{1}{\theta}-t\right)\right] dx \\ &= \frac{e^{-1}}{2\theta} \cdot \frac{1}{\left(t+\frac{1}{\theta}\right)} \cdot \exp\left[\theta\left(t+\frac{1}{\theta}\right)\right] \\ &\quad + \frac{e}{2\theta} \cdot \frac{1}{\left(\frac{1}{\theta}-t\right)} \cdot \exp\left[-\theta\left(\frac{1}{\theta}-t\right)\right] \\ &= \frac{e^{t\theta}}{2(0t+1)} + \frac{e^{t\theta}}{2(1-0t)} = \frac{e^{t\theta}}{1-\theta^2 t^2} \\ &= e^{t\theta} (1-\theta^2 t^2)^{-1} \\ &= [1 + \theta t + \frac{\theta^2 t^2}{2!} + \dots] [1 + \theta^2 t^2 + \theta^4 t^4 + \dots] \\ &= 1 + \theta t + \frac{3\theta^2 t^2}{2!} + \dots \end{aligned}$$

$$\therefore E(X) = \mu' = \text{Coefficient of } t \text{ in } M_X(t) = \theta$$

$$\mu_2' = \text{Coefficient of } \frac{t^2}{2!} \text{ in } M_X(t) = 3\theta^2$$

$$\text{Hence } \text{Var}(X) = \mu_2' - \mu_1'^2 = 3\theta^2 - \theta^2 = 2\theta^2$$

**Example 6.39.** If the moments of a variate  $X$  are defined by

$$E(X') = 0.6 ; r = 1, 2, 3, \dots$$

show that  $P(X=0) = 0.4, P(X=1) = 0.6, P(X \geq 2) = 0.$

[Delhi Univ. B.Sc. (Maths Hons.), 1985]

**Solution.** The m.g.f. of variate  $X$  is :

$$\begin{aligned} M_X(t) &= \sum_{r=0}^{\infty} \frac{t^r}{r!} \mu_r' = 1 + \sum_{r=1}^{\infty} \frac{t^r}{r!} (0.6) \\ &= 0.4 + 0.6 \sum_{r=0}^{\infty} \frac{t^r}{r!} = 0.4 + 0.6 e^t \end{aligned} \quad \dots(i)$$

$$\begin{aligned} \text{But } M_X(t) &= E(e^{tX}) = \sum_{x=0}^{\infty} e^{tx} P(X=x) \\ &= P(X=0) + e^t \cdot P(X=1) + \sum_{x=2}^{\infty} e^{tx} \cdot P(X=x) \end{aligned} \quad \dots(ii)$$

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

$$P(X=0) = 0.4 ; P(X=1) = 0.6 ; P(X \geq 2) = 0.$$

**Remark.** In fact (i) is the m.g.f. of Bernoulli variate  $X$  with  $P(X=0) = q = 0.4$  and  $P(X=1) = p = 0.6$  [See § 7.1.2] and  $P(X \geq 2) = 0.$

**Example 6.40.** Find the moment generating function of the random variable whose moments are

$$\mu_r' = (r+1)! 2^r$$

**Solution.** The m.g.f. is given by

$$\begin{aligned} M_X(t) &= \sum_{r=0}^{\infty} \frac{t^r}{r!} \mu_r' = \sum_{r=0}^{\infty} \frac{t^r}{r!} (r+1)! 2^r \\ &= \sum_{r=0}^{\infty} (r+1) (2t)^r \\ M_X(t) &= 1 + 2 \cdot (2t) + 3 (2t)^2 + 4 (2t)^3 + \dots \\ &= (1 - 2t)^{-2} \end{aligned}$$

Aliter. The R.H.S. is an arithmetic-geometric series with ratio  $r = (2t)$

$$\text{Let } S = 1 + 2r + 3r^2 + 4r^3 + \dots$$

$$\text{Then } rS = r + 2r^2 + 3r^3 + \dots$$

$$\therefore (1-r)S = 1 + r + r^2 + \dots = \frac{1}{(1-r)}$$

$$\Rightarrow S = \frac{1}{(1-r)^2} = (1-r)^{-2} = (1-2t)^{-2}$$

**Remark.** This is the m.g.f. of Chi-square ( $\chi^2$ ) distribution with parameter (degrees of freedom)  $n = 2$  [c.f. Chapter 13].

**Example 6.41.** If  $\mu'_r$  is the  $r$ th moment about origin, prove that

$$\mu'_r = \sum_{j=1}^r \binom{r-1}{j-1} \mu_{r-j} \kappa_j,$$

where  $\kappa_j$  is the  $j$ th cumulant.

**Solution.** Differentiating both sides of (6.62a) in § 6.11, page 6.72 w.r.t.  $t$ , we get

$$\begin{aligned} & \kappa_1 + \kappa_2 t + \kappa_3 \frac{t^2}{2!} + \dots + \kappa_r \frac{t^{r-1}}{(r-1)!} + \dots \\ &= \frac{\mu'_1 + \mu'_2 t + \mu'_3 \frac{t^2}{2!} + \dots + \mu'_r \frac{t^{r-1}}{(r-1)!} + \dots}{1 + \mu'_1 t + \mu'_2 \frac{t^2}{2!} + \dots + \mu'_r \frac{t^r}{r!} + \dots} \\ \Rightarrow & \left[ \kappa_1 + \kappa_2 t + \kappa_3 \frac{t^2}{2!} + \dots + \kappa_r \frac{t^{r-1}}{(r-1)!} + \dots \right] \\ & \quad \times \left[ 1 + \mu'_1 t + \mu'_2 \frac{t^2}{2!} + \dots + \mu'_r \frac{t^r}{r!} + \dots \right] \\ &= \mu'_1 + \mu'_2 t + \mu'_3 \frac{t^2}{2!} + \dots + \mu'_r \frac{t^{r-1}}{(r-1)!} + \dots \end{aligned}$$

Comparing the coefficient of  $\frac{t^{r-1}}{(r-1)!}$  on both sides, we get

$$\begin{aligned} \mu'_r &= \kappa_1 \cdot \mu'_{r-1} + (r-1) \kappa_2 \cdot \mu'_{r-2} + \binom{r-1}{2} \kappa_3 \cdot \mu'_{r-3} + \dots + \kappa_r \\ &= \binom{r-1}{0} \mu'_{r-1} \kappa_1 + \binom{r-1}{1} \mu'_{r-2} \kappa_2 + \binom{r-1}{2} \mu'_{r-3} \kappa_3 \\ &\quad + \dots + \binom{r-1}{r-1} \mu'_0 \kappa_r \\ &= \sum_{j=1}^r \binom{r-1}{j-1} \mu'_{r-j} \kappa_j, \end{aligned}$$

which is the required result.

**6.12. Characteristic Function.** In some cases m.g.f. does not exist, since the integral  $\int_{-\infty}^{\infty} e^{tx} f(x) dx$  or the series  $\sum_x e^{tx} p(x)$  does not converge absolutely for real values of  $t$  for some distributions. For example, for the continuous probability distribution

$$dF(x) = C \frac{1}{(1+x^2)^m} dx ; m > 1, -\infty < x < \infty,$$

the m.g.f. does not exist, since the integral

$$M_X(t) = C \int_{-\infty}^{\infty} e^{tx} \frac{1}{(1+x^2)^m} dx,$$

does not converge absolutely for finite positive values of  $m$  because the function  $e^x$  dominates the function  $x^{2m}$  so that  $e^x/x^{2m} \rightarrow \infty$  as  $x \rightarrow \infty$ .

Again, for the discrete probability distribution

$$\left. \begin{aligned} f(x) &= \frac{6}{\pi^2 x^2}; x = 1, 2, 3, \dots \\ &= 0, \text{ elsewhere} \end{aligned} \right\}$$

$$M_X(t) = \sum_x e^{itx} f(x) = \frac{6}{\pi^2} \sum_{x=1}^{\infty} \left( \frac{e^{itx}}{x^2} \right)$$

The series is not convergent (by D'Alembert's Ratio Test) for  $t > 0$ . Thus there does not exist a positive number  $h$  such that  $M_X(t)$  exists for  $-h < t < h$ . Hence  $M_X(t)$  does not exist in this case also.

A more serviceable function than the m.g.f. is what is known as characteristic function and is defined as

$$\begin{aligned} \phi_X(t) &= E(e^{itX}) = \int e^{itx} f(x) dx \\ &\quad \text{(for continuous probability distributions)} \\ &= \sum_x e^{itx} f(x) \\ &\quad \text{(for discrete probability distributions)} \end{aligned} \quad \dots(6-64)$$

If  $F_X(x)$  is the distribution function of a continuous random variable  $X$ , then

$$\phi_X(t) = \int_{-\infty}^{\infty} e^{itx} dF(x) \quad \dots(6-64a)$$

Obviously  $\phi(t)$  is a complex valued function of real variable  $t$ . It may be noted that

$$|\phi(t)| = \left| \int e^{itx} f(x) dx \right| \leq \int |e^{itx}| f(x) dx = \int f(x) dx = 1,$$

$$\text{since } |e^{itx}| = |\cos tx + i \sin tx|^{1/2} = (\cos^2 tx + \sin^2 tx)^{1/2} = 1$$

Since  $|\phi(t)| \leq 1$ , characteristic function  $\phi_X(t)$  always exists.

Yet another advantage of characteristic function lies in the fact that it uniquely determines the distribution function, i.e., if the characteristic function of a distribution is given, the distribution can be uniquely determined by the theorem, known as the Uniqueness Theorem of Characteristic Functions (c.f. Theorem 6.27 page 6.90).

**6.12.1. Properties of Characteristic Functions.** For all real 't', we have

$$(i) \phi(0) = \int_{-\infty}^{\infty} dF(x) = 1 \quad \dots(6-64b)$$

$$(ii) |\phi(t)| \leq 1 = \phi(0) \quad \dots(6-64c)$$

(iii)  $\phi(t)$  is continuous everywhere, i.e.,  $\phi(t)$  is a continuous function of  $t$  in  $(-\infty, \infty)$ . Rather  $\phi(t)$  is uniformly continuous in 't'

$$\text{Proof. For } h \neq 0, |\phi_X(t+h) - \phi_X(t)| = \left| \int_{-\infty}^{\infty} [e^{i(t+h)x} - e^{itx}] dF(x) \right|$$

$$\begin{aligned} &\leq \int_{-\infty}^{\infty} \left| e^{itx} (e^{ihx} - 1) \right| dF(x) \\ &= \int_{-\infty}^{\infty} \left| e^{ihx} - 1 \right| dF(x) \quad \dots(*) \end{aligned}$$

The last integral does not depend on  $t$ . If it tends to zero as  $h \rightarrow 0$  then  $\phi_X(t)$  is uniformly continuous in ' $t$ '.

$$\text{Now } \left| e^{ihx} - 1 \right| \leq \left| e^{ihx} \right| + 1 = 2$$

$$\therefore \int_{-\infty}^{\infty} \left| e^{ihx} - 1 \right| dF(x) \leq 2 \int_{-\infty}^{\infty} dF(x) = 2.$$

Hence by Dominated Convergence Theorem (D.C.T.), taking the limit inside the integral sign in (\*), we get

$$\lim_{h \rightarrow 0} \left| \phi_X(t+h) - \phi_X(t) \right| \leq \int_{-\infty}^{\infty} \lim_{h \rightarrow 0} \left| e^{ihx} - 1 \right| dF(x) = 0$$

$$\Rightarrow \lim_{h \rightarrow 0} \phi_X(t+h) = \phi_X(t) \quad \forall t.$$

Hence  $\phi_X(t)$  is uniformly continuous in ' $t$ '.

(iv)  $\phi_X(-t)$  and  $\phi_X(t)$  are conjugate functions, i.e.,  $\phi_X(-t) = \overline{\phi_X(t)}$ , where  $\bar{a}$  is the complex conjugate of  $a$ .

$$\text{Proof. } \phi_X(t) = E(e^{itX}) = E[\cos tX + i \sin tX]$$

$$\begin{aligned} \Rightarrow \overline{\phi_X(t)} &= E[\cos tX - i \sin tX] \\ &= E[\cos(-t)X + i \sin(-t)X] \\ &= E(e^{-itX}) = \phi_X(-t). \end{aligned}$$

### 6.12.2. Theorems on Characteristic Function.

**Theorem 6.20.** If the distribution function of a r.v.  $X$  is symmetrical about zero, i.e.,

$$1 - F(x) = F(-x) \Rightarrow f(-x) = f(x),$$

then  $\phi_X(t)$  is real valued and even function of  $t$ .

**Proof.** By definition we have

$$\begin{aligned} \phi_X(t) &= \int_{-\infty}^{\infty} e^{itx} f(x) dx = \int_{-\infty}^{\infty} e^{-ity} f(-y) dy \quad (x = -y) \\ &= \int_{-\infty}^{\infty} e^{-ity} f(y) dy \quad [\because f(-y) = f(y)] \\ &= \phi_X(-t) \quad \dots(*) \end{aligned}$$

$\Rightarrow \phi_X(t)$  is an even function of  $t$ .

Also  $\overline{\phi_X(t)} = \phi_X(-t)$  [c.f. Property (iv) § 6.12.1.]

$\therefore \phi_X(t) = \phi_X(-t) = \phi_X(t)$  (From \*)

Hence  $\phi_X(t)$  is a real valued function of  $t$ .

**Theorem 6.21.** If  $X$  is some random variable with characteristic function  $\phi_X(t)$ , and if  $\mu'_r = E(X')$  exists, then

$$\mu' = (-i)^r \left[ \frac{\partial^r}{\partial t^r} \phi(t) \right]_{t=0} \quad \dots(6-65)$$

**Proof.**  $\phi(t) = \int_{-\infty}^{\infty} e^{itx} f(x) dx$

Differentiating (under the integral sign) 'r' times w.r.t.  $t$ , we get

$$\begin{aligned} \frac{\partial^r}{\partial t^r} \phi(t) &= \int_{-\infty}^{\infty} (ix)^r \cdot e^{itx} f(x) dx = (i)^r \int_{-\infty}^{\infty} x^r e^{itx} f(x) dx \\ \therefore \left[ \frac{\partial^r}{\partial t^r} \phi(t) \right]_{t=0} &= (i)^r \left[ \int_{-\infty}^{\infty} x^r e^{itx} f(x) dx \right]_{t=0} \\ &= (i)^r \int_{-\infty}^{\infty} x^r f(x) dx = i^r E(X^r) = i^r \mu'_r \end{aligned}$$

Hence  $\mu'_r = \left( \frac{1}{i} \right)^r \left[ \frac{\partial^r}{\partial t^r} \phi(t) \right]_{t=0} = (-i)^r \left[ \frac{\partial^r}{\partial t^r} \phi(t) \right]_{t=0}$

The theorems, viz., 6-17, 6-18 and 6-19 on m.g.f. can be easily extended to the characteristic functions as given below.

**Theorem 6-22.**  $\phi_{cx}(t) = \phi_x(ct)$ ,  $c$ , being a constant.

**Theorem 6-23.** If  $X_1$  and  $X_2$  are independent random variables, then

$$\phi_{x_1+x_2}(t) = \phi_{x_1}(t) \cdot \phi_{x_2}(t) \quad \dots(*)$$

More, generally for independent random variables  $X_i$ ;  $i = 1, 2, \dots, n$ , we have

$$\phi_{x_1+x_2+\dots+x_n}(t) = \phi_{x_1}(t) \cdot \phi_{x_2}(t) \dots \phi_{x_n}(t)$$

**Important Remark.** Converse of (\*) is not true, i.e.,

$\phi_{x_1+x_2}(t) = \phi_{x_1}(t) \cdot \phi_{x_2}(t)$  does not imply that  $X_1$  and  $X_2$  are independent.

For example, let  $X_1$  be a standard Cauchy variate with p.d.f.

$$f(x) = \frac{1}{\pi(1+x^2)}, -\infty < x < \infty$$

Then  $\phi_{x_1}(t) = e^{-|t|}$  (c.f. Chapter 8)

Let  $X_2 = X_1$ , i.e.,  $P(X_1 = X_2) = 1$ .  $\dots(**)$

Then  $\phi_{x_2}(t) = e^{-|t|}$

$$\begin{aligned} \text{Now } \phi_{x_1+x_2}(t) &= \phi_{2x_1}(t) = \phi_{x_1}(2t) = e^{-2|t|} \\ &= \phi_{x_1}(t) \cdot \phi_{x_2}(t) \end{aligned}$$

i.e., (\*) is satisfied but obviously  $X_1$  and  $X_2$  are not independent, because of (\*\*).

In fact, (\*) will hold even if we take  $X_1 = aX$  and  $X_2 = bX$ ,  $a$  and  $b$  being real numbers so that  $X_1$  and  $X_2$  are connected by the relation :

$$\frac{X_1}{a} = X = \frac{X_2}{b} \Rightarrow aX_2 = bX_1.$$

As another example let us consider the joint p.d.f. of two random variables  $X$  and  $Y$  given by

$$\begin{aligned} f(x, y) &= \frac{1}{4a^2} [1 + xy(x^2 - y^2)]; |x| \leq a, |y| \leq a, a > 0 \\ &= 0, \text{ elsewhere} \end{aligned}$$

Then the marginal p.d.f.'s of  $X$  and  $Y$  are given by

$$g(x) = \int_{-a}^a f(x, y) dy = \frac{1}{2a}; |x| \leq a \quad (\text{on simplification})$$

$$h(y) = \int_{-a}^a f(x, y) dx = \frac{1}{2a}; |y| \leq a$$

Then

$$\Phi_X(t) = \int_{-\infty}^{\infty} e^{itx} g(x) dx = \frac{1}{2a} \int_{-a}^a e^{itx} dx$$

$$= \frac{e^{iat} - e^{-iat}}{2ait} = \frac{\sin at}{at}$$

Similarly

$$\Phi_Y(t) = \frac{\sin at}{at}$$

$$\therefore \Phi_X(t) \Phi_Y(t) = \left( \frac{\sin at}{at} \right)^2 \quad \dots(*)$$

The p.d.f.  $k(z)$  of the random variable  $Z = X + Y$  is given by the convolution of p.d.f.'s of  $X$  and  $Y$ , viz.,

$$\begin{aligned} k(z) &= \int f(u, z-u) du \\ &= \frac{1}{4a^2} \int [1 + u(z-u)(u^2 - (z-u)^2)] du \\ &= \frac{1}{4a^2} \int (1 + 3z^2u^2 - 2zu^3 - z^3u) du, \end{aligned}$$

the limits of integration for  $u$  being in terms of  $z$  and are given by (left as an exercise to the reader)

$$-a \leq u \leq z+a; u \leq 0$$

$$z-a \leq u \leq a; u > 0$$

and

Thus

$$k(z) = \begin{cases} \frac{1}{4a^2} \int_{-a}^{z+a} (1 + 3z^2u^2 - 2zu^3 - z^3u) du = \frac{2a+z}{4a^2}; & -2a \leq z \leq 0 \\ \frac{1}{4a^2} \int_{z-a}^a (1 + 3z^2u^2 - 2zu^3 - z^3u) du = \frac{2a-z}{4a^2}; & 0 < z \leq 2a \\ 0, \text{ elsewhere} & \end{cases}$$

Now

$$\begin{aligned} \phi_{X+Y}(t) &= \Phi_Z(t) = \int_{-2a}^{2a} e^{itz} k(z) dz \\ &= \int_{-2a}^0 \left( \frac{2a+z}{4a^2} \right) e^{itz} dz + \int_0^{2a} \left( \frac{2a-z}{4a^2} \right) e^{itz} dz \end{aligned}$$

$$\begin{aligned}
 &= \int_0^{2a} \left( e^{-itz} + e^{itz} \right) \left( \frac{2a-z}{4a^2} \right) dz \\
 &\quad [\text{Changing } z \text{ to } -z \text{ in the first integral}] \\
 &= \frac{1}{2a^2} \int_0^{2a} (2a-z) \cos tz dz \\
 &= \frac{2 - 2 \cos 2at}{4a^2 t^2} = \frac{1 - \cos 2at}{2a^2 t^2} \\
 &= \left( \frac{\sin at}{at} \right) \\
 &= \phi_X(t) \cdot \phi_Y(t)
 \end{aligned}$$

[From (\*)]

But  $g(x) \cdot h(y) \neq f(x, y)$   
 $\Rightarrow X$  and  $Y$  are not independent.

However

$$\phi_{X_1, X_2}(t_1, t_2) = E(e^{it_1 X_1 + it_2 X_2}) = \phi_{X_1}(t_1) \cdot \phi_{X_2}(t_2)$$

implies that  $X_1$  and  $X_2$  are independent.

(For proof see Theorem 6-28)

**Theorem 6-24.** Effect of Change of Origin and Scale on Characteristic Function. If  $U = \frac{X-a}{h}$ ,  $a$  and  $h$  being constants, then

$$\phi_U(t) = e^{-iat/h} \phi_X(t/h)$$

In particular if we take  $a = E(X) = \mu$  (say) and  $h = \sigma_X = \sigma$  then the characteristic function of the standard variate

$$Z = \frac{X - E(X)}{\sigma_X} = \frac{X - \mu}{\sigma},$$

is given by  $\phi_Z(t) = e^{-i\mu/\sigma} \phi_X(t/\sigma)$  ... (6-66)

**Definition.** A random variable  $X$  is said to be a Lattice variable or be lattice distributed, if for some  $h > 0$ ,

$$P\left[\frac{X}{h} \text{ is an integer}\right] = 1,$$

$h$  is called a mesh.

**Theorem 6-25.** If  $|\phi_X(s)| = 1$  for some  $s \neq 0$ , then for some real  $a$ ,  $X - a$  is a Lattice variable with mesh  $h = 2\pi/|s|$ .

**Proof.** Consider any fixed  $t$ . We can write

$\phi_X(t) = |\phi_X(t)| e^{i\alpha t}$ , (a dependent on  $t$ ), since any complex number  $z$  can be written as  $z = |z| e^{i\theta}$ .

$$\begin{aligned}
 \therefore |\phi_X(t)| &= e^{-i\alpha t} \phi_X(t) = \phi_{X-a}(t) \\
 &= E[\cos t(X-a) + i \sin t(X-a)] = E[\cos t(X-a)]
 \end{aligned}$$

since left-hand side being real, we must have  $E[\sin t(X-a)] = 0$ .

$$\therefore 1 - |\phi_X(t)| = E[1 - \cos t(X-a)] \quad \dots (*)$$

If  $|\phi_X(s)| = 1, s \neq 0$  then for some  $a$  dependent on  $s$ , we have from (\*)

$$E[1 - \cos s(X - a)] = 0 \quad \dots (**)$$

But since  $1 - \cos s(X - a)$  is a non-negative random variable, (\*\*)

$$\Rightarrow P[1 - \cos s(X - a) = 0] = 1$$

$$\Rightarrow P[\cos s(X - a) = 1] = 1$$

$$\Rightarrow P[s(X - a) = 2n\pi] = 1$$

$$\Rightarrow P\left[(X - a) = \frac{2n\pi}{|s|}\right] = 1, \text{ for some } n = 0, 1, 2, \dots$$

Thus  $(X - a)$  is a Lattice variable with mesh  $\hbar = \frac{2\pi}{|s|}$ .

**6-12-3. Necessary and Sufficient Conditions for a Function  $\phi(t)$  to be Characteristic Function.** Properties (i) to (iv) in § 6-12-1 are merely the necessary conditions for a function  $\phi(t)$  to be the characteristic function of an r.v.  $X$ . Thus if a function  $\phi(t)$  does not satisfy any one of these four conditions, it cannot be the characteristic function of an r.v.  $X$ . For example, the function

$$\phi(t) = \log(1+t),$$

cannot be the c.f. of r.v.  $X$  since  $\phi(0) = \log 1 = 0 \neq 1$ .

These conditions are, however, not sufficient. It has been shown (c.f. Methods of Mathematical Statistics by H.Cramer) that if  $\phi(t)$  is near  $t = 0$  of the form,

$$\phi(t) = 1 + O(t^2 + \delta), \delta > 0 \quad \dots (*)$$

where  $O(t')$  divided by  $t'$  tends to zero as  $t \rightarrow 0$ , then  $\phi(t)$  cannot be the characteristic function unless it is identically equal to one. Thus, the functions

$$(i) \quad \phi(t) = e^{t^4} = 1 + O(t^4)$$

$$(ii) \quad \phi(t) = \frac{1}{1+t^4} = 1 + O(t^4)$$

being of the form (\*) are not characteristic functions, though both satisfy all the necessary conditions.

We give below a set of sufficient but not necessary conditions, due to Polya for a function  $\phi(t)$  to be the characteristic function :

$\phi(t)$  is a characteristic function if

- (1)  $\phi(0) = 1$ ,
- (2)  $\phi(t) = \phi(-t)$
- (3)  $\phi(t)$  is continuous
- (4)  $\phi(t)$  is convex for  $t > 0$ , i.e., for  $t_1, t_2 > 0$ ,

$$2\phi\left[\frac{t_1+t_2}{2}\right] \leq \phi(t_1) + \phi(t_2)$$

- (5)  $\lim_{t \rightarrow \infty} \phi(t) = 0$ ;

Hence by Polya's conditions the functions  $e^{-|t|}$  and  $[1 + |t|]^{-1}$  are characteristic functions. However, Polya's conditions are only sufficient and not necessary for a characteristic function. For example, if  $X \sim N(\mu, \sigma^2)$ ,

$$\phi(t) = e^{it\mu - t^2\sigma^2/2}, \quad [\text{c.f. } \S \text{ 8-5}]$$

and  $\phi(-t) \neq \phi(t)$ .

Various necessary and sufficient conditions are known, the simplest seems to be the following, due to Cramer.

"In order that a given, bounded and continuous function  $\phi(t)$  should be the characteristic function of a distribution, it is necessary and sufficient that  $\phi(0) = 1$  and that the function

$$\phi(x, A) = \int_0^A \int_0^A \phi(t-u) e^{(t-u)x} dt du$$

is real and non-negative for all real  $x$  and all  $A > 0$ .

#### 6.12.4. Multi-variate Characteristic Function. Then

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} \text{ and } t = \begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_n \end{bmatrix}, \quad t \text{ real}$$

be  $n \times 1$  column vectors. Then characteristic function of  $X$  is defined as

$$\phi_X(t) = E(e^{it'X}) = E[e^{i(t_1X_1 + t_2X_2 + \dots + t_nX_n)}] \quad \dots(6-67)$$

We may also write it as

$$\phi_{x_1, x_2, \dots, x_n}(t_1, t_2, \dots, t_n) \text{ or } \phi_X(t_1, t_2, \dots, t_n)$$

#### Some Properties.

$$(i) \quad \phi_X(0, 0, \dots, 0) = 1$$

$$(ii) \quad \phi_{-X}(t) = \overline{\phi_X(t)}$$

$$(iii) \quad |\phi_X(t)| \leq 1$$

(iv)  $\phi_X(t)$  is uniformly continuous in  $n$ -dimensional Euclidian space.

(v) If  $f_X(\cdot)$  is p.d.f. of  $X$ ,

$$\phi_X(t) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f_X(x_1, x_2, \dots, x_n) e^{i \sum t_i x_i} dx_1 dx_2 \dots dx_n$$

(vi)  $\phi_X(t) = \phi_Y(t)$  for all  $t$ , then  $X$  and  $Y$  have the same distribution.

$$(vii) \quad \text{If } \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left| \phi_X(t_1, t_2, \dots, t_n) \right| dt_1 dt_2 \dots dt_n < \infty,$$

then  $X$  is absolutely continuous and has a uniformly continuous p.d.f.

$$f_X(x_1, x_2, \dots, x_n) = \frac{1}{(2\pi)^n} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{i \sum t_i x_i} \phi_X(t_1, t_2, \dots, t_n) dt_1 dt_2 \dots dt_n$$

(viii) The random variables  $X_1, X_2, \dots, X_n$  are (mutually) independent iff

$$\phi_{x_1, x_2, \dots, x_n}(t_1, t_2, \dots, t_n) = \phi_{X_1}(t_1) \phi_{X_2}(t_2) \dots \phi_{X_n}(t_n)$$

**Remark. Multivariate Moment Generating Function.** Similarly, the m.g.f. of vector  $X = (X_1, X_2, \dots, X_n)'$  is given by :

$$M_X(t) = E(e^{tX}) = E(e^{t_1 X_1 + t_2 X_2 + \dots + t_n X_n}) \quad \dots(6-68)$$

We may also write :

$$M_X(t) = M_{X_1, X_2, \dots, X_n}(t_1, t_2, \dots, t_n) = E(e^{t_1 X_1 + t_2 X_2 + \dots + t_n X_n})$$

In particular, for two variates  $X_1$  and  $X_2$

$$M_X(t) = M_{X_1, X_2}(t_1, t_2) = E(e^{t_1 X_1 + t_2 X_2}) = \sum_{r=0}^{\infty} \sum_{s=0}^{\infty} \frac{t_1^r t_2^s}{r! s!} E(X_1^r X_2^s), \quad \dots(6-69)$$

provided it exists for  $-h_1 < t_1 < h_1$  and  $-h_2 < t_2 < h_2$ , where  $h_1$  and  $h_2$  are positive.

$$M_{X_1, X_2}(t_1, 0) = E(e^{t_1 X_1}) = M_{X_1}(t_1) \quad \dots(6-69a)$$

$$M_{X_1, X_2}(0, t_2) = E(e^{t_2 X_2}) = M_{X_2}(t_2) \quad \dots(6-69b)$$

If  $M(t_1, t_2)$  exists, the moments of all orders of  $X$  and  $Y$  exist and are given by:

$$E(X_2') = \left[ \frac{\partial^r M(t_1, t_2)}{\partial t_2^r} \right]_{t_1=t_2=0} = \frac{\partial^r M(0, 0)}{\partial t_2^r} \quad \dots(6-70)$$

$$E(X_1') = \left[ \frac{\partial^r M(t_1, t_2)}{\partial t_1^r} \right]_{t_1=t_2=0} = \frac{\partial^r M(0, 0)}{\partial t_1^r} \quad \dots(6-70a)$$

$$E(X_1^r X_2^s) = \left[ \frac{\partial^{r+s} M(t_1, t_2)}{\partial t_1^r \partial t_2^s} \right]_{t_1=t_2=0} = \frac{\partial^{r+s} M(0, 0)}{\partial t_1^r \partial t_2^s} \quad \dots(6-70b)$$

Cumulant generating function of  $X = (X_1, X_2)'$  is given by :

$$K_{X_1, X_2}(t_1, t_2) = \log M_{X_1, X_2}(t_1, t_2). \quad \dots(6-71)$$

**Example 6-42.** For a distribution, the cumulants are given by

$$\kappa_r = n [(r-1)!], \quad n > 0$$

Find the characteristic function. (Delhi Univ. B.Sc. (Stat. Hons.), 1990)

**Solution.** The cumulant generating function  $K(t)$ , if it exists, is given by

$$\begin{aligned} K(t) &= \sum_{r=1}^{\infty} \frac{(it)^r}{r!} \kappa_r = \sum_{r=1}^{\infty} \frac{(it)^r}{r!} n [(r-1)!] = n \sum_{r=1}^{\infty} \frac{(it)^r}{r} \\ &= n \left[ it + \frac{(it)^2}{2} + \frac{(it)^3}{3} + \dots \right] = n [\log(1-it)] \\ &= -n \log(1-it) = \log(1-it)^{-n} \end{aligned}$$

Also we have

$$K(t) = \log \phi(t) = \log(1-it)^{-n}$$

$$\therefore \phi(t) = (1-it)^{-n}$$

**Remark.** This is the characteristic function of the gamma distribution : (c.f. § 8-3-1)

$$f(x) = \frac{e^{-x} x^{n-1}}{\Gamma(n)}; n > 0, 0 < x < \infty.$$

**Example 6.43.** The moments about origin of a distribution are given by

$$\mu_r' = \frac{\Gamma(v+r)}{\Gamma(v)}.$$

Find the characteristic function.

(Madurai Kamaraj Univ. B.Sc., 1990)

**Solution.** We have

$$\begin{aligned}\phi(t) &= \sum_{r=0}^{\infty} \frac{(it)^r}{r!} \mu_r' = \sum_{r=0}^{\infty} \frac{(it)^r}{r!} \cdot \frac{\Gamma(v+r)}{\Gamma(v)} \\ &= \sum_{r=0}^{\infty} \frac{(it)^r}{r!} \cdot \frac{(v+r-1)!}{(v-1)!} \\ &= \sum_{r=0}^{\infty} (it)^r \cdot {}^{v+r-1}C_r = \sum_{r=0}^{\infty} (-1)^r \cdot {}^{-v}C_r (it)^r \\ [\because {}^{-v}C_r &= (-1)^{r(v+r-1)} C_r \Rightarrow (-1)^r \cdot {}^{-v}C_r = {}^{v+r-1}C_r] \\ \therefore \phi(t) &= \sum_{r=0}^{\infty} {}^{-v}C_r (-it)^r = (1-it)^{-v}\end{aligned}$$

**Example 6.44.** Show that

$$e^{itx} = 1 + (e^{it} - 1)x^{(1)} + (e^{it} - 1)^2 \cdot \frac{x^{(2)}}{2!} + \dots + (e^{it} - 1)^r \cdot \frac{x^{(r)}}{r!} + \dots$$

where  $x^{(r)} = x(x-1)(x-2)\dots(x-r+1)$ . Hence show that

$\mu_{(r)'} = [D^r \phi(t)]_{t=0}$ , where  $D = \frac{d}{dx}(e^{it})$  and  $\mu_{(r)'}$  is the  $r^{\text{th}}$  factorial moment.

**Solution.** We have

$$\begin{aligned}\text{R.H.S.} &= 1 + (e^{it} - 1)x^{(1)} + (e^{it} - 1)^2 \cdot \frac{x^{(2)}}{2!} + \dots + (e^{it} - 1)^r \cdot \frac{x^{(r)}}{r!} + \dots \\ &= 1 + (e^{it} - 1)(xC_1) + (e^{it} - 1)^2 (xC_2) \\ &\quad + (e^{it} - 1)^3 (xC_3) + \dots + (e^{it} - 1)^r (xC_r) \\ &= [1 + (e^{it} - 1)]^x = e^{itx} = \text{L.H.S.}\end{aligned}$$

By def.

$$\phi(t) = \int_{-\infty}^{\infty} e^{itx} f(x) dx$$

$$\begin{aligned}
 &= \int_{-\infty}^{\infty} \left[ 1 + (e^u - 1)x^{(0)} + (e^u - 1)^2 \cdot \frac{x^{(2)}}{2!} + \dots + (e^u - 1)^r \cdot \frac{x^{(r)}}{r!} + \dots \right] f(x) dx \\
 &= 1 + (e^u - 1) \int_{-\infty}^{\infty} x^{(1)} f(x) dx + \frac{(e^u - 1)^2}{2!} \int_{-\infty}^{\infty} x^{(2)} f(x) dx + \dots \\
 &\quad + \frac{(e^u - 1)^r}{r!} \int_{-\infty}^{\infty} x^{(r)} f(x) dx + \dots \\
 \therefore \quad [D' \phi(t)]_{t=0} &= \left[ \frac{d' \phi(t)}{d(e^u)^r} \right]_{t=0} = \int_{-\infty}^{\infty} x^{(r)} f(x) dx = \mu_r'
 \end{aligned}$$

where  $\mu_r'$  is the  $r$ th factorial moment.

**Theorem 6-26. (Inversion Theorem).** Lemma. If  $(a-h, a+h)$  is the continuity interval of the distribution function  $F(x)$ , then

$$F(a+h) - F(a-h) = \lim_{T \rightarrow \infty} \frac{1}{\pi} \int_{-T}^T \frac{\sin ht}{t} e^{-iu} \phi(t) dt,$$

$\phi(t)$  being the characteristic function of the distribution.

**Corollary.** If  $\phi(t)$  is absolutely integrable over  $R^1$ , i.e., if

$$\int_{-\infty}^{\infty} |\phi(t)| dt < \infty,$$

then the derivative of  $F(x)$  exists, which is bounded, continuous on  $R^1$  and is given by

$$f(x) = F'(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ix} \phi(t) dt, \quad \dots(6-72)$$

for every  $x \in R^1$ .

**Proof.** In the above lemma replacing  $a$  by  $x$  and on dividing by  $2h$ , we have

$$\begin{aligned}
 \frac{F(x+h) - F(x-h)}{2h} &= \frac{1}{2\pi} \lim_{T \rightarrow \infty} \int_{-T}^T \frac{\sin ht}{ht} e^{-iu} \phi(t) dt \\
 &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\sin ht}{ht} e^{-iu} \phi(t) dt \\
 \therefore \lim_{h \rightarrow 0} \frac{F(x+h) - F(x-h)}{2h} &= \frac{1}{2\pi} \lim_{h \rightarrow 0} \int_{-\infty}^{\infty} \frac{\sin ht}{ht} e^{-iu} \phi(t) dt
 \end{aligned}$$

Since  $\int_{-\infty}^{\infty} |\phi(t)| dt < \infty$ ,

the integrand on the right hand side is bounded by an integrable function and hence by Dominated Convergence Theorem, we get

$$\lim_{h \rightarrow 0} \frac{F(x+h) - F(x-h)}{2h} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \lim_{h \rightarrow 0} \left( \frac{\sin ht}{ht} \right) e^{-iu} \phi(t) dt$$

By mean value theorem of differential calculus, we have

$$\lim_{h \rightarrow 0} \frac{F(x+h) - F(x-h)}{2h} = F'(x) = f(x),$$

where  $f(\cdot)$  is the p.d.f. corresponding to  $\phi(t)$ . Thus

$$f(x) = F'(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \phi(t) dt,$$

as desired.

**Remark.** Consider the function  $\mathcal{F}_c$  defined by

$$\mathcal{F}_c = \int_{-c}^c e^{-itx} \phi(t) dt$$

Now if  $F'(x) = f(x)$  exists, then

$$\begin{aligned} \lim_{c \rightarrow \infty} \frac{\mathcal{F}_c}{2c} &= \lim_{c \rightarrow \infty} \frac{1}{2c} \int_{-c}^c \phi(t) e^{-itx} dt \\ &= \lim_{c \rightarrow \infty} \left\{ \frac{1}{2c} \cdot 2\pi f(x) \right\} = 0 \end{aligned}$$

Hence  $\frac{\mathcal{F}_c}{2c} \rightarrow 0$  at all points where  $F(x)$  is continuous. In other words, if the probability distribution is continuous

$$\frac{\mathcal{F}_c}{2c} \rightarrow 0 \text{ as } c \rightarrow \infty$$

If, however, the frequency function is discontinuous, i.e., distribution is discrete, consider one point of discontinuity say, the frequency  $f_j$  at  $x = x_j$ . Then the contribution of  $x_j$  to  $\phi(t)$  is  $f_j e^{itx_j}$  and hence its contributions to  $\mathcal{F}_c$  will be

$$\begin{aligned} &\int_{-c}^c f_j e^{itx_j} e^{-itx} dt \\ \therefore \lim_{c \rightarrow \infty} \frac{\mathcal{F}_c}{2c} &= \lim_{c \rightarrow \infty} \frac{1}{2c} f_j \int_{-c}^c e^{it(x_j - x)} dt \\ &= \lim_{c \rightarrow \infty} \frac{1}{2c} f_j \left[ \frac{e^{it(x_j - x)}}{it(x_j - x)} \right]_c \\ &= \begin{cases} 0 & \text{for } x \neq x_j \\ f_j & \text{for } x = x_j \end{cases} \end{aligned}$$

Hence if  $\mathcal{F}_c/2c \rightarrow 0$  at a point, there is no discontinuity in the distribution function at that point, but if it tends to a positive number  $f_j$ , the distribution is discontinuous at that point and the frequency is  $f_j$ . This gives us a criterion whether a given characteristic function represents a continuous distribution or not.

**Theorem 6-27. Uniqueness Theorem of Characteristic Functions.** Characteristic function uniquely determines the distribution, i.e., a necessary and sufficient condition for two distributions with p.d.f.'s  $f_1(\cdot)$  and  $f_2(\cdot)$  to be identical is that their characteristic functions  $\phi_1(t)$  and  $\phi_2(t)$  are identical.

**Proof.** If  $f_1(\cdot) = f_2(\cdot)$ , then from the definition of characteristic function, we get

$$\phi_1(t) = \int_{-\infty}^{\infty} e^{itx} f_1(x) dx = \int_{-\infty}^{\infty} e^{itx} f_2(x) dx = \phi_2(t)$$

Conversely if  $\phi_1(t) = \phi_2(t)$ , then from corollary to Theorem 6-26, we get

$$f_1(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \phi_1(t) dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \phi_2(t) dt = f_2(x)$$

**Remark.** This is one of the most fundamental theorems in the distribution theory. It implies that corresponding to a distribution there is only one characteristic function and corresponding to a given characteristic function, there is only one distribution. This *one to one correspondence between characteristic functions and the p.d.f.'s enables us to identify the form of the p.d.f. from that of characteristic function.*

**Theorem 6-28.** *Necessary and sufficient condition for the random variables  $X_1$  and  $X_2$  to be independent is that their joint characteristic function is equal to the product of their individual characteristic functions, i.e.,*

$$\phi_{X_1, X_2}(t_1, t_2) = \phi_{X_1}(t_1) \phi_{X_2}(t_2) \quad \dots (*)$$

**Proof.** (i) *Condition is Necessary.* If  $X_1$  and  $X_2$  are independent then we have to show that (\*) holds. By def.,

$$\begin{aligned} \phi_{X_1, X_2}(t_1, t_2) &= E(e^{it_1 X_1 + it_2 X_2}) = E(e^{it_1 X_1} e^{it_2 X_2}) \\ &= E(e^{it_1 X_1}) E(e^{it_2 X_2}) (\because X_1, X_2 \text{ are independent}) \\ &= \phi_{X_1}(t_1) \phi_{X_2}(t_2), \end{aligned}$$

as required.

(ii) *Condition is sufficient.* We have to show that if (\*) holds, then  $X_1$  and  $X_2$  are independent.

Let  $f_{X_1, X_2}(x_1, x_2)$  be the joint p.d.f. of  $X_1$  and  $X_2$  and  $f_1(x_1)$  and  $f_2(x_2)$  be the marginal p.d.f.'s of  $X_1$  and  $X_2$  respectively. Then by definition (for continuous r.v.'s), we get

$$\begin{aligned} \phi_{X_1}(t_1) &= \int_{-\infty}^{\infty} e^{it_1 x_1} f_1(x_1) dx_1 \\ \phi_{X_2}(t_2) &= \int_{-\infty}^{\infty} e^{it_2 x_2} f_2(x_2) dx_2 \\ \therefore \phi_{X_1}(t_1) \phi_{X_2}(t_2) &= \left[ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{it_1 x_1} f_1(x_1) dx_1 \right] \left[ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{it_2 x_2} f_2(x_2) dx_2 \right] \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{i(t_1 x_1 + t_2 x_2)} f_1(x_1) f_2(x_2) dx_1 dx_2 \quad \dots (**). \end{aligned}$$

by Fubini's theorem, since the integrand is bounded by an integrable function.

Also by def<sup>n</sup>

$$\Phi_{X_1, X_2}(t_1, t_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{i(t_1 x_1 + t_2 x_2)} f(x_1, x_2) dx_1 dx_2$$

If (\*) holds, we get from (\*\*)

$$\Phi_{X_1, X_2}(t_1, t_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{i(t_1 x_1 + t_2 x_2)} f_1(x_1) f_2(x_2) dx_1 dx_2$$

Hence by uniqueness theorem of characteristic functions, we get

$$f(x_1, x_2) = f_1(x_1) f_2(x_2),$$

which implies that  $X_1$  and  $X_2$  are independent.

**Remarks.** 1. For discrete r.v.'s, the result is established by using summation instead of integration.

2. The result can be generalised to the case of more than two variables.

Necessary and sufficient condition for the mutual independence of random variables  $X_i$ , ( $i = 1, 2, 3, \dots, n$ ) is that

$$\phi_{X_1, X_2, \dots, X_n}(t_1, t_2, \dots, t_n) = \phi_{X_1}(t_1) \phi_{X_2}(t_2) \dots \phi_{X_n}(t_n)$$

3. In terms of moment generating functions, the necessary and sufficient condition for the r.v.'s  $X_1, X_2, \dots, X_n$  to be mutually independent is that

$$M_{X_1, X_2, \dots, X_n}(t_1, t_2, \dots, t_n) = M_{X_1}(t_1) M_{X_2}(t_2) \dots M_{X_n}(t_n)$$

provided m.g.f.'s exist.

**Theorem 6.29. Hally-Bray Theorem.** If the sequence of distribution functions  $\{F_n(x)\}$  converges to the distribution function  $F(x)$  at all the points of continuity of the latter and  $g(x)$  is bounded continuous function over the line  $R^1(-\infty, \infty)$ , then

$$\lim_{n \rightarrow \infty} \int_{-\infty}^{\infty} g(x) dF_n(x) = \int_{-\infty}^{\infty} g(x) dF(x) \quad \dots(*)$$

**Corollary.** If  $F_n(x) \rightarrow F(x)$ , then the corresponding sequence of characteristic functions  $\phi_n(t)$  of  $F_n(x)$  converges to the characteristic function  $\phi(t)$  of  $F$  at every point 't'.

**Proof.**  $\cos ix$  and  $\sin ix$  are continuous and bounded functions of  $x$  for all  $t$  and hence from (\*), we get

$$\lim_{n \rightarrow \infty} \int_{-\infty}^{\infty} \cos ix dF_n(x) = \int_{-\infty}^{\infty} \cos ix dF(x)$$

and

$$\lim_{n \rightarrow \infty} \int_{-\infty}^{\infty} \sin ix dF_n(x) = \int_{-\infty}^{\infty} \sin ix dF(x)$$

$$\therefore \lim_{n \rightarrow \infty} \int_{-\infty}^{\infty} (\cos ix + i \sin ix) dF_n(x) = \int_{-\infty}^{\infty} (\cos ix + i \sin ix) dF(x).$$

$$\Rightarrow \lim_{n \rightarrow \infty} \int_{-\infty}^{\infty} e^{itx} dF_n(x) = \int_{-\infty}^{\infty} e^{itx} dF(x)$$

$$\Rightarrow \phi_n(t) \rightarrow \phi(t) \text{ as } n \rightarrow \infty$$

**Theorem 6.30.** Continuity Theorem for Characteristic Functions.  
For a sequence of distribution functions  $\{F_n(x)\}$  with the corresponding sequence of characteristic functions  $\{\phi_n(t)\}$ , a necessary and sufficient condition that  $F_n(x) \rightarrow F(x)$  at all points of continuity of  $F$  is that for every real  $t$ ,  $\phi_n(t) \rightarrow \phi(t)$ , which is continuous at  $t=0$  and  $\phi(t)$  is the characteristic function corresponding to  $F$ .

**Example 6.45.** Let  $F_n(x)$  be the distribution function defined by

$$F_n(x) = 0 \text{ for } x \leq -n$$

$$= \frac{x+n}{2n} \text{ for } -n < x < n$$

$$= 1 \text{ for } x \geq n$$

Is the limit  $F_n(x)$  a distribution function? If not, why?

**Solution.**

$$\phi_n(t) = \int_{-n}^n e^{itx} \frac{1}{2n} dx \quad (\because f_n(x) = F_n'(x))$$

$$= \frac{1}{2n} \left[ \frac{e^{itn} - e^{-itn}}{it} \right] = \frac{\sin nt}{nt}$$

$$\phi(t) = \lim_{n \rightarrow \infty} \phi_n(t) = \lim_{n \rightarrow \infty} \frac{\sin nt}{nt}$$

$$= \begin{cases} 1 & \text{if } t=0 \\ 0 & \text{if } t \neq 0 \end{cases}$$

i.e.,  $\phi(t)$  is discontinuous at  $t=0$ .

and  $\lim_{n \rightarrow \infty} F_n(x) = \frac{1}{2}$

Hence  $F(x)$  is not a distribution function.

**Example 6.46.** Find the distribution for which characteristic function is

(a)  $\phi(t) = (q + pe^{it})^n$ , (b)  $\phi(t) = e^{-t^2/2}$

**Solution.** (a)  $\phi(t) = (q + pe^{it})^n = \sum_{j=0}^n {}^n C_j p^j q^{n-j} e^{itj}$

We have

$$\mathcal{F}_c = \int_{-c}^c e^{-itx} \phi(t) dt = \int_{-c}^c \left\{ e^{-itx} \sum_{j=0}^n {}^n C_j p^j q^{n-j} e^{itj} \right\} dt$$

$$= \sum_{j=0}^n {}^n C_j p^j q^{n-j} \int_{-c}^c e^{-it(x-j)} dt$$

(i) If  $x \neq j$ ,

$$\begin{aligned}\mathcal{F}_c &= \sum_{j=0}^n {}^n C_j p^j q^{n-j} \left| \frac{e^{-it(x-j)}}{-i(x-j)} \right|_{-c}^c = \sum_{j=0}^n {}^n C_j p^j q^{n-j} \left[ \frac{e^{ic(x-j)} - e^{-ic(x-j)}}{i(x-j)} \right] \\ &= \sum_{j=0}^n \left[ {}^n C_j p^j q^{n-j} \cdot \frac{2i \sin(c(x-j))}{(x-j)} \right]\end{aligned}$$

$$\therefore \lim_{c \rightarrow \infty} \frac{\mathcal{F}_c}{2c} \rightarrow 0 \quad \forall x.$$

Hence there is no discontinuity in the distribution function when  $x \neq j$ .

(ii) If  $x = j$ ,

$$\mathcal{F}_c = \sum_{j=0}^n \left[ {}^n C_j p^j q^{n-j} \int_{-c}^c dt \right] = 2c \sum_{j=0}^n {}^n C_j p^j q^{n-j} = 2c(q+p)^n = 2c$$

Since  $\frac{\mathcal{F}_c}{2c} \rightarrow 1$  at  $x = j$ , the distribution function is discontinuous and its frequency is  ${}^n C_j p^j q^{n-j}$ .

$$(b) \text{ Let } \mathcal{F}_c = \int_{-c}^c e^{-itx - \frac{1}{2}\sigma^2 t^2} dt$$

$$\begin{aligned}\therefore |\mathcal{F}_c| &\leq \int_{-c}^c \left| e^{-itx - \frac{1}{2}\sigma^2 t^2} \right| dt \leq \int_{-c}^c e^{-\frac{1}{2}\sigma^2 t^2} dt \\ &\leq \int_{-\infty}^{\infty} e^{-\frac{1}{2}\sigma^2 t^2} dt = \frac{\sqrt{2\pi}}{\sigma}\end{aligned}$$

$$\therefore \lim_{c \rightarrow \infty} \frac{\mathcal{F}_c}{2c} = 0 \quad \forall x.$$

Hence the distribution function is continuous for all  $x$ .

$$\begin{aligned}f(x) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} e^{-t^2 \sigma^2/2} dt \\ &= \frac{1}{2\pi} e^{-x^2/2 \sigma^2} \int_{-\infty}^{\infty} \exp \left\{ -\frac{1}{2} \left( t\sigma + \frac{ix}{\sigma} \right)^2 \right\} dt\end{aligned}$$

$$\text{Put } t\sigma + \frac{ix}{\sigma} = \xi, \text{ i.e., } \sigma dt = d\xi$$

$$\therefore f(x) = \frac{1}{2\pi} \exp \left\{ -\frac{x^2}{2\sigma^2} \right\} \int_{-\infty}^{\infty} e^{-\xi^2/2} \frac{d\xi}{\sigma}$$

Hence

$$f(x) = \frac{1}{2\pi} \exp \left\{ -\frac{x^2}{2\sigma^2} \right\} \frac{\sqrt{2\pi}}{\sigma} = \frac{1}{\sigma\sqrt{2\pi}} \exp \left\{ -\frac{x^2}{2\sigma^2} \right\}, \quad -\infty < x < \infty$$

which is the p.d.f. of normal distribution.

**Example 6.47.** Find the density function  $f(x)$  corresponding to the characteristic function defined as follows :

$$\phi(t) = \begin{cases} 1 - |t|, & |t| \leq 1 \\ 0, & |t| > 1 \end{cases}$$

[Delhi Univ, B.Sc. (Maths Hons.), 1989]

**Solution.** By Inversion Theorem, the p.d.f. of  $X$  is given by :

$$\begin{aligned} f(x) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ixt} \phi(t) dt \\ &= \frac{1}{2\pi} \int_{-1}^0 e^{-ixt} (1+t) dt + \frac{1}{2\pi} \int_0^1 e^{-ixt} (1-t) dt * \end{aligned}$$

Now

$$\begin{aligned} \int_{-1}^0 e^{-ixt} (1+t) dt &= \left[ \frac{e^{-ixt}}{-ix} (1+t) \right]_{-1}^0 + \frac{1}{ix} \int_{-1}^0 e^{-ixt} dt \\ &= -\frac{1}{ix} + \frac{1}{ix} \left[ \frac{e^{-ixt}}{-ix} \right]_{-1}^0 \\ &= -\frac{1}{ix} + \frac{1}{(ix)^2} (e^{ix} - 1) \end{aligned}$$

Similarly,

$$\int_0^1 e^{-ixt} (1-t) dt = \frac{1}{ix} + \frac{1}{(ix)^2} (e^{-ix} - 1)$$

$$\begin{aligned} \therefore f(x) &= \frac{1}{2\pi} \left[ \frac{1}{(ix)^2} \left\{ e^{ix} - 1 + e^{-ix} - 1 \right\} \right] \\ &= \frac{1}{\pi x^2} \left[ 1 - \frac{e^{ix} + e^{-ix}}{2} \right] = \frac{1}{\pi} \cdot \frac{1 - \cos x}{x^2}, \quad -\infty < x < \infty \end{aligned}$$

### EXERCISE 6 (c)

- Define m.g.f. of a random variable. Hence or otherwise find the m.g.f. of:

\* : for  $-1 < t < 0$ ,  $|t| = -t$  and for  $0 < t < 1$ ,  $|t| = +t$

$$(i) Y = aX + b, \quad (ii) Y = \frac{X - m}{\sigma}.$$

[Sri Venkat Univ. B.Sc., Sept. 1990; Kerala Univ. B.Sc., Sept. 1992]

2. The random variable  $X$  takes the value  $n$  with probability  $1/2^n$ ,  $n = 1, 2, 3, \dots$ . Find the moment generating function of  $X$  and hence find the mean and variance of  $X$ .

3. Show that if  $\bar{X}$  is mean of  $n$  independent random variables, then

$$M_{\bar{X}}(t) = \left[ M_X\left(\frac{t}{n}\right) \right]^n$$

4. (a) Define moments and moment generating function (m.g.f.) of a random variable  $X$ . If  $M(t)$  is the m.g.f. of a random variable  $X$  about the origin, show that the moment  $\mu'_r$  is given by

$$\mu'_r = \left[ \frac{d^r M(t)}{dt^r} \right]_{t=0} \quad [\text{Baroda Univ. B.Sc., 1992}]$$

(b) If  $\mu'_r$  is the  $r$ th order moment about the origin and  $\kappa_j$  is the cumulant of  $j$ th order, prove that

$$\frac{\delta \mu'_r}{\delta \kappa_j} = \binom{r-1}{j-1} \mu'_{r-j}$$

(c) If  $\mu'_r$  is the  $r$ th moment about the origin of a variable  $X$  and if  $\mu'_r = r!$ , find the m.g.f. of  $X$ .

5. (a) A random variable 'X' has probability function

$$p(x) = \frac{1}{2^x}; x = 1, 2, 3, \dots$$

Find the M.G.F., mean and variance.

(b) Show that the m.g.f. of r.v.  $X$  having the p.d.f.

$$f(x) = \frac{1}{3}, -1 < x < 2$$

= 0, elsewhere,

is  $M(t) = \frac{e^{2t} - e^{-t}}{3t}, t \neq 0$

$$= 1, t = 0 \quad [\text{Gujarat Univ. B.Sc., Oct. 1991}]$$

(i) A random variable 'X' has the density function :

$$f(x) = \frac{1}{2\sqrt{x}}, 0 < x < 1$$

$$= 0, \text{ elsewhere}$$

Obtain the moment generating function and hence the mean and variance.

6.  $X$  is a random variable and  $p(x) = ab^x$ , where  $a$  and  $b$  are positive,  $a+b=1$  and  $x$  taking the values 0, 1, 2, ... Find the moment generating function of  $X$ . Hence show that

$$m_2 = m_1(2m_1 + 1)$$

$m_1$  and  $m_2$  being the first two moments.

7. Find the characteristic function of the following distributions and their variances :

$$(i) \quad dF(x) = ae^{-ax} dx, (a > 0, x > 0)$$

$$(ii) \quad dF(x) = \frac{1}{2} e^{-|x|} dx, (-\infty < x < \infty)$$

$$(iii) P(X=j) \text{ lies in } \binom{n}{j} p^j q^{n-j}, (0 < p < 1, q = 1 - p, j = 0, 1, 2, \dots, n).$$

8. Obtain the m.g.f. of the random variable  $X$  having p.d.f.,

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

Determine  $\mu'_1$ ,  $\mu_2$ ,  $\mu_3$  and  $\mu_4$ .

$$\text{Ans. } \left( \frac{e^t - 1}{t} \right)^2, \quad \mu'_1 = 1, \quad \mu_2 = 7.$$

9. (a) Define cumulants and obtain the first four cumulants in terms of central moments.

(b) If  $X$  is a variable with zero mean and cumulants  $\kappa_r$ , show that the first two cumulants  $k_1$  and  $k_2$  of  $X^2$  are given by  $k_1 = \kappa_2$  and  $k_2 = 2\kappa_2^2 + \kappa_4$ .

10. Show that the  $r$ th cumulant for the distribution

$$f(x) = ce^{-cx}, \text{ where } c \text{ is positive and } 0 \leq x < \infty$$

$$\text{is } \frac{1}{c^r} \cdot (r-1)!$$

11. If  $X$  is a random variable with cumulants  $\kappa_r$ ;  $r = 1, 2, \dots$ . Find the cumulants of

(i)  $cX$ , (ii)  $c + X$ , where  $c$  is a constant.

12. (a) Define the characteristic function of a random variable. Show that the characteristic function of the sum of two independent variables is equal to the product of their characteristic functions.

(b) If  $X$  is a random variable having cumulants  $\kappa_r$ ;  $r = 1, 2, \dots$  given by

$$\kappa_r = (r-1)! pa^{-r}; p > 0, a > 0,$$

find the characteristic function of  $X$ .

(c) Prove that the characteristic function of a random variable  $X$  is real if and only if  $X$  has a symmetric distribution about 0.

13. Define  $\phi(t)$ , the characteristic function of a random variable. Find the characteristic function of a random variable  $X$  defined as follows :

$$f(x) = \begin{cases} 0, & x < 0 \\ 1, & 0 \leq x \leq 1 \\ 0, & x > 1 \end{cases}$$

$$\text{Ans. } e^{(e-1)t}/it$$

14. For the joint distribution of two-dimensional random variable  $(X, Y)$  given by

$$f(x, y) = \frac{1}{4a^2} [1 + xy(x^2 - y^2)] \quad |x| \leq a; |y| \leq a, a > 0 \\ = 0, \text{ elsewhere,}$$

show that the characteristic function of  $X + Y$  is equal to the product of the characteristic functions of  $X$  and  $Y$ . Comment on the result.

Hint. See remark to Theorem 6.22, page 6.81.

15. Let  $K(t_1, t_2) = \log_e M(t_1, t_2)$  where  $M(t_1, t_2)$  is the m.g.f. of  $X$  and  $Y$ . Show that :

$$\frac{\partial K(0, 0)}{\partial t_1} = E(X); \quad \frac{\partial^2 K(0, 0)}{\partial t_1^2} = \text{Var } X; \quad \frac{\partial^2 K(0, 0)}{\partial t_1 \partial t_2} = \text{Cov}(X, Y)$$

[Delhi Univ. B.A.(Stat. Hons.), Spl. Course 1987]

### OBJECTIVE TYPE QUESTIONS

I. Comment on the following, giving examples, if possible :

- (i) M.g.f. of a r.v. always exist.
- (ii) Characteristic function of a r.v. always exists.
- (iii) M.g.f. is not affected by change of origin or/and scale.
- (iv)  $\phi_{X+Y}(t) = \phi_X(t) \cdot \phi_Y(t)$  implies  $X$  and  $Y$  are independent.
- (v)  $\phi_X(t) = \phi_Y(t)$  implies  $X$  and  $Y$  have the same distribution.
- (vi)  $\phi(0) = 1$  and  $|\phi(t)| \leq 1$ .
- (vii) Variance of a r.v. is 5 and its mean does not exist.
- (ix) It is possible to find a r.v. whose first  $k$  moments exist but  $(k+1)^{\text{th}}$  moment does not exist.

(x) If a r.v.  $X$  has a symmetrical distribution about origin then

- (a)  $\phi_X(t)$  is even valued function of  $t$ .
- (b)  $\phi_X(t)$  is complex valued function of  $t$ .

II: (a) Can the following be the characteristic functions of any distribution ? Give reasons.

(i)  $\log(1+t)$ , (ii)  $\exp(-t^2)$ , (iii)  $1/(1+t^4)$ .

(b) Prove that  $\phi(t) = \exp(-t^\alpha)$ , cannot be a characteristic function unless  $\alpha = 2$ .

III. State the relations, if any, between the following :

- (i)  $E(X')$  and  $\phi_X(t)$ .
- (ii)  $M_X(t)$  and  $M_{X-a}(t)$ ,  $a$  being constant.
- (iii)  $M_X(t)$  and  $M_{(X-a)/h}(t)$ ,  $a$  and  $h$  being constants.

(iv)  $\phi_X(t)$  and p.d.f. of  $X$ .

(v)  $\mu$  and  $\mu'$ .

(vi) First four cumulants in terms of first four moments about mean.

**IV.** Let  $X_1, X_2, \dots, X_n$  be  $n$  i.i.d. (independent and identically distributed) r.v.'s with m.g.f.  $M(t)$ . Then prove that

$$M_{\bar{X}}(t) = [M(t/n)]^n,$$

where

$$\bar{X} = \sum_{i=1}^n X_i / n$$

**V.** If  $X_1, X_2, \dots, X_n$  are independent r.v.'s then prove that

$$M_{\sum_{i=1}^n c_i X_i}(t) = \prod_{i=1}^n M_{X_i}(c_i t).$$

**VI.** Fill in the blanks :

(i) If  $\int_{-\infty}^{\infty} |\phi_X(t)| dt < \infty$ , then p.d.f. of  $X$  is given by ...  
 $\leftrightarrow$

(ii)  $X_1$  and  $X_2$  are independent if and only if ...  
 (Give result in terms of characteristic functions.)

(iii) If  $X_1$  and  $X_2$  are independent then

$$\phi_{X_1 - X_2}(t) = \dots$$

(iv)  $\phi(t)$  is ... defined and is ... for all  $t$  in  $(-\infty, \infty)$ .

**VII.** Examine critically the following statements :

(a) Two distributions having the same set of moments are identical.

(b) The characteristic function of a certain non-degenerate distribution is  $e^{-t^2}$ .

**6.13. Chebychev's Inequality.** The role of standard deviation as a parameter to characterise variance is precisely interpreted by means of the well known Chebychev's inequality. The theorem discovered in 1853 was later on discussed in 1856 by Bienayme.

**Theorem 6.31.** If  $X$  is a random variable with mean  $\mu$  and variance  $\sigma^2$ , then for any positive number  $k$ , we have

$$P\{|X - \mu| \geq k\sigma\} \leq 1/k^2 \quad \dots (6.73)$$

$$\text{or} \quad P\{|X - \mu| < k\sigma\} \geq 1 - (1/k^2) \quad \dots (6.73a)$$

**Proof. Case (i).**  $X$  is a continuous r.v. By def.,

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

$$= \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx, \text{ where } f(x) \text{ is p.d.f. of } X.$$

$$\begin{aligned}
 &= \int_{-\infty}^{\mu-k\sigma} (x-\mu)^2 f(x) dx + \int_{\mu-k\sigma}^{\mu+k\sigma} (x-\mu)^2 f(x) dx + \int_{\mu+k\sigma}^{\infty} (x-\mu)^2 f(x) dx \\
 &\geq \int_{-\infty}^{\mu-k\sigma} (x-\mu)^2 f(x) dx + \int_{\mu+k\sigma}^{\infty} (x-\mu)^2 f(x) dx \quad \dots(*)
 \end{aligned}$$

We know that :

$$x \leq \mu - k\sigma \text{ and } x \geq \mu + k\sigma \Leftrightarrow |x - \mu| \geq k\sigma \quad \dots(**)$$

Substituting in (\*), we get

$$\begin{aligned}
 \therefore \sigma^2 &\geq k^2 \sigma^2 \left[ \int_{-\infty}^{\mu-k\sigma} f(x) dx + \int_{\mu+k\sigma}^{\infty} f(x) dx \right] \\
 &= k^2 \sigma^2 [P(X \leq \mu - k\sigma) + P(X \geq \mu + k\sigma)] \quad [\text{From } (**)] \\
 &= k^2 \sigma^2 \cdot P(|X - \mu| \geq k\sigma) \quad [\text{From } (**)] \\
 \Rightarrow P(|X - \mu| \geq k\sigma) &\leq 1/k^2, \quad \dots(***)
 \end{aligned}$$

which establishes (6-73)

Also since

$$P(|X - \mu| \geq k\sigma) + P(|X - \mu| < k\sigma) = 1, \text{ we get}$$

$$P(|X - \mu| < k\sigma) = 1 - P(|X - \mu| \geq k\sigma) \geq 1 - (1/k^2) \quad [\text{From } (***)]$$

which establishes (6-73a).

**Case (ii).** In case of discrete random variable, the proof follows exactly similarly on replacing integration by summation.

**Remark.** In particular, if we take  $k\sigma = c > 0$ , then (\*) and (\*\*) give respectively

$$\begin{aligned}
 &P\left\{|X - \mu| \geq c\right\} \leq \frac{\sigma^2}{c^2} \text{ and } P\left\{|X - \mu| < c\right\} \geq 1 - \frac{\sigma^2}{c^2} \\
 \Rightarrow P\left\{|X - E(X)| \geq c\right\} &\leq \frac{\text{Var}(X)}{c^2} \quad \dots(6-73b) \\
 \text{and } P\left\{|X - E(X)| < c\right\} &\geq 1 - \frac{\text{Var}(X)}{c^2}.
 \end{aligned}$$

**6.13.1. Generalised Form of Bienayme-Chebychev's Inequality.** Let  $g(X)$  be a non-negative function of a random variable  $X$ . Then for every  $k > 0$ , we have

$$P\{g(X) \geq k\} \leq \frac{E[g(X)]}{k} \quad \dots(6-74)$$

[Bangalore Univ. B.Sc., 1992]

**Proof.** Here we shall prove the theorem for continuous random variable. The proof can be adapted to the case of discrete random variable on replacing integration by summation over the given range of the variable.

Let  $S$  be the set of all  $X$  where  $g(X) \geq k$ , i.e.,

$$S = \{x : g(x) \geq k\}$$

then  $\int_S dF(x) = P(X \in S) = P[g(X) \geq k], \dots (*)$

where  $F(x)$  is the distribution function of  $X$ .

Now  $E[g(X)] = \int_{-\infty}^{\infty} g(x) dF(x) \geq \int_S g(x) dF(x)$   
 $\geq k \cdot P[g(X) \geq k] \quad [\because \text{on } S, g(X) \geq k \text{ and using } (*)]$   
 $\Rightarrow P[g(X) \geq k] \leq \frac{E[g(X)]}{k}$

**Remarks 1.** If we take  $g(X) = (X - E(X))^2 = (X - \mu)^2$  and replace  $k$  by  $k^2 \sigma^2$  in (6.74), we get

$$P\{(X - \mu)^2 \geq k^2 \sigma^2\} \leq \frac{E(X - \mu)^2}{k^2 \sigma^2} = \frac{\sigma^2}{k^2 \sigma^2} = \frac{1}{k^2}$$

$$\Rightarrow P\{|X - \mu| \geq k \sigma\} \leq 1/k^2, \quad \dots (6.74a)$$

which is Chebychev's inequality.

**2. Markov's Inequality.** Taking  $g(X) = |X|$  in (6.74) we get, for any  $k > 0$

$$P[|X| \geq k] \leq \frac{E|X|}{k}, \quad \dots (6.75)$$

which is Markov's inequality.

Rather, taking  $g(X) = |X|^r$  and replacing  $k$  by  $k^r$  in (6.74), we get a more generalised form of Markov's inequality, viz.,

$$P[|X|^r \geq k^r] \leq \frac{E|X|^r}{k^r} \quad \dots (6.75a)$$

**3.** If we assume the existence of only second-order moments of  $X$ , then we cannot do better than Chebychev's inequality (6.73). However, we can sometimes improve upon the results of Chebychev's inequality if we assume the existence of higher order moments. We give below (without proof) one such inequality which assumes the existence of moments of 4th order.

**Theorem 6.31a.**  $E|X|^4 < \infty, E(X) = 0$  and  $E(X^2) = \sigma^2$

$$P(|X| > k\sigma) \geq \frac{\mu_4 - \sigma^4}{\mu_4 + \sigma^4 k^4 - 2k^2 \sigma^4} \quad \dots (6.76)$$

If  $X \sim U[0, 1]$ , [c.f. Chapter 8], with p.d.f.  $p(x) = 1, 0 < x < 1$  and  $= 0$ , otherwise, then

$$E(X^r) = 1/(r+1); (r = 1, 2, 3, 4) \quad \dots (*)$$

$$E(X) = 1/2, E(X^2) = 1/3, E(X^3) = 1/4, E(X^4) = 1/5 \quad \dots (*)$$

$$\text{Var}(X) = E(X^2) - [E(X)]^2 = 1/12$$

$$\mu_4 = E(X - \mu)^4 = E(X - \frac{1}{2})^4 = 1/80$$

[On using binomial expansion with (\*)] Chebychev's inequality (6.73a) with  $k = 2$  gives :

$$P \left[ \left| X - \frac{1}{2} \right| < 2 \frac{1}{\sqrt{12}} \right] \geq 1 - \frac{1}{4} = 0.75$$

With  $k = 2$ , (6.76) gives:

$$\begin{aligned} P \left[ \left| X - \frac{1}{2} \right| > 2 \frac{1}{\sqrt{12}} \right] &\leq \frac{1/80 - 1/44}{1/80 + 1/4 - 1/8} = \frac{4}{49} \\ \Rightarrow P \left[ \left| X - \frac{1}{2} \right| \leq 2 \frac{1}{\sqrt{12}} \right] &\geq 1 - \frac{4}{49} = \frac{45}{49} = 0.92, \end{aligned}$$

which is a much better lower bound than the lower bound given by Chebychev's inequality.

**6.14. Convergence in probability.** We shall now introduce a new concept of convergence, viz., convergence in probability or stochastic convergence which is defined as follows:-

A sequence of random variables  $X_1, X_2, \dots, X_n, \dots$  is said to converge in probability to a constant  $a$ , if for any  $\epsilon > 0$ ,

$$\lim_{n \rightarrow \infty} P(|X_n - a| < \epsilon) = 1 \quad \dots(6.77)$$

or its equivalent

$$\lim_{n \rightarrow \infty} P(|X_n - a| \geq \epsilon) = 0 \quad \dots(6.77a)$$

and we write

$$X_n \xrightarrow{P} a \text{ as } n \rightarrow \infty \quad \dots(6.77b)$$

If there exists a random variable  $X$  such that  $X_n - X \xrightarrow{P} a$  as  $n \rightarrow \infty$ , then we say that the given sequence of random variables *converges in probability to the random variable X*.

**Remark. 1.** If a sequence of constants  $a_n \rightarrow a$  as  $n \rightarrow \infty$ , then regarding the constant as a random variable having a one-point distribution at that point, we can say that as  $a_n \xrightarrow{P} a$  as  $n \rightarrow \infty$ .

**2.** Although the concept of convergence in probability is basically different from that of ordinary convergence of sequence of numbers, it can be easily verified that the following simple rules hold for convergence in probability as well.

If  $X_n \xrightarrow{P} \alpha$  and  $Y_n \xrightarrow{P} \beta$  as  $n \rightarrow \infty$ , then

$$(i) \quad X_n \pm Y_n \xrightarrow{P} \alpha \pm \beta \text{ as } n \rightarrow \infty$$

$$(ii) \quad X_n Y_n \xrightarrow{P} \alpha \beta \text{ as } n \rightarrow \infty$$

$$(iii) \quad \frac{X_n}{Y_n} \xrightarrow{P} \frac{\alpha}{\beta} \text{ as } n \rightarrow \infty, \text{ provided } \beta \neq 0.$$

**6-14-1. (Chebychev's Theorem).** As an immediate consequence of Chebychev's inequality, we have the following theorem and convergence in probability.

If  $X_1, X_2, \dots, X_n$  is a sequence of random variables and if mean  $\mu_n$  and standard deviation  $\sigma_n$  of  $X_n$  exists for all  $n$  and if  $\sigma_n \rightarrow 0$  as  $n \rightarrow \infty$ , then

$$X_n - \mu_n \xrightarrow{P} 0 \text{ as } n \rightarrow \infty$$

**Proof.** We know, for any  $\epsilon > 0$

$$P \left\{ |X_n - \mu_n| \geq \epsilon \right\} \leq \frac{\sigma_n^2}{\epsilon^2} \rightarrow 0 \text{ as } n \rightarrow \infty$$

$$\text{Hence } X_n - \mu_n \xrightarrow{P} 0 \text{ as } n \rightarrow \infty$$

**6-15. Weak Law of Large Numbers.** Let  $X_1, X_2, \dots, X_n$  be a sequence of random variables and  $\mu_1, \mu_2, \dots, \mu_n$  be their respective expectations and let

$$B_n = \text{Var}(X_1 + X_2 + \dots + X_n) < \infty$$

$$\text{Then } P \left\{ \left| \frac{X_1 + X_2 + \dots + X_n}{n} - \frac{\mu_1 + \mu_2 + \dots + \mu_n}{n} \right| < \epsilon \right\} \geq 1 - \eta \quad \dots(6-78)$$

for all  $n > n_0$ , where  $\epsilon$  and  $\eta$  are arbitrary small positive numbers, provided

$$\lim_{n \rightarrow \infty} \frac{B_n}{n^2} = 0$$

**Proof.** Using Chebychev's Inequality (6-73b), to the random variable  $(X_1 + X_2 + \dots + X_n)/n$ , we get for any  $\epsilon > 0$ ,

$$\begin{aligned} P \left\{ \left| \left( \frac{X_1 + X_2 + \dots + X_n}{n} \right) - \left( E \frac{X_1 + X_2 + \dots + X_n}{n} \right) \right| < \epsilon \right\} &\geq 1 - \frac{B_n}{n^2 \epsilon^2}, \\ \left[ \text{since } \text{Var} \left( \frac{X_1 + X_2 + \dots + X_n}{n} \right) = \frac{1}{n^2} \text{Var}(X_1 + X_2 + \dots + X_n) = \frac{B_n}{n^2} \right] \\ \Rightarrow P \left\{ \left| \frac{X_1 + X_2 + \dots + X_n}{n} - \frac{\mu_1 + \mu_2 + \dots + \mu_n}{n} \right| < \epsilon \right\} &\geq 1 - \frac{B_n}{n^2 \epsilon^2}. \end{aligned}$$

So far, nothing is assumed about the behaviour of  $B_n$  for indefinitely increasing values of  $n$ . Since  $\epsilon$  is arbitrary, we assume  $\frac{B_n}{n^2 \epsilon^2} \rightarrow 0$ , as  $n$  becomes indefinitely large. Thus, having chosen two arbitrary small positive numbers  $\epsilon$  and  $\eta$ , number  $n_0$  can be found so that the inequality

$$\frac{B_n}{n^2 \epsilon^2} < \eta,$$

will hold for  $n > n_0$ . Consequently, we shall have

$$P \left\{ \left| \frac{X_1 + X_2 + \dots + X_n}{n} - \frac{\mu_1 + \mu_2 + \dots + \mu_n}{n} \right| < \epsilon \right\} \geq 1 - \eta$$

for all  $n > n_0(\varepsilon, \eta)$ .

This conclusion leads to the following important result, known as the (Weak) Law of Large Numbers:

"With the probability approaching unity or certainty as near as we please, we may expect that the arithmetic mean of values actually assumed by  $n$  random variables will differ from the arithmetic mean of their expectations by less than any given number, however small, provided the number of variables can be taken sufficiently large and provided the condition

$$\frac{B_n}{n^2} \rightarrow 0 \text{ as } n \rightarrow \infty$$

is fulfilled".

**Remarks.** 1. Weak law of large numbers can also be stated as follows:

$$\bar{X}_n \xrightarrow{P} \bar{\mu}_n$$

provided  $\frac{B_n}{n^2} \rightarrow 0$  as  $n \rightarrow \infty$ , symbols having their usual meanings.

2. For the existence of the law we assume the following conditions:

- (i)  $E(X_i)$  exists for all  $i$ ,
- (ii)  $B_n = \text{Var}(X_1 + X_2 + \dots + X_n)$  exists, and
- (iii)  $B_n/n^2 \rightarrow 0$  as  $n \rightarrow \infty$ .

Condition (i) is necessary, without it the law itself cannot be stated. But the conditions (ii) and (iii) are not necessary, (iii) is however a sufficient condition.

3. If the variables  $X_1, X_2, \dots, X_n$  are independent and identically distributed, i.e., if  $E(X_i) = \mu$  (say), and  $\text{Var}(X_i) = \sigma^2$  (say) for all  $i = 1, 2, \dots, n$  then

$$B_n = \text{Var}(X_1 + X_2 + \dots + X_n) = \sum_{i=1}^n \text{Var}(X_i)$$

the covariance terms vanish, since variables are independent.

$$\therefore B_n = n\sigma^2 \quad \dots (*)$$

$$\text{Hence} \quad \lim_{n \rightarrow \infty} \frac{B_n}{n^2} = \lim_{n \rightarrow \infty} (\sigma^2/n) = 0$$

Thus, the law of large number holds for the sequence  $\{X_n\}$  of i.i.d.r.v.'s and we get

$$P \left\{ \left| \frac{X_1 + X_2 + \dots + X_n}{n} - \mu \right| < \varepsilon \right\} > 1 - \eta \quad \forall n > n_0$$

$$\text{i.e.,} \quad P \{ |\bar{X}_n - \mu| < \varepsilon \} \rightarrow 1 \text{ as } n \rightarrow \infty$$

$$\Rightarrow P \{ |\bar{X}_n - \mu| \geq \varepsilon \} \rightarrow 0 \text{ as } n \rightarrow \infty$$

where  $\bar{X}_n$  is the mean of the  $n$  random variables  $X_1, X_2, \dots, X_n$ . This result implies that  $\bar{X}_n$  converges in probability to  $\mu$ , i.e.,

$$\bar{X}_n \xrightarrow{P} \mu$$

**Note.** If  $\bar{X}_n$  is the mean of  $n$  i.i.d. r.v.'s  $X_1, X_2, \dots, X_n$  with

$$E(X_i) = \mu; \text{Var}(X_i) = \sigma^2, \text{ then} \quad [\text{On using (*)}]$$

$$E(\bar{X}_n) = \mu; \text{and } \text{Var}(\bar{X}_n) = \text{Var}\left(\sum_{i=1}^n x_i / n\right) \quad \dots(6.80)$$

**Theorem 6.32.** If the variables are uniformly bounded then the condition,

$$\lim_{n \rightarrow \infty} \frac{B_n}{n^2} = 0$$

is necessary as well as sufficient for WLLN to hold.

**Proof.** Let  $\xi_i = X_i - a_i$ , where  $E(X_i) = a_i$ ; then  $E(\xi_i) = 0$ , ( $i = 1, 2, \dots, n$ ).

Since  $X_i$ 's are uniformly bounded, there exists a positive number  $c < \infty$  such that  $|\xi_i| < c$ .

$$\text{If } p = P[|\xi_1 + \xi_2 + \dots + \xi_n| \leq n\epsilon]$$

$$\text{then } 1-p = P[|\xi_1 + \xi_2 + \dots + \xi_n| > n\epsilon]$$

$$\text{Let } U_n = \xi_1 + \xi_2 + \dots + \xi_n,$$

$$\text{then } E(U_n) = \sum_{i=1}^n E(\xi_i) = 0$$

$$\text{and } \text{Var}(U_n) = E(U_n^2) = B_n \text{ (say).}$$

$$\Rightarrow B_n = \int_0^\infty U_n^2 dF(U_n), \text{ where } F(U_n) \text{ is d.f. of } U_n$$

$$= \int_{U_n^2 \leq n^2 \epsilon^2} U_n^2 dF + \int_{U_n^2 > n^2 \epsilon^2} U_n^2 dF$$

$$\leq n^2 \epsilon^2 \int_{|U_n| \leq n\epsilon} dF + n^2 c^2 \int_{|U_n| > n\epsilon} dF$$

$$\leq n^2 \epsilon^2 p + n^2 c^2 (1-p)$$

$$\therefore \frac{B_n}{n^2} \leq \epsilon^2 p + c^2 (1-p)$$

If the law of large numbers holds,

$$1-p = P[|\xi_1 + \xi_2 + \dots + \xi_n| > n\epsilon] \rightarrow 0 \text{ as } n \rightarrow \infty.$$

Hence as  $n \rightarrow \infty$ ,  $(1-p) \rightarrow 0$ , and

$$\frac{B_n}{n^2} < \epsilon^2 p + c^2 \delta, \epsilon \text{ and } \delta \text{ being arbitrarily small positive numbers.}$$

$$\text{Hence } \frac{B_n}{n^2} \rightarrow 0 \text{ as } n \rightarrow \infty.$$

**6.15.1. Bernoulli's Law of Large Numbers.** Let there be  $n$  trials of an event, each trial resulting in a success or failure. If  $X$  is the number of successes in  $n$  trials with constant probability  $p$  of success for each trial, then  $E(X) = np$  and  $\text{Var}(X) = npq$ ,  $q = 1-p$ . The variable  $X/n$  represents the proportion of successes or the relative frequency of successes, and

$$E(X/n) = \frac{1}{n} E(X) = p, \text{ and } \text{Var}(X/n) = \frac{1}{n^2} \text{Var}(X) = \frac{pq}{n}$$

Then

$$P\left\{\left|\frac{X}{n} - p\right| < \epsilon\right\} \rightarrow 1 \text{ as } n \rightarrow \infty \quad \dots(6.81)$$

$$\Rightarrow P\left\{\left|\frac{X}{n} - p\right| \geq \epsilon\right\} \rightarrow 0 \text{ as } n \rightarrow \infty \quad \dots(6.81a)$$

for any assigned  $\epsilon > 0$ . This implied that  $(X/n)$  converges in probability to  $p$  as  $n \rightarrow \infty$ .

**Proof.** Applying Chebychev's Inequality [Form (6.73 b)] to the variable  $X/n$ , we get for any  $\epsilon > 0$ ;

$$\begin{aligned} & P\left\{\left|\frac{X}{n} - E\left(\frac{X}{n}\right)\right| \geq \epsilon\right\} \frac{\text{Var}(X/n)}{\epsilon^2} \\ \Rightarrow & P\left\{\left|\frac{X}{n} - p\right| < \epsilon\right\} \rightarrow 1 \text{ as } n \rightarrow \infty \end{aligned}$$

since the maximum value of  $pq$  is at  $p = q = 1/2$  i.e.,  $\max(pq) = 1/4$  i.e.,  $pq \leq 1/4$ .

Since  $\epsilon$  is arbitrary, we get

$$\begin{aligned} & P\left\{\left|\frac{X}{n} - p\right| \geq \epsilon\right\} \rightarrow 0 \text{ as } n \rightarrow \infty \\ \Rightarrow & P\left\{\left|\frac{X}{n} - p\right| < \epsilon\right\} \rightarrow 1 \text{ as } n \rightarrow \infty \end{aligned}$$

**6.15.2. Markoff's Theorem.** *The law of large numbers holds if for some  $\delta > 0$ , all the mathematical expectations*

$$E(|X_i|^{1+\delta}); i = 1, 2, \dots \quad \dots(6.82)$$

*exist and are bounded.*

**6.15.3. Khinchin's Theorem.** *If  $X_i$ 's are identically and independently distributed random variables, the only condition necessary for the law of large numbers to hold is that  $E(X_i); i = 1, 2, \dots$  should exist.*

**Theorem 6.33.** *Let  $\{X_n\}$  be any sequence of r.v.'s. Write :*

$$Y_n = [S_n - E(S_n)]/n \text{ where } S_n = X_1 + X_2 + \dots + X_n.$$

*A necessary and sufficient condition for the sequence  $\{X_n\}$  to satisfy the W.L.L.N. is that*

$$E\left\{\frac{Y_n^2}{1 + Y_n^2}\right\} \rightarrow 0 \text{ as } n \rightarrow \infty. \quad \dots(6.83)$$

**Proof.** If Part: Let us assume that (6.83) holds. We shall prove  $\{X_n\}$  satisfies W.L.L.N.

For real numbers  $a, b ; a \geq b > 0$  we have:

$$\therefore a \geq b \Rightarrow a + ab \geq b + ab \quad \dots (*)$$

Let us define the event  $A = \{ |Y_n| \geq \varepsilon \}$ .

$$w \in A_x \Rightarrow |Y_n| \geq \varepsilon \Rightarrow |Y_n|^2 \geq \varepsilon^2 > 0$$

$\therefore$  Taking  $a = Y_n^2$  and  $b = \varepsilon^2$  in (\*), we define another event  $B$  as follows:

$$B_n = \left\{ \left( \frac{Y_n^2}{1+Y_n^2} \right) \left( \frac{1+\varepsilon^2}{\varepsilon^2} \right) \geq 1 \right\} = \left\{ \frac{Y_n^2}{1+Y_n^2} \geq \frac{\varepsilon^2}{1+\varepsilon^2} \right\}$$

Since  $w \in A \Rightarrow w \in B$ ,  $A \subseteq B \Rightarrow P(A) \leq P(B)$

$$\begin{aligned} \therefore P[|Y_n| \geq \varepsilon] &\leq P\left[\frac{y_n^2}{1+Y_n^2} \frac{\varepsilon^2}{1+\varepsilon^2}\right] \\ &\leq \frac{E[Y_n^2/(1+Y_n^2)]}{\varepsilon^2/(1+\varepsilon^2)} \end{aligned}$$

[By Markov's Inequality (6.75)]

$\rightarrow 0$  as  $n \rightarrow \infty$  [By assumption (6.83)]

$$\therefore P[|Y_n| \geq \varepsilon] \rightarrow 0 \text{ as } n \rightarrow \infty$$

$$\Rightarrow \lim_{n \rightarrow \infty} P\left[\left|\frac{S_n - E(S_n)}{n}\right| \geq \varepsilon\right] \rightarrow 0$$

$\Rightarrow$  WI.LN holds for the sequence  $\{X_n\}$  of r.v.'s.

Conversely, if  $\{X_n\}$  satisfies WLLN, we shall establish (6.83). Let us assume that  $X_i$ 's are continuous and let  $Y_n$  have p.d.f.  $f_n(y)$ . Then

$$\begin{aligned} E\left\{\frac{Y_n^2}{1+Y_n^2}\right\} &= \int_{-\infty}^{\infty} \frac{y^2}{1+y^2} \cdot f_n(y) dy \\ &= \left( \int_A + \int_{A^c} \right) \frac{y^2}{1+y^2} f_n(y) dy \end{aligned}$$

where  $A = \{|Y| \geq \varepsilon\}$  and  $A^c = \{|y| < \varepsilon\}$

$$\begin{aligned}
 E\left(\frac{Y_n^2}{1+Y_n^2}\right) &\leq \int_A 1 \cdot f_n(y) dy + \int_{A^c} y^2 \cdot f_n(y) dy \quad \left| \because \frac{y^2}{1+y^2} < 1 \text{ and } \frac{y^2}{1+y^2} < y^2 \right| \\
 &\leq P(A) + \varepsilon^2 \int_{A^c} f_n(y) dy \quad (\because \text{On } A^c : |y| < \varepsilon) \\
 &= P(A) + \varepsilon^2 \cdot P(A^c) \\
 &\leq P(A) + \varepsilon^2 \quad (\because P(A^c) < 1) \\
 \Rightarrow E\left[\frac{Y_n^2}{1+Y_n^2}\right] &\leq P\{|Y_n| \geq \varepsilon\} + \varepsilon^2 \quad \dots (**)
 \end{aligned}$$

But since  $\{X_n\}$  satisfies WLLN, we have :

$$\lim_{n \rightarrow \infty} P(|Y_n| \geq \varepsilon) \rightarrow 0$$

and since  $\varepsilon^2$  is arbitrarily small positive number, we get on taking limits in (\*\*),

$$\lim_{n \rightarrow \infty} E\left[\frac{Y_n^2}{1+Y_n^2}\right] \rightarrow 0$$

**Corollary.** Let  $X_1, X_2, \dots, X_n$  be sequence of independent r.v.'s such that  $\text{Var}(X_i) < \infty$  for  $i = 1, 2, \dots$  and

$$\frac{B_n}{n^2} = \frac{\text{Var}\left(\sum_{i=1}^n X_i\right)}{n^2} = \frac{\text{Var}(S_n)}{n^2} \rightarrow 0 \text{ as } n \rightarrow \infty$$

Then WLLN holds.

**Proof.** We have :

$$\begin{aligned}
 \frac{Y_n^2}{1+Y_n^2} &\leq Y_n^2 = \left[\frac{S_n - E(S_n)}{n}\right]^2 \\
 \Rightarrow E\left[\frac{Y_n^2}{1+Y_n^2}\right] &\leq \frac{1}{n^2} \cdot E[S_n - E(S_n)]^2 = \frac{\text{Var } S_n}{n^2} = \frac{B_n}{n^2} \\
 \therefore \lim_{n \rightarrow \infty} E\left[\frac{Y_n^2}{1+Y_n^2}\right] &\leq \lim_{n \rightarrow \infty} \frac{B_n}{n^2} \rightarrow 0 \quad (\text{By assumption})
 \end{aligned}$$

Hence by the above theorem WLLN holds for the sequence  $\{X_n\}$  of r.v.'s.

**Remark.** The result of Theorem 6-33 holds even if  $E(X_i)$  does not exist. In this case we simply define  $Y_n = [S_n/n]$  rather than  $[S_n - E(S_n)]/n$ .

**Example 6-48.** A symmetric die is thrown 600 times. Find the lower bound for the probability of getting 80 to 120 sixes.

**Solution.** Let  $S$  be total number of successes.

Then

$$E(S) = np = 600 \times \frac{1}{6} = 100$$

$$V(S) = npq = 600 \times \frac{1}{6} \times \frac{5}{6} = \frac{500}{6}$$

Using Chebychev's inequality, we get

$$P(|S - E(S)| < k\sigma) \geq 1 - \frac{1}{k^2}$$

$$\Rightarrow P(|S - 100| < k\sqrt{500/6}) \geq 1 - \frac{1}{k^2}$$

$$\Rightarrow P(100 - k\sqrt{500/6} < S < 100 + k\sqrt{500/6}) \geq 1 - \frac{1}{k^2}$$

Taking  $k = \frac{20}{\sqrt{500/6}}$ , we get

$$P(80 \leq S \leq 120) \geq 1 - \frac{1}{400 \times (6/500)} = \frac{19}{24}$$

**Example 6.49.** Use Chebychev's inequality to determine how many times a fair coin must be tossed in order that the probability will be at least 0.90 that the ratio of the observed number of heads to the number of tosses will lie between 0.4 and 0.6.

[Madras Univ. B.Sc (Stat.) Oct 1991; Delhi Univ. B.Sc. (Stat Hons.) 1989]

**Solution.** As in the proof of Bernoulli's Law of Large Numbers, we get for any  $\epsilon > 0$ ,

$$P\left\{\left|\frac{X}{n} - p\right| \geq \epsilon\right\} \leq \frac{1}{4n\epsilon^2}$$

$$\Rightarrow P\left\{\left|\frac{X}{n} - p\right| < \epsilon\right\} \geq 1 - \frac{1}{4n\epsilon^2}$$

Since  $p = 0.5$  (as the coin is unbiased) and we want the proportion of successes  $X/n$  to lie between 0.4 and 0.6, we have

$$\left|\frac{X}{n} - p\right| \leq 0.1$$

Thus choosing  $\epsilon = 0.1$ , we have

$$P\left\{\left|\frac{X}{n} - p\right| < 0.1\right\} \geq 1 - \frac{1}{4n(0.1)^2} = 1 - \frac{1}{0.04n}$$

Since we want this probability to be 0.90, we fix

$$1 - \frac{1}{0.04n} = 0.90$$

$$\Rightarrow 0.10 = \frac{1}{0.04n}$$

$$\Rightarrow n = \frac{1}{0.10 \times 0.04} = 250$$

Hence the required number of tosses is 250.

**Example 6-50.** For geometric distribution  $p(x) = 2^{-x}$ ;  $x = 1, 2, 3, \dots$ , prove that Chebychev's inequality gives

$$P(|X - 2| \leq k) > \frac{1}{k^2}$$

while the actual probability is  $\frac{15}{16}$ . [Nagpur Univ. B.Sc. (Stat.), 1989]

**Solution.**  $E(X) = \sum_{x=1}^{\infty} x \cdot \frac{1}{2^x} = \frac{1}{2} + \frac{2}{2^2} + \frac{3}{2^3} + \frac{4}{2^4} + \dots$

$$= \frac{1}{2}(1 + 2A + 3A^2 + 4A^3 + \dots), \quad (A = 1/2)$$

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

$$E(X^2) = \sum_{x=1}^{\infty} x^2 \cdot \frac{1}{2^x} = \frac{1}{2^2} + \frac{4}{2^3} + \frac{9}{2^4} + \dots$$

$$= \frac{1}{4}[1 + 4A + 9A^2 + \dots] = \frac{1}{4}(1 + A)(1 - A)^{-3} = 6$$

[See Example 6-17]

$$\therefore \text{Var}(X) = E(X^2) - [E(X)]^2 = 6 - 4 = 2$$

Using Chebychev's inequality, we get

$$P(|X - E(X)| \leq k\sigma) > 1 - \frac{1}{k^2}$$

With  $k = \sqrt{2}$ , we get

$$P(|X - 2| \leq \sqrt{2}, \sqrt{2}) > 1 - \frac{1}{2} = \frac{1}{2}$$

$$\Rightarrow P(|X - 2| \leq 2) > \frac{1}{2}$$

And the actual probability is given by

$$P(|X - 2| \leq 2) = P(0 \leq X \leq 4) = P\{X = 1, 2, 3 \text{ or } 4\}$$

$$= \frac{1}{2} + (\frac{1}{2})^2 + (\frac{1}{2})^3 + (\frac{1}{2})^4 = \frac{15}{16}$$

**Example 6-51.** Does there exist a variate  $X$  for which

$$P[\mu_x - 2\sigma \leq X \leq \mu_x + 2\sigma] = 0.6 \quad \dots(*)$$

[Delhi Univ. B.Sc.(Maths Hons.) 1983]

**Solution.** We have :

$$P[\mu_x - 2\sigma \leq X \leq \mu_x + 2\sigma] = P[|X - \mu_x| \leq 2\sigma]$$

$$\geq 1 - \frac{1}{4} = 0.75 \quad (\text{Using Chebychev's Inequality})$$

Since lower bound for the probability is 0.75, there does not exist a r.v.  $X$  for which (\*) holds.

**Example 6.52.** (a) For the discrete variate with density

$$f(x) = \frac{1}{8} I_{(-1)}(x) + \frac{6}{8} I_{(0)}(x) + \frac{1}{8} I_{(1)}(x)$$

evaluate  $P[|X - \mu_x| \geq 2\sigma_x]$ . [Delhi Univ. B.Sc.(Maths Hons.) 1989]

(b) Compare this result with that obtained on using Chebychev's inequality.

**Hint.** (a) Here  $X$  has the probability distribution :

$$\begin{array}{cccccc} x : & -1 & 0 & 1 & \therefore E(X) = -1 \times \frac{1}{8} + 1 \times \frac{1}{8} = 0 \\ p(x) : & 1/8 & 6/8 & 1/8 & EX^2 = 1 \times \frac{1}{8} + 1 \times \frac{1}{8} = \frac{1}{4} \\ \therefore \text{Var}(X) = E(X^2) - [E(X)]^2 = 1/4 & \Rightarrow \sigma_x = 1/2 \\ P[|X - \mu_x| \geq 2\sigma_x] = P[|X| \geq 1] & = 1 - P(|X| < 1) \\ & = 1 - P[-1 < X < 1] = 1 - P(X = 0) = 1/4 \end{array}$$

(b)  $P[|X - \mu_x| \geq 2\sigma_x] \leq \frac{1}{4}$  (By Chebychev's Inequality)

In this case both results are same.

**Note.** This example shows that, in general, Chebychev's inequality cannot be improved.

**Example 6.53.** Two unbiased dice are thrown. If  $X$  is the sum of the numbers showing up, prove that

$$P(|X - 7| \geq 3) \leq \frac{35}{54}.$$

Compare this with the actual probability. [Karnataka Univ. B.Sc., 1988]

**Solution.** The probability distribution of the r.v.  $X$  (the sum of the numbers on the two dice) is as given below :

$X$	Favourable cases (distinct)	Probability ( $p$ )
2	(1, 1)	1/36
3	(1, 2), (2, 1)	2/36
4	(1, 3), (3, 1), (2, 2)	3/36
5	(1, 4), (4, 1), (2, 3), (3, 2)	4/36
6	(1, 5), (5, 1), (2, 4), (4, 2), (3, 3)	5/36
7	(1, 6), (6, 1), (2, 5), (5, 2), (3, 4), (4, 3)	6/36
8	(2, 6), (6, 2), (3, 5), (5, 3), (4, 4)	5/36
9	(3, 6), (6, 3), (4, 5), (5, 4)	4/36
10	(4, 6), (6, 4), (5, 5)	3/36
11	(5, 6), (6, 5)	2/36
12	(6, 6)	1/36

$$\begin{aligned} E(X) &= \sum_x p \cdot x \\ &= \frac{1}{36} (2 + 6 + 12 + 20 + 30 + 42 + 40 + 36 + 30 + 22 + 12) \end{aligned}$$

$$= \frac{1}{36} (252) = 7$$

$$\begin{aligned} E(X^2) &= \sum_x p \cdot x^2 \\ &= \frac{1}{36} [4 + 18 + 48 + 100 + 180 + 294 + 320 + 324 + 300 \end{aligned}$$

$$+ 242 + 144]$$

$$= \frac{1}{36} (1974) = \frac{1974}{36} = \frac{329}{6}$$

$$\text{Var}(X) = E(X^2) - [E(X)]^2 = \frac{329}{6} - (7)^2 = \frac{35}{6}$$

By Chebychev's inequality, for  $k > 0$ , we have

$$P(|X - \mu| \geq k) \leq \frac{\text{Var } X}{k^2}$$

$$\Rightarrow P(|X - 7| \geq 3) \leq \frac{35/6}{9} = \frac{35}{54} \quad (\text{Taking } k = 3)$$

**Actual Probability :**

$$\begin{aligned} P(|X - 7| \geq 3) &= 1 - P(|X - 7| < 3) \\ &= 1 - P(4 < X < 10) \\ &= 1 - [P(X = 5) + P(X = 6) + P(X = 7) \\ &\quad P(X = 8) + P(X = 9)] \\ &= 1 - \frac{1}{36} [4 - 5 + 6 + 5 + 4] = 1 - \frac{24}{36} = \frac{1}{3} \end{aligned}$$

**Example 6-54.** If  $X$  is the number scored in a throw of a fair die, show that the Chebychev's inequality gives

$$P(|X - \mu| > 2.5) < 0.47,$$

where  $\mu$  is the mean of  $X$ , while the actual probability is zero.

[Kerala Univ. B.Sc., Oct. 1989]

**Solution.** Here  $X$  is a random variable which takes the values 1, 2, ..., 6, each with probability 1/6. Hence

$$E(X) = \frac{1}{6} (1 + 2 + \dots + 6) = \frac{1}{6} \cdot \frac{6 \times 7}{2} = \frac{7}{2}$$

$$E(X^2) = \frac{1}{6} (1^2 + 2^2 + \dots + 6^2) = \frac{1}{6} \cdot \frac{6 \times 7 \times 13}{6} = \frac{91}{6}$$

$$\therefore \text{Var}(X) = E(X^2) - [E(X)]^2 = \frac{91}{6} - \frac{49}{4} = \frac{35}{12} = 2.9167$$

For  $k > 0$ , Chebychev's inequality gives

$$P[|X - E(X)| > k] < \frac{\text{Var } X}{k^2}$$

Choosing  $k = 2.5$ , we get

$$P[|X - \mu| > 2.5] < \frac{2.9167}{6.25} = 0.47$$

The actual probability is given by

$$p = P[|X - 3.5| > 2.5]$$

$\Rightarrow P[X \text{ lies outside the limits } (3.5 - 2.5, 3.5 + 2.5), \text{i.e., } (1, 6)]$

But since  $X$  is the number on the dice when thrown, it cannot lie outside the limits of 1 and 6.

$$\therefore p = p(\varphi) = 0$$

**Example 6.55.** If the variable  $X_p$  assumes the value  $2^{p-2\log p}$  with probability  $2^{-p}$ ;  $p = 1, 2, \dots$ , examine if the law of large numbers holds in this case.

**Solution.** Putting  $p = 1, 2, 3, \dots$ , we get

$$2^{1-2\log 1}, 2^{2-2\log 2}, 2^{3-2\log 3}, \dots$$

as the values of the variables  $X_1, X_2, X_3, \dots$  respectively, and

$$\frac{1}{2}, \frac{1}{2^2}, \frac{1}{2^3}, \dots$$

their corresponding probabilities. Therefore,

$$\begin{aligned} E(X_k) &= 2^{1-2\log 1} \cdot \frac{1}{2} + 2^{2-2\log 2} \cdot \frac{1}{2^2} + \dots = \sum_{p=1}^{\infty} 2^{p-2\log p} \cdot 2^{-p} \\ &= \sum_{p=1}^{\infty} \frac{1}{2^{2\log p}} \end{aligned}$$

Let  $U = 2^{2\log p}$ , then

$$\log U = 2 \log p \log 2 = \log p \cdot \log 2^2 = \log 4 \cdot \log p = \log p^{\log 4}$$

$$\therefore E(X_k) = \sum_{p=1}^{\infty} \frac{1}{p^{\log 4}} = 1 + \frac{1}{2^{\log 4}} + \frac{1}{3^{\log 4}} + \dots$$

which is a convergent series.

$\left[ \because \sum_{n=1}^{\infty} \frac{1}{n^p}$  is convergent if and only if  $p > 1 \right]$

Therefore, the mathematical expectation of the variables  $X_1, X_2, \dots$ , exists. Thus by Khinchin's theorem, the law of large numbers holds in this case.

**Example 6.56.** Let  $X_1, X_2, \dots, X_n$  be i.i.d. variables with mean  $\mu$  and variance  $\sigma^2$  and as  $n \rightarrow \infty$ ,

$$(X_1^2 + X_2^2 + \dots + X_n^2)/n \xrightarrow{P} c,$$

for some constant  $c$ ;  $(0 \leq c \leq \infty)$ . Find  $c$ . [Delhi Univ. B.Sc. (Stat. Hons.), 1989]

**Solution.** We are given  $E(X_i) = \mu$ ;  $\text{Var}(X_i) = \sigma^2$ ;  $i = 1, 2, \dots, n$ .

$$\therefore E(\bar{X}_n^2) = \text{Var}(\bar{X}_n) + [E(\bar{X}_n)]^2 = \sigma^2 + \mu^2 \text{ (finite)}; i = 1, 2, \dots, n.$$

Since  $E(\bar{X}_n^2)$  is finite; by Khinchine's Theorem WLLN holds for the sequence  $\bar{X}_n^2$  of i.i.d. r.v.'s so that

$$(X_1^2 + X_2^2 + \dots + X_n^2)/n \xrightarrow{P} E(X_i^2), \text{ as } n \rightarrow \infty$$

$$\Rightarrow (X_1^2 + X_2^2 + \dots + X_n^2)/n \xrightarrow{P} \sigma^2 + \mu^2 = c, \text{ as } n \rightarrow \infty$$

$$\text{Hence } c = \sigma^2 + \mu^2.$$

**Example 6.57.** How large a sample must be taken in order that the probability will be at least 0.95 that  $\bar{X}_n$  will be lie within 0.5 of  $\mu$ .  $\mu$  is unknown and  $\sigma = 1$ . [Delhi Univ. B.Sc. (Maths Hons.) 1988]

**Solution.** We have:  $E(\bar{X}_n) = \mu$  and  $\text{Var}(\bar{X}_n) = \sigma^2/n$

[c.f. § 6.15, = n (6.80)]

Applying Chebychev's inequality to the r.v.  $\bar{X}_n$  we get, for any  $c > 0$

$$P(|\bar{X}_n - E(\bar{X}_n)| < c) \geq 1 - \frac{\text{Var}(\bar{X}_n)}{c^2}$$

$$\Rightarrow P(|\bar{X}_n - \mu| < c) \geq 1 - \frac{\sigma^2}{nc^2} \quad \dots (*)$$

We want  $n$  so that

$$P(|\bar{X}_n - \mu| < 0.5) \geq 0.95 \quad \dots (**)$$

Comparing (\*) and (\*\*) we get :

$$c = 0.5 = 1/2 \quad \text{and} \quad 1 - \frac{\sigma^2}{nc^2} = 0.95 \quad \text{and} \quad \sigma = 1 \quad (\text{Given})$$

$$\therefore 1 - \frac{4}{n} = 0.95 \Rightarrow \frac{4}{n} = 0.05 = \frac{1}{20} \Rightarrow n = 80.$$

Hence  $n \geq 80$ .

**Example 6.58.** (a) Let  $X_i$  assume that values  $i$  and  $-i$  with equal probabilities. Show that the law of large numbers cannot be applied to the independent variables  $X_1, X_2, \dots$ , i.e.,  $X_i$ 's.

(b) If  $X_i$  can have only two values with equal probabilities  $i^\alpha$  and  $-i^\alpha$ , show that the law of large numbers can be applied to the independent variables  $X_1, X_2, \dots$ , if  $\alpha < \frac{1}{2}$ .

**Solution.** (a) We have

$$P(X_i = i) = \frac{1}{2}, \quad P(X_i = -i) = \frac{1}{2}$$

$$E(X_i) = \frac{1}{2}(i) + \frac{1}{2}(-i) = 0; \quad i = 1, 2, 3, \dots$$

$$\text{Var}(X_i) = E(X_i^2) = \frac{i^2}{2} + \frac{i^2}{2} = i^2 \quad [\because E(X_i) = 0] \quad \dots (*)$$

$$B_n = V(X_1 + X_2 + \dots + X_n) = V(X_1) + V(X_2) + \dots + V(X_n)$$

$$= (1^2 + 2^2 + \dots + n^2) = \frac{n(n+1)(2n+1)}{6} \quad \dots[\text{From } (*)]$$

$\therefore \frac{B_n}{n^2} \rightarrow \infty$  as  $n \rightarrow \infty$ . Hence we cannot draw any conclusion whether WLLN holds or not. Here, we need to apply further tests. (See Theorem 6.33 page 6-104)

$$(b) \quad E(X_i) = \frac{i^\alpha}{2} + \left( \frac{-i^\alpha}{2} \right) = 0$$

$$E(X_i^2) = \frac{(i^\alpha)^2}{2} + \frac{(-i^\alpha)^2}{2} = i^{2\alpha}$$

$$V(X_i) = E(X_i^2) - |E(X_i)|^2 = i^{2\alpha}$$

$$B_n = V(X_1 + X_2 + \dots + X_n) = \sum_{i=1}^n V(X_i) = \sum_{i=1}^n i^{2\alpha}$$

$$= 1^{2\alpha} + 2^{2\alpha} + \dots + n^{2\alpha} = \int_0^n x^{2\alpha} dx$$

[From Euler Maclaurin's Formula]

$$= \left| \frac{x^{2\alpha+1}}{2\alpha+1} \right|_0^n = \frac{n^{2\alpha+1}}{2\alpha+1}$$

$$\therefore \frac{B_n}{n^2} = \frac{n^{2\alpha+1}}{2\alpha+1} \rightarrow 0 \text{ if } 2\alpha+1 < 0 \Rightarrow \alpha < -\frac{1}{2}$$

Hence the result.

**Example 6.59.** Let  $\{X_k\}$  be mutually independent and identically distributed random variables with mean  $\mu$  and finite variance. If  $S_n = X_1 + X_2 + \dots + X_n$ , prove that the law of large numbers does not hold for the sequence  $\{S_n\}$ .

**Solution.** The variables now are  $S_1, S_2, \dots, S_n$ .

$$B_n = V(S_1 + S_2 + \dots + S_n)$$

$$= V(X_1 + (X_1 + X_2) + (X_1 + X_2 + X_3) + \dots + (X_1 + X_2 + \dots + X_n))$$

$$= V[nX_1 + (n-1)X_2 + \dots + 2X_{n-1} + X_n]$$

$$= n^2 V(X_1) + (n-1)^2 V(X_2) + \dots + 2^2 V(X_{n-1}) + 1^2 V(X_n)$$

(Covariance terms vanish since variables are independent.)

$$\text{Let } V(X_i) = \sigma^2 \text{ for all } i.$$

(Since the variables are identically distributed.)

$$\therefore B_n = (1^2 + 2^2 + \dots + n^2) \sigma^2 + \frac{n(n+1)(2n+1)}{6} \sigma^2$$

$$\therefore \frac{B_n}{n^2} = \frac{(n+1)(2n+1)}{6n} \cdot \sigma^2 \rightarrow \infty \text{ as } n \rightarrow \infty$$

**Example 6.60.** Examine whether the weak law of large numbers holds for the sequence  $\{X_k\}$  of independent random variables defined as follows:

$$P[X_k = \pm 2^k] = 2^{-(2k+1)}$$

$$P[X_k = 0] = 1 - 2^{-2k} \quad [\text{Delhi Univ. B.Sc. (Maths Hons.), 1988}]$$

**Solution.** We have

$$\begin{aligned} E(X_k) &= 2^k \cdot 2^{-(2k+1)} + (-2^k) \cdot 2^{-(2k+1)} + 0 \times (1 - 2^{-2k}) \\ &= 2^{-(2k+1)} (2^k - 2^k) = 0 \end{aligned}$$

$$\begin{aligned} E(X_k^2) &= (2^k)^2 \cdot 2^{-(2k+1)} + (-2^k)^2 \cdot 2^{-(2k+1)} + 0^2 \times (1 - 2^{-2k}) \\ &= 2^{2k} \cdot 2^{-(2k+1)} + 2^{k^2} \cdot 2^{-(2k+1)} \\ &= \frac{1}{2} + \frac{1}{2} = 1 \end{aligned}$$

$$\therefore \text{Var}(X_k) = E(X_k^2) - E(X_k)^2 = 1 - 0 = 1$$

$$B_n = \text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{Var}(X_i)$$

[ $\because X_{it}$ 's ( $i = 1, 2, \dots, n$ ) are independent]

$$= \sum_{i=1}^n (1) = n$$

$$\therefore \lim_{n \rightarrow \infty} \frac{B_n}{n^2} = \lim_{n \rightarrow \infty} \frac{1}{n} \rightarrow 0$$

Hence (Weak) Law of large numbers, holds for the sequence of independent r.v.'s  $\{X_k\}$ .

**Example 6.61.** Let  $X_1, X_2, \dots, X_n$  be jointly normal with  $E(X_i) = 0$  and  $E(X_i^2) = 1$  for all  $i$  and  $\text{Cov}(X_i, X_j) = \rho$  if  $|j-i| = 1$  and  $= 0$ , otherwise. Examine if WLLN holds for the sequence  $\{X_n\}$ . ... (\*)

**Solution.** We have:

$$\text{Var}(X_i) = (X_i^2) - [E(X_i)]^2 = 1, \quad (i = 1, 2, \dots, n).$$

$$E(S_n) = E\left(\sum_{i=1}^n X_i\right) = 0$$

$$\begin{aligned} \text{Var}(S_n) &= \text{Var}\left(\sum_{i=1}^n X_i\right) \\ &= \sum_{i=1}^n \text{Var} X_i + 2 \sum_{i \leq j=1}^n \text{Cov}(X_i, X_j) \\ &= n + 2 \cdot (n-1) \rho \end{aligned}$$

[On using (\*)]

Since  $X_1, X_2, \dots, X_n$  are jointly normal,

$$S_n = \sum_{i=1}^n X_i \sim N(0, \sigma^2) \text{ where } \sigma^2 = n + 2(n-1)\rho > 0,$$

Taking  $Y_n = \frac{1}{n} \sum_{i=1}^n X_i = \frac{S_n}{n}$ , we have:

$$\begin{aligned} E\left[\frac{Y_n^2}{1+Y_n^2}\right] &= E\left[\frac{S_n^2}{n^2+S_n^2}\right] \\ &= \int_{-\infty}^{\infty} \left(\frac{x^2}{n^2+x^2}\right) \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2} dx \\ &= \frac{2}{\sqrt{2\pi}} \int_0^{\infty} \frac{\sigma^2 y^2}{n^2 + \sigma^2 y^2} \cdot e^{-y^2/2} dy ; \quad \left(\frac{x}{\sigma} = y\right) \\ &= \frac{2}{\sqrt{2\pi}} \left\{ n + 2(n-1)\rho \right\} \cdot \int_0^{\infty} \frac{y^2}{n^2 + y^2 \{n + 2(n-1)\rho\}} \cdot e^{-y^2/2} dy \\ &\leq \frac{n+2(n-1)\rho}{n^2} \cdot \frac{2}{\sqrt{2\pi}} \int_0^{\infty} y^2 \cdot e^{-y^2/2} dy \\ &\rightarrow 0 \text{ as } n \rightarrow \infty \end{aligned}$$

$$\therefore \lim_{n \rightarrow \infty} E\left[\frac{Y_n^2}{1+Y_n^2}\right] \rightarrow 0$$

Hence by Theorem 6-33, WLLN holds for  $\{X_n\}$ , i.e.,  $\frac{S_n}{n} \xrightarrow{P} 0$  as  $n \rightarrow \infty$ .

**6-16. Borel-Cantelli Lemma. (Zero-One Law).** Let  $A_1, A_2, \dots$  be a sequence of events. Let  $A$  be the event "that an infinite number of  $A_n$  occur. That is  $\omega \in A$  if  $\omega \in A_n$  for an infinite number of values of  $n$  (but not necessarily every  $n$ ). But the set of such  $\omega$  is precisely  $\limsup A_n$ , i.e.,  $\overline{\limsup A_n}$ . Thus the event  $A$ , that an infinite number of  $A_n$  occur is just  $\overline{\limsup A_n}$ . Sometimes we are interested in the probability that an infinite number of the events  $A_n$  occur. Often this question is answered by means of the Borel-Cantelli lemma or its converse.

**Theorem 6-34. (Borel-Cantelli Lemma).** Let  $A_1, A_2, \dots$  be a sequence of events on the probability space  $(S, B, P)$  and let  $A = \overline{\limsup A_n}$ .

$$\text{if } \sum_{n=1}^{\infty} P(A_n) < \infty, \text{ then } P(A) = 0.$$

In other words, this states that if  $\sum P(A_n)$  converges then with probability one, only a finite number of  $A_1, A_2, \dots$  can occur.

**Proof.** Since  $A = \overline{\limsup A_n} = \overline{\bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m}$

we have  $A \subset \bigcup_{m=n}^{\infty} A_m$ , for every  $n$ .

Thus for each  $n$ ,

$$P(A) \leq \sum_{m=n}^{\infty} P(A_m)$$

Since  $\sum_{n=1}^{\infty} P(A_n)$  is convergent (given),  $\sum_{m=n}^{\infty} P(A_m)$ , being the remainder

term of a convergent series, tends to zero as  $n \rightarrow \infty$ .

$$\therefore P(A) \leq \sum_{m=n}^{\infty} P(A_m) \rightarrow 0 \text{ as } n \rightarrow \infty.$$

Thus  $P(A) = 0$  as required.

The result just proved does not require events  $A_1, A_2, \dots$  considered to be independent. For the converse result it is necessary to make this further assumption.

**Theorem 6.35. Borel-Cantelli Lemma (Converse).** Let  $A_1, A_2, \dots$  be independent events on  $(S, \mathcal{B}, P)$ ,  $A$  equal to  $\overline{\lim} A_n$ .

If  $\sum_{n=1}^{\infty} P(A_n) = \infty$ , then  $P(A) = 1$ .

**Proof.** Writing, in usual notation,  $\overline{A}_n$  for the complement  $S - A_n$  of  $A_n$ , we have for any  $m, n$  ( $m > n$ ).

$$\begin{aligned} \bigcap_{k=n}^{\infty} \overline{A}_k &\subset \bigcap_{k=n}^m \overline{A}_k \\ P\left(\bigcap_{k=n}^{\infty} \overline{A}_k\right) &\leq P\left(\bigcap_{k=n}^m \overline{A}_k\right) \\ &= \prod_{k=n}^m P(\overline{A}_k), \end{aligned}$$

because of the fact that if  $(A_n, A_{n+1}, \dots, A_m)$  are independent events, so are  $(\overline{A}_n, \overline{A}_{n+1}, \dots, \overline{A}_m)$ .

Hence

$$\begin{aligned} P\left(\bigcap_{k=n}^{\infty} \overline{A}_k\right) &\leq \prod_{k=n}^m [1 - P(A_k)] \\ &\leq \prod_{k=n}^m e^{-P(A_k)} \\ &= e^{-\sum_{k=n}^m P(A_k)} \quad \forall m \end{aligned}$$

[ $\because 1 - x \leq e^{-x}$  for  $x \geq 0$ ]

Since

$$\sum_{k=1}^{\infty} P(A_k) = \infty, \quad \sum_{k=n}^m P(A_k) \rightarrow \infty \text{ as } m \rightarrow \infty$$

and hence  $e^{-\sum_{k=n}^m P(A_k)} \rightarrow 0$  as  $m \rightarrow \infty$ .

$$\therefore P\left(\bigcap_{k=n}^{\infty} \overline{A}_k\right) = 0 \quad \dots (*)$$

$$\text{But } A = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k$$

$$\therefore \overline{A} = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} \overline{A}_k \quad (\text{De-Morgan's Law})$$

$$\Rightarrow P(\overline{A}) \leq \sum_{n=1}^{\infty} P\left(\bigcap_{k=n}^{\infty} \overline{A}_k\right) = 0 \quad [\text{From (*)}]$$

Hence  $P(A) = 1 - P(\overline{A}) = 1$ , as required.

If  $A_1, A_2, \dots$  are independent events it follows from Theorems 6-34 and 6-35 that the probability that an infinite number of them occur is either zero (when  $\sum_{n=1}^{\infty} P(A_n) < \infty$ ) or one [when  $\sum_{n=1}^{\infty} P(A_n) = \infty$ ]. This statement is a

special case of so-called "Zero One law" which we now state.

**Theorem 6-36. (Zero One Law).** If  $A_1, A_2, \dots$  are independent and if  $E$  belongs to the  $\sigma$ -field generated by the class  $(A_n, A_{n+1}, \dots)$  for every  $n$ , then  $P(E)$  is zero or one.

**Example 6-62.** What is the probability that in a sequence of Bernoulli trials with probability of success  $p$  for each trial, the pattern SFS appears infinitely often?

**Solution.** Let  $A_k$  be the event that the trial number  $k, k+1, k+2$  produce the sequence SFS ( $k = 0, 1, 2, \dots$ ). The events  $A_k$  are not mutually independent but the sequence  $A_1, A_4, A_7, A_{10}, \dots$  contains only mutually independent events (since no two depend on the outcome of the same trials).

$$p_k = P(A_k) = P(SFS) = p^2 q, \quad (q = 1 - p)$$

is independent of  $k$ , and hence the series

$$p_1 + p_4 + p_7 + \dots, \text{diverges}$$

Hence by B.C.T. (converse) the pattern SFS appears infinitely often with probability one.

**Example 6-63.** A bag contains one black ball and white balls. A ball is drawn at random. If a white ball is drawn, it is returned to the bag together with an additional white ball. If the black ball is drawn, it alone is returned to the bag.

Let  $A_n$  denote the event that the black ball is not drawn in the first  $n$  trials. Discuss the converse to Borel-Cantelli Lemma with reference to events  $A_n$ .

**Solution.**  $A_n =$  The event that blackball is not drawn in the first  $n$  trials. ...(\*)

= The event that each of the first  $n$  trials resulted in the draw of a white ball.

$$\Rightarrow P(A_n) = P(E_1 \cap E_2 \cap \dots \cap E_n),$$

where  $E_i$  is the event of drawing a white ball in the  $i$ th trial.

$$\therefore P(A_n) = P(E_1) P(E_2 | E_1) \dots P(E_3 | E_1 \cap E_2) \dots P(E_n | E_1 \cap E_2 \dots \cap E_{n-1})$$

$$= \frac{m}{m+1} \times \frac{m+1}{m+2} \times \dots \times \frac{m+n-1}{m+n} = \frac{m}{m+n}$$

(Since if first ball drawn is white it is returned together with an additional white ball, i.e., for the second draw the box contains 1 b,  $m+1$  W balls and

$$\therefore P(E_2 | E_1) = \frac{m+1}{m+2}, \text{ and so on.}$$

$$\begin{aligned} \sum_{n=1}^{\infty} PA(A_n) &= \sum_{n=1}^{\infty} \frac{m}{m+n} = m \sum_{n=1}^{\infty} \frac{1}{m+n} \\ &= m \left[ \frac{1}{m+1} + \frac{1}{m+2} + \frac{1}{m+3} + \dots \right] \\ &= m \left[ \sum_{n=1}^{\infty} \frac{1}{n} - \left( \sum_{n=1}^m \frac{1}{n} \right) \right] \quad \dots (***) \end{aligned}$$

But,  $\sum_{n=1}^{\infty} \frac{1}{n}$  is divergent, ( $\because \sum_{n=1}^{\infty} \frac{1}{n^p}$  is convergent, iff  $p > 1$ )

and,  $\sum_{n=1}^{\infty} \frac{1}{n}$  is finite,

$\therefore R.H.S.$  of  $(**)$  is divergent.

$$\text{Hence } \sum_{n=1}^{\infty} P(A_n) = \infty$$

From the definition of  $A_n$  in  $(*)$  it is obvious that  $A_n \downarrow$ .

$$\therefore A = \overline{\lim}_{n \rightarrow \infty} A_n = \limsup_{n \rightarrow \infty} A_n = \varphi$$

$$\therefore P(A) = P(\varphi) = 0$$

This result is *inconsistent* with converse of Borel-Cantelli Lemma, the reason being that the events  $A_n$  ( $n = 1, 2, \dots$ ) considered here are not independent,

$$\therefore P(A_i \cap A_j) = P(A_i) = \frac{m}{m+i} \neq P(A_i)P(A_j),$$

since for  $(j > i) A_j \subset A_i$  as  $A_n \downarrow$

### EXERCISE 6 (d)

1. State and prove Chebychev's inequality.
2. (a) A random variable  $X$  has a mean value of 5 and variance of 3.
  - What is the least value of  $\text{Prob}(|X - 5| < 3)$ ?
  - What value of  $h$  guarantees that  $\text{Prob}(|X - 5| < h) \geq 0.99$ ?
  - What is the least value of  $\text{Prob}(|X - 5| < 7.5)$ ?

(b) A random variable  $X$  takes the values  $-1, 1, 3, 5$  with associated probabilities  $1/6, 1/6, 1/6$  and  $1/2$ . Find by direct computation  $P(|X - 3| \geq 1)$ . Find an upper bound to this probability by applying Chebychev's inequality.

(c) If  $X$  denote the sum of the numbers obtained when two dice are thrown, use Chebychev's inequality to obtain an upper bound for  $P(|X - 7| > 4)$ . Compare this with the actualy probability.

3. (a) An unbaised coin is tossed 100 times. Show that the probability that the number of heads will be between 30 and 70 is greater than 0.93.

(b) Within what limits will the number of heads lie, with 95 p.c. probability, in 1000 tosses of a coin which is practically unbised?

(c) A symmetric die is thrown 720 times. Use Chebychev's inequality to find the lower bound for the probability of getting 100 to 140 sixes.

(d) Use Chebychev's inequality to determine how many times a fair coin must be tossed in order that the probability will be at least 0.95 that the ratio of the number of heads to the number of tosses will be between 0.45 and 0.55.

[Delhi Univ. B.Sc. (Stat. Hons.), 1988]

4. (a) If you wish to estimate the proportion of engineers and scientists who have studied probability theory and you wish your estimate to be correct within 2% with probability 0.95 or more, how large a sample would you take (i) if you have no idea what the true proportion is, (ii) if you are confident that the true proportion is less than 0.2?

[Burdwan Univ. (B.Sc. (Hons.), 1992)]

**Hint.** (i)  $\epsilon = 2\%$  or  $0.02$

$$P(|f_n - p| < 0.02) \geq 1 - \frac{1}{4n(0.02)^2} = 0.95$$

$$\therefore 0.05 = \frac{1}{4n(0.02)^2} \Rightarrow n = 12,500$$

$$(ii) p < 0.2, P(|f_n - p| \leq \epsilon) \geq 1 - \frac{p(1-p)}{n\epsilon^2}$$

$$\text{Now } p(1-p) < 0.16, \text{ therefore } 1 - \frac{0.16}{n\epsilon^2} = 0.95$$

$$\text{Hence } n = 50 \times 50 \times 20 \times 0.16 = 8000$$

(b) Let the sample mean of a random variable  $X$  be  $\bar{X}$  s.d.s. Then if at least 99 per cent of the values of  $X$  fall within  $K$  standard deviations from the mean, find  $K$ .

5. (a) If  $X$  is a r.v. such that  $E(X) = 3$  and  $E(X^2) = 13$ , use Chebychev's inequality to determine a lower bound for  $P(-2 < X < 8)$ .

[Delhi Univ. B.Sc. (Maths Hons.), 1990]

**Hint.**  $\mu_x = 3, \sigma_x^2 = 4 \Rightarrow \sigma_x = 2$ . Chebychev's inequality gives

$$P(|X - 3| < 2k) \geq 1 - 1/k^2 \Rightarrow P(3 - 2k < \bar{X} < 3 + 2k) \geq 1 - 1/k^2$$

Now taking  $k = 2.5$ , we get  $P(-2 < \bar{X} < 8) \geq 21/25$ .

(b) State and prove Chebychev's inequality. Use it to prove that in 2000 throws with a coin the probability that the number of heads lies between 900 and 1100 is at least 19/20. [Delhi Univ. B.Sc. (Maths Hons.), 1989]

6. (a) A random variable  $X$  has the density function  $e^{-x}$  for  $x \geq 0$ . Show that Chebychev's inequality gives  $P(|X - 1| > 2) < 1/4$  and show that the actual probability is  $e^{-3}$ .

(b) Let  $X$  have the p.d.f.:

$$f(x) = \frac{1}{2\sqrt{3}}, -\sqrt{3} < x < \sqrt{3}\\ = 0, \text{ elsewhere.}$$

Find the actual probability  $P(|X - \mu| \geq \frac{3}{2}\sigma)$  and compare it with the upper bound obtained by Chebychev's inequality.

7. If  $X$  has the distribution with p.d.f.

$$f(x) = e^{-x}, 0 \leq x < \infty,$$

use Chebychev's inequality to obtain a lower bound to the probability of the inequality  $-1 \leq X \leq 3$ , and compare it with actual value.

8. Explain the concept of "convergence in probability".

If  $X_1, X_2, \dots, X_n$  be r.v.s. with means  $\mu_1, \mu_2, \dots, \mu_n$  and standard deviations  $\sigma_1, \sigma_2, \dots, \sigma_n$  and if  $\sigma_n \rightarrow 0$  as  $n \rightarrow \infty$ , show that  $\bar{X}_n - \bar{\mu}_n$  converges to zero stochastically.

Hence show that if  $m$  is the number of successes in  $n$  independent trials, the probability of success at the  $i$ th trial being  $p_i$  then  $m/n$  converges in probability to  $(p_1 + p_2 + \dots + p_n)/n$ .

9. (a) If  $X_n$  takes the values 1 and 0 with corresponding probabilities  $p_n$  and  $1 - p_n$ , examine whether the weak law of large numbers can be applied to the sequence  $\{X_n\}$  where the variables  $X_n, n = 1, 2, 3, \dots$  are independent.

(b)  $\{X_i\}_{i=1}^n$  is a sequence of independent random variables with expected value of  $X_i$  equal to  $m_i$  and variance of  $X_i$  equal to  $\sigma_i^2$ . If  $\frac{1}{n^2} \sum_{i=1}^n \sigma_i^2$  tends to zero as  $n$  tends to infinity, show that the weak law of large numbers holds good to the sequence. [Bombay Univ. B.Sc. (Stat.), 1992]

10.  $\{X_k\}, k = 1, 2, \dots$  is a sequence of independent random variables each taking the values -1, 0, 1. Given that

$$P(X_k = 1) = \frac{1}{k} = P(X_k = -1), P(X_k = 0) = 1 - \frac{2}{k}.$$

Examine if the law of large numbers holds for this sequence.

11. (a) Derive Chebychev's inequality and show how it leads to the weak law of large numbers. Mention some important particular cases wherein the weak law of large numbers holds good.

(b) State and prove the weak law of large numbers. Deduce as a corollary Bernoulli theorem and comment on its applications.

(c) Examine whether the weak law of large numbers holds good for the sequence  $X_n$  of independent random variables where

$$P\left(X_n = \frac{1}{\sqrt{n}}\right) = \frac{2}{3}, \quad P\left(X_n = -\frac{1}{\sqrt{n}}\right) = \frac{1}{3}$$

(d)  $\{X_n\}$  is a sequence of independent random variables such that

$$P\left(X_n = \frac{1}{\sqrt{n}}\right) = p_n, \quad P\left(X_n = 1 + \frac{1}{\sqrt{n}}\right) = 1 - p_n$$

Examine whether the weak law of large numbers is applicable to the sequence  $\{X_n\}$ .

(e) If  $X$  is a random variable and  $E(X^2) < \infty$ , then prove that

$$P\{|X| \geq a\} \leq \frac{1}{a^2} E(X^2), \text{ for all } a > 0.$$

Use Chebyshev's inequality to show that for  $n > 36$ , the probability that in  $n$  throws of a fair die, the number of sixes lies between  $\frac{1}{6}n - \sqrt{n}$  and  $\frac{1}{6}n + \sqrt{n}$  is at least  $31/36$ .

[Calcutta Univ. B.Sc. (Maths Hons.), 1991]

12. Let  $\{X_n\}$  be a sequence of mutually independent random variables such that  $X_n = \pm 1$  with probability  $\frac{1-2^{-n}}{2}$

and  $X_n = \pm 2^{-n}$  with probability  $2^{-n-1}$

Examine whether the weak law of large numbers can be applied to the sequence  $\{X_n\}$ .

13. Examine whether the law of large numbers holds for the sequence  $\{X_k\}$  of independent random variables defined by  $P(X_k = \pm k^{-1/2}) = \frac{1}{2}$ .

14. (a) State Khinchin's theorem.

(b) Let  $X_1, X_2, X_3, \dots$  be a sequence of independent and identically distributed r.v.'s, each uniform on  $[0, 1]$ . For the geometric mean

$$G_n = (X_1 X_2 \dots X_n)^{1/n}$$

show that  $G_n \xrightarrow{P} c$  for some finite number  $c$ . Find  $c$ .

**Hint.**  $X \sim U[0, 1]$ , let  $Y = -\log X$ ;

Then  $F_Y(y) = 1 - e^{-y}$ ;  $\Rightarrow f_Y(y) = e^{-y}; y \geq 0$ .

$\therefore Y_i = -\log X_i$ , ( $i = 1, 2, \dots, n$ ) are i.i.d. r.v.'s with  $E(Y_i) = 1$ .

By Khinchin's theorem

$$\sum_{i=1}^n Y_i/n = -\left(\sum_{i=1}^n \log X_i/n\right) = -\log G_n \xrightarrow{P} E(Y_i) = 1 \Rightarrow G_n \xrightarrow{P} e^{-1} = c.$$

(c) Let  $X_1, X_2, \dots$  be i.i.d. r.v.'s with common p.d.f.

$$f(x) = \frac{1+\delta}{x^{2+\delta}}, x \geq 1, \delta > 0 \\ = 0, x < 1$$

Discuss if WLLN holds for the sequence  $\{X_n\}$ .

**Hint.**  $E X_i = (1 + \delta)/\delta < \infty$  (finite). Hence by Khinchin's theorem

$$S_n/n = \sum X_i/n \xrightarrow{P} (1 + \delta)/\delta \text{ as } n \rightarrow \infty.$$

**15.** Let  $\{X_n\}$  be any sequence of r.v.'s. Write  $Y_n = \frac{1}{n} \sum_{i=1}^n X_i$ .

Prove that a necessary and sufficient condition for the sequence  $\{X_n\}$  to satisfy the weak law of large numbers is that

$$E \left[ \frac{Y_n^2}{1 + Y_n^2} \right] \rightarrow 1 \text{ as } n \rightarrow \infty.$$

**Hint.** See Remark to Theorem 6-33.

**16.** State and prove Weak Law of Large Numbers. Determine whether it holds for the following sequence of independent random variables:

$$P(X_n = +1) = (1 - 2^{-n})/2 = P(X_n = -1)$$

[Delhi Univ. B.Sc. (Maths Hons.), 1989]

**17.** Let  $X_1, X_2, \dots$  be i.i.d. standard Cauchy variates. Show that the WLLN does not hold for the sequence  $\{X_n\}$ .

**Hint.** Use Theorem 6-33.

$$E \left( \frac{Y_n^2}{1 + Y_n^2} \right) = E \left( \frac{S_n^2}{n^2 + S_n^2} \right) = E \left( \frac{(S_n/n)^2}{1 + (S_n/n)^2} \right) \\ = \int_{-\infty}^{\infty} \frac{x^2}{1+x^2} \cdot \frac{1}{\pi} \frac{1}{1+x^2} dx$$

$\left[ \because \frac{S_n}{n} = \frac{X_1 + X_2 + \dots + X_n}{n} \text{ is also a standard Cauchy variate. See Remark 4, § 8-9-1} \right]$

$$= \frac{2}{\pi} \int_0^{\infty} \sin^2 \theta d\theta = \frac{1}{2} \quad (x = \tan \theta)$$

$$\Rightarrow \lim_{n \rightarrow \infty} E \left[ \frac{Y_n^2}{1 + Y_n^2} \right] \rightarrow 0 \Rightarrow \text{WLLN does not hold for } \{X_n\}.$$

**18. (a)** Examine if the WLLN holds for the sequence  $\{X_n\}$  of i.i.d. r.v.'s with

$$P[X_i = (-1)^{k-1} \cdot k] = \frac{6}{\pi^2 k^2}; \quad k = 1, 2, 3, \dots, i = 1, 2, 3, \dots$$

[Delhi Univ. B.Sc. (Maths Hons.), 1990]

$$\begin{aligned}
 \text{Hint. } E(X_i) &= \sum_{k=1}^{\infty} (-1)^{k+1} \cdot k \cdot \frac{6}{\pi^2 k^2} = \frac{6}{\pi^2} \sum_{k=1}^{\infty} (-1)^{k+1} \cdot (1/k) \\
 &= \frac{6}{\pi^2} \left[ 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \dots \right] \\
 &= \frac{6}{\pi^2} \cdot \log_e 2
 \end{aligned}$$

[ $\because$  The series in bracket is convergent by Leibnitz test for alternating series].

Hence by Khinchin's theorem, WLLN holds for the sequence  $\{X_i\}$  of i.i.d. r.v.'s.

(b) The r.v.'s  $X_1, X_2, \dots, X_n$  have equal expectations and finite variation. Is the weak law of large numbers applicable to this sequence if all the co-variances  $\sigma_{ij}$  are negative? [Delhi Univ. B.Sc. (Maths Hons.), 1987]

$$\begin{aligned}
 \text{Hint. } \frac{B_n}{n^2} &= \frac{\text{Var}(X_1 + X_2 + \dots + X_n)}{n^2} = \frac{1}{n^2} \left[ \sum_{i=1}^n \sigma_i^2 + 2 \sum_{i < j=1}^n \sigma_{ij} \right] \\
 &< \frac{1}{n^2} \left( \sum_{i=1}^n \sigma_i^2 \right) \rightarrow 0 \text{ as } n \rightarrow \infty \\
 &\quad (\because \sigma_i^2 \text{ are finite})
 \end{aligned}$$

Hence WLLN holds.

19. State and prove Borel Cantelli Lemma.

In a sequence of Bernoulli trials let  $A_n$  be the event that a run of  $n$  consecutive successes occurs between the  $2^n$ th and  $2^{n+1}$ th trials. Show that if  $p \geq \frac{1}{2}$ , there is probability one that infinitely many  $A_n$  occur; if  $p < \frac{1}{2}$ , then with probability one only finitely many  $A_n$  occur.

20. Let  $X_1, X_2, \dots$  be independent r.v.'s and  $S_n = \sum_{k=1}^n X_k$ . If  $\sum_{n=1}^{\infty} \sigma^2_{X_n}$  converges, prove that the series  $\sum (X_n - E X_n)$  converges in probability.

If  $\frac{1}{b_n} \sum_{k=1}^n \sigma^2_{X_k} \rightarrow 0$ , then prove that

$$\frac{S_n - ES_n}{b_n} \xrightarrow{P} 0.$$

Deduce Chebychev's inequality.

#### 6.17. Probability Generating Function

**Definition.** If  $a_0, a_1, a_2, \dots$  is a sequence of real numbers and if

$$A(s) = a_0 + a_1 s + a_2 s^2 + \dots = \sum_{i=0}^{\infty} a_i s^i \quad \dots(6.84)$$

converges in some interval  $-s_0 < s < s_0$ , when the sequence is infinite then the function  $A(s)$  is known as the generating function of the sequence  $\{a_i\}$ .

The variable  $s$  has no significance of its own and is introduced to identify  $a_i$  as the co-efficient of  $s^i$  in the expansion of  $A(s)$ . If the sequence  $\{a_i\}$  is bounded, then the comparison with the geometric series shows that  $A(s)$  converges at least for  $|s| < 1$ .

In the particular case when  $a_i$  is the probability that an integral valued discrete variable  $X$  takes the value  $i$ ,

i.e.,  $a_i = p_i = P(X = i)$ ;  $i = 0, 1, 2, \dots$  with  $\sum p_i = 1$ , then the probability generating function, abbreviated as p.g.f., of r.v.  $X$  is defined as :

$$P(s) = E(s^X) = \sum_{x=0}^{\infty} s^x \cdot p_x \quad \dots(6.85)$$

**Remarks.** 1. Obviously we have  $P(1) = \sum_x p_x = 1$ .

Thus a function  $P(s)$  defined in (6.85) is a p.g.f. iff  $p_x \geq 0 \forall x$  and  $\sum_x p_x = 1$

2. *Relation between p.g.f. and m.g.f.*

Taking  $s = e^t$  in (6.85), we get

$$P(e^t) = E(e^{xt}) = M_X(t). \quad \dots(6.86)$$

i.e., from p.g.f. we can obtain m.g.f. on replacing  $s$  by  $e^t$ .

3. *Bivariate probability generating function.* The joint p.g.f. of two random variables  $X_1$  and  $X_2$  is a function of two variables  $s_1$  and  $s_2$  defined by :

$$P_{X_1, X_2}(s_1, s_2) = E(s_1^{x_1} \cdot s_2^{x_2}) = \sum_{x_1} \sum_{x_2} s_1^{x_1} s_2^{x_2} \cdot p(x_1, x_2) \quad \dots(6.87)$$

Marginal p.g.f.'s can be obtained from (6.87) as given below.

$$P_{X_1}(s_1) = E(s_1^{x_1}) = P_{X_1, X_2}(s_1, 1); \quad P_{X_2}(s_2) = E(s_2^{x_2}) = P_{X_1, X_2}(1, s_2) \quad \dots(6.88)$$

4. Two r.v.'s are independent if and only if :

$$P_{X_1, X_2}(s_1, s_2) = P_{X_1}(s_1) \cdot P_{X_2}(s_2). \quad \dots(6.89)$$

The above concepts can be generalised to  $n$  random variables

**Theorem 6.37.** If  $X$  is a random variable which assumes only integral values with probability distribution

$$P(X = k) = p_k; \quad k = 0, 1, 2, \dots \text{ and } P(X > k) = q_k, \quad k \geq 0$$

so that  $q_k = p_{k+1} + p_{k+2} + \dots = 1 - \sum_{i=0}^k p_i$  and two generating functions are

$$P(s) = p_0 + p_1 s + p_2 s^2 + \dots$$

$$Q(s) = q_0 + q_1 s + q_2 s^2 + \dots$$

$$\text{then for } -1 < s < 1, \quad Q(s) = \frac{1 - P(s)}{1 - s} \quad \dots(6.90)$$

**Proof.** We have

$$q_{k-1} - q_k = p_k, \quad k \geq 1$$

$$\sum_{k=1}^{\infty} q_{k-1} s^k - \sum_{k=1}^{\infty} q_k s^k = \sum_{k=1}^{\infty} p_k s^k$$

$$\Rightarrow sQ(s) - Q(s) + q_0 = P(s) - p_0$$

$$\Rightarrow Q(s) = \frac{(p_0 + q_0) - P(s)}{(1-s)}$$

$$\text{But } p_0 + q_0 = p_0 + p_1 + p_2 + \dots = 1$$

Hence the theorem.

**Theorem 6-38.** For a random variable  $X$ , which assumes only integral values, the expectation  $E(X)$  can be calculated either from the probability distribution  $P(X=i) = p_i$  or in terms of

$$q_k = p_{k+1} + p_{k+2} + \dots$$

$$\text{Thus } E(X) = \sum_{i=1}^{\infty} i p_i = \sum_{k=0}^{\infty} q_k$$

In terms of the generating functions

$$E(X) = P'(1) = Q(1) \quad \dots(6-91)$$

**Proof.**  $\hat{P}(s) = \sum_{k=0}^{\infty} p_k s^k$ . If  $E(X)$  exists, then

$$E(X) = \sum_{k=1}^{\infty} k p_k$$

$$P'(s) = \sum_{k=1}^{\infty} k p_k s^{k-1} \Rightarrow P'(1) = \sum_{k=1}^{\infty} k p_k$$

$$\therefore E(X) = P'(1)$$

We know that

$$Q(s)[1-s] = 1 - P(s)$$

Differentiating both sides w.r.t.  $s$ , we get

$$Q'(s)[1-s] - Q(s) = -P'(s) \quad \dots(*)$$

$$\therefore Q(1) = P'(1)$$

$$\text{Hence } E(X) = P'(1) = Q(1)$$

**Theorem 6-39.** If  $E(X^2) = \sum k^2 p_k$  exists, then

$$E(X^2) = P''(1) + P'(1) = 2Q'(1) + Q(1)$$

$$\text{and hence } V(X) = 2Q'(1) + Q(1) - \{Q(1)\}^2 = P''(1) + P'(1) - \{P'(1)\}^2$$

$$\dots(6-92)$$

**Proof.**  $P(s) = \sum_{k=0}^{\infty} p_k s^k$ ,  $P'(s) = \sum k p_k s^{k-1}$

$$sP'(s) = \sum k p_k s^k. \text{ Differentiating again, we get}$$

$$P'(s) + sP''(s) = \sum k^2 p_k s^{k-1}$$

$$\therefore P'(1) + P''(1) = \sum k^2 p_k = E(X^2) \quad \dots(*)$$

$$\therefore Q''(s)[1-s] - 2Q'(s) = -P''(s) \quad [\text{Differentiating } (*) \text{ again}]$$

Putting  $s = 1$ , we get

$$2Q'(1) = P''(1). \text{ Substituting in } (*), \text{ we get}$$

$$E(X^2) = P'(1) + P''(1) = Q(1) + 2Q'(1)$$

$$\therefore \text{Var}(X) = E(X^2) - \{E(X)\}^2 = P''(1) + P'(1) - \{P'(1)\}^2 \\ = 2Q'(1) + Q(1) - \{Q(1)\}^2$$

**6.17.1. Probability Generating Function for the sum of independent variables (Convolutions).** If  $X$  and  $Y$  are non-negative independent, integral valued discrete random variables with respective probability generating functions,

$$P(s) = \sum_{k=0}^{\infty} p_k s^k, \quad p_k = P(X=k)$$

$$R(s) = \sum_{k=0}^{\infty} r_k s^k, \quad r_k = P(Y=k),$$

it is possible to deduce the probability generating function for the variable  $Z = X + Y$ , which is also clearly integral valued, in terms of  $P(s)$  and  $Q(s)$ .

Let  $w_k$  denote  $P(Z=k)$ . The event  $Z = k$  is the union of the following mutually exclusive events,

$$(X=0 \cap Y=k), (X=1 \cap Y=k-1), (X=2 \cap Y=k-2), \dots, (X=k \cap Y=0)$$

and

since the variables  $X$  and  $Y$  are independent, each joint probability is the product of the appropriate individual probabilities. Therefore the distribution  $w_k = P(Z=k)$  is given by

$$w_k = p_0 r_k + p_1 r_{k-1} + p_2 r_{k-2} + \dots + p_k r_0 \text{ for all integral } k \geq 0$$

The new sequence of probabilities  $\{w_k\}$  defined in terms of the sequences  $\{p_k\}$  and  $\{r_k\}$  is called the convolution of these sequences and is denoted by

$$\{w_k\} = \{p_k\} * \{r_k\} \quad \dots(6.93)$$

**Theorem 6-40.**

$$\mu'(r) = E \left[ X(X-1) \dots (X-r+1) \right] = \left[ \frac{\partial^r}{\partial s^r} P(s) \right]_{s=1}$$

**Proof.** Differentiating (6.85) partially  $r$  times w.r.t.  $s$ , we get

$$\frac{\partial^r P(s)}{\partial s^r} = \sum_x x(x-1)(x-2) \dots (x-r+1) s^{x-r} p_x$$

$$\left[ \frac{d^r P(s)}{ds^r} \right]_{s=1} = \sum_x x(x-1)(x-2)\dots(x-r+1)p_x = \mu'(r)$$

**Theorem 6.41.** If  $\{p_k\}$  and  $\{r_k\}$  are the sequences with the generating functions  $P(s)$ ,  $R(s)$  and  $\{w_k\}$  is their convolution, then  $W(s) = P(s)R(s)$ , where  $W(s) = \sum w_k s^k$  is the generating function of the sum  $X + Y$ .

**Proof.** Since the co-efficient of  $s^k$  in the product  $P(s)R(s)$  is

$$p_0 r_k + p_1 r_{k-1} + \dots + p_{k-1} r_1 + p_k r_0 = w_k,$$

it follows that the probability generating function for  $Z$ , namely,

$$W(s) = \sum_{k=0}^{\infty} w_k s^k \text{ is equal to } P(s)R(s).$$

**Cor.** If  $X_1, X_2, \dots, X_n$  are independent integral-valued discrete variables with respective probability generating functions  $P_1(s), P_2(s), \dots, P_n(s)$  and if  $Z = X_1 + X_2 + \dots + X_n$ , the probability generating function for  $Z$  is given by

$$P_Z(s) = \prod_{i=1}^n P_i(s)$$

In particular, when  $X_1, X_2, \dots, X_n$  all have a common distribution and hence common probability generating function  $P(s)$ , we have

$$P_Z(s) = [P(s)]^n$$

**Example 6.64.** Can  $P(s) = 2/(1+s)$  be the p.g.f. of a r.v.  $X$ ? Give reasons.

**Solution.** We have  $P(1) = 2/2 = 1$ .

$$\begin{aligned} \text{Also } P(s) &= \sum_{r=0}^{\infty} p_r s^r = 2(1+s)^{-1} \\ &= 2(1-s+s^2-s^3+\dots) \\ &= 2 \sum_{r=0}^{\infty} (-1)^r \cdot s^r \quad \dots(*) \\ \Rightarrow \quad p_r &= 2(-1)^r \end{aligned}$$

Hence  $p^{2k+1} < 0$ , i.e.,  $p_1, p_3, p_5, \dots$  are negative.

Since some co-efficient in (\*) are negative,  $P(s)$  cannot be the p.g.f. of a r.v.  $X$ .

**Example 6.65.** If  $P(s)$  is the probability generating function for  $X$ , find the generating function for  $(X-a)/b$ .

**Solution.**  $P(s) = E(s^X)$

$$\text{P.G.F. for } \frac{X-a}{b} = E(s^{(x-a)/b}) = e^{-ax/b} \cdot E(s^{x/b})$$

$$= s^{-a/b} \cdot E[(s^{1/b})^x] = s^{-a/b} P(s^{1/b})$$

**Example 6.66.** Let  $X$  be a random variable with generating function  $P(s)$ . Find the generating function of (a)  $X+1$  (b)  $2X$ .

$$\text{Solution. (a)} \quad P(s) = \sum_{k=0}^{\infty} p_k s^k = E(s^X)$$

$$\therefore \text{P.G.F. of } X+1 = E(s^{X+1}) = s \cdot E(s^X) = s \cdot P(s)$$

$$(b) \quad \text{P.G.F. of } 2X = E(s^{2X}) = E[(s^2)^X] = P(s^2)$$

**Example 6.67.** Find the generating function of (a)  $P(X \leq n)$ , (b)  $P(X < n)$ , and (c)  $P(X = 2n)$ . [Delhi Univ. M.Sc. (O.R.), 1989]

**Solution.** (a) Let  $X$  be an integral valued random variable with the probability distribution

$$P(X = n) = p_n \text{ and } P(X \leq n) = q_n$$

so that  $q_n = p_0 + p_1 + p_2 + \dots + p_n; n = 0, 1, 2, \dots$

$$\therefore q_n - q_{n-1} = p_n, n \geq 1$$

$$\Rightarrow \sum_{n=1}^{\infty} q_n s^n - \sum_{n=1}^{\infty} q_{n-1} s^n = \sum_{n=1}^{\infty} p_n s^n$$

$$\Rightarrow Q(s) - q_0 - sQ(s) = P(s) - p_0$$

$$\Rightarrow Q(s) = \frac{P(s) + q_0 - p_0}{1-s}, [\because q_0 = p_0]$$

$$(b) \quad \text{Let } q_n = P(X < n) = p_0 + p_1 + \dots + p_{n-1}, q_n - q_{n-1} = p_{n-1}, n \geq 2$$

$$\Rightarrow \sum_{n=2}^{\infty} q_n s^n - \sum_{n=2}^{\infty} q_{n-1} s^n = \sum_{n=2}^{\infty} p_{n-1} s^n = s \sum_{n=1}^{\infty} p_n s^n$$

$$\Rightarrow Q(s) - q_1 s - sQ(s) = sP(s) - sp_0 \quad [\because q_0 = 0]$$

$$\Rightarrow Q(s)[1-s] = sP(s) - sp_0 + q_1 s \quad [\because q_1 = p_0]$$

$$\text{Hence} \quad Q(s) = \frac{sP(s)}{1-s}$$

$$(c) \quad \text{Let } P(X = 2n) = p_{2n}$$

$$\therefore Q(s) = \sum_{n=0}^{\infty} p_{2n} s^n = p_0 + p_2 s + p_4 s^2 + \dots$$

$$2Q(s) = 2p_0 + 2p_2 s + 2p_4 s^2 + \dots$$

$$= (p_0 + p_1 s^{1/2} + p_2 s + p_3 s^{3/2} + p_4 s^2 + \dots) \\ + (p_0 - p_1 s^{1/2} + p_2 s - p_3 s^{3/2} + \dots)$$

$$= \sum_{k=0}^{\infty} p_k (s^{1/2})^k + \sum_{k=0}^{\infty} p_k (-s^{1/2})^k$$

$$= P(s^{1/2}) + P(-s^{1/2})$$

$$Q(s) = \frac{P(s^{1/2}) + P(s^{-1/2})}{2}$$

**Example 6-68.** Let  $\{X_k\}$  be mutually independent, each assuming the values  $0, 1, 2, \dots, a-1$  with probability  $\frac{1}{a}$ .

Let  $S_n = X_1 + X_2 + \dots + X_n$ . Show that the generating function of  $S_n$  is

$$P(s) = \left[ \frac{1-s^a}{a(1-s)} \right]^n$$

and hence

$$P(S_n = j) = \frac{1}{a^n} \sum_{v=0}^{\infty} (-1)^{v+j+av} \binom{n}{v} \binom{-n}{j-av}$$

(Rajasthan Univ. M.Sc. 1992)

**Solution.** As the  $\{X_k\}$  are mutually independent variables and each  $X_i$  assumes the same values  $0, 1, 2, \dots, a-1$ , with the probability  $\frac{1}{a}$ , therefore each will have the same generating function and the generating function of  $S_n$  will be the  $n$ th convolution of generating function of  $X_1$ . Now

$$P_{X_1}(s) = \sum_{k=0}^{a-1} p_k s^k = \frac{1}{a} [s^0 + s^1 + \dots + s^{a-1}] = \frac{1-s^a}{a(1-s)}$$

$$\therefore P_{S_n}(s) = \left[ \frac{1-s^a}{a(1-s)} \right]^n$$

Now the probability  $S_n = j$  is the co-efficient of  $s^j$  in

$$\frac{1}{a^n} (1-s^a)^n (1-s)^{-n}$$

If we take  $(v+1)$ th term from  $(1-s^a)^n$ , then it will have the power of  $s$  equivalent to  $s^{av}$  and hence to get the power of  $s$  as  $j$ , we must take the term from  $(1-s)^{-n}$  having the power of  $s$  as  $j-av$ .

$\therefore$  Required probability

$$= \frac{1}{a^n} \sum_{v=0}^{\infty} \binom{n}{v} (-1)^v \binom{-n}{j-av} (-1)^{j-av}$$

$$= \frac{1}{a^n} \sum_{v=0}^{\infty} (-1)^{j+v-av} \binom{n}{v} \binom{-n}{j-av}$$

$$= \frac{1}{a^n} \sum_{v=0}^{\infty} (-1)^{j+v+av} \binom{n}{v} \binom{-n}{j-av}$$

$$[\because (-1)^{2av} = 1 \Rightarrow (-1)^{av} = (-1)^{-av}]$$

**Example 6-69.** A random variable  $X$  assumes the value  $\lambda_1, \lambda_2, \dots$  with probabilities  $u_1, u_2, \dots$ , show that

$$p_k = \frac{1}{k!} \sum_{j=0}^{\infty} u_j e^{-\lambda_j} (\lambda_j)^k; \quad \lambda_j > 0, \quad \sum u_j = 1$$

is a probability distribution. Find its generating function and prove that its mean equals  $E(X)$  and variance equals  $V(X) + E(X)$ .

**Solution.**

$$\begin{aligned} \sum_{k=0}^{\infty} p_k &= \sum_{k=0}^{\infty} \left[ \frac{1}{k!} \sum_{j=0}^{\infty} u_j e^{-\lambda_j} (\lambda_j)^k \right] \\ &= \sum_{j=0}^{\infty} u_j e^{-\lambda_j} \sum_{k=0}^{\infty} (\lambda_j)^k / k! = \sum_{j=0}^{\infty} u_j e^{-\lambda_j} e^{\lambda_j} \\ &= \sum_{j=0}^{\infty} u_j = 1. \end{aligned}$$

Hence  $p_k$  represents a probability distribution.

Let  $P(s)$  be the generating function of  $p_k$ , then

$$\begin{aligned} P(s) &= \sum_{k=0}^{\infty} p_k s^k = \sum_{k=0}^{\infty} \left[ \frac{1}{k!} \sum_{j=0}^{\infty} u_j e^{-\lambda_j} (\lambda_j)^k s^k \right] \\ &= \sum_{j=0}^{\infty} u_j e^{-\lambda_j} \sum_{k=0}^{\infty} (s\lambda_j)^k / k! \\ &= \sum_{j=0}^{\infty} u_j e^{-\lambda_j} e^{s\lambda_j} = \sum_{j=0}^{\infty} u_j e^{\lambda_j(s-1)} \end{aligned} \quad (\text{Fubini's Theorem})$$

$$\text{Thus } P(s) = \sum_{j=0}^{\infty} u_j \left[ 1 + \lambda_j(s-1) + \frac{\lambda_j^2(s-1)^2}{2!} + \dots \right]$$

$$P'(1) = \sum_{j=0}^{\infty} u_j \lambda_j^0 = E(X)$$

$$P''(s) = \sum_{j=0}^{\infty} u_j [\lambda_j^2 + \lambda_j^3(s-1) + \dots]$$

$$P''(1) = \sum_{j=0}^{\infty} \lambda_j^2 u_j = E(X^2)$$

$$\begin{aligned} V(p_k) &= P''(1) + P'(1) - [P'(1)]^2 = E(X^2) + E(X) - [E(X)]^2 \\ &= E(X) + V(X) \end{aligned}$$

### EXERCISE 6(e)

1. (a) Define the probability generating function (p.g.f.) of a random variable.

(b)  $X$  is a positive integral valued variable, such that  $P(X=n) = p_n$ ,  $n = 0, 1, 2, \dots$ . Define the probability generating function  $G(s)$  and the moment generating function  $M(t)$  for  $X$  and show that,  $M(t) = G(e^t)$ . Hence or otherwise prove that

$$E(X) = G'(1), \text{ var}(X) = G''(1) + G'(1) - [G'(1)]^2$$

(c) If  $X$  and  $Y$  are non-negative integral valued independent random variables with  $P(s)$  and  $Q(s)$  as their probability generating functions, show that their sum  $X + Y$  has the p.g.f.  $P(s)Q(s)$ .

2. A test of the strength of a wire consists of bending and unbending until it breaks. Considering bending and unbending as two operations, let  $X$  denote the random variable corresponding to the number of operations necessary to break the wire. If  $P(X = r) = (1 - p)p^{r-1}$ ;  $r = 1, 2, 3, \dots$  and  $0 < p < 1$ , find the probability generating function of  $X$ .

3. Define the generating function  $A(s)$  of the sequence  $\{a_j\}$ . Let  $a_j$  be the number of ways in which the score  $j$  can be obtained by throwing a die any number of times. Show that the g.f. of  $\{a_j\}$  is  $(1 - s - s^2 - s^3 - s^4 - s^5 - s^6)^{-1} - 1$ .

4. Four tickets are drawn, one at a time with replacement, from a set of ten tickets numbered respectively 1, 2, 3, ..., 10, in such a way that at each draw each ticket is equally likely to be selected. What is the probability that the total of the numbers on the four drawn tickets is 20?

**Hint.** If  $X_i$  denotes the number on the  $i$ th ticket then, for  $i = 1, 2, 3, 4$ , we observe that  $X_i$  is an integral-valued variate with possible values 1, 2, 3, ..., 10, each having associated probability  $1/10$ . Here each  $X_i$  has

$$\text{g.f. } = \frac{1}{10}s + \frac{1}{10}s^2 + \dots + \frac{1}{10}s^{10} = \frac{1}{10}s(1-s^{10})(1-s)^{-1}$$

and, since the  $X_i$ 's are independent, it follows that the total of the numbers on the drawn tickets

$$Z = X_1 + X_2 + X_3 + X_4$$

has probability generating function

$$\left\{ \frac{1}{10}s(1-s^{10})(1-s)^{-1} \right\}^4 = \frac{1}{10^4}s^4(1-s^{10})^4(1-s)^{-4}$$

The required probability is the coefficient of  $s^{16}$  in

$$\frac{1}{10^4}(1-s^{10})^4(1-s)^{-4} = \frac{63}{10,000}$$

5. Find the generating functions of (a)  $P(X \geq n)$ , (b)  $P(X > n+1)$ .

6. (a) Obtain the generating function of  $q_n$ , the probability that in  $n$  tosses of an ideal coin, no run of three heads occurs.

(b) Let  $X$  be a non-negative integral-valued random variable with probability generating function  $P(s) = \sum_{n=0}^{\infty} p_n s^n$ . After observing  $X$ , conduct  $X$  binomial trials with probability  $p$  of success, and let  $Y$  denote the corresponding resulting number of successes.

Determine (i) the probability generating function of  $Y$  and, (ii) probability generating function of  $X$  given that  $Y = X$ .

7. In a sequence of Bernoulli trials, let  $U_n$  be the probability that the first combination  $SF$  occurs at trials number  $(n - 1)$  and  $n$ .

Find the generating function, mean and variance.

Hint.  $U_2 = P(SF) = pq, U_3 = P(SSF) + P(FSF) = pq(p+q)$

$$U_4 = (SSSF) + P(FSSF) + P(FFSF) = pq(p^2 + pq + q^2)$$

In general  $U_n = pq \sum_{k=0}^{n-2} p^k q^{n-2-k}$

$$= pq^{n-1} \left[ 1 + \frac{p}{q} + \left(\frac{p}{q}\right)^2 + \dots + \left(\frac{p}{q}\right)^{n-2} \right]$$

$$= pq \cdot \frac{q^{n-1} - p^{n-1}}{q - p}.$$

8. (a) In a sequence of Bernoulli trials, let  $U_n$  be the probability of an even number of successes. Prove the recursion formula

$$U_n = qU_{n-1} + (1 - U_{n-1})p$$

From this derive the generating function and hence the explicit formula for  $U_n$ .

(b) A series of independent Bernoulli trials is performed until an uninterrupted run of  $r$  successes is obtained for the first time where  $r$  is a given positive integer. Assuming that the probability of a success in any trial is  $p = 1 - q$ , show that the probability generating function of the number of trials is

$$F(s) = \frac{p^r s^r (1 - ps)}{1 - s + qp^r s^{r+1}}$$

### ADDITIONAL EXERCISES ON CHAPTER VI

1. (a) A horizontal line of length '5' units is divided into two parts. If the first part is of length  $X$ , find  $E(X)$  and  $E[X(5-X)]$ .

(b) Show that

$$E(X - \mu)^3 = E(X^3) - 3\mu\sigma^2 - \mu^3$$

where  $\mu$  and  $\sigma^2$  are the mean and variance of  $X$  respectively.

2. (a) Two players  $A$  and  $B$  alternately roll a pair of fair dice.  $A$  wins if he gets six points before  $B$  gets seven points and  $B$  wins if he gets seven points before  $A$  gets six points. If  $A$  takes the first turn, find the probability that  $B$  wins and the expected number of trials for  $A$  to win.

(b) A box contains  $2^n$  tickets among which "C" tickets bear the numbers  $i$  ( $i = 0, 1, 2, \dots, n$ ). A group of  $m$  tickets is drawn. Let  $S$  denote the sum of their numbers. Find the expectation and variance of  $S$ .

Ans.  $\frac{1}{2}mn, \frac{1}{4}mn - \{mn(m-1)/4(2^n - 1)\}$

3. In an objective type examination, consisting of 50 questions, for each question there are four answers of which only one is correct. A candidate scores 1 if he picks up the correct answer and  $-1/3$  otherwise. If a candidate makes only a random choice in respect of each of the 50 questions, find his expected score and the variance of his score.

4. (a) A florist, in order to satisfy the needs of a number of regular and sophisticated customers, stocks a highly perishable flower. A dozen flowers cost Rs. 3 and sell for Rs. 10. Any flowers not sold the day they are stocked are worthless. Demand in dozens of flowers is as follows :

Demand	0	1	2	3	4	5
Probability	0.1	0.2	0.3	0.2	0.1	0.1

(i) How many flowers should the florist stock daily in order to maximise the expected value of his net profit ?

(ii) Assuming that failure to satisfy any one customer's request will result in future lost profits amounting to Rs. 5.10 (goodwill cost), in addition to the lost profit on the immediate sale; how many flowers should the florist stock ?

(iii) What is the smallest goodwill cost of stocking five dozen flowers ?

**Hint.** For  $i = 0, 1, 2, 3, 4, 5$ , let  $X_i$  be the random variable giving the florist's net profit, when he decides to stock ' $i$ ' dozen flowers. Determine the probability function for each and the mean of each and pick up that ' $i$ ' for which it is maximum.

**Ans.** (i) 3 dozen, (ii) 4 dozen and (iii) Rs. 2

5. Consider a sequence of Bernoulli trials with a constant probability  $p$  of success in a single trial. Let  $X_k$  denote the number of failures following the  $(k - 1)$ th and preceding the  $k$ th success and let  $S_r = \sum_{k=1}^r X_k$ .

Derive the probability distribution of  $X_k$ . Hence derive the probability distribution of  $S_r$ . Find  $E(S_r)$  and  $\text{Var}(S_r)$ .

6. In the simplest type of weather forecasting — "rain" or "no rain" in the next 24 hours — suppose the probability of raining is  $p (> \frac{1}{2})$ , and that a forecaster scores a point if his forecast proves correct and zero otherwise. In making  $n$  independent forecasts of this type, a forecaster, who has no genuine ability, predicts "rain" with probability  $\lambda$  and "no rain" with probability  $(1 - \lambda)$ . Prove that the probability of the forecast being correct for any one day is

$$[1 - p + (2p - 1)\lambda]$$

Hence derive the expectation of the total score ( $S_n$ ) of the forecaster for the  $n$  days, and show that this attains its maximum value for  $\lambda = 1$ . Also, prove that

$$\text{Var}(S_n) = n [p - (2p - 1)\lambda] [1 - p + (2p - 1)\lambda]$$

and thereby deduce that, for fixed  $n$ , this variance is maximum for  $\lambda = \frac{1}{2}$ .

**Hint.**  $P(X_i = 1) = 1 - P(X_i = 0) = p\lambda + q(1 - \lambda)$ ,  $S_n = \sum_{i=1}^n X_i$ ,

and the  $X_i$ 's being independent, the stated results follow.

7. In the simplest type of weather forecasting — rain or no rain in the next 24 hours — suppose the probability of raining is  $p (> \frac{1}{2})$ , and that a forecaster scores a point if his forecast proves correct and zero otherwise. In making  $n$  independent forecasts of this type, a forecaster who has no genuine ability decides to allocate at random  $r$  days to a "rain" forecast and the rest to "no rain". Find the expectation of his total score  $\{S_n\}$  for the  $n$  days and show that this attains its maximum value for  $r = n$ . What is the variance of  $S_n$ ?

**Hint.** Let  $X_i$  be a random variable such that

$$\begin{aligned} X_i &= 1 \text{ if forecast is correct for } i\text{th day} \\ &= 0 \text{ if forecast is incorrect for } i\text{th day}, (i = 1, 2, \dots, n) \end{aligned}$$

Then  $P(X_i = 1) = \frac{r}{n}$ ,  $P(X_i = 0) = 1 - \frac{r}{n} = \frac{n-r}{n}$

$$\therefore E(X_i) = p\left(\frac{r}{n}\right) + q\left(\frac{n-r}{n}\right) \text{ and } S_n = \sum_{i=1}^n X_i$$

But the  $X_i$ 's are correlated random variables, so that for  $i \neq j$ ,

$$E(X_i X_j) = P(X_i = 1 \cap X_j = 1)$$

$$\begin{aligned} &= p \cdot \frac{r}{n} \left[ p\left(\frac{r-1}{n-1}\right) + q\left(\frac{n-r}{n-1}\right) \right] \\ &\quad + q\left(\frac{n-r}{n}\right) \left[ p\left(\frac{r}{n-1}\right) + q\left(\frac{n-r-1}{n-1}\right) \right] \end{aligned}$$

Hence  $E(S_n) = np - (n-r)(p-q) < np$ , for  $p > q$  and  $V(S_n) = npq$ .

8. Let  $n_1$  letters 'A' and  $n_2$  letters 'B' be arranged at random in a sequence. A run is a succession of like letters preceded and followed by none or an unlike letter. Let  $W$  be the total number of runs of 'A's and 'B's. Obtain expressions for Prob  $\{W = r\}$ , where  $r$  is a given positive even integer and also when  $r$  is odd.

Compute the expectation of  $W$ .

9. An urn contains  $K$  varieties of objects in equal numbers. The objects are drawn one at a time and replaced before the next drawing. Show that the probability that  $n$  and no less drawings will be required to produce objects of all varieties is

$$\sum_{r=0}^{k-1} (-1)^r {}^{k-1}C_r \left(\frac{k-1-r}{k}\right)^{n-1}$$

Hence or otherwise, find the expected number of drawings in a simple form.

10. An urn contains  $a$  white and  $b$  black balls. After a ball is drawn, it is to be returned to the urn if it is white, but if it is black, it is to be replaced by a white ball from another urn. Show that the probability of drawing a white ball after the foregoing operation has been repeated  $x$  times is

$$1 - \frac{b}{a+b} \left( 1 - \frac{1}{a+b} \right)^x$$

11. A box contains  $k$  varieties of objects, the number of objects of each variety being the same. These objects are drawn one at a time and put back before the next drawing. Denoting by  $n$  the smallest number of drawings which produce objects of all varieties, find  $E(n)$  and  $V(n)$ .

12. There is a lot of  $N$  objects from which objects are taken at random one by one with replacement. Prove that the expected value and variance of the least number of drawings needed to get  $n$  different objects are respectively given by

$$N \left[ \frac{1}{N} + \frac{1}{N-1} + \dots + \frac{1}{N-n+1} \right]$$

and

$$N \left[ \frac{1}{(N-1)^2} + \frac{2}{(N-2)^2} + \dots + \frac{n-1}{(N-n+1)^2} \right].$$

13. A large population consists of equal number of individuals of  $c$  different types. Individuals are drawn at random one by one until at least one individual of each type has been found, whereupon sampling ceases. Show that the mean number of individuals in the sample is

$$c \left( 1 + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{c} \right)$$

and the variance of the number is

$$c^2 \left( 1 + \frac{1}{2^2} + \frac{1}{3^2} + \dots + \frac{1}{c^2} \right) - c \left( 1 + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{c} \right)$$

14. (a) A point  $P$  is taken at random in a line  $AB$  of length  $2a$ , all positions of the point being equally likely. Show that the expected value of the area of the rectangle  $AP \cdot AB$  is  $2a^2/3$  and the probability of the area exceeding  $a^2/2\sqrt{2}$ .

- (b) A point is chosen at random on a circle of radius  $a$ . Show that the expectation of its distance from another fixed point also on the circle is  $4a/\pi$ .

- (c) Two points  $P$  and  $Q$  are selected at random in a square of side  $a$ . Prove that

$$E(|PQ|^2) = a^2/3$$

15. If the roots  $x_1, x_2$  of the equation  $x^2 - ax + b = 0$  are real and  $b$  is positive but otherwise unknown, prove that

$$E(x_1) = \frac{1}{6}a \text{ and } E(x_2) = \frac{5}{6}a$$

16. (*Banach's Match-box Problem*). A certain mathematician always carries two match boxes (initially containing  $N$  match-sides). Each time he wants a match-stick, he selects a box at random, inevitably a moment comes when he finds

a box empty. Show that the probability that there are exactly  $r$  match-sticks in one box when the other box becomes empty is

$$2^{N-r} C_N \times \frac{1}{2^{2N-r}}$$

Prove also that the expected number of matches is

$$2^N C_N \times \frac{2N+1}{2^{2N}} - 1$$

17.  $n$  couples procreate independently with no limits on family size. Births are single and independent and for the  $i$ th couple, the probability of a baby is  $p_i$ . The sex ratio  $S$  is defined as

$$S = \frac{\text{Mean number of all boys}}{\text{Mean number of all children}}$$

Show that if all couples,

(i) Stop procreating on birth of a boy, then

$$S = n \left/ \sum_{i=1}^n \frac{1}{p_i} \right.$$

(ii) Stop procreating on birth of a girl, then

$$S = 1 - \left[ n \left/ \sum_{i=1}^n \frac{1}{q_i} \right. \right], \text{ where } q_i = 1 - p_i$$

(iii) Stop procreating when they have children of both sexes, then

$$S = \left[ \sum_{i=1}^n \frac{1}{q_i} - \sum_{i=1}^n p_i \right] \left/ \left[ \sum_{i=1}^n \frac{1}{p_i q_i} - n \right] \right.$$

18. Show that if  $X$  is a random variable such that  $P(a \leq X \leq b) = 1$ , then  $E(X)$  and  $\text{Var}(X)$  exist, and  $a \leq E(X) \leq b$  and  $\text{Var}(X) \leq (b-a)^2/4$ .

19.  $(X, Y)$  is a two-dimensional discrete random variable with the possible values 0 and 1 for  $X$ , and also 0 and 1 for  $Y$ , and with the joint probabilities given by

		$X$	0	1
		0	$p_{00}$	$p_{10}$
$Y$	0	$p_{01}$	$p_{11}$	
	1			

Find the characteristic functions  $\psi_1(t)$ ,  $\psi_2(t)$  and  $\xi(t_1, t_2)$  for  $X$ ,  $Y$  and  $(X, Y)$  respectively and show that  $\psi(t_1, t_2) = \psi_1(t_1)\psi_2(t_2)$  when  $p_{00}p_{11} = p_{01}p_{10}$ .

20. For a given sequence  $\{X_n\}$  of r.v.'s,

$$\varphi_n(t, X_n) = (\sin nt)/nt,$$

determine the distribution function of  $X_n$ . Hence show that even though the sequence of characteristic functions  $\varphi_n(t)$  converges to a limit  $\varphi(t)$ , the sequence of distribution functions does not converge to a distribution function. What is the condition that is violated here? [Indian Civil Services, 1984]

21. Two continuous variates  $X$  and  $Y$  have a joint p.d.f. with a joint characteristic function  $\varphi(t_1, t_2)$ : If  $g_x(x)$  is the marginal density, show that  $\mu'_r(x)$ , the  $r$ th simple moment for the conditional distribution of  $Y$  given  $X=x$ , satisfies the equation

$$\mu'_r(x) \cdot g(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\partial^r \varphi(t_1, 0)}{\partial t_2^r} e^{-it_1 x} dt_1$$

[Indian Civil Services, 1987]

22. Prove that the real part of a characteristic function is again a characteristic function. Prove further that if  $\psi_1(t) = a_1(t) + ib_1(t)$  and  $\psi_2(t) = a_2(t) + ib_2(t)$  are characteristic functions, then  $a_1(t)a_2(t) - b_1(t)b_2(t)$  is a characteristic function.

23. Show that for any distribution

$$\int_{-\infty}^{\infty} \left( 1 - \frac{x^2}{t^2} \right) dF(x) \leq \int_{-t}^t dF(x)$$

and hence deduce  $P(|X - E(X)| > k\sigma) \leq \frac{1}{k^2}$ , where  $k > 0$  and  $\text{Var}(X) = \sigma^2$ .

24. (a) Let  $X$  be a random variable with moment generating function  $M(t)$ ,  $-h < t < h$ . Prove that

$$P(X \geq a) \leq e^{-at} M(t), \quad 0 < t < h$$

and that

$$P(X \leq a) \leq e^{-at} M(t), \quad -h < t < 0.$$

$$(b) \text{ Let } f(x, y) = xe^{-x(y+1)}; \quad x > 0, y > 0 \\ = 0, \text{ elsewhere}$$

Find moment generating function of  $Z = XY$ .

$$\begin{aligned} \text{Hint. } M_{XY}(t) &= \int_0^\infty \int_0^\infty e^{txy} f(x, y) dx dy \\ &= \int_0^\infty x e^{-x} \left[ \int_0^\infty e^{-(1-t)xy} dy \right] dx; \quad 1-t > 0 \\ &= \int_0^\infty x e^{-x} \frac{1}{(1-t)x} dx = \frac{1}{1-t}; \quad t < 1. \end{aligned}$$

25. The probability of obtaining a 6 with a biased die is  $p$ , where  $(0 < p < 1)$ . Three players  $A$ ,  $B$  and  $C$  roll this die in order,  $A$  starting. The first

one to throw a 6 wins. Find the probability of winning for  $A$ ,  $B$  and  $C$ .

If  $X$  is a random variable which takes the value  $r$  if the game finishes at the  $r$ th throw, determine the probability generating function of  $X$  and hence, or otherwise, evaluate  $E(X)$  and  $\text{Var}(X)$ .

**Hint.** Probabilities for the wins of  $A$ ,  $B$  and  $C$  are  $p/(1-q^3)$ ,  $pq(1-q^3)$  and  $pq^2/(1-q^3)$  respectively.

$$P(X=r) = pq^{r-1}, \text{ for } r \geq 1.$$

The probability generating function of  $X$  is  $P(s) = p s / (1 - qs)$ ,

whence  $E(X) = 1/p$  and  $\text{Var}(X) = 1/qp^2$ .

26. Define convergence in probability. Let  $X_1, X_2, \dots$  be i.i.d. variates with  $f(x) = e^{-(x-1)}$ ,  $x \geq 1$ . Show that  $Y_n \rightarrow 1$  in probability where  $Y_n = \min(X_k); 1 \leq k \leq n$ . (Indian Civil Services, 1982)

27. Let  $\{X_n, n = 1, 2, \dots\}$  be a sequence of standardised variates and  $\text{Corr}(X_m, X_n) = \exp[-|m-n|\alpha]$ ,  $\alpha > 0$  and  $m = n$ . Show that W.L.L.N. holds for this sequence. (Indian Civil Services, 1988)

28. From the probability generating function (p.g.f.) of two random variables  $X$  and  $Y$  given by

$$P(s, t) = \exp[-\lambda - \mu - b + \lambda s + \mu t + bst],$$

- (i) obtain the marginal p.g.f.'s and identify them,
- (ii) obtain the p.g.f. of  $X + Y$  and  $P(X + Y) = 0$ , and
- (iii) interpret the case  $b = 0$ .