

Modern Probability Theory

An Introductory Textbook

(SECOND EDITION)

B. Ramdas Bhat
Professor of Statistics
Karnatak University
Dharwad



WILEY EASTERN LIMITED
New Delhi Bangalore Bombay Calcutta Madras Hyderabad

Copyright ©1981, WILEY EASTERN LIMITED
Second Edition, 1985
First Reprint, 1986

WILEY EASTERN LIMITED

4835/24, Ansari Road, Daryaganj, New Delhi-110002
4654/21, Daryaganj, New Delhi-110002
6 Shri B. P. Wadia Road, Basavangudi, Bangalore-560004
Abid House, Dr. Bhadkamkar Marg, Bombay-400007
40/8, Ballygunge, Circular Road, Calcutta-700019
Post Box No. 8604 Thiruvanniyur, Madras-600041
Post Box No. 1050, Himayath Nagar, P. O. Hyderabad-500029

This book or any part thereof may not be
reproduced in any form without the
written permission of the publisher

This book is not to be sold outside the
country to which it is consigned by
Wiley Eastern Limited

10 9 8 7 6 5 4 3 2 1

Rs. 40-00

Published by Mohinder Singh Sejwal for Wiley Eastern Limited,
4835/24, Ansari Road, Daryaganj, New Delhi - 110 002
and printed by Sunil Dutt at Gopsons Papers Pvt. Ltd.
A-28, Sector IX, NOIDA. Printed in India.

To
MICHEL LOEVE

Preface to the Second Edition

The second revised edition of the book incorporates many suggestions made by my friends and colleagues. Numerous errors of the earlier edition have been weeded out. A number of problems have been added at the end of different chapters. Chapters 10 and 11 on Laws of Large Numbers and Central Limit Theorem have been enlarged and completely rewritten.

I gratefully acknowledge the suggestions and corrections received from the large number of teachers and students using the book, in particular from Profs M.S. Prasad, M.S. Chikkagoudar, Dr. N.R. Mohan and others. I am looking forward to receiving similar suggestions from the readers in future also.

Dharwad
May, 1985

B. RAMDAS BHAT

Preface to the First Edition

While teaching the subject of probability theory to post-graduate students of statistics I could find books dealing with either from a very low mathematical level or from a highly sophisticated mathematical level. Books written for students with only the knowledge of graduate level mathematics are hard to find. To fill this gap the present book is written.

It forms the contents of a one-year probability theory which the author has taught for a number of years. The background needed is a good knowledge of mathematical analysis including Riemann integration and some familiarity with probability theory. Even though the book does not assume any previous knowledge of measure theory and develops probability theory in a self-contained manner those who had previous knowledge of measure theory can cover the contents of the book quickly, probably in about half the time, i.e., in one semester.

Chapter 1 of the book introduces set theory and defines *event* as a set of outcomes of an experiment. In this chapter set-theoretical background is provided for defining probability on a class of events, which is done in Chapter 3. Chapter 2 discusses properties of another important concept, viz., random variable as a measurable function. A salient feature of the book is to distinguish "definition" of probability from the "method" of its calculation. Axiomatic definition is given and its properties as a finite measure are studied. Distribution functions of r.v.'s and general distributions are introduced in Chapter 4. These play an important role in the study of statistics. Distribution functions which are mixtures of discrete and continuous distributions are highlighted in the book.

Expectations are defined similar to Lebesgue-Stieltjes integral. Chapter 5 is self-contained and does not assume any previous knowledge of the theory of Lebesgue integration. Chapter 6 is concerned not only with the definitions of the various modes of convergence of a sequence of r.v.'s, viz., convergence in probability, convergence almost surely, convergence in r th mean, and convergence in distribution, but discusses their interrelationships. There is a section on the interchange of limits and expectations. Double and iterated integrals are defined and Fubini theorem stated in the last section. Characteristic function is an important tool of probability analysis. Some basic properties which are used in later chapters, particularly in Chapters 8 and 11 are derived in Chapter 7. Weak convergence of distributions, which is related to weak convergence of probability measures is an important concept amenable to sophisticated

mathematical analysis. This is touched briefly in Chapter 8.

Concept of independence occurring in probability theory does not have a parallel in measure theory. How this condition of independence affects the probability measure defined on the probability space induced by a sequence of r.v.'s is discussed in Chapter 9. Convergence and stability properties of a series of independent r.v.'s and laws of large numbers are the topics of Chapter 10. Convergence to normal law, the so-called central limit theorem, is derived for the univariate and p -variate cases in Chapter 11.

A special feature of the book which makes it different from the usual texts, is the treatment of conditional probability. It is defined on the space of measurable subsets of a measurable set. Conditional expectation is defined and its properties studied in Chapter 12. As an illustration, simple properties of a martingale are deduced. Markov chain with finite state space is analysed in the last Chapter 13 as an introduction to the study of Markov processes. The latter is a subject of intense research activity during the last three decades and has been analysed using highly sophisticated mathematical tools. If the student's interest is kindled to take up further study in this area, the purpose of writing this chapter is fulfilled.

A notable feature of this book is the collection of theoretical and numerical problems given at the end of each chapter. Complementary topics not covered in the chapter are also mentioned here. A few of these have been taken from the question papers set at different university examinations. The author gratefully acknowledges his indebtedness to these universities. Phrases such as 'how', 'why', 'prove' have been inserted at different places to guard the inexperienced student against assuming a few statements to be true intuitively, rather than deriving the missing steps logically. Emphasis, throughout the book, is on rigorous derivation of various results, starting from the definitions and assumptions. It is hoped that such derivations would train a student in analytical and logical thinking.

I don't propose to list the books and the authors whose work inspired me to write the present work, because it is bound to be incomplete. My teachers at Berkeley, particularly Prof. M. Loève, are responsible for rousing my interest in the subject of probability theory. If some readers feel that the present book is an enlarged version of Loève's classical treatise, there is some justification for it.

I am indebted to the University Grants Commission for providing financial assistance to write this book. I am indebted to my friends, colleagues and students for their valuable criticism and comments.

Dharwad
December, 1980

B. RAMDAS BHAT

List of Abbreviations

$\in$	element of
$\notin$	not an element of
$\Rightarrow$	logical implication
$\Leftrightarrow$	logical equivalence
$\subseteq$	such that
$\exists$	there exists
$\forall$	each, all or every
$\therefore$	therefore
$\because$	since
$(*)$	end of the proof
P	convergence in probability
$\rightarrow$	
a.s.	convergence almost surely
$\xrightarrow{d(\alpha)}$	convergence in law (or distribution)
$\xrightarrow{r}$	convergence in r th mean
$\xrightarrow{w}$	weak convergence
$\xrightarrow{c}$	complete convergence
a.a.	almost all
$A, B, C, \text{ etc.}$	sets
$\mathcal{A}, \mathcal{B}, \mathcal{C}, \text{ etc.}$	classes
A^c	complement of A
a.e.	almost everywhere
a.s.	almost surely
$\mathcal{B}_n$	Borel field in R^n
ch. fn.	characteristic function
d.f.	distribution function
iff	if and only if
$I(A), I_A$	indicator function of A
i.d.	infinitely divisible
i.i.d.	independent identically distributed
M.C.	Markov chain

m.g.f.	moment generating function
R^k	k -dimensional real space
r.v.	random variable (or vector)
$\cup$	union
$\cap$	intersection
u.t.a.c.	up to additive constant
w.r.t.	with respect to
X^{-1}	inverse of X
L.H.S.	left hand side
R.H.S.	right hand side
$\leftrightarrow$	communicating states of a Markov chain

Contents

<i>Preface to the Second Edition</i>	<i>vii</i>
<i>Preface to the First Edition</i>	<i>ix</i>
<i>List of Abbreviations</i>	<i>xi</i>
1. Sets and Classes of Events	1
1.1 The Event 1	
✓ 1.2 Algebra of Sets 3	
✓ 1.3 Fields and σ -Fields 9	
1.4 Class of Events 17	
Complements and Problems 18	
2. Random Variables	20
2.1 Functions and Inverse Functions 20	
2.2 Random Variables 29	
2.3 Limits of Random Variables 33	
Complements and Problems 40	
✓ 3. Probability Space	42
3.1 Definition of Probability 42	
3.2 Some Simple Properties 47	
3.3 Discrete Probability Space 50	
3.4 General Probability Space 54	
3.5 Induced Probability Space 59	
3.6 Other Measures 61	
Complements and Problems 65	
4. Distribution Functions	69
4.1 Distribution Function of a Random Variable 69	
✓ 4.2 Decomposition of D.F.'s 72	
4.3 Distribution Functions of Vector Random Variables 74	
4.4 Correspondence Theorem 77	
Complements and Problems 79	
5. Expectation and Moments	81
5.1 Definition of Expectation 81	
5.2 Properties of Expectation 87	

5.3	Moments, Inequalities	93	
	Complements and Problems	100	
6.	Convergence of Random Variables		105
6.1	Convergence in Probability	105	
6.2	Convergence Almost Surely	110	
6.3	Convergence in Distribution	114	
6.4	Convergence in r th Mean	117	
6.5	Convergence Theorems for Expectations	119	
6.6	Fubini's Theorem	124	
	Complements and Problems	127	
7.	Characteristic Functions		132
7.1	Definition and Simple Properties	132	
7.2	Some Simple Properties	135	
7.3	Inversion Formula	138	
7.4	Characteristic Functions and Moments	145	
7.5	Bochner's Theorem	150	
	Complements and Problems	152	
8.	Convergence of Distribution Functions		156
8.1	Weak Convergence	156	
8.2	Convergence of Distribution Functions and Characteristic Functions	162	
8.3	Convergence of Moments	164	
	Complements and Problems	166	
9.	Independence		169
9.1	Definition	169	
9.2	Multiplication Properties	173	
9.3	Zero-One Laws	178	
	Complements and Problems	181	
10.	Laws of Large Numbers		185
10.1	Convergence of a Series of Independent Random Variables	185	
10.2	Kolmogorov Inequalities and A.S. Convergence	189	
10.3	Stability of Independent R.V.'s	193	
	Complements and Problems	202	
11.	Central Limit Theorem		207
11.1	Introduction	207	
11.2	I.I.D. Case	208	
11.3	Variable Distributions	210	
11.4	Multivariate Central Limit Theorem	216	
	Complements and Problems	218	

CONTENTS

xv

12. Conditioning	223
12.1 Radon-Nikodym Theorem	223
12.2 Conditional Expectation	228
12.3 Properties of Conditional Expectation	230
12.4 Martingales	235
Complements and Problems	238
✓ 13. Finite Markov Chain	240
13.1 Definition	240
13.2 Calculation of n -Step Transition Probabilities	243
13.3 Recurrence	247
13.4 Limiting Properties of Transition Probabilities	252
Complements and Problems	255
Appendix A	258
Appendix B	260
Appendix C	262
<i>Bibliography</i>	265
<i>Index</i>	267

1. Sets and Classes of Events

1.1 THE EVENT

In everyday life we come across many phenomena, the nature of which cannot be predicted in advance or many experiments, whose outcomes may not be known precisely. However we may know that the outcome has to be one of the several possibilities. The weight of a newborn baby cannot be known before the birth, except that it may lie in a certain range. When a coin is tossed, we know that the outcome has to be either a head or a tail; but we do not know the outcome of a particular throw in advance.

By a *statistical experiment* or simply *experiment* we mean not only an experiment such as the tossing of a coin or the observation of the number of defects in a certain sample of N items chosen from the daily production, in which the possible outcomes are finite; but also the observation of a phenomenon, such as the weight of a new born baby, or the weather condition of a certain region, where the number of possible outcomes is infinite. The concept of experiment is fundamental to the study of probability theory, because it is concerned with assigning chances to the outcomes of the statistical experiment or to possible state of nature and studying them.

Let us denote by ω , the typical outcome of an experiment E . ω is called a *sample point*. The totality of all outcomes of E will be denoted by Ω and is called the *sample space*. A collection of outcomes of E in which we are interested will be called an *event*. Thus an event is a *subset* of Ω . The number of sample points may be finite, countable or uncountable as illustrated by the following examples.

Example 1.1. An industrial engineer may be interested in determining the proportion of defectives produced by a certain machine. He may choose n items from the daily production and classify each one of them as defective (D) or not (G). The number of possible outcomes of this experiment is 2^n . The typical outcome may be represented by an n -tuple consisting of D 's and G 's as $(DDG...GD)$, which may be denoted by ω . "The number of defectives in a sample of 5 does not exceed 2" is an event. It consists of sample points $(DDGGG)$, $(DGDGG)$, $(DGGDG)$, $(DGGGD)$, $(GDDGG)$, $(GDGDG)$, $(GDGGD)$, $(GGDDG)$, $(GGDGD)$, $(GGGDD)$, $(DGGGG)$, $(GDGGG)$, $(GGDGG)$, $(GGGDG)$, $(GGGGD)$, $(GGGGG)$.

Example 1.2. A textile manufacturer may observe the number of faults in a yard of cloth chosen from the daily production. The possible outcome ω may be represented by a number $0, 1, 2, \dots$. Thus $\Omega = \{0, 1, 2, \dots\}$ contains countably infinite number of points. "The number of faults per yard exceeds 5" is an event, containing the sample points $6, 7, \dots$. "The number of faults is 3" is an event containing only one sample point 3.

Example 1.3. The managing director of a company may be interested in knowing the profit (or loss), ω , which is likely to occur to his company in the coming year. Since ω may be positive or negative and may take any value, Ω may be taken to be $\{-\infty < \omega < \infty\}$. 'Profit will exceed a ' is an event consisting of all those ω which exceed a . It may be written as $\{\omega : \omega > a\}$. Similarly $\{\omega : a < \omega < b\}$ is an event consisting of all those ω which lie in the range (a, b) , representing the event that profit will lie in the range (a, b) .

Events will be denoted by Latin letters $A, B, C, \dots$, etc. Many a time the events will consist of all those sample points ω which obey a certain relation, as is the case in Example 1.3. $\{\omega : \omega \text{ satisfying a relation } R\}$ is an event consisting of all those ω 's which satisfy the relation R . The event consisting of a single outcome ω_0 will be denoted by $\{\omega_0\}$. It is a *singleton* and is also called a *simple event*. If the event contains ω_1 and ω_2 it is denoted by $\{\omega_1, \omega_2\}$ and is a *double-ton*. For convenience, we shall consider the subset of Ω containing no point, viz., the empty set ϕ , to be an event and shall call it an *impossible event*. Similarly the whole space Ω will be called a *sure event*.

Any collection of events is a *class of events*. Classes will be denoted by script letters $\mathcal{A}, \mathcal{B}, \mathcal{C}$, etc. In example 1.1 the sample space Ω contains only a finite number of sample points. Hence the number of all possible subsets of Ω is also finite. If Ω contains n points, there are 2^n subsets of Ω . The class of all subsets of Ω is called the *power set* of Ω . Note that the power set is a class. The number of points in a set is called its *cardinality*.

Example 1.4. Let $\Omega = \{\omega_1, \omega_2, \omega_3\}$. Then the power set consists of the 8 events $\{\phi\}, \{\omega_1\}, \{\omega_2\}, \{\omega_3\}, \{\omega_1, \omega_2\}, \{\omega_1, \omega_3\}, \{\omega_2, \omega_3\}, \{\omega_1, \omega_2, \omega_3\}$. (What is the power set of Ω when $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$?)

It must be emphasized here that the experimenter may be interested only in a particular class of outcomes and not in all events. In Example 1.1, suppose the experimenter is interested only in the total number of defective items in the observed sample. If ω_i denotes the outcome of observing i defectives, Ω will contain only 6 sample points, $\omega_0, \omega_1, \dots, \omega_5$. $\{\omega_i\}$ is a simple event in the present case, but a composite event containing $\binom{5}{i}$ sample points in the case of Example 1.1. The power set of Ω in the present case is only a subclass of the power set of Example 1.1.

We shall use the following abbreviations which have become standard in mathematical texts.

Symbol	Meaning	Example
$\in$	element of	$\omega \in A, A \in \mathcal{A}$
$\notin$	not an element of	$\omega \notin A$
$\Rightarrow$	logical implication	$a > b, b > c \Rightarrow a > c$
$\Leftrightarrow$	logical equivalence	$a \geq b \text{ and } b \geq a \Leftrightarrow a = b$
$\ni$	such that	
$\exists$	there exists	
iff	if and only if	
$\forall$	each, all or every	

1.2. ALGEBRA OF SETS

(a) SET-OPERATIONS

Since events are sets of sample points it is essential that one becomes familiar with the algebra of sets, in order to understand manipulations involving events. In the following we assume that Ω is given. The important set operations are: (i) Complementation; (ii) Inclusion and Equality; and (iii) Union and Intersection.

(i) *Complementation*: To every set A we can associate another set A^c consisting of all points of Ω not contained in A . A^c is called the *complement* of A . Symbolically,

$$A^c = \{\omega : \omega \notin A\} \quad (1.1)$$

Evidently, $\Omega^c = \phi$ and $\phi^c = \Omega$.

Example 1.5. Suppose a coin is tossed 3 times. The possible outcomes may be denoted by

$$\begin{array}{cccccccc} HHH, & HHT, & HTH, & THH, & HTT, & THT, & TTH, & TTT \\ \omega_1, & \omega_2, & \omega_3, & \omega_4, & \omega_5, & \omega_6, & \omega_7, & \omega_8 \end{array}$$

If A is the event that at least one head turns up,

$$A = \{\omega_1, \omega_2, \dots, \omega_7\}$$

then $A^c = \{TTT\} = \{\omega_8\}$ represents the event that no head turns up.

If B is the event that exactly one head turns up, then

$$B = \{\omega_5, \omega_6, \omega_7\}, B^c = \{\omega_1, \dots, \omega_4, \omega_8\},$$

latter representing the event that the number of heads turning up is not

equal to one. In general, complements of events are events (see section 1.4).

(ii)a *Inclusion*: If all points of a set A are also points of another set B , then we say that A is a *subset* of B or, A is included in or *contained in* B . This is denoted by $A \subset B$ or $B \supset A$. Symbolically,

$$A \subset B \Leftrightarrow (\omega \in A \Rightarrow \omega \in B). \quad (1.2)$$

Evidently (i) $A \subset A$, (reflexivity);

$$\begin{aligned} \text{(ii) } A \subset B, B \subset C &\Rightarrow (\omega \in A \Rightarrow \omega \in B \Rightarrow \omega \in C) \\ &\Rightarrow A \subset C \text{ (transitivity)}. \end{aligned} \quad (1.3)$$

In Example 1.5, B is included in A .

If Ω represents the collection of all students, A , who pass in first class and B , who pass, then $A \subset B$. It may be noted that

$$(A^c)^c = \{\omega : \omega \notin A^c\} = \{\omega : \omega \in A\} = A \quad (1.4)$$

and if $A \subset B$, then

$$A^c \supset B^c, \quad (1.5)$$

Students who do not pass form a subset of those who do not pass in first class.

We shall see later that a subset of an event may not be an event.

(ii)b *Equality*: If $A \subset B$ and $B \subset A$, then A and B are said to be *equal*, denoted by $A = B$. Thus in any problem, in order to establish the equality of two sets A and B we have to prove that

$$(\omega \in A) \Leftrightarrow (\omega \in B).$$

Since inclusion relation is reflexive and transitive, equality relation is also reflexive and transitive. It is also symmetric, i.e., $A = B \Leftrightarrow B = A$.

(iii) *Union and Intersection*: If A and B are two sets, then the set of all points ω which belong to either A or B is called *A union B* and is denoted by $A \cup B$. The set of all points which belong to both A and B is called *A intersection B* and is denoted by $A \cap B$. Symbolically

$$A \cup B = \{\omega : \omega \in A \text{ or } \omega \in B\} = \{\omega : \omega \text{ belongs to at least one of the sets } A \text{ or } B\} \quad (1.6)$$

$$A \cap B = \{\omega : \omega \in A \text{ and } \omega \in B\} = \{\omega : \omega \in \text{both } A \text{ and } B\}. \quad (1.7)$$

If $A \subset B$,

$$A \cup B = B, A \cap B = A \text{ (Why?)}. \quad (1.8)$$

Example 1.6. In Example 1.5, if

$$A = \{\text{all the 3 tosses result in the same outcome}\},$$

$$= \{\omega_1, \omega_3\},$$

and $B = \{\text{the 3 tosses have at most one tail}\},$

$$= \{\omega_1, \omega_2, \omega_3, \omega_4\},$$

then $A \cup B = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5\},$

$$A \cap B = \{\omega_1\}.$$

If $A \cap B = \phi$ then A and B are said to be *disjoint* or *mutually exclusive*. In this case and only in this case $A \cup B$ will be denoted by $A + B$. Thus, $A + A^c = \Omega$. We use this notation throughout the book.

If $A \subset B$, $B \cap A^c$ will be denoted by $B - A$ and is called the *proper difference* of B and A . $B - A$ and A are disjoint and $(B - A) + A = B$. We may note that $A^c = \Omega - A$. Many a time $A \cap B$ is written as AB omitting $\cap$.

We may note that

$$AB \subset A, A - AB = AB^c \subset B^c;$$

$$AB \subset B, B - AB = BA^c \subset B.$$

Hence AB^c and BA^c are disjoint. $AB^c + BA^c = A \Delta B$ is called the *symmetric difference* of A and B . Δ is called the *symmetric difference operation*. While the proper difference is defined only if $A \subset B$, symmetric difference is always defined. (Mark $A \Delta B$ on a Venn-diagram!) We may note that $AB + AB^c = A$.

Unions, intersections and differences of events are events (see sec. 1.4).

Example 1.7. Suppose Ω is the real line R consisting of all real points ω ,

$${}_{\omega}\Omega = \{\omega : -\infty < \omega < \infty\}.$$

Define

$$A = \{\omega : \omega \in (-\infty, a)\} = \{\omega < a\},$$

$$B = \{\omega : \omega \in (c, d)\} = \{c < \omega < d\},$$

Then

$$A \cap B = \phi, \text{ if } a < c < d,$$

$$= (c, a), \text{ if } c < a < d,$$

$$= (c, d), \text{ if } c < d < a.$$

Since $A \cup B = \{\omega : \text{either } \omega < a \text{ or } c < \omega < d\}$, it will not be an interval if $a < c < d$, even though A and B are intervals.

If $a < c$, $A^c \supset B$, and $A^c \cup B = A^c$, $A^c \cap B = B$. Note that $A^c = [a, \infty)$, $B^c = (-\infty, c] \cup [d, \infty)$. Thus the complement of an interval need not be an interval.

The definition of union of an arbitrary possibly uncountable number of sets is given by

$$\bigcup_{i \in I} A_i = \{\omega \in \Omega; \omega \in A_i \text{ for some } i \in I\}, \quad (1.9)$$

where I is an arbitrary index set assumed nonempty. If I is finite we

have a finite union. If I is countable and is given by $1, 2, \dots$, we have a countable union denoted by $\bigcup_{i=1}^{\infty} A_i$. Similarly the intersection of $A_i (i \in I)$ is

$$\bigcap_{i \in I} A_i = \{\omega \in \Omega : \omega \in A_i \text{ for all } i \in I\}. \quad (1.10)$$

The operations of $\cup$ and $\cap$ are reflexive, commutative, associative and distributive as can be easily verified (How?).

$$(i) \quad A \cup A = A, A \cap A = A; \text{ (Reflexive)}$$

$$(ii) \quad A \cup B = B \cup A, A \cap B = B \cap A; \text{ (Commutative)}$$

$$(iii) \quad (A \cup B) \cup C = A \cup B \cup C = A \cup (B \cup C) \\ = (A \cup C) \cup B;$$

$$(A \cap B) \cap C = A \cap B \cap C = A \cap (B \cap C) \\ = (A \cap C) \cap B; \text{ (Associative)}$$

$$(iv) \quad A \cap (B \cup C) = (A \cap B) \cup (A \cap C);$$

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C). \text{ (Distributive)}$$

From (iv) we conclude that $A(B+C) = AB + AC$ and more generally

$$A(\sum_i B_i) = \sum_i AB_i$$

By the definition of A^c ,

$$A \cup A^c = \Omega, A \cap A^c = \phi,$$

Now

$$(A \cap B)^c = \{\omega : \omega \notin A \cap B\}, \\ = \{\omega : \omega \text{ does not belong to both } A \text{ and } B\}, \\ = \{\omega : \text{either } \omega \notin A \text{ or } \omega \notin B\}, \\ = A^c \cup B^c. \quad (1.11)$$

Replacing A by A^c and B by B^c in (1.11), we have

$$(A^c \cap B^c)^c = (A^c)^c \cup (B^c)^c = A \cup B$$

and hence taking complementation of both sides,

$$A^c \cap B^c = (A \cup B)^c. \quad (1.12)$$

This may be proved directly also. In general, if $A_i, i \in I$ are arbitrary sets

$$(\bigcup_{i \in I} A_i)^c = \{\omega : \omega \text{ does not belong to even one } A_i, i \in I\}, \\ = \{\omega : \omega \notin A_i \text{ for each one of } i \in I\}, \\ = \bigcap_{i \in I} A_i^c \quad (1.13)$$

$$\text{and } (\bigcap_{i \in I} A_i)^c = \{\omega : \omega \text{ does not belong to each one of } A_i\text{'s}\}, \\ = \{\omega : \omega \text{ does not belong to at least one } A_i\},$$

$$= \bigcup_{i \in I} A_i^c. \quad (1.14)$$

(1.13) and (1.14) are known as *de Morgan rules* (or formulae) for deriving complements.

LEMMA 1.1: Given a class $\{A_i, i = 1, 2, \dots, n\}$ of n sets there exists a class $\{B_i, i = 1, 2, \dots, n\}$ of disjoint sets such that,

$$\bigcup_{i=1}^n A_i = \bigcup_{i=1}^n B_i. \quad (1.15)$$

Proof: The Lemma will be proved by induction. Evidently

$$A_1 \cup A_2 = A_1 + A_1^c A_2 = B_1 + B_2, \text{ say,} \quad (1.16)$$

where B_1 and B_2 are disjoint.

Thus the Lemma is true for $n = 2$. Suppose it is true for all $n \leq m \leq 2$.

Then

$$\begin{aligned} \bigcup_{i=1}^{m+1} A_i &= \left(\bigcup_{i=1}^m A_i \right) \cup A_{m+1}, \\ &= \left(\sum_{i=1}^m B_i \right) \cup A_{m+1} \text{ (by induction hypothesis),} \\ &= \sum_{i=1}^m B_i + (\sum_{i=1}^m B_i)^c A_{m+1}, \quad (\text{by 1.16}) \\ &= \sum_{i=1}^m B_i + B_{m+1}, \text{ say,} \end{aligned}$$

where B_{m+1} and $\sum B_i$ are disjoint and hence B_{m+1} and B_i are disjoint for $i = 1, 2, \dots, m$ (Why?). Hence the Lemma holds for $n = m + 1$ and by induction, for arbitrary n . (*)¹

Note that $B_i \subset A_i$ for all i .

$$\text{COROLLARY 1.1: } \bigcup_{i=1}^{\infty} A_i = A_1 + A_1^c A_2 + A_1^c A_2^c A_3 + \dots \quad (1.17)$$

This is the extension of Lemma 1.1 to the countable case. It tells us how to get a countable class of disjoint sets, starting from a countable class of arbitrary sets, such that their unions are equal. Thus, in future, we may assume that we are given a countable class of disjoint sets, without loss of generality. (Is the class $\{B_i\}$ on the RHS of (1.17) unique?)

Proof: Suppose $\omega \in \bigcup_{i=1}^{\infty} A_i$. Then ω belongs to some A_i . Thus ω may belong to A_1 or ω may belong to A_1^c . In the latter case ω has to belong to A_2 or A_2^c . Thus either $\omega \in A_1^c A_2$ or $\omega \in A_1^c A_2^c$. In the latter case ω has to belong to A_3 or A_3^c . Continuing in this manner, ω has to belong

¹Will signify the end of the proof.

to either $A_1, A_1^c A_2, A_1^c A_2^c A_3, \dots, A_1^c A_2^c \dots, A_{n-1}^c A_n$, etc. Hence $\omega \in \text{RHS}$ of (1.17).

Conversely, if $\omega \in \text{RHS}$ of (1.17), $\omega \in A_1^c A_2^c \dots, A_{k-1}^c A_k$ for some k . But $A_1^c A_2^c \dots, A_{k-1}^c A_k \subset A_k$. Hence $\omega \in A_k$ for some k . Thus $\omega \in \bigcup_{k=1}^{\infty} A_k$, which establishes the equivalence of the two sets on the two sides of (1.17). (*)

(b) SEQUENCES, LIMITS

A sequence of sets $\{A_n\}$ is said to be *monotone increasing* if $A_n \subseteq A_{n+1}$ for each n . Since in this case $\bigcup_{k=1}^n A_k = A_n$, $\bigcup_{k=1}^{\infty} A_k = A$ is called the *limit* of the sequence. Symbolically this is written as $A_n \uparrow A$. If $A_n \supseteq A_{n+1}$ for each n , the sequence $\{A_n\}$ of sets is said to be *monotone decreasing*. Then $\bigcap_{k=1}^n A_k = A_n$ and $\bigcap_{k=1}^{\infty} A_k = A$ is called the *limit* of $\{A_n\}$. This is written as $A_n \downarrow A$.

Example 1.8. Let $A_n = \{\omega: 0 < \omega < 1 - 1/n\} \subset (-\infty, \infty) = \Omega$. Then $A_n \uparrow A = \{\omega: 0 < \omega < 1\}$. Let $B_n = \{\omega: 0 < \omega < 1 + 1/n\}$. Then $B_n \downarrow B = \{\omega: 0 < \omega \leq 1\}$. It may be noted that A is the union of A_n 's and B is the intersection of B_n 's.

If $A_n \downarrow$, then $A_n^c \uparrow$. If $A_n \downarrow A$, then by de Morgan rule it follows that $A_n^c \uparrow A^c$ (Why?).

So far we defined the limit of a monotone sequence of sets. Now we shall proceed to define it for an arbitrary sequence. For any sequence $\{A_n\}$, define

$$\begin{aligned} B_n &= \inf_{k \geq n} A_k = \bigcap_{k=n}^{\infty} A_k, \\ &= \{\omega: \omega \text{ belonging to all } A_k \text{'s except perhaps } A_1, \dots, A_{n-1}\}; \\ C_n &= \sup_{k \geq n} A_k = \bigcup_{k=n}^{\infty} A_k, \\ &= \{\omega: \omega \text{ belonging to at least one of } A_n, A_{n+1}, \dots\}. \end{aligned}$$

Then $\{B_n\}$ is a monotone increasing sequence (Why?) with limit

$$B = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k = \liminf A_n = \underline{\lim} A_n$$

and $\{C_n\}$ is a monotone decreasing sequence with limit

$$C = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k = \limsup A_n = \overline{\lim} A_n$$

Evidently B is the set of all points which belong to all but a finite number of A_n 's. By de Morgan rule, $C^c = (\liminf A_n^c) = \{\omega: \omega \text{ belonging}$

to only finite A_n 's. Thus C is the set of all points which belong to infinitely many A_n 's. If ω belongs to all but a finite number of A_n 's, it belongs to infinitely many of them. Hence,

$$\lim A_n \subseteq \overline{\lim} A_n. \quad (1.18)$$

If $\lim A_n = \overline{\lim} A_n = A$, say, then limit of $\{A_n\}$ is said to exist and A is called the *limit*. If $\{A_n\}$ is monotone, decreasing or increasing, then the limit as defined now coincides with the one given earlier. If $\{A_n\}$ is not monotone and A exists, then we write $A_n \rightarrow A$. We also say that $\{A_n\}$ converges to A . It should be noted that $\lim A_n$ and $\overline{\lim} A_n$ will always exist, even though $\lim A_n$ may not exist.

Example 1.9. Let A_n be the set of points (x, y) of the Cartesian plane lying within the rectangle bounded by the two axes and the lines $x = n$ and $y = 1/n$, i.e.

$$A_n = \{(x, y): 0 \leq x < n, 0 \leq y < 1/n\}, (x, y) \in R^2.$$

Evidently $\{A_n\}$ is not monotone. But $C_n \subset \{(x, y): 0 \leq x < \infty, 0 \leq y < 1/n\}$ and

$$C = \overline{\lim} A_n = \{(x, y): 0 \leq x < \infty, y = 0\}.$$

Similarly, $B_n = \bigcap_{k=n}^{\infty} A_k \subset \{(x, y): 0 \leq x < n, y = 0\}$ and

$$B = \lim A_n = \{(x, y): 0 \leq x < \infty, y = 0\} = \overline{\lim} A_n = C$$

is the limit of $\{A_n\}$.

Example 1.10. Consider the sequence $\{A_n\}$, where

$$A_n = \{\omega: 0 < \omega < b + (-1)^n/n\}, \quad (b > 1)$$

Then
$$C_n = \begin{cases} [\omega: 0 < \omega < b + (1/n)], & \text{if } n \text{ is even,} \\ [\omega: 0 < \omega < b + (1/(n+1))], & \text{if } n \text{ is odd,} \end{cases}$$

and
$$\overline{\lim} A_n = \{\omega: 0 < \omega \leq b\}.$$

Similarly,
$$B_n = \begin{cases} [\omega: 0 < \omega < b - (1/n)] & \text{if } n \text{ is odd,} \\ [\omega: 0 < \omega < b - (1/(n+1))] & \text{if } n \text{ is even,} \end{cases}$$

and hence
$$\lim A_n = \{\omega: 0 < \omega < b\}.$$

Then $\overline{\lim} A_n \neq \lim A_n$ and $\{A_n\}$ does not converge.

1.3 FIELDS AND σ -FIELDS

(a) CLOSURE

Suppose $\mathcal{A}$ is a class of sets. If by performing a certain operation on one or more elements of $\mathcal{A}$, we obtain an element of the same class, we

say that the class is *closed* under that operation. For example, if

$$A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A},$$

($A^c \in \mathcal{A}$ whenever A belongs to $\mathcal{A}$) then $\mathcal{A}$ is said to be closed under complementation. If

$$A, B \in \mathcal{A} \Rightarrow A \cup B \in \mathcal{A}, \quad (1.19)$$

then $\mathcal{A}$ is said to be closed under unions. We may note that by induction (1.19) implies that

$$A_1, A_2, \dots, A_n \in \mathcal{A} \Rightarrow \bigcup_{i=1}^n A_i \in \mathcal{A}, \quad \forall n < \infty. \quad (1.20)$$

Thus (1.19) implies closure of $\mathcal{A}$ under finite unions. Similarly, if

$$A, B \in \mathcal{A} \Rightarrow A \cap B \in \mathcal{A} \quad (1.21)$$

$$\text{then } A_1, A_2, \dots, A_n \in \mathcal{A} \Rightarrow \bigcap_{i=1}^n A_i \in \mathcal{A}, \quad \forall n < \infty. \quad (1.22)$$

Hence (1.21) implies that $\mathcal{A}$ is closed under finite intersections. Thus, in order to establish closure under finite intersection (or union) it is sufficient if we establish (1.21) (or 1.19).

Example 1.11. Let $\mathcal{C}$ be the class of all intervals of the form (x, ∞) , $x \in R$, considered as subset of the real line. Now (1.19) and (1.21) are satisfied for $\mathcal{C}$, since,

$$(x, \infty) \cup (y, \infty) = (u, \infty),$$

$$(x, \infty) \cap (y, \infty) = (v, \infty),$$

where $u = \min(x, y)$ and $v = \max(x, y)$. Hence $\mathcal{C}$ is closed under finite unions and finite intersections. But

$$(x, \infty)^c = (-\infty, x] \notin \mathcal{C}.$$

Hence $\mathcal{C}$ is not closed under complementation.

Example 1.12. Consider the class $\mathcal{C}$ of all intervals of the form (a, b) , $a, b \in R$ ($a < b$) and ϕ . Then,

$$(a, b) \cap (c, d) = \phi, \quad \text{if } a < b < c < d \text{ or } c < d < a < b.$$

$$= (c, b), \quad \text{if } a < c < b < d,$$

$$= (a, d), \quad \text{if } c < a < d < b,$$

$$= (c, d), \quad \text{if } a < c < d < b,$$

$$= (a, b), \quad \text{if } c < a < b < d.$$

Thus, $\mathcal{C}$ is closed under finite intersections. But it is not closed under complementations or unions, because

$$(a, b)^c = (-\infty, a] \cup [b, \infty) \notin \mathcal{C}$$

and $(a, b) \cup (c, d)$ is not an interval, if $a < b < c < d$, or $c < d < a < b$.

(b) FIELD, MINIMAL FIELD, PARTITION

A nonempty class of sets $\mathcal{A}$, closed under complementation and finite intersections, is called a field.

LEMMA 1.2: *A field is closed under finite unions. Conversely, a class closed under complementations and finite unions is a field.*

Proof: Suppose $\mathcal{A}$ is a field. Then

$$(i) \quad A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}, \quad (1.23)$$

$$(ii) \quad A_1, A_2, \dots, A_n \in \mathcal{A} \Rightarrow \bigcap_{i=1}^n A_i \in \mathcal{A}. \quad (1.24)$$

But $A_1, A_2, \dots, A_n \in \mathcal{A} \Rightarrow A_1^c, A_2^c, \dots, A_n^c \in \mathcal{A}$, (by 1.23)

$$\Rightarrow \bigcap_{i=1}^n A_i^c \in \mathcal{A}, \quad (\text{by 1.24})$$

$$\Rightarrow \left(\bigcap_{i=1}^n A_i^c \right)^c \in \mathcal{A}, \quad (\text{by 1.23})$$

$$\Rightarrow \bigcup_{i=1}^n A_i \in \mathcal{A}, \quad (\text{by de Morgan rule}).$$

Hence $\mathcal{A}$ is closed under finite unions also.

Conversely, suppose $\mathcal{A}$ is a class such that (1.23) holds true for every $A \in \mathcal{A}$ and

$$(iii) \quad A_1, A_2, \dots, A_n \in \mathcal{A} \Rightarrow \bigcup_{i=1}^n A_i \in \mathcal{A}.$$

Then proceeding as above and using de Morgan rule, we can prove that $\mathcal{A}$ is closed under finite intersections also. Hence we have the Lemma. (*)

From the Lemma, a field is sometimes defined as a class closed under complementations, finite intersections and/or finite unions. Evidently, if $\mathcal{A}$ is a field,

$$A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A},$$

$$\Rightarrow A \cup A^c \in \mathcal{A}, \quad \text{i.e. } \Omega \in \mathcal{A}, \quad (1.25)$$

$$\Rightarrow A \cap A^c \in \mathcal{A}, \quad \text{i.e. } \phi \in \mathcal{A}. \quad (1.26)$$

Hence we have the following Lemma.

LEMMA 1.3: *Every field contains the empty set ϕ and the whole space Ω .*

The class containing only ϕ and Ω is a field. It is the smallest field and is contained in every other field. It is called the *degenerate* or *trivial*.

field. The power set consisting of every subset of Ω is also a field and is the largest field. If a field contains A , it has to contain A^c , and hence contains the class

$$\{A, A^c, \phi, \Omega\}. \quad (1.27)$$

But this class is a field, contained in every field containing A . Therefore it is the smallest field containing A . (What is the smallest field containing A, B ?)

Example 1.13. Let A_1, A_2, A_3 be such that $A_i \cap A_j = \phi$ for all $i, j = 1, 2, 3$ ($i \neq j$) and $\sum_{i=1}^3 A_i = \Omega$. Thus Ω is subdivided with 3 mutually exclusive and *exhaustive* sets. To determine the field containing the class A_1, A_2, A_3 , it is sufficient if we add to the above class the sets $\phi, A_1 + A_2, A_2 + A_3, A_1 + A_3$ and Ω . Since $A_1 + A_2 = A_3^c, A_1 + A_3 = A_2^c, A_2 + A_3 = A_1^c$, the resulting class

$$\{\phi, A_1, A_2, A_3, A_1 + A_2, A_2 + A_3, A_1 + A_3, \Omega\} \quad (1.28)$$

is closed under complementations and finite unions; hence it is a field. It is the smallest field containing $\{A_1, A_2, A_3\}$.

Consider an arbitrary class $\mathcal{C}$ of sets. The smallest field containing $\mathcal{C}$ is called the *minimal field* containing $\mathcal{C}$ or the field *generated* by $\mathcal{C}$. It is contained in every field containing $\mathcal{C}$. If $\mathcal{C}_i$ ($i \in I$) denote the fields containing $\mathcal{C}$, then $\mathcal{C}_0 = \bigcap_{i \in I} \mathcal{C}_i$ is the minimal field containing $\mathcal{C}$ denoted by $\mathcal{F}(\mathcal{C})$. This follows from the following theorem 1.1 and the fact that $\mathcal{C}$ is contained in $\mathcal{C}_0$ (see also Neveu (1965), p. 14.)

THEOREM 1.1: *The intersection of arbitrary number of fields is a field.*

Proof: Let $\mathcal{C}_i$ ($i \in I$) be the fields indexed by points of a nonempty set I which may be finite, countable or uncountable. Let $\mathcal{C}_0 = \bigcap_{i \in I} \mathcal{C}_i$, the class of all sets common to all $\mathcal{C}_i$. Since each one of the $\mathcal{C}_i$ contains ϕ and Ω , $\mathcal{C}_0$ contains ϕ and Ω . Suppose $A \in \mathcal{C}_0$, then $A \in \mathcal{C}_i, \forall i \in I$. Hence $A^c \in \mathcal{C}_i$, for all i . Thus $A^c \in \mathcal{C}_0$, and $\mathcal{C}_0$ is closed under complementation.

If $A_k \in \mathcal{C}_0, (k = 1, 2, \dots, n)$, then $A_k \in \mathcal{C}_i, \forall i \in I$. Then $\bigcap_{k=1}^n A_k \in \mathcal{C}_i$ for all $i \in I$ and hence it belongs to $\bigcap_{i \in I} \mathcal{C}_i = \mathcal{C}_0$ implying that $\mathcal{C}_0$ is closed under finite intersections. Hence $\mathcal{C}_0$ is a field. (*)

We may note that even though the intersections of two or more fields is a field, the unions of two fields may not be a field; e.g. $\{A, A^c, \phi, \Omega\}, \{B, B^c, \phi, \Omega\}$ are two fields, but their union

$$\{A, A^c, B, B^c, \phi, \Omega\} \quad (1.29)$$

is not a field.

Let $\{A_i\}$ $i = 1, 2, \dots, n$ be a class of disjoint sets such that $\sum_{i=1}^n A_i = \Omega$. Thus, $\{A_i\}$ are mutually exclusive and exhaustive. Then $\{A_i\}$ is called a *partition* of Ω . In this case,

$$A_1^c = A_2 + A_3 + \dots + A_n, (A_1 + A_2)^c = A_3 + A_4 + \dots + A_n,$$

and so on. Hence the class

$$\{\phi, A_1, A_2, \dots, A_n, A_1 + A_2, A_1 + A_3, \dots, A_{n-1} + A_n, A_1 + A_2 + A_3, \dots, A_{n-2} + A_{n-1} + A_n, \dots, A_1 + A_2 + \dots + A_n = \Omega\}, \quad (1.30)$$

containing ϕ , Ω and the unions of A_i 's taken one set at a time, 2 sets at a time, etc., which is therefore closed under finite unions, is also closed under complementations. Thus it is a field. It is the minimal field containing

$$\{A_i, i = 1, 2, \dots, n\}.$$

It contains $\binom{n}{0} + \binom{n}{1} + \dots + \binom{n}{n} = 2^n$ sets.

Partition $\mathcal{A}'$ is said to be *finer* than partition $\mathcal{A}$, or $\mathcal{A}$, *coarser* than $\mathcal{A}'$, if the minimal field $\mathcal{F}(\mathcal{A}')$ containing $\mathcal{A}'$ contains the minimal field $\mathcal{F}(\mathcal{A})$ containing $\mathcal{A}$. For example, $\{A, A^c\}$ is coarser than $\{A, BA^c, (A \cup B)^c\}$; $\{AB, AB^c, BA^c, (A \cup B)^c\}$ is finer than both these partitions. This may be verified by obtaining the minimal fields containing these partitions.

We have seen that it is very easy to obtain $\mathcal{F}(\mathcal{C})$, the minimal field containing a class $\mathcal{C}$, if $\mathcal{C}$ is a partition. In general, to obtain $\mathcal{F}(\mathcal{C})$ we may proceed along the following steps:

(i) Obtain $\mathcal{C}_1 = \{\phi, \Omega, A, A^c\}$, such that either $A \in \mathcal{C}$ or $A^c \in \mathcal{C}$ where $A \subset \Omega$. Evidently $\mathcal{C}_1$ is closed under complementation and contains $\mathcal{C}$.

(ii) Obtain the class $\mathcal{C}_2$ containing $\bigcap_{k=1}^n B_k$, where $B_k \in \mathcal{C}_1$, $k = 1, 2, \dots, n$, B_k 's and n being arbitrary. Now $\mathcal{C}_2$ is closed under finite intersections, but not under complementation.

(iii) Obtain $\mathcal{C}_3$, the class of all finite unions of pair-wise disjoint subsets belonging to $\mathcal{C}_2$. Since they also contain complements, $\mathcal{C}_3$ is a field and is the minimal field containing $\mathcal{C}$. (Prove!)

Example 1.14. Let $\mathcal{C} = \{A, B\}$. Then

$$\mathcal{C}_1 = \{\phi, \Omega, A, A^c, B, B^c\},$$

$$\mathcal{C}_2 = \{\mathcal{C}_1, AB, AB^c, A^cB, A^cB^c\},$$

$$\mathcal{C}_3 = \{\mathcal{C}_2, \overbrace{AB + A^cB^c}^{AB^c + A^cB}, \overbrace{AB^c + A^cB}^{AB + A^cB^c}, \overbrace{AB^c + A^cB^c}^{AB + A^cB} + \overbrace{AB + A^cB^c}^{AB^c + A^cB} + \overbrace{AB^c + A^cB^c}^{AB + A^cB^c}\},$$

$\mathcal{C}_3$ is the smallest field containing $\mathcal{C}$. It coincides with the minimal field containing the partition $\{AB, AB^c, A^cB, A^cB^c\}$. The minimal field containing $\{A, B, C\}$ is the minimal field containing the partition, $\{ABC, A^cBC, AB^cC, ABC^c, A^cB^cC, AB^cC^c, A^cBC^c, A^cB^cC^c\}$ (Why?).

(c) σ -FIELD, BOREL FIELD

Closure under finite operations does not imply closure under countable operations. In Example 1.11, we have seen that the class $\mathcal{C}$ of all intervals of form (x, ∞) , $x \in R$ is closed under finite intersection. But it is not closed under countable intersections, because

$$\bigcap_{n=1}^{\infty} (x - 1/n, \infty) = [x, \infty) \notin \mathcal{C}.$$

A nonempty class of sets which is closed under complementations and countable unions (or countable intersections) is called a σ -field. Evidently it is a field closed under countable operations of union and intersection (How?). It possesses all the properties of a field, hence contains ϕ and Ω . If $\mathcal{C}$ contains only a finite number of sets and is a field, it is also a σ -field. However a field containing an infinite number of sets may not be a σ -field.

Example 1.15. Let $\Omega = \{1, 2, 3, 4, \dots\}$ and $\mathcal{C}$ be the class of subsets A of Ω such that either A contains a finite number of points or A^c contains a finite number of points. Evidently $\mathcal{C}$ is closed under complementation. It is also closed under finite unions, because $A \cup B$ will contain a finite number of points if each one of A and B is finite and $(A \cup B)^c = A^c \cap B^c$ will contain a finite number of points if either A^c is finite or B^c is finite. Thus either $A \cup B$ contains a finite number of points or $(A \cup B)^c$ contains a finite number of points. Hence $A \cup B \in \mathcal{C}$. Thus $\mathcal{C}$ is a field.

But $\mathcal{C}$ is not a σ -field. Let

$$A_i = \{2i\} \quad (i = 1, 2, \dots).$$

Then $\bigcup_{i=1}^{\infty} A_i = \{2, 4, 6, \dots\}$, is neither finite nor its complement is finite. Hence it does not belong to $\mathcal{C}$.

By Theorem 1.1, intersection of arbitrary number of fields is a field. Following similar steps we can prove Theorem 1.2.

THEOREM 1.2: The intersection of an arbitrary number of σ -fields is a σ -field.

Given a class $\mathcal{C}$, the minimal σ -field containing the class $\mathcal{C}$ will be denoted by $\sigma(\mathcal{C})$. It is the intersection of all σ -fields containing $\mathcal{C}$. It is also called the σ -field generated by $\mathcal{C}$. If $\mathcal{C}$ is finite, the minimal field

$\mathcal{F}(\mathcal{C})$ containing $\mathcal{C}$ coincides with the minimal σ -field $\sigma(\mathcal{C})$ containing $\mathcal{C}$.

To obtain $\sigma(\mathcal{C})$ from a given class $\mathcal{C}$, we follow the same steps (i)–(iii) used in the procedure of obtaining $\mathcal{F}(\mathcal{C})$, except that in step 2 we allow n to be infinite.

Example 1.16. Consider the class $\mathcal{C}$ of all intervals of the form $(-\infty, x)$, $x \in R$, as subsets of the real line R . This class is closed under finite intersections, but not under complementation nor under countable intersections. Let $\sigma(\mathcal{C}) = \mathcal{B}$ be the minimal σ -field containing $\mathcal{C}$. Then $\mathcal{B}$ contains intervals of the form $[x, \infty)$; which are complements of sets of the form $(-\infty, x)$. It contains intervals of the form

$$\left. \begin{aligned} (-\infty, a] &= \bigcap_{n=1}^{\infty} (-\infty, a + 1/n), \text{ (by countable intersection),} \\ (a, \infty) &= (-\infty, a]^c, \text{ (by complementation)} \\ (a, b) &= (-\infty, b) \cap (a, \infty) \text{ (} a < b \text{),} \\ (a, b], [a, b), \text{ etc., for } a, b \in R. \end{aligned} \right\} (1.31)$$

$\mathcal{B}$ is called the *Borel field* of subsets of the real line. The sets of $\mathcal{B}$ are called *Borel sets*. Borel field and Borel sets play a very important role in the study of probability theory.

LEMMA 1.4: Let $\mathcal{C}_1$ be the class of all intervals of the form (a, b) ($a < b$), $a, b \in R$, but arbitrary. Then $\sigma(\mathcal{C}_1) = \mathcal{B}$.

Proof: By (1.31), $(a, b) \in \mathcal{B}$, for all a, b . Hence $\mathcal{C}_1 \subset \mathcal{B}$. By definition of minimal σ -field, $\sigma(\mathcal{C}_1) \subset \mathcal{B}$.

To prove inclusion in the reverse direction, we note that, by definition of $\sigma(\mathcal{C}_1)$,

$$\begin{aligned} \bigcup_{n=1}^{\infty} (-n, x) &\in \sigma(\mathcal{C}_1), \forall x \\ \Rightarrow (-\infty, x) &\in \sigma(\mathcal{C}_1), \forall x \\ \Rightarrow \mathcal{C} &\subset \sigma(\mathcal{C}_1) \end{aligned}$$

where $\mathcal{C}$ is defined in Example 1.16. Hence

$$\sigma(\mathcal{C}) \subset \sigma(\mathcal{C}_1).$$

But $\sigma(\mathcal{C}) = \mathcal{B}$ and therefore $\sigma(\mathcal{C}_1) = \sigma(\mathcal{C}) = \mathcal{B}$. (*)

In the same manner as above, one can prove that the Borel field is the minimal field containing any one of the following classes,

$$\begin{aligned} \mathcal{C}_2 &= \{(-\infty, x], x \in R\}, \\ \mathcal{C}_3 &= \{(a, b], a < b, a, b \in R\}, \\ \mathcal{C}_4 &= \{[a, b), a < b, a, b \in R\}, \\ \mathcal{C}_5 &= \{[a, b], a < b, a, b \in R\}, \\ \mathcal{C}_6 &= \{[x, \infty), x \in R\}, \text{ etc.} \end{aligned} \quad (1.32)$$

Thus, $\mathcal{B}$ contains all subsets of R of the above form and their complements, countable unions and intersections. However $\mathcal{B}$ is not the class of all subsets of R . Examples of subsets of R which are not Borel sets can be had from any book on measure theory.

If $x \in R$, singleton $\{x\}$ is a Borel set, since

$$\{x\} = (-\infty, x] \cap [x, \infty).$$

Hence any countable subset of R , containing a countable union of singletons is a Borel set. Thus $\{0, 1, 2, \dots\}$ is a Borel set. But an uncountable union of such singletons, viz., a subset containing an uncountable number of points of the real line may not be a Borel set.

Example 1.17. The Borel field $\mathcal{B}_2$ of subsets of the real plane R^2 is defined as the minimal σ -field containing all open rectangles of the form

$$B = \{(x, y) : a < x < b, c < y < d\}, a < b, c < d, \quad (1.33)$$

for, a, b, c, d belonging to R and arbitrary. Taking

$$B_n = \left\{ (x, y) : a < x < b + \frac{1}{n}, c < y < d \right\} \in \mathcal{B}_2.$$

$$\cap B_n = \{(x, y) : a < x \leq b, c < y < d\} \in \mathcal{B}_2.$$

Thus, all half-open rectangles with at least one-side open and at least one side closed are Borel sets. Similarly all closed rectangles are also Borel sets. Since infinite rectangles of the form

$$\{x < b, c < y < d\} = \bigcup_n \{-n < x < b, c < y < d\},$$

$$\text{and } \{a < x, c < y < d\} = \bigcup_n \{a < x < n, c < y < d\},$$

they also belong to $\mathcal{B}_2$. In fact, all these classes of rectangles generate $\mathcal{B}_2$. (Why?)

Similarly Borel field $\mathcal{B}_n$ of subsets of R^n , the n -dimensional Euclidean space with typical point $(x_1, x_2, \dots, x_n)$, is generated by n -dimensional open rectangles of the form

$$B = \{(x_1, \dots, x_n) : a_i < x_i < b_i, i = 1, 2, \dots, n\}, \quad (1.34)$$

$a_i \in R, b_i \in R$, arbitrary, or by rectangles, closed on one or more of its 2^n sides. For example, it is generated by rectangles of the following form

$$B = \{(x_1, \dots, x_n) : a_i < x_i \leq b_i, i = 1, 2, \dots, n\}. *$$

σ -field is also called as σ -algebra by some authors (see Halmos, P.R. (1950)). Theorem 1.3 is useful in determining whether a certain class is a σ -field or not. A field $\mathcal{F}$ is said to be a *monotone field*, if it is closed under monotone operations, i.e. if $\lim A_n \in \mathcal{F}$, whenever $\{A_n\}$ is a monotone sequence of sets of $\mathcal{F}$, i.e.

$$A_n \in \mathcal{F}, A_n \uparrow A \Rightarrow A \in \mathcal{F},$$

and

$$A_n \in \mathcal{F}, A_n \downarrow A \Rightarrow A \in \mathcal{F}.$$

THEOREM 1.3: *A σ -field is a monotone field and conversely.*

Proof: Let $\mathcal{F}$ be a σ -field and $A_n \in \mathcal{F}$. If $A_n \uparrow A$, then $A = \bigcup_{n=1}^{\infty} A_n$ is a countable union of sets of $\mathcal{F}$. Hence $A \in \mathcal{F}$. Similarly if $A_n \downarrow A$, $A \in \mathcal{F}$. Hence $\mathcal{F}$ is a monotone field.

Conversely, let $\mathcal{F}$ be a monotone field and let $A_1, A_2, \dots$ be sets belonging to $\mathcal{F}$. Then $\bigcup_{k=1}^n A_k$ and $\bigcap_{k=1}^n A_k$ belong to $\mathcal{F}$, since $\mathcal{F}$ is a field. These are monotone sequences whose limits $\bigcup_{k=1}^{\infty} A_k$ and $\bigcap_{k=1}^{\infty} A_k$ must belong to $\mathcal{F}$. Thus, $\mathcal{F}$ is a σ -field. (*)

1.4 CLASS OF EVENTS

In section 1.1 an event is defined as a collection of outcomes of an experiment E , in which one is interested. Since Ω is defined as the collection of all outcomes of E , if A is an event, A^c will also be an event. It denotes the non-occurrence of A . If $\mathcal{A}$ is the class of all events associated with the experiment E , it is closed under complementation.

If A and B are events, then $A \cup B$ denotes the event that at least one of the events A or B occurs; $A \cap B$ denotes the event that both the events A and B occur simultaneously. Thus $\mathcal{A}$ is closed under finite unions and intersections and is a field. If $A_1, A_2, \dots, A_n$ are events,

$\bigcup_{i=1}^n A_i$ will denote the events that at least one of the events A_i will occur;

$\bigcap_{i=1}^n A_i$ will denote the event that all the events A_i occur simultaneously.

Since the union and intersection can be extended to the case when n is infinite, $\mathcal{A}$ will naturally be a σ -field if E has an infinite number of outcomes.

Thus, with each experiment E , we can associate Ω , the space of all outcomes and $\mathcal{A}$, the σ -field of events. For different experiments we shall have different Ω and different $\mathcal{A}$. This becomes a handicap when we have to consider outcomes of more than one experiment simultaneously. This handicap is avoided by considering Ω to be the collection of outcomes of all experiments "under consideration." However $\mathcal{A}$, the σ -field of events, will be associated with only a particular experiment in which one is interested. For example, Ω may consist of outcomes of a dice-throwing and a coin-tossing experiment. The events of interest may be based on the outcomes of coin-tossing experiment only. In this case an event is a subset of Ω and not conversely.

With every experiment having a finite number of possible outcomes, the class of all events will be a field. If the field is degenerate, the only events are ϕ and Ω . These are called respectively as *impossible* and *sure* events in line with our proposal to associate chance with the events in the later chapters. If the number of possible outcomes is infinite, then the class of events is a σ -field. The space Ω of all outcomes of experiments together with the specification of the σ -field $\mathcal{A}$ of events is sometimes called a *measurable space* and is denoted by the pair $(\Omega, \mathcal{A})$.

Let $\mathcal{A}$ be a class of events and $B \in \mathcal{A}$. Then for $A \in \mathcal{A}$, $B \cap A$ is an event. The class of all such events

$$\{B \cap A : A \in \mathcal{A}\} = B \cap \mathcal{A}, \text{ say,}$$

is a σ -field of subsets of B (Why?). It is the restriction of $\mathcal{A}$ to subsets of B and hence possesses the same properties as those of $\mathcal{A}$.

COMPLEMENTS AND PROBLEMS

1. *Countability.* Show that (i) every infinite subset of a countable set is countable and (ii) the union of a countable number of countable sets is a countable set. Deduce that the set of all rational numbers is countable.
2. Show that (i) $(A \Delta B) \cup (A \Delta C) = A \cup B \cup C - A \cap B \cap C$.
(ii) $(A - B) \cap C = (A \cap C) - (B \cap C)$.
3. If $A_n \rightarrow A$, show that $A_n^c \rightarrow A^c$.
4. Prove that

$$\overline{\lim} (A_n \cup B_n) = \overline{\lim} A_n \cup \overline{\lim} B_n.$$

What can you say for $\underline{\lim} (A_n \cup B_n)$, $\overline{\lim} (A_n \cap B_n)$, $\underline{\lim} (A_n \cap B_n)$?

5. Let $\Omega = R^2$, A_n = interior of the circle with centre at $((-1)^n/n, 0)$ and radius 1. Find $\overline{\lim} A_n$, $\underline{\lim} A_n$, and $\lim A_n$, if exists.
6. Prove that (i) $(\cup A_i) \cap B = \cup (A_i \cap B)$,
(ii) $A_i \downarrow$ and $\cap A_i = \phi$, then $A_1 = \bigcup_{i=1}^{\infty} (A_i - A_{i+1})$.
7. If $A_n = A$, $n = 1, 3, 5, \dots$
 $= B$, $n = 2, 4, 6, \dots$

show that $\overline{\lim} A_n = A \cup B$, $\underline{\lim} A_n = A \cap B$. When does $\lim A_n$ exist?

8. Examine the following sequences of sets for convergence. If convergent, derive the limit.
(a) $A_{2n} = (0, 1/2n)$, $A_{2n+1} = [-1, 1/(2n+1)]$;
(b) $A_n = [\text{the set of rationals in } (1 - 1/(n+1), 1 + 1/n)]$;

$$(c) \quad A_n = (2 - 1/n, 2 + 2/n), n \text{ odd}, \\ = \phi, \quad n \text{ even}.$$

9. Halmos (1956) calls a σ -field as Boolean σ -algebra with operations of addition and multiplication suitably defined. Boolean σ -algebra $\mathcal{A}$ is said to be *separable* or of countable type if there exists a countable family of subsets of Ω which generates $\mathcal{A}$. Taking open intervals with rational end-points show that Borel field is separable.
10. *Minimal monotone class* containing $\mathcal{C}$ is the smallest class closed under monotone operations and containing $\mathcal{C}$. If $\mathcal{M}_0$ is the minimal monotone class containing $\mathcal{C}$, then show that $\mathcal{M}_0 = \sigma(\mathcal{C})$.
11. Show that a class $\mathcal{A}$ of subsets of Ω is a field if
- $A, B \in \mathcal{A} \Rightarrow A \cup B \in \mathcal{A}$,
 - $B \subset A \Rightarrow A - B \in \mathcal{A}$,
 - $\Omega \in \mathcal{A}$.

The class $\mathcal{A}$ is a σ -field if in addition

$$(iv) \quad A_i \in \mathcal{A}, (i = 1, 2, \dots) \Rightarrow A = \bigcup_1^\infty A_i \in \mathcal{A}.$$

12. If $\{A_1, \dots, A_m\}$ and $\{B_1, \dots, B_n\}$ are two partitions of Ω , show that
- $\{A_i \cap B_j\}$ is a partition of Ω ,
 - $\sigma\{\sigma\{A_i\} \cup \sigma\{B_j\}\} = \sigma\{A_i \cap B_j\},$
 $= \sigma(\{A_i\} \cup \{B_j\}).$

Can you extend the above result when $\{A_i\}$ and $\{B_j\}$ are (i) countable partitions (ii) arbitrary classes?

13. Let $\mathcal{C}$ be an arbitrary class of subsets of Ω and let $\mathcal{A}$ be the collection of all finite unions $\bigcup_{i=1}^n A_i$ ($n = 1, 2, \dots$) where each $A_i = \bigcap_{j=1}^n B_{ij}$, $B_{ij} \in \mathcal{C}$ or $B_{ij}^c \in \mathcal{C}$. Show that $\mathcal{A}$ is the minimal field over $\mathcal{C}$. This gives us a procedure of obtaining minimal field over a class $\mathcal{C}$. Extend this procedure to arrive at the minimal σ -field.
14. $\bar{R} = R \cup \{-\infty\} \cup \{+\infty\}$ is called the *extended real line*. One can define Borel field $\mathcal{B}$ in $\bar{R}$ similarly to $\mathcal{B}$ in R . Study the relationship between $\mathcal{B}$ and $\bar{\mathcal{B}}$.
15. Borel sets are subsets of R . But the converse is not true. There exist subsets of R which are not obtainable by countable operations on intervals. Give an example of such a set. Also give an example of a non-countable Borel set which is not an interval. (see Halmos 1956)
16. Let B be a Borel set in R . Let $B' = \{a + x : x \in B\}$, $a \in R$. Show that B' is also a Borel set.
17. If $\mathcal{C}$ generates a σ -field $\mathcal{A}$, the σ -field $B \cap \mathcal{A}$ of subsets of B is generated by $B \cap \mathcal{C}$.

2. Random Variables

2.1 FUNCTIONS AND INVERSE FUNCTIONS

(a) POINT FUNCTION AND SET FUNCTION

Suppose Ω is the sample space with sample points ω . Sometimes we are interested in a value $X(\omega)$ associated with ω , and not in ω itself. $X(\omega)$ may be observable while ω is not.

Example 2.1. Consider the sample space in Example 1.5, consisting of 8 possible outcomes of the experiment of tossing a coin 3 times. We may be interested only in the number of times head turns up. Thus, with the outcome $HHH = \omega_1$ we will associate a number $X(\omega_1) = 3$, representing the number of heads in ω_1 . With the outcome $HHT = \omega_2$ we will associate $X(\omega_2) = 2$, and so on.

A function X on a space Ω to a space Ω' assigns to each point $\omega \in \Omega$ a unique point in Ω' denoted by $X(\omega)$. $X(\omega)$ is the *image* of the *argument* ω under X . It is also called the *value* of X at ω . X is also a mapping from Ω to Ω' , denoted by $\Omega \xrightarrow{X} \Omega'$. ω is mapped on to $X(\omega) = \omega' \in \Omega'$ by X .

X also establishes correspondence relation between points of Ω with points of Ω' . Ω is called the *domain* of X and Ω' is called the *range*. The set $\Omega'' = [X(\omega) : \omega \in \Omega]$ which is a subset of Ω' , is called the *strict range* of X . If $\Omega' = \Omega''$, then we say that X is a mapping from Ω *onto* Ω' , otherwise X is said to be a mapping from Ω *into* Ω' . The symbols $f(x)$, $X(\omega)$, etc., will be used to denote functions, even though they really denote the values of functions.

Example 2.2. Let $\Omega = [0, \pm 1, \pm 2, \dots]$, $\Omega' = [0, 1, 2, \dots]$ and $X(\omega) = \omega^2$. Then the strict range of X is $[0, 1, 4, 9, \dots] = \Omega''$. Hence X is a mapping of Ω 'into' Ω' and 'onto' Ω'' . If $X(\omega) = \omega + 2$, then X is a mapping of Ω 'onto' Ω .

We consider one-valued functions only. Thus, $\omega_1 = \omega_2 \Rightarrow X(\omega_1) = X(\omega_2)$. Many-valued functions such as $X(\omega) = \pm \sqrt{\omega}$, $\omega \in R$ are outside our purview. But, in general $X(\omega_1) = X(\omega_2) \nRightarrow \omega_1 = \omega_2$ because, there may be more than one ω which is mapped on the same image ω' . If $X(\omega_1) = X(\omega_2) \Rightarrow \omega_1 = \omega_2$ whatever be $\omega_1, \omega_2 \in \Omega$, then X is said to be a *one-to-one* (1-1) function. In Example 2.2, $X(\omega) = \omega + 2$ is a

1-1 function, as there corresponds a unique image to a point $\omega \in \Omega$; but $X(\omega) = \omega^2$ is not a 1-1 function, since $\omega_1 = +2$, $\omega_2 = -2$ have the same image $X(\omega_1) = 4 = X(\omega_2)$. If Ω is the real line ($-\infty < \omega < \infty$) and $\Omega' = (0 < \omega' < \infty)$, the function $X(\omega) = \exp(\omega)$ is a 1-1 'onto' function from Ω to Ω' , and is a 1-1 'into' function from Ω to Ω .

If the range space is the real line R or its subset, the function is said to be a *numerical* or *real-valued* function. In Example 2.1 with each ω we associate a numerical value. Hence, we have a mapping from Ω to R . If the arguments of X are points of space Ω , we have a *point function*. In the above examples, we considered only point functions. If the arguments of a function are sets of a certain class, then we have a *set function*. If $\mathcal{A}$ is a class of sets, and with each $A \in \mathcal{A}$, we associate a value $\mu(A)$, say, then μ is a set function. μ may represent such entities as weight, length, measure, etc., associated with the sets of $\mathcal{A}$. For example with an interval (a, b) we may associate its length $b-a$; with the disjoint union $(a, b) \cup (c, d)$ we may associate, $\mu((a, b) \cup (c, d)) = (b-a) + (d-c)$, and so on.

Two real valued functions X and Y on Ω are said to be equal iff

$$X(\omega) = Y(\omega), \forall \omega \in \Omega,$$

and we write $X = Y$. Similarly we write $X \lesseqgtr Y$, iff $X(\omega) \lesseqgtr Y(\omega)$, for all $\omega \in \Omega$. Thus, to verify the equality (or inequality) of two functions, we have to verify the equality (or inequality) of the values of these functions at all argument points ω . If $X(\omega) = c$, a constant for all ω , X is *degenerate*.

(b) INVERSE FUNCTION

We have seen that for a point $\omega' \in \Omega'$ there may exist one or more than one points of Ω , whose image under X is ω' . The set of all points $\omega \in \Omega$ whose image under X is ω' is called the *inverse image* of $\{\omega'\}$, denoted as $X^{-1}(\{\omega'\})$. Thus,

$$X^{-1}(\{\omega'\}) = \{\omega \in \Omega : X(\omega) = \omega'\}. \quad (2.1)$$

Hence forward, sets of points or classes of sets will be denoted either by $\{\cdot\}$ or by $[\cdot]$.

In general, let $B' \subset \Omega'$. The set of all points of Ω for which $X(\omega) \in B'$ is called the inverse image of B' under X , denoted by $X^{-1}(B')$:

$$X^{-1}(B') = \{\omega : X(\omega) \in B'\}. \quad (2.2)$$

Thus, with every point function X , we associate a set function X^{-1} , whose domain is a class $\mathcal{B}'$ of subsets of Ω' and whose range is a class $\mathcal{B}$, (say) of subsets of Ω . X^{-1} is called *inverse function* (or mapping) of X . We shall denote

$$\begin{aligned} X(B) &= [X(\omega) : \omega \in B], B \subset \Omega, \\ X^{-1}(\mathcal{B}') &= [X^{-1}(B') : B' \in \mathcal{B}']. \end{aligned} \quad (2.3)$$

Evidently $X^{-1}(\Omega') = [\omega : X(\omega) \in \Omega'] = \Omega$.

Example 2.3. Consider the function $X(\omega) = \omega^2$ from R to R . Let B' be the set of points contained in the interval $(1, 2)$. Then $X^{-1}(B') = (1, \sqrt{2}) \cup (-\sqrt{2}, -1)$. If the range and the domain of X is the same and is $(0, \infty)$, then $X^{-1}(B') = (1, \sqrt{2})$.

Example 2.4. Consider the real valued function I_A or $I(A)$ defined on Ω as follows:

$$\left. \begin{aligned} I(A)(\omega) &= I_A(\omega) = 1, \text{ if } \omega \in A, \\ I_A(\omega) &= 0, \text{ if } \omega \in A^c. \end{aligned} \right\} \quad (2.4)$$

Then I_A or $I(A)$ is called the *indicator* function (or *characteristic* function by some authors) of A . The strict range of I_A is

$$I_A(\Omega) = \{I_A(\omega) : \omega \in \Omega\} = \{0, 1\}.$$

If $B \subset R$, the range space, then

$$\left. \begin{aligned} I_A^{-1}(B) &= \phi, \text{ if } B \text{ does not contain '0' or '1',} \\ &= A, \text{ if } B \text{ contains '1' but not '0',} \\ &= A^c, \text{ if } B \text{ contains '0' but not '1',} \\ &= \Omega, \text{ if } B \text{ contains both '0' and '1'.} \end{aligned} \right\} \quad (2.5)$$

$$\text{Thus } I_A^{-1}(\mathcal{B}) = \{\phi, A, A^c, \Omega\} = \sigma\{A\}. \quad (2.6)$$

Evidently, cI_A takes the value c on A and 0 on A^c , and hence

$$(cI_A^{-1})(\mathcal{B}) = I_A^{-1}(\mathcal{B}) = \sigma\{A\}.$$

I_Ω takes the value, '1' for all $\omega \in \Omega$. If $X(\omega) = c$ for all $\omega \in \Omega$, then $X = cI_\Omega$ and in this case

$$X^{-1}(B') = \phi \text{ or } \Omega, \text{ whatever be } B' \subset R, \quad (2.7)$$

$$\text{i.e., } I_\Omega^{-1}(\mathcal{B}) = \{\phi, \Omega\} = (cI_\Omega)^{-1}(\mathcal{B}). \quad (2.8)$$

As the indicator function is the simplest non-trivial function, we give below some of its important properties which can be easily established. (Prove!)

$$\left. \begin{aligned} \text{(i)} \quad & A \subset B \Leftrightarrow I(A) \leq I(B), \\ & A = B \Leftrightarrow I(A) = I(B), \\ & I(A) = I^2(A) = I^n(A), I_\Omega = 1. \\ \text{(ii)} \quad & I(A^c) = 1 - I(A), I(B - A) = I(B) - I(A). \\ \text{(iii)} \quad & I(AB) = I(A) \cdot I(B), I\left(\bigcap_{i=1}^n A_i\right) = \prod_{i=1}^n I(A_i) \\ & \quad \quad \quad = \min\{I(A_1), \dots, I(A_n)\}, \\ \text{(iv)} \quad & I(A \cup B) = I(A) + I(B) - I(A) \cdot I(B) = \max\{I(A), I(B)\}, \\ & I(A + B) = I(A) + I(B), \\ & I\left(\bigcup_i A_i\right) = \sum_i I(A_i) - \sum_{i < j} I(A_i) \cdot I(A_j) \\ & \quad \quad \quad + \sum_{i < j < k} I(A_i) \cdot I(A_j) \cdot I(A_k) - \dots, \\ & \quad \quad \quad = \max\{I(A_1), I(A_2), \dots, I(A_n)\}. \end{aligned} \right\} \quad (2.9)$$

In order to establish the last relation in (2.9) (iv) we may note that if ω belongs to k of A_i 's, then on the right hand side it will contribute $\binom{k}{1}$ to the first term, $\binom{k}{2}$ to the second term, and so on. Since $\binom{k}{1} - \binom{k}{2} + \dots - \binom{k}{k} = 1$, we have the result.

For the numerical valued function X defined in Example 2.1 we can determine $X^{-1}(B)$ for various sets $B \in \mathcal{B}$, the Borel field. (How?)

If A, B are subsets of Ω and $A \cap B = \phi$, this does not imply that $X(A) \cap X(B) = \phi$. (Give counter-example!). However, if A and B are subsets of Ω' and $A \cap B = \phi$, this implies that $X^{-1}(A) \cap X^{-1}(B) = \phi$. In general, we have the following:

LEMMA 2.1: *Inverse mapping preserves all set relations.*

Proof: (i) Let $B \subset C \subset \Omega'$. Then

$$X^{-1}(B) = [\omega : X(\omega) \in B] \subset [\omega : X(\omega) \in C] = X^{-1}(C).$$

(ii) Let $B_k \subset \Omega'$. Then

$$\begin{aligned} \omega \in X^{-1}(\cap_k B_k) &\Leftrightarrow X(\omega) \in \cap_k B_k, \\ &\Leftrightarrow X(\omega) \in B_k, \forall k, \\ &\Leftrightarrow \omega \in X^{-1}(B_k), \forall k, \\ &\Leftrightarrow \omega \in \cap_k X^{-1}(B_k). \end{aligned}$$

$$\text{Hence} \quad X^{-1}(\cap_k B_k) = \cap_k X^{-1}(B_k). \quad (2.10)$$

Similarly, we can prove that

$$X^{-1}(\cup_k B_k) = \cup_k X^{-1}(B_k). \quad (\text{How?})$$

$$\begin{aligned} \text{(iii)} \quad \omega \in X^{-1}(B^c) &\Leftrightarrow X(\omega) \in B^c \Leftrightarrow X(\omega) \notin B, \\ &\Leftrightarrow \omega \notin X^{-1}(B). \end{aligned}$$

$$\text{Thus,} \quad X^{-1}(B^c) = (X^{-1}(B))^c. \quad (2.11)$$

It may be noted here that the unions and intersections considered above need not be countable. (*)

Evidently,

$$X^{-1}(\Omega') = [\omega : X(\omega) \in \Omega'] = \Omega, X^{-1}(\phi) = \phi.$$

If $A \cap B = \phi$, then $X^{-1}(A \cap B) = X^{-1}(A) \cap X^{-1}(B) = \phi$. Hence, $X^{-1}(A)$ and $X^{-1}(B)$ are disjoint. If $\mathcal{C}'$ is a certain class of subsets of Ω' and $X^{-1}(\mathcal{C}') = [X^{-1}(C) : C \in \mathcal{C}']$, from the above Lemma 2.1 the question arises whether $X^{-1}(\mathcal{C}')$ is a σ -field, if $\mathcal{C}'$ is a σ -field. A partial answer to this question is contained in the following Corollary 2.1 and the complete answer in Corollary 2.2.

COROLLARY 2.1: If $\mathcal{A}$ is a class of subsets of Ω and is a σ -field, then the class $\mathcal{B}$ of all sets whose inverse images belong to $\mathcal{A}$ is also a σ -field.

Proof: By definition of $\mathcal{B}$, since $\mathcal{A}$ is a σ -field, using Lemma 2.1,

$$B_1, B_2, \dots \in \mathcal{B} \Rightarrow X^{-1}(B_1), X^{-1}(B_2), \dots \in \mathcal{A},$$

$$\Rightarrow \bigcap_{i=1}^{\infty} X^{-1}(B_i) \in \mathcal{A},$$

$$\Rightarrow X^{-1}\left(\bigcap_i B_i\right) \in \mathcal{A},$$

$$\Rightarrow \bigcap_i B_i \in \mathcal{B}.$$

Thus, $\mathcal{B}$ is closed under countable intersections. Similarly, by Lemma 2.1 it is closed under complementation also. Hence $\mathcal{B}$ is a σ -field. (*)

COROLLARY 2.2: (i) If $\mathcal{C}$ is a field (or a σ -field) of subsets of Ω' , $X^{-1}(\mathcal{C})$ is a field (or a σ -field) of subsets of Ω . (ii) Inverse image of the minimal σ -field over any class $\mathcal{C}$ is the minimal σ -field over $X^{-1}(\mathcal{C})$,

$$\text{i.e.} \quad \sigma\{X^{-1}(\mathcal{C})\} = X^{-1}\{\sigma(\mathcal{C})\}. \quad (2.12)$$

Proof: (i) Let $\mathcal{C}$ be a σ -field on Ω' and $\mathcal{A}$ be the class of inverse images of the elements of $\mathcal{C}$. Then, since $\mathcal{C}$ is a σ -field

$$X^{-1}(B_1), X^{-1}(B_2), \dots \in \mathcal{A} \Rightarrow B_1, B_2, \dots \in \mathcal{C},$$

$$\Rightarrow \bigcap_i B_i \in \mathcal{C},$$

$$\Rightarrow X^{-1}\left(\bigcap_i B_i\right) \in \mathcal{A},$$

$$\Rightarrow \bigcap_i X^{-1}(B_i) \in \mathcal{A},$$

using Lemma 2.1. Thus, $\mathcal{A}$ is closed under countable intersections. Since $(X^{-1}(B))^c = X^{-1}(B^c)$, $\mathcal{A}$ is closed under complementations and hence it is a σ -field. Thus, the inverse of a σ -field is a σ -field and in particular, inverse of a field is a field.

(ii) Let $\sigma(\mathcal{C})$ be the minimal σ -field containing $\mathcal{C}$, a class of subsets of Ω' . From (i) $X^{-1}(\sigma(\mathcal{C}))$ is a σ -field of subsets of Ω . It contains $X^{-1}(\mathcal{C})$ and hence $\sigma(X^{-1}(\mathcal{C}))$. Thus,

$$X^{-1}(\sigma(\mathcal{C})) \supset \sigma(X^{-1}(\mathcal{C})). \quad (2.13)$$

To prove the inclusion in the reverse direction, from Corollary 2.1, the class of all subsets $\mathcal{B}$ of Ω' whose inverse images are elements of the σ -field $\sigma(X^{-1}(\mathcal{C}))$, is a σ -field. Hence

$$X^{-1}(\mathcal{C}) \subset \sigma(X^{-1}(\mathcal{C})) \Rightarrow \mathcal{C} \subset \mathcal{B},$$

$$\Rightarrow \sigma(\mathcal{C}) \subset \mathcal{B},$$

$$\Rightarrow X^{-1}(\sigma(\mathcal{C})) \subset X^{-1}(\mathcal{B}).$$

But by definition of $\mathcal{B}$, $X^{-1}(\mathcal{B})$ equals $\sigma(X^{-1}(\mathcal{C}))$. Hence

$$X^{-1}(\sigma(\mathcal{C})) \subset \sigma(X^{-1}(\mathcal{C})). \quad (2.14)$$

(2.13) and (2.14) imply (2.12). (*)

(c) MEASURABLE FUNCTION, BOREL FUNCTION, INDUCED σ -FIELD

An important class of functions is that of real valued functions X on Ω to R . Let $\mathcal{B}$ be the Borel field of subsets of R , and $\mathcal{A}$ be a σ -field of subsets of Ω . If $X^{-1}(B) \in \mathcal{A}$ for all Borel sets $B \in \mathcal{B}$, X is said to be a $\mathcal{A}$ -measurable function, or a function measurable with respect to $\mathcal{A}$. If Ω is also the real line R or its subset and if X is measurable with respect to the Borel field $\mathcal{B}$ on the domain, then X is called a Borel-function. Thus, X is a Borel function if $X^{-1}(B) \in \mathcal{B}$, for Borel sets B in the range space, and the Borel field $\mathcal{B}$ in the domain.

Example 2.5. The indicator function I_A is a real valued function. By (2.5), $I_A^{-1}(B) \in \{\phi, A, A^c, \Omega\}$. Since $\mathcal{A}$ is a σ -field in Ω , if $A \in \mathcal{A}$, A^c also belongs to $\mathcal{A}$ and hence, $I_A^{-1}(B) \in \mathcal{A} \Leftrightarrow A \in \mathcal{A}$. Thus I_A is $\mathcal{A}$ -measurable iff $A \in \mathcal{A}$. A function taking a constant value for $\omega \in \Omega$ is measurable with respect to any σ -field in Ω (Why?). We can verify that, if X can take three distinct values c_0, c_1, c_2 such that

$$\begin{aligned} X(\omega) &= c_0, \text{ if } \omega \in A_0, \\ &= c_1, \text{ if } \omega \in A_1, \\ &= c_2, \text{ if } \omega \in A_2, \end{aligned} \quad (2.15)$$

then, $\{A_0, A_1, A_2\}$ forms a partition of Ω (Why?). Then X will be $\mathcal{A}$ -measurable iff $\{A_0, A_1, A_2\} \subset \mathcal{A}$ (see Example (2.6)).

Let X be a real valued function on Ω and $\mathcal{B}$ be the Borel field. By Cor. 2.2, the class of sets $X^{-1}(\mathcal{B}) = \{X^{-1}(B) : B \in \mathcal{B}\}$ is a σ -field. It is called the σ -field induced by X . It is also denoted by $\mathcal{B}(X)$, later on: X will be $\mathcal{A}$ -measurable iff it induces a σ -field which is a sub- σ -field of $\mathcal{A}$. If $\mathcal{A} \subset \mathcal{A}'$ and X is $\mathcal{A}$ -measurable then it is also $\mathcal{A}'$ -measurable. If $\mathcal{A}'' \subset \mathcal{A}$ and X is $\mathcal{A}$ -measurable then X may not be $\mathcal{A}''$ -measurable. X would be $\mathcal{A}''$ -measurable if $X^{-1}(\mathcal{B}) \subset \mathcal{A}''$ also. The smallest σ -field with respect to which X is measurable is $X^{-1}(\mathcal{B})$, the σ -field induced by X . Since it is contained in every σ -field with respect to which X is measurable, it is the intersection of all such σ -fields.

By (2.5), $I_A^{-1}(\mathcal{B}) = \{\phi, A, A^c, \Omega\}$ which is the σ -field induced by I_A . The σ -field induced by the constant is the degenerate σ -field. Therefore, a constant function is measurable with respect to every σ -field of subsets of Ω . Conversely if $\{\phi, \Omega\}$ is the σ -field induced by X , it is degenerate (Why?).

Example 2.6. Consider the function X defined by (2.15). Then

$$\begin{aligned}
X^{-1}(B) &= \phi, \text{ if } B \text{ does not contain } c_0 \text{ or } c_1 \text{ or } c_2, \\
&= A_0, \text{ if } B \text{ contains } c_0 \text{ only,} \\
&= A_1, \text{ if } B \text{ contains } c_1 \text{ only,} \\
&= A_2, \text{ if } B \text{ contains } c_2 \text{ only,} \\
&= A_0 + A_1, \text{ if } B \text{ contains } c_0 \text{ and } c_1, \\
&= A_0 + A_2, \text{ if } B \text{ contains } c_0 \text{ and } c_2, \\
&= A_1 + A_2, \text{ if } B \text{ contains } c_1 \text{ and } c_2, \\
&= \Omega, \text{ if } B \text{ contains } c_0, c_1 \text{ and } c_2.
\end{aligned}$$

Hence

$$X^{-1}(\mathcal{B}) = \{\phi, A_0, A_1, A_2, A_0 + A_1, A_0 + A_2, A_1 + A_2, \Omega\}, \quad (2.16)$$

which is the minimal σ -field containing the partition $\{A_0, A_1, A_2\}$. In general, if

$$X(\omega) = c_i, \text{ if } \omega \in A_i \quad (i = 1, 2, \dots, n), \quad (2.17)$$

then $\{A_1, A_2, \dots, A_n\}$ forms a partition of Ω . The σ -field induced by X is the minimal σ -field containing this partition. (Why?) Now

$$X = \sum_{i=1}^n c_i I_{A_i} \quad (2.18)$$

in terms of indicator functions of disjoint sets of the partition $\{A_i\}$. Since we are concerned with one-valued functions only, any function taking finite distinct values $\{c_1, c_2, \dots, c_n\}$ can be written in the form (2.18), where

$$A_i = \{\omega : X(\omega) = c_i\} = X^{-1}(\{c_i\}), \quad (i = 1, 2, \dots, n). \quad (2.19)$$

Hence A_i 's are evidently disjoint and form a partition of Ω .

A finite linear combination of indicators of sets (not necessarily disjoint) is called a simple function. If Ω is finite, all real valued functions will be simple. (Why?) *Ugma*

LEMMA 2.2: Any simple function X can be written in the form *Ugma*

$$X = \sum_{k=1}^n x_k I_{A_k}, \quad (2.20)$$

where x_k 's are distinct numerical constants and $\{A_1, A_2, \dots, A_n\}$ is a partition of Ω .

Proof: Let X be a linear function of I_{B_1} and I_{B_2} only, say

$$X = c_1 I_{B_1} + c_2 I_{B_2}. \quad (2.21)$$

Then $X(\omega) = 0$, if $\omega \notin B_1$ and $\omega \notin B_2$, i.e., $\omega \in B_1^c B_2^c = A_1$,
 $= c_1$, if $\omega \in B_1$ and $\omega \notin B_2$, i.e., $\omega \in B_1 B_2^c = A_2$,
 $= c_2$, if $\omega \notin B_1$ and $\omega \in B_2$, i.e., $\omega \in B_1^c B_2 = A_3$,
 $= c_1 + c_2$, if $\omega \in B_1$ and $\omega \in B_2$, i.e., $\omega \in B_1 B_2 = A_4$.

Hence $X(\omega) = x_i$ for $\omega \in A_i$, where $x_1 = 0$, $x_2 = c_1$, $x_3 = c_2$ and $x_4 = c_1 + c_2$. Moreover $\{A_1, A_2, A_3, A_4\}$ forms a partition of Ω . Thus, the Lemma is proved for the linear combination of 2 indicators.

Suppose in general X is a linear combination of m indicators,

$$X = \sum_{i=1}^m c_i I_{B_i}. \quad (2.22)$$

$$\text{Then} \quad X(\omega) = c_{j_1} + c_{j_2} + \dots + c_{j_k}, \quad (2.23)$$

if $\omega \in B_{j_1} \cap B_{j_2} \cap \dots \cap B_{j_k} \cap B_{j_{k+1}}^c \cap \dots \cap B_{j_m}^c = A_i$ say,

where $(j_1, j_2, \dots, j_m)$ is a permutation of $(1, 2, \dots, m)$ and $k = 0, 1, \dots, m$. Thus, X can have at most 2^m values which are distinct. The sets A_i being disjoint, $\{A_i\}$ forms a partition of Ω . The Lemma follows by taking for n , the number of distinct values of X . (*)

Thus, a simple function takes only a finite number of values. $x_1, x_2, \dots, x_n$, (say) and we have seen from the earlier discussion that conversely a function taking a finite number of values is simple. $\{A_i\}$ is the partition on which X takes distinct values. It is called the *partition induced by X* . In fact

$$A_i = \{\omega : X(\omega) = x_i\} = X^{-1}\{x_i\}, \quad i = 1, 2, \dots, n.$$

We may note that the partitions induced by X and cX ($c \neq 0$) are the same. Hence forward we shall assume that a simple function is of the form (2.20), with x_k 's distinct and A_k 's disjoint forming a partition of Ω .

LEMMA 2.3: *The σ -field induced by a simple function (2.20) is the minimal σ -field containing the partition $\{A_1, A_2, \dots, A_n\}$.*

Proof: Let B be any Borel subset of the real line containing

$$[x_{j_1}, x_{j_2}, \dots, x_{j_k}] \subset [x_1, x_2, \dots, x_n],$$

Then $X(\omega) \in B$ iff $X(\omega) = x_{j_1}$ or x_{j_2} or ... x_{j_k} .

But $X(\omega) = x_{j_i}$ if $\omega \in A_{j_i}$. Hence

$$[\omega : X(\omega) \in B] = X^{-1}(B) = \sum_{i=1}^k A_{j_i} \in \sigma\{A_1, A_2, \dots, A_n\}.$$

Since the sets of $\sigma\{A_i\}$ are sums of sets of $\{A_i\}$, Lemma follows. (*)

COROLLARY 2.3: *The σ -field induced by $X = \sum_{i=1}^m c_i I_{B_i}$ is (equal to) the minimal σ -field containing $B_1, B_2, \dots, B_m$.*

Corollary follows from Lemmas 2.2 and 2.3 since each one of the A_i 's belongs to $\sigma\{B_1, B_2, \dots, B_m\}$.

A countable linear combination of indicator functions is called an elementary function. A few authors do not distinguish simple and ele-

mentary functions and call them both as simple functions. But we shall restrict ourselves to the above definition.

Without loss of generality, we can assume that an elementary function is the linear combination of a countable number of indicator functions of disjoint sets.

$$\text{If } X = \sum_{i=1}^{\infty} x_i I_{A_i},$$

then $A_i = X^{-1}(\{x_i\}) = [\omega : X(\omega) = x_i] = [X = x_i]$, say,

and $\{A_i\}$ forms a partition of Ω . (Why?)

If B is a Borel set and the domain $R \equiv \Omega$, the indicator function of B is a Borel function. A linear combination of indicator functions of Borel sets B_i is also $\mathcal{B}$ -measurable and is hence a Borel function. This follows from the fact that if $B_i \in \mathcal{B}$, ($i = 1, 2, \dots$) $\{B_i\} \subset \mathcal{B}$ and $\sigma\{B_i\} \subset \mathcal{B}$. Similarly, all the results we prove for $\mathcal{A}$ -measurable functions will hold in particular for Borel functions.

If $X(\omega) = \omega + 2$, $\omega \in R$, $X^{-1}(a, b) = (a-2, b-2) \in \mathcal{B}$. In general, $X^{-1}(B) \in \mathcal{B}$ for any Borel set B in the range space (see Lemma 2.5), and hence X is a Borel function. If $X(\omega) = \omega^2$, $\omega \in R$, $X^{-1}(a, b)$ ($b > a > 0$) equals $[\omega : \omega^2 \in (a, b)] = (-\sqrt{b}, -\sqrt{a}) \cup (\sqrt{a}, \sqrt{b}) \in \mathcal{B}$. In general, $X^{-1}(B) \in \mathcal{B}$ for any Borel set $B \subset R$, the range space and $\mathcal{B}$ in the domain. Hence X is a Borel function. Similarly, we can establish that any 1-1 monotone function $f(\omega)$, which has an inverse point function f^{-1} , is a Borel function (Prove!).

(d) FUNCTION OF A FUNCTION

If X is a function or a mapping from Ω to Ω' and X' is a function from Ω' to Ω'' , then $X'(X(\omega))$ is a function from Ω to Ω'' , denoted by $X'X$ or $X'(X)$. It is said to be a function of a function or a *composition of two functions* X and X' . Its inverse $(X'X)^{-1}$ is a function on the subsets of Ω'' to the subsets of Ω such that for any $B \subset \Omega''$.

$$\begin{aligned} (X'X)^{-1}(B) &= [\omega : X'(X(\omega)) \in B] \\ &= [\omega : X(\omega) \in X'^{-1}(B)], \\ &= X^{-1}(X'^{-1}(B)). \end{aligned} \quad (2.24)$$

Thus, $(X'X)^{-1} \equiv X^{-1}X'^{-1}$.

LEMMA 2.4: *Borel function of $\mathcal{A}$ -measurable function X is $\mathcal{A}$ -measurable and induces a sub- σ -field of that induced by X .*

Proof: If $f(X)$ is the Borel function of X and B is any Borel set,

$$[f(X) \in B] = [X \in f^{-1}(B)].$$

But, if B is a Borel set, $f^{-1}(B)$ is a Borel set B' and hence

$$\{X \in f^{-1}(B)\} = \{\omega : X(\omega) \in B'\} \in X^{-1}(\mathcal{B}) \subset \mathcal{A}.$$

Thus $f(X)$ is $\mathcal{A}$ -measurable and induces a sub- σ -field of $X^{-1}(\mathcal{B})$, which is the σ -field induced by X . (*)

In the Example 1.5, suppose we associate with each ω , a number representing (No. of heads—No. of tails) called $X(\omega)$. We can obtain the σ -field induced by X and X^2 and verify (!) that the σ -field induced by X^2 is a sub- σ -field of that induced by X . We may note that X^2 is a Borel function of a $\mathcal{A}$ -measurable function X .

2.2 RANDOM VARIABLES

(a) ECONOMICAL DEFINITION

So far, in the definition of a function we have not assumed that Ω is the sample space. Here onwards we shall restrict ourselves only to the case where Ω is the sample space. Let $\mathcal{A}$ be the σ -field of events associated with a certain fixed experiment. Any real valued, $\mathcal{A}$ -measurable function defined on Ω is called a *random variable* (r.v.). Thus X is a r.v. iff $X^{-1}(\mathcal{B})$, the σ -field induced by X , is contained in $\mathcal{A}$ or equivalently, $X^{-1}(B)$ is an event for every Borel set B in the range space of X .

Example 2.7. From (2.5), if $A \in \mathcal{A}$, $I_A^{-1}(B) \in \mathcal{A}$, i.e., if A is an event, I_A is a random variable. By Lemma 2.3, a simple function $X = \sum_{i=1}^n x_i I_{A_i}$ induces the σ -field containing the partition $\{A_1, A_2, \dots, A_n\}$. If this partition is contained in $\mathcal{A}$, i.e., if each one of the A_i 's is an event, then X is a r.v. called a *simple r.v.* Since singletons $\{x\}$, ($x \in R$), belong to $\mathcal{B}$, $[X = x] = [\omega : X(\omega) = x] = X^{-1}(\{x\}) \in \mathcal{A}$.

From Lemma 2.4, Borel function $g(X)$ of a r.v. X is a r.v. and induces a σ -field, which is a sub- σ -field of that which is induced by X . For a given X , define two non-negative functions

$$\begin{aligned} X^+(\omega) &= X(\omega), \text{ if } X(\omega) \geq 0; \\ &= 0, \text{ if } X(\omega) < 0, \end{aligned} \quad (2.25)$$

and

$$\begin{aligned} X^-(\omega) &= -X(\omega), \text{ if } X(\omega) < 0, \\ &= 0, \text{ if } X(\omega) \geq 0, \end{aligned} \quad (2.26)$$

called respectively, the *positive* and *negative parts* of X . Then X^+ and X^- are Borel functions of X (Why?) and will be random variables if X is a r.v. We may note that X^+ and X^- are non-negative functions, such that for all ω ,

$$X(\omega) = X^+(\omega) - X^-(\omega), \quad |X(\omega)| = X^+(\omega) + X^-(\omega).$$

These functions play an important role in the theory of integration to be developed in later chapters.

To verify whether a certain function is a random variable, it is not

necessary to determine whether, $X^{-1}(B) \in \mathcal{A}$ for every $B \in \mathcal{B}$. It is sufficient if we verify $X^{-1}(\mathcal{C}) \subset \mathcal{A}$, where $\mathcal{C}$ is any class of subsets of R given in equation (1.32), which generates $\mathcal{B}$. Hence we have the following 'economical definition' of a r.v.

LEMMA 2.5: X is a r.v. iff $X^{-1}(\mathcal{C}) \subset \mathcal{A}$, where $\mathcal{C}$ is any class of subsets of R , which generates $\mathcal{B}$.

Proof: We have to prove that

$$X^{-1}(\mathcal{C}) \subset \mathcal{A} \Leftrightarrow X^{-1}(\mathcal{B}) \subset \mathcal{A}. \quad (2.27)$$

Since $\mathcal{C} \subset \mathcal{B}$ and $X^{-1}(\mathcal{B}) \subset \mathcal{A}$, $X^{-1}(\mathcal{C}) \subset \mathcal{A}$. To prove the converse, since $\mathcal{A}$ is a σ -field and

$$\begin{aligned} X^{-1}(\mathcal{C}) \subset \mathcal{A} &\Rightarrow \sigma(X^{-1}(\mathcal{C})) \subset \mathcal{A}, \\ &\Rightarrow X^{-1}(\sigma(\mathcal{C})) \subset \mathcal{A}, \text{ (by Cor. (2.2))} \\ &\Rightarrow X^{-1}(\mathcal{B}) \subset \mathcal{A}. \end{aligned}$$

Hence we have the Lemma. (*)

From the Lemma 2.5, to verify that X is a r.v., it is sufficient to verify that $X^{-1}(-\infty, x) \in \mathcal{A}$ (say), for all $x \in R$. Thus some authors define a r.v. X to be a real valued function for which $X^{-1}(-\infty, x) = [\omega: X(\omega) < x] = [X < x]$, (say), is an event, whatever be $x \in R$.

Sometimes it is more convenient to verify that $[X > x] = X^{-1}(x, \infty)$ are events $\forall x \in R$, for establishing that X is a r.v. We can prove directly that X^+ , X^- and $|X|$ are r.v.'s, if X is a r.v. (Prove!).

(b) VECTOR RANDOM VARIABLE

With any $\omega \in \Omega$ associate $Z(\omega) = (X(\omega), Y(\omega))$, a point in the 2-dimensional Euclidean R^2 . Z defines a function from Ω to R^2 . In R^2 , consider the class $\mathcal{C}$ of all rectangles bounded by the lines $x = a, x = b, y = c, y = d, a < b, c < d$, arbitrary. The minimal σ -field containing $\mathcal{C}$ is Borel field $\mathcal{B}_2$ in R^2 . We have seen in section (1.3) that it is also the minimal σ -field containing all infinite rectangles of the form $[(x, y): x < a, y < c, a, c \in R]$. Z will be called a 2-dimensional r.v. if $Z^{-1}(\mathcal{B}_2) \subset \mathcal{A}$. $Z^{-1}(\mathcal{B}_2)$ is the σ -field induced by Z . In analogy with the economical definition, (Lemma 2.5) Z will be a r.v., iff $Z^{-1}(B_2)$ is an event for any rectangle B_2 in R^2 .

Example 2.8. Let $Z = (I_A, I_B)$. Then

$$\begin{aligned} Z(\omega) &= (0, 0), \text{ if } \omega \in A^c B^c, \\ &= (1, 0), \text{ if } \omega \in AB^c, \\ &= (0, 1), \text{ if } \omega \in A^c B, \\ &= (1, 1), \text{ if } \omega \in AB. \end{aligned}$$

$$\begin{aligned}
Z^{-1}(B_2) &= \phi, \text{ if } B_2 \text{ contains none of the above 4 points,} \\
&= A^c B^c, \text{ if } B_2 \text{ contains } (0, 0) \text{ only;} \\
&= AB^c, \text{ if } B_2 \text{ contains } (1, 0) \text{ only,} \\
&= AB^c + A^c B^c, \text{ if } B_2 \text{ contains } (0, 0) \text{ and } (1, 0) \text{ only, etc.}
\end{aligned}$$

$$\begin{aligned}
\text{Hence } Z^{-1}(\mathcal{B}_2) &= \{\phi, A^c B^c, AB^c, AB^c + A^c B^c, \dots, \Omega\}, \\
&= \sigma\{A^c B^c, AB^c, A^c B, AB\}, \\
&= \sigma\{A, B\}.
\end{aligned} \tag{2.28}$$

Thus, Z will be r.v. if A and B are events.

LEMMA 2.6: $Z^{-1}(\mathcal{B}_2)$ is the smallest σ -field with respect to which X and Y are measurable.

Proof: Symbolically we have to prove that

$$Z^{-1}(\mathcal{B}_2) = \sigma[X^{-1}(\mathcal{B}) \cup Y^{-1}(\mathcal{B})]. \tag{2.29}$$

Consider the rectangle

$$\begin{aligned}
B &= \{(x, y) : a < x < b, c < y < d\}, a, b, c, d \in R, \\
Z^{-1}(B) &= [\omega : a < X(\omega) < b, c < Y(\omega) < d], \\
&= [\omega : a < X(\omega) < b] \cap [\omega : c < Y(\omega) < d], \\
&= X^{-1}(a, b) \cap Y^{-1}(c, d).
\end{aligned}$$

Taking $c = -\infty, d = +\infty$, we have $Y^{-1}(c, d) = Y^{-1}(R) = \Omega$.
Hence the class

$$\{Z^{-1}(B) : B \text{ a rectangle}\} \supset \{X^{-1}(a, b) : a < b, a, b \in R\}. \tag{2.31}$$

But $\sigma\{X^{-1}(a, b), a, b \in R\} = X^{-1}(\mathcal{B})$, the σ -field induced by X , and $\sigma\{Z^{-1}(B), B \text{ a rectangle}\} = Z^{-1}(\mathcal{B}_2)$, the σ -field induced by Z . Hence from (2.31),

$$Z^{-1}(\mathcal{B}_2) \supset X^{-1}(\mathcal{B}). \tag{2.32}$$

Similarly, $Z^{-1}(\mathcal{B}_2) \supset Y^{-1}(\mathcal{B})$. Therefore, $Z^{-1}(\mathcal{B}_2) \supset X^{-1}(\mathcal{B}) \cup Y^{-1}(\mathcal{B})$ and since $Z^{-1}(\mathcal{B}_2)$ is a σ -field,

$$Z^{-1}(\mathcal{B}_2) \supset \sigma[X^{-1}(\mathcal{B}) \cup Y^{-1}(\mathcal{B})]. \tag{2.33}$$

To prove the reverse inclusion, we see that

$$Z^{-1}(B) = X^{-1}(a, b) \cap Y^{-1}(c, d) \in \sigma[X^{-1}(\mathcal{B}) \cup Y^{-1}(\mathcal{B})].$$

Since the R.H.S. of (2.33) is a σ -field and since

$$\{Z^{-1}(B), B \text{ a rectangle}\} \subset \sigma[X^{-1}(\mathcal{B}) \cup Y^{-1}(\mathcal{B})],$$

$$\sigma\{Z^{-1}(B), B \text{ a rectangle}\} \subset \sigma[X^{-1}(\mathcal{B}) \cup Y^{-1}(\mathcal{B})].$$

$$\text{i.e., by (2.10), } Z^{-1}(\mathcal{B}_2) \subset \sigma[X^{-1}(\mathcal{B}) \cup Y^{-1}(\mathcal{B})]. \tag{2.34}$$

From (2.33) and (2.34) we have the Lemma 2.5. (*)

The Lemma implies that (X, Y) is a vector r.v. iff X and Y are r.v.'s.

Example 2.9. To obtain the σ -field induced by (I_A, I_B) , consider

$$I_A^{-1}(\mathcal{B}) = \{\phi, A, A^c, \Omega\}, I_B^{-1}(\mathcal{B}) = \{\phi, B, B^c, \Omega\},$$

$$I_A^{-1}(\mathcal{B}) \cup I_B^{-1}(\mathcal{B}) = \{\phi, A, B, A^c, B^c, \Omega\},$$

$$\sigma[I_A^{-1}(\mathcal{B}) \cup I_B^{-1}(\mathcal{B})] = \sigma\{AB, A^cB, AB^c, A^cB^c\} = \sigma\{A, B\}$$

which is the σ -field induced by (I_A, I_B) as can be verified from Example (2.8).

$\{AB, A^cB, AB^c, A^cB^c\}$ is a partition of Ω and is called the partition induced by (I_A, I_B) . In general, if $X = \sum_{j=1}^m x_j I_{A_j}$, $Y = \sum_{k=1}^n y_k I_{B_k}$ are two simple functions, the σ -field induced by (X, Y) is the minimal σ -field containing

$$\{(A_j \cap B_k), j = 1, 2, \dots, m, k = 1, 2, \dots, n\}. \quad (\text{Why?})$$

A Borel function on R^2 to R (or R^2) is a function such that the inverse of a Borel set in R (or R^2) is a Borel set in R^2 .

LEMMA 2.7: Any Borel function of a vector r.v. (X, Y) is a r.v.

Proof follows from Lemma 2.4, since a Borel function of a vector r.v. (X, Y) induces a σ -field which is a sub- σ -field of that induced by (X, Y) . It follows from this Lemma 2.7 that if (X, Y) is a vector random variable, $X + Y, X/Y (Y \neq 0), XY$, etc., which are Borel functions (Prove!) of (X, Y) are r.v.'s.

If X and Y are r.v.'s one can also directly prove that $X + Y$ is a r.v., without using the above Lemma 2.7 as follows: By Lemma (2.5), $X + Y$ would be a r.v. if for all $x \in R$,

$$[X + Y < x] = [\omega: X(\omega) + Y(\omega) < x] \in \mathcal{A}.$$

But $[X + Y < x]$ iff for some r , rational, $X < (x - r)$ and $Y < r$. Symbolically,

$$[X + Y < x] = \bigcup_{r, \text{rational}} \{[X < x - r] \cap [Y < r]\}. \quad (2.35)$$

Since X and Y are r.v.'s $[X < x - r] \in \mathcal{A}$ and $[Y < r] \in \mathcal{A}$ and hence $[X < x - r] \cap [Y < r] \in \mathcal{A}$. R.H.S. of (2.35) is a countable union, since rationals are countable. Thus, $[X + Y < x]$ is a countable union of sets of $\mathcal{A}$. Since $\mathcal{A}$ is a σ -field, $[X + Y < x]$ is an event, implying that $X + Y$ is a r.v. Similarly, one can prove directly that $XY, X/Y (Y \neq 0)$ are r.v.'s. (How?)

Thus, we see that 2-dimensional vector r.v.'s do not offer any more conceptual difficulty than the one-dimensional r.v.'s. Hence we can define a k -dimensional vector r.v. $Z = (X_1, X_2, \dots, X_k)$, as a mapping from Ω to R^k , the k -dimensional Euclidean space, such that $Z^{-1}(B)$ is an event, where B is a k -dimensional rectangle (say) defined by

$$B = \{(x_1, x_2, \dots, x_k): a_i < x_i < b_i, a_i, b_i \in R, i = 1, 2, \dots, k\} \quad (2.36)$$

If $\mathcal{B}_k$ is the Borel field in R^k , i.e., the minimal σ -field containing all the k -dimensional rectangles of the form (2.36), $Z^{-1}(\mathcal{B}_k)$ is the σ -field induced by Z . As in Lemma 2.6 we can prove (How?) that $Z^{-1}(\mathcal{B}_k)$ is the minimal σ -field containing the σ -fields

$$X_i^{-1}(\mathcal{B}), (i=1, 2, \dots, k),$$

$$\text{i.e.,} \quad Z^{-1}(\mathcal{B}_k) = \sigma\left(\bigcup_{i=1}^k X_i^{-1}(\mathcal{B})\right). \quad (2.37)$$

$Z^{-1}(\mathcal{B}_k)$ is also sometimes denoted by $\mathcal{B}(X_1, X_2, \dots, X_k)$.

The Lemma 2.6 extended to k -dimensions, also implies that Z will be a random vector iff each one of its components is a r.v.

If $X_1, X_2, \dots, X_k$ are r.v.'s, $\Sigma X_i, \Sigma X_i^2, (X_1 X_2 \dots X_k), \max(X_i), \min(X_i)$, etc., which are Borel functions of $(X_1, X_2, \dots, X_k)$, are r.v.'s. If X and Y are r.v.'s, $X - Y$ is a r.v. Hence the set

$$[X - Y = 0] = [\omega : X(\omega) = Y(\omega)] \in \mathcal{A}.$$

A r.v. X taking a finite or countable set of distinct values $x_1, x_2, \dots$ is said to be a *discrete r.v.* Evidently, $[X = x_n] \in \mathcal{A}$ for all n and these sets generate the σ -field induced by X . *Discrete random vector* is defined similarly. (Give an example!)

A *complex valued random variable* is a function on Ω to the complex plane whose real and imaginary parts are r.v.'s. Thus, $Z = X + iY$ is a complex r.v. iff (X, Y) is a random vector. If a r.v. can take the values $\pm \infty$, then it is said to be *extended real valued* or simply *extended r.v.* Thus, it is a mapping from Ω to $[-\infty, \infty] = \bar{R}$, the extended real line, such that the inverse images of all sets of the form $(-\infty, x), \{-\infty\}$ and $\{\infty\}$ belong to $\mathcal{A}$ (refer section 13.3, for an example).

2.3 LIMITS OF RANDOM VARIABLES

(a) σ -FIELD INDUCED BY A SEQUENCE OF R.V.'S

In the previous section 2.2 we have defined the σ -field induced by $(X_1, X_2, \dots, X_n)$ to be the minimal σ -field $\mathcal{A}_n$ containing the σ -fields induced by each one of the r.v.'s $X_1, X_2, \dots, X_n$, i.e.

$$\mathcal{A}_n = \sigma\left[\bigcup_{i=1}^n X_i^{-1}(\mathcal{B})\right].$$

Evidently $\{\mathcal{A}_n\}$ is a monotone increasing sequence of σ -fields with limit $\bigcup_{n=1}^{\infty} \mathcal{A}_n$. This union of σ -fields may not be a σ -field. The minimal σ -field $\mathcal{C}_1$ containing $\bigcup \mathcal{A}_n$ is called the σ -field induced by the infinite sequence $Z = (X_1, X_2, \dots)$. The range space of Z is R^∞ , the set of all infinite sequences of real numbers.

The Borel field $\mathcal{B}_\infty$ of subsets of R^∞ is defined as the minimal σ -field

containing all *cylinder sets* with finite dimensional base of the form

$$B = [(x_1, x_2, \dots) : a_{i_j} < x_{i_j} < b_{i_j}, j = 1, 2, 3, \dots, k; \\ -\infty < x_{i_j} < \infty, j = k+1, k+2, \dots] \quad (2.38)$$

where $(i_1, i_2, \dots)$ is any permutation of $(1, 2, 3, \dots)$, $a_i \in R$, $b_i \in R$ and k being arbitrary but finite. Evidently,

$$Z^{-1}(B) = \{\omega : a_{i_j} < X_{i_j}(\omega) < b_{i_j}; j = 1, 2, 3, \dots, k, -\infty < X_{i_j}(\omega) < \infty, \\ j = k+1, k+2, \dots\} \quad (2.39) \\ = \bigcap_{j=1}^k [\omega : a_{i_j} < X_{i_j}(\omega) < b_{i_j}]$$

Since $Z^{-1}(B)$ belongs to $\sigma\{\bigcup_{j=1}^k X_{i_j}^{-1}(\mathcal{B})\}$ for some k , $Z^{-1}(B) \in \mathcal{A}_n$ for some $n = 1, 2, \dots$. Thus, $Z^{-1}(B) \in \bigcup_{n=1}^{\infty} \mathcal{A}_n$ and hence to $\mathcal{C}_1$. Since $\mathcal{C}_1$ is a σ -field, the minimal σ -field containing $\{Z^{-1}(B)\}$, for every cylinder set B is contained in $\mathcal{C}_1$. Using (2.12),

$$\mathcal{B}(X_1, X_2, \dots) = Z^{-1}(\mathcal{B}_{\infty}) \subset \mathcal{C}_1. \quad (2.40)$$

Conversely, we can see that $\mathcal{A}_n \subset Z^{-1}(\mathcal{B}_{\infty})$ for any n (Why?) and hence $\bigcup \mathcal{A}_n \subset Z^{-1}(\mathcal{B}_{\infty})$ implying that $\mathcal{C}_1 \subset Z^{-1}(\mathcal{B}_{\infty})$. Thus,

$$\mathcal{C}_1 = Z^{-1}(\mathcal{B}_{\infty}). \quad (2.41)$$

From 2.41, we see that the σ -field induced by an infinite sequence of r.v.'s is the minimal σ -field containing σ -fields induced by each one of them (Why?). It is also the σ -field containing the inverse images of all cylinder sets with finite dimensional base. But the latter sets are events, i.e., elements of $\mathcal{A}$. Hence $\mathcal{C}_1 \subset \mathcal{A}$.

Let $\mathcal{C}_n$ be the σ -field induced by $(X_n, X_{n+1}, \dots)$. Then

$$\mathcal{C}_n = \sigma\left(\bigcup_{k=n}^{\infty} X_k^{-1}(\mathcal{B})\right) \supset \mathcal{C}_{n+1}; n = 1, 2, \dots$$

Thus $\mathcal{C}_n$ is a monotone decreasing sequence of σ -fields tending to the limit $\mathcal{T} = \bigcap_{n=1}^{\infty} \mathcal{C}_n$. By theorem 1.2, being the intersection of σ -fields, $\mathcal{T}$ is a σ -field, called the *tail σ -field* of the sequence $\{X_n\}$. Evidently the tail σ -field $\mathcal{T}$ is contained in $\mathcal{C}_n$ for all n . The events of $\mathcal{T}$ are called *tail events*. A function (r.v.) which induces a σ -field which is a sub- σ -field of $\mathcal{T}$ is called a *tail-measurable function* (r.v.). If $\mathcal{T}$ is degenerate then the only tail-measurable functions are degenerate functions (r.v.'s) taking constant values.

(b) LIMITS OF RANDOM VARIABLES

Let $\{X_n\}$ be a sequence of r.v.'s on $(\Omega, \mathcal{A})$. Let us define functions Y_n and Z_n on Ω as follows:

$$Y_n(\omega) = \inf_{k \geq n} X_k(\omega) = \text{greatest lower bound of } X_k(\omega) \text{ for } k \geq n,$$

$$Z_n(\omega) = \sup_{k \geq n} X_k(\omega) = \text{least upper bound of } X_k(\omega) \text{ for } k \geq n.$$

Then, as $n \rightarrow \infty$,

$$Y_n(\omega) \uparrow \sup Y_n(\omega) = \liminf X_n(\omega) = \underline{\lim} X_n(\omega),$$

$$Z_n(\omega) \downarrow \inf Z_n(\omega) = \limsup X_n(\omega) = \overline{\lim} X_n(\omega).$$

$\lim X_n(\omega)$ and $\overline{\lim} X_n(\omega)$ will exist for all $\omega \in \Omega$. Moreover $\overline{\lim} X_n(\omega) \geq \lim X_n(\omega)$, $\forall \omega$. We say that $\{X_n\}$ converges to X on A , denoted $X_n(\omega) \rightarrow X(\omega)$, $\omega \in A$, if

$$\overline{\lim} X_n(\omega) = \lim X_n(\omega) = X(\omega) < \infty, \forall \omega \in A. \quad (2.42)$$

The set A of all ω for which $X_n(\omega) \rightarrow X(\omega)$ is called the *set of convergence* of the sequence $\{X_n\}$. Evidently, if for a given ω , $X_n(\omega) \rightarrow X(\omega) < \infty$, then for given $\varepsilon > 0$, there exists $n(\varepsilon, \omega)$ such that for $n > n(\varepsilon, \omega)$, $|X_n(\omega) - X(\omega)| < \varepsilon$. If $n(\varepsilon, \omega)$ is independent of $\omega \in A$, then the convergence is said to be *uniform on A*. If A is the set of convergence, then A^c is the set of all ω where $\{X_n\}$ does not converge. The set of all ω where $\lim X_n(\omega) = \overline{\lim} X_n(\omega) = X(\omega) = \pm \infty$, is the set of divergence. If Ω is the set of convergence, then $\{X_n\}$ is said to *converge everywhere*. If the set of convergence is ϕ , $\{X_n\}$ converges nowhere.

Example 2.9. Let $\{A_n\}$ be a sequence of sets and $\{I(A_n)\}$ be the sequence of their indicator functions. Then

$$\inf_{k \geq n} I(A_k)(\omega) = 1, \text{ if } \omega \in A_k \text{ for all } k \geq n \\ = 0, \text{ otherwise.}$$

Hence,

$$\inf_{k \geq n} I(A_k) = I(\inf_{k \geq n} A_k), \forall n.$$

Similarly,

$$\sup_{k \geq n} I(A_k)(\omega) = I(\sup_{k \geq n} A_k)(\omega) = 1, \text{ if } \omega \in A_k \text{ for some } k \geq n, \\ = 0, \text{ otherwise.}$$

$$\liminf I(A_n)(\omega) = I(\liminf A_n)(\omega) = 1,$$

if $\omega \in$ all but finite number of A_n 's.

$$\limsup I(A_n)(\omega) = I(\limsup A_n)(\omega) = 1,$$

if $\omega \in$ infinitely many A_n 's. If $\lim A_n$ exists, then

$$I(\lim A_n) = \lim I(A_n).$$

LEMMA 2.8: $Y_n, Z_n, \lim X_n, \overline{\lim} X_n$ are extended r.v.'s and $\lim X_n$ (when it exists) is also an extended r.v.

Proof: Consider the function Y_n . Evidently it is real valued or $-\infty$. For any $x \in R$,

$$\begin{aligned}
 [Y_n < x] &= \{\omega : \text{at least one of } X_k(\omega) \text{ is less than } x, k \geq n\}, \\
 &= \bigcup_{k \geq n} [X_k < x].
 \end{aligned}
 \quad (2.43)$$

Since X_k 's are r.v.'s $[X_k < x] \in \mathcal{A}$. Hence $[Y_n < x] \in \mathcal{A}, \forall x \in R$. Thus, Y_n is an extended r.v. whatever be n .

Similarly Z_n is real valued or $+\infty$ with

$$[Z_n > x] = \bigcup_{k \geq n} [X_k > x] \in \mathcal{A}, \quad (2.44)$$

for all $x \in R$ and all n . Hence Z_n is an extended r.v. But supremum and infimum of extended r.v.'s are extended r.v.'s. Hence $\sup Y_n$ and $\inf Z_n$ are extended r.v.'s, i.e., $\lim X_n$ and $\lim X_n$ are extended r.v.'s. If $\lim X_n$ exists on Ω , then it is equal to $\lim X_n$ (say) and is hence an extended r.v.

If A is the set of convergence of $\{X_n\}$, limit of $\{X_n\}$ will be finite on A only and

$$A = \{\omega : \lim X_n(\omega) = \lim X_n(\omega) < \infty\}. \quad (2.45)$$

If $V = \lim X_n - \lim X_n$, then V is an extended real valued non-negative random variable on the set where it exists and hence

$$A = V^{-1}(\{0\}) \in \mathcal{A}.$$

Note that V does not exist on the set of divergence, i.e., if

$$\lim X_n = \lim X_n = \pm \infty.$$

Define

$$\begin{aligned}
 X(\omega) &= \lim X_n(\omega), \text{ say, for } \omega \in A, \\
 &= \infty, \text{ for } \omega \in \text{set of divergence.}
 \end{aligned}$$

Then for $B \in \mathcal{B}$, since $\lim X_n$ is an extended r.v.,

$$X^{-1}(B) = (\lim X_n)^{-1}(B) \cap A \in \mathcal{A}. \quad (2.46)$$

Since the set of divergence is measurable, X is an extended r.v. (*).

It may be noted that, since Y_n is a function of $(X_n, X_{n+1}, \dots)$ it is $\mathcal{C}_n$ -measurable. If $T_n = \sup_{k \geq n} Y_k$, then T_n is also $\mathcal{C}_n$ -measurable. But $T_n = \lim X_k$. Thus, $\lim X_k$ is measurable with respect to $\mathcal{C}_n$ for all $n = 1, 2, \dots$. Since $\mathcal{C}_n \downarrow \mathcal{T}$, $\lim X_n$ is $\mathcal{T}$ -measurable. Similarly, $\lim X_n$ and $\lim X_n$ are also $\mathcal{T}$ -measurable. Since V is $\mathcal{T}$ -measurable, $A \in \mathcal{T}$. Thus, the set of convergence is a tail event.

We have seen that if a r.v. X takes only a finite number of values $x_1, x_2, \dots, x_n$, then it is a simple r.v. If it takes a countable number of values $x_1, x_2, \dots$, then it can be represented by

$$X = \sum_{k=1}^{\infty} x_k I(A_k),$$

and is called an *elementary r.v.*, where

$$A_k = \{\omega : X(\omega) = x_k\} \in \mathcal{A} (k = 1, 2, \dots).$$

The partition $\{A_k\}$ induced by X contains infinite number of events and so also the σ -field $\sigma\{A_k\}$ induced by X . A discrete r.v. is a finite or an elementary r.v.

(c) CONSTRUCTIVE DEFINITION

The following Theorem 2.1 gives us a *constructive definition* of a r.v. It tells us that we can get an arbitrary, r.v. as the pointwise limit (finite) of either simple r.v.'s or elementary r.v.'s.

THEOREM 2.1: *The class of r.v.'s = The class of finite limits of sequences of simple r.v.'s = The class of uniform (finite) limits of sequences of elementary r.v.'s.*

Proof: (i) Let $\mathcal{R}$ be the class of all r.v.'s and $\mathcal{S}$ be the class of finite limits of sequences of simple r.v.'s. Let S be an element of $\mathcal{S}$. S is a limit of a sequence of simple r.v.'s. By Lemma 2.8, finite limits of r.v.'s are r.v.'s. Hence S is a r.v. and $S \in \mathcal{R}$.

Conversely, if X is any r.v., define a real valued function X_n by

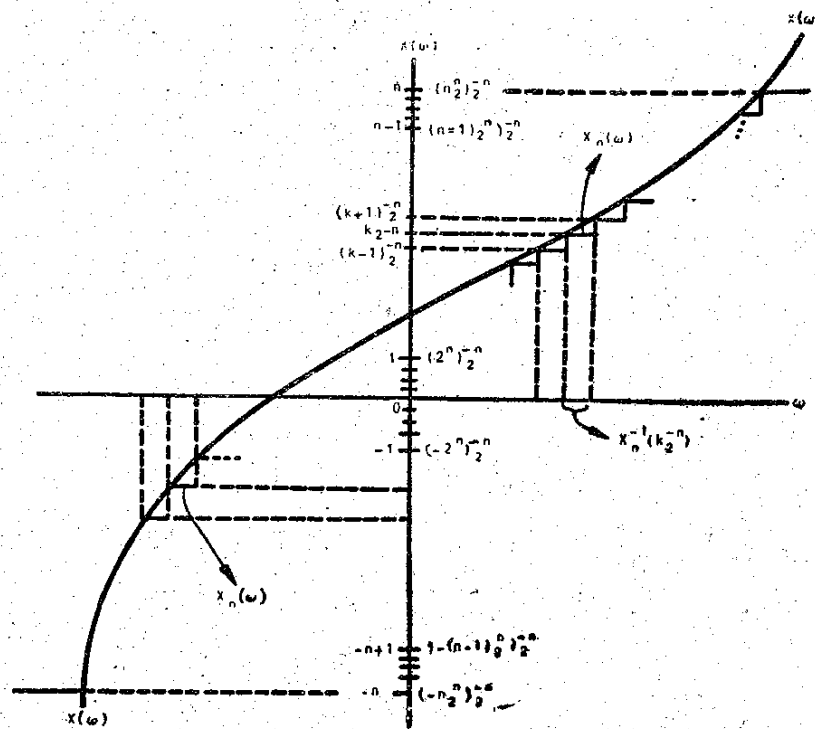


Fig. 2.1.

$$\begin{aligned}
X_n(\omega) &= n, \text{ if } X(\omega) \geq n, \\
&= k2^{-n}, \text{ if } k2^{-n} \leq X(\omega) < (k+1)2^{-n}, \\
(k &= -n2^n, -n2^n - 1, \dots, n2^n - 1), \\
&= -n, \text{ if } X(\omega) < -n.
\end{aligned}$$

$$\begin{aligned}
\text{i.e., } X_n &= -nI[X < -n] + nI[X \geq n] \\
&+ \sum_{k=-n2^n}^{n2^n-1} k2^{-n}I[k2^{-n} \leq X < (k+1)2^{-n}]. \quad (2.47)
\end{aligned}$$

See Fig. (2.1), where Ω is taken as the real line. Since X is a r.v., the sets $[X < -n]$, $[X \geq n]$, $[k2^{-n} \leq X < (k+1)2^{-n}]$ are events. Thus, X_n is a simple r.v. Moreover, for $|X(\omega)| < n$, $|X_n(\omega) - X(\omega)| < 2^{-n}$, and $|X_n(\omega)| = n$ for $|X(\omega)| \geq n$. Therefore, $X_n(\omega) \rightarrow X(\omega)$ as $n \rightarrow \infty$ for all ω , implying that if $X \in \mathcal{R}$, then $X \in \mathcal{S}$.

(ii) Let $\mathcal{E}$ be the class of uniform limits of elementary r.v.'s. Then by Lemma 2.8, if E is an element of $\mathcal{E}$, then $E \in \mathcal{R}$. To prove the converse, define

$$X_n = \sum_{-\infty}^{\infty} k2^{-n}I[k2^{-n} \leq X < (k+1)2^{-n}]. \quad (2.48)$$

Now X_n is an elementary r.v. $|X_n(\omega) - X(\omega)| < 2^{-n}$ for all ω . Thus, X is the uniform limit of X_n 's. Hence any X belonging to $\mathcal{R}$ is also an element of $\mathcal{E}$. The theorem is proved. (*)

COROLLARY 2.4: *If $X \geq 0$, there exists a monotone increasing sequence of non-negative simple function $\{X_n\}$ converging to X .*

Proof: Define

$$\begin{aligned}
X_n(\omega) &= k2^{-n}, \text{ if } k2^{-n} \leq X(\omega) < (k+1)2^{-n}, (k=0, 1, \dots, n2^n-1), \\
&= n, \text{ if } X(\omega) \geq n,
\end{aligned}$$

so that

$$X_n = \sum_{k=0}^{n2^n-1} k2^{-n}I[k2^{-n} \leq X < (k+1)2^{-n}] + nI[X \geq n] \quad (2.49)$$

is a simple r.v. Then $X_n(\omega) \geq 0$ for all n and ω . Moreover,

$$\begin{aligned}
X_{n+1}(\omega) &= 2k2^{-(n+1)}, \text{ if } 2k2^{-(n+1)} \leq X(\omega) < (2k+1)2^{-(n+1)}, \\
&= (2k+1)2^{-(n+1)}, \text{ if } (2k+1)2^{-(n+1)} \leq X(\omega) < (2k+2)2^{-(n+1)}, \\
&\quad (k=0, \dots, (n+1)2^{n+1}-1), \\
&= n+1, \text{ if } X(\omega) \geq n+1.
\end{aligned}$$

Comparing with $X_n(\omega)$, we see that $X_{n+1}(\omega) \geq X_n(\omega)$ for all n and ω . Since $X_n(\omega) \leq X(\omega)$ and $|X_n(\omega) - X(\omega)| < 2^{-n}$ for $X(\omega) \leq n$, as $n \rightarrow \infty$ $0 \leq X_n(\omega) \uparrow X(\omega)$ for all $\omega \in \Omega$. Hence we have the Corollary. (*)

Theorem 2.1 and Corollary 2.4 will play a very important role in the theory of integration, where we build up the properties of integrals

starting with simple functions and then deriving it for non-negative functions, later, for arbitrary functions. With any r.v. X we may associate two non-negative r.v.'s defined by (2.25) and (2.26) or equivalently by

$$X^+ = XI[X \geq 0], \quad (2.50)$$

$$X^- = -XI[X < 0].$$

If X is a r.v., $[\omega: X(\omega) \geq 0]$ and $[\omega: X(\omega) < 0]$ are events and hence the indicators of these sets are r.v.'s. Thus, X^+ is the product of two r.v.'s and is hence a r.v. Similarly X^- is also a r.v. Since they are non-negative, there exist monotone increasing sequences of simple functions converging to them. We may note that since

$$X = X^+ - X^-, \quad (2.51)$$

$$|X| = X^+ + X^-,$$

X and $|X|$ are linear functions of X^+ and X^- .

In this section so far we were concerned with $\mathcal{A}$ -measurable real valued functions on Ω . If Ω is the real line with $\mathcal{B}$ as the σ -field, then the above theorem implies that any Borel measurable real valued function, i.e., Borel function is a limit of simple Borel functions. But simple Borel functions are *step functions* which increase or decrease in steps only. Thus, Borel functions are step functions or limits of step functions.

THEOREM 2.2: *Continuous real valued functions on R are Borel functions.*

Proof. Let $g(\cdot)$ be a continuous function from R to R . Consider $g(X)$, where X is $\mathcal{B}$ -measurable function from R to R . By Theorem (2.1), there exists a sequence $\{X_n\}$, of elementary functions converging uniformly to X . Since X_n takes only a countable number of distinct values, $g(X_n)$ also takes only a countable number of distinct values and is therefore, an elementary function from R to R . Since g is continuous, as

$$X_n(\omega) \rightarrow X(\omega), \quad g(X_n(\omega)) \rightarrow g(X(\omega)).$$

Thus, $g(X)$ is a limit of elementary Borel-measurable functions and is hence Borel-measurable.

Take X to be the identity function with $X(\omega) \equiv \omega$. This is Borel function. Then $g(X(\omega)) = g(\omega)$. Since $g(X)$ is shown to be Borel measurable for any measurable function X , $g(\cdot)$ is a Borel function. (*)

Since continuous functions are Borel functions, and by Lemma 2.8 the class of Borel functions is closed under the formation of finite limits, limits of continuous functions are also Borel functions. But continuous functions or their limits are called *Baire functions*. Theorem 2.2 implies that a Baire function is a Borel function. It may be noted that limits of continuous functions need not be continuous.

COMPLEMENTS AND PROBLEMS

1. Let $f: \Omega \rightarrow \Omega'$. Prove that

(i) $A \subset f^{-1}(f(A))$, if $A \subset \Omega$,

(ii) $B \supset f(f^{-1}(B))$, if $B \subset \Omega'$.

By a suitable example show that the equality relation may not hold good.

2. Show that every function X on Ω is measurable with respect to the power set of Ω .

3. We have seen that if Y is a function of X , Y induces a sub- σ -field of that induced by X . Is the converse true? Can you prove that there exists a Borel function g such that $Y = g(X)$, in the case X is real valued? (Hint: Borel field has a countable number of generators).

4. Show that a necessary and sufficient condition that a given function is measurable is that its positive and negative parts are measurable.

5. Show that $|X|$ is $\mathcal{A}$ -measurable does not imply that X is $\mathcal{A}$ -measurable, by giving a counter example.

6. Obtain the positive and negative parts of the function $f: R \rightarrow R$ defined by $f(x) = ax^2 + bx + c$.

7. (a) If A_1, A_2 are measurable sets and a function X is defined by

$$\begin{aligned} X(\omega) &= -1, & \omega \in A_1, \\ &= +1, & \omega \in A_1^c A_2, \\ &= 0, & \omega \in A_1^c A_2^c. \end{aligned}$$

Examine whether X is measurable.

(b) Let $\Omega = \{a, b, c, d\}$, $\mathcal{A} = \{\phi, \Omega, \{a, b\}, \{c, d\}\}$.

$X(a) = X(b) = -1$, $X(c) = 1$, $X(d) = 2$. Examine whether X is $\mathcal{A}$ -measurable.

(c) Let $\Omega = \{-2, -1, 0, 1, 2\}$. If $X(i) = i$, ($i = 0, \pm 1, \pm 2$) obtain the σ -fields induced by X and X^2 and verify that the latter induces a sub- σ -field of the former.

8. Let f and g be Borel functions on $(R, \mathcal{B})$. Define for $A \in \mathcal{B}$,

$$\begin{aligned} h(\omega) &= f(\omega), & \omega \in A, \\ &= g(\omega), & \omega \in A^c \end{aligned}$$

Show that h is a Borel function.

9. Using the constructive definition of measurability of a function, show that an elementary function is measurable. Show further that the positive and negative parts of an elementary function are elementary.

10. Let T be a mapping such that a point $(x_1, \dots, x_n) \in R^n$ is mapped onto the point $(x^{(1)}, \dots, x^{(n)}) \in R^n$ where $x^{(1)} < x^{(2)} < \dots < x^{(n)}$, is the ordered set obtained from $(x_1, \dots, x_n)$. Determine (i) the strict range

of T ; (ii) $T^{-1}(y_1, \dots, y_n)$; (iii) $T^{-1}(B)$ where B is a Borel subset of the strict range of T .

11. Let T' be the statistic (r.v.) such that a point $(x_1, \dots, x_n) \in R^n$ is mapped into the point $(\sum_{i=1}^n x_i, \sum_{i=1}^n x_i^2, \dots, \sum_{i=1}^n x_i^n) \in R^n$. Show that the statistic T introduced in exercise 10 and the statistic T' induce the same σ -field.
12. Later, we shall have occasions to take limits of probabilities of an uncountable class of events, $A_{x'}$, as $x' \downarrow x$, say, through a continuum of points x' of an interval. Also, we shall have occasion to consider limits of r.v.'s $X_{t'}$ indexed by a continuum of points t' of a certain interval t . The question naturally arises whether such a limiting process preserves measurability. The answer is definitely in the affirmative, if the uncountable class (of events or r.v.'s) can be replaced by a countable class. If $x' \in R$ and $x' \downarrow x$, then x' may be restricted to the set of rationals converging to x . Since rationals are countable dense we have a sequence of points $x_n' \downarrow x$. Similarly even though the index set t' is uncountable, if they correspond to points of an interval, we may replace t' by an equivalent sequence t_n' of rationals converging to t . Thus, $\lim X_{t'}$ as $t' \downarrow t$, $\lim X_{t'}$, $\lim X_{t'}$, etc. will be r.v.'s, if the index set T to which t' and t belong is a subset of the real line or in general has a countable dense subset in it.

3. Probability Space

3.1 DEFINITION OF PROBABILITY

(a) CONCEPT OF PROBABILITY

When an experiment is performed it may result in one of possibly several outcomes. Because of 'chance' factors, we may not be able to predict the outcome precisely. However, we may be able to offer our opinion about the possible outcomes, by giving a quantitative measure of the 'chance' that a particular outcome would occur. Thus, with every event A consisting of one or more outcomes of an experiment, we may associate a numerical quantity, called the *probability of A* , denoted by $P(A)$, or PA , which will measure the 'chance' that the event A will occur. We take $0 \leq P(A) \leq 1$, so that ' $P(A) = 0$ ' implies that there is no 'chance' that A will occur and ' $P(A) = 1$ ' implies that A will certainly occur. If $P(A) = \frac{1}{2}$, then it will mean that there is 50% 'chance' that A would occur.

The purpose of introducing a quantitative measure of 'chance' is manifold. It allows us to compare the chances of occurrence of two or more events. Out of six cards held by the two opponents in a bridge game, which distribution is most likely among (3, 3), (4, 2), (5, 1) and (6, 0)? In a journey of 500 miles, whether travel by road is more risky than travel by air? Questions such as these can be answered if we have a precise quantitative measure of the 'chances' associated with the occurrence of the concerned events.

Another advantage of defining probabilities of events in quantitative terms is that it now becomes amenable to further mathematical analysis. We can talk of the possible gain to a gambler or expected profit from a risky venture. We can calculate the insurance premium to be charged to cover a certain risk.

We are all familiar with the intuitive meaning of 'chance'. In fact, in everyday life, almost everything that we do or see is subject to chance. There is hardly anything that is deterministic. If we want to allow this chance factor to be taken into our judgement, we are led to the study of *stochastic* (or probabilistic) variables. Old classical laws of physics such as Newton's laws of motion which are deterministic are now being replaced by laws which allow chance or probability to enter explicitly or

implicitly in the model. Quantum mechanics or models for nuclear reactions are some such models.

The concept of probability is similar to that of distance, velocity or volume. It is clear intuitively but requires a clear-cut definition and some unit to measure it. The unit of measurement of probability is 'unity', for convenience. It may be expressed as a percentage by multiplying by 100. There is no reason why we should not multiply by any other number, *c*.

As stated by Feller, determining probability of an event may be compared to the determination of distance between two points. Several methods may be adopted to determine the same, not all of which give the same value. But this does not deter us from assigning the distance between two points or studying its properties. Similarly, there are several methods suggested to obtain the 'value' of the probability of an event. At times these methods may lead to different values raising controversies. Controversies are aggravated by confusing these 'methods' for 'definitions' of probability. Feller (1968) has differentiated between the notion of probability which is fundamental and idealistic and the method of calculating their values, which may vary with the situation at hand. But the confusion caused by calling a 'method' a 'definition' still persists. In this book we shall be concerned with the concept of probability as defined by a set of mathematical axioms. Defining probability in this manner is similar to the definition of velocity or volume, which are also mathematical entities. We assume that its actual value is determined by some relevant method, without discussing as to how this value is obtained.

In this section we shall give a brief account of the important 'definitions of probability'. These are nothing but different procedures of determining probabilities of events which have evolved over time. But these different procedures will not complicate the later development, which is based on the properties common to all the approaches. These common properties have been codified by Kolmogorov to form the axioms of the modern probability theory.

(b) CLASSICAL METHOD

This method of assigning probabilities is applicable only in the case of an experiment with finite number (N , say) of outcomes, $\omega_1, \omega_2, \dots, \omega_N$. It will be assumed here that all the outcomes are 'equally likely' or have the same chance of occurrence. Consequently, probability $P(\{\omega_i\})$ assigned to each one of the singleton events $\{\omega_i\}$ is the same. Since one of the outcomes is sure to happen, total probability

$$P(\Omega) = P(\{\omega_1\} + \{\omega_2\} + \dots + \{\omega_N\}) = \sum_{i=1}^N P(\{\omega_i\}) = 1. \quad (3.1)$$

Here we have made an assumption of additivity that if a certain event A contains the points

$$\{\omega_{i_1}, \omega_{i_2}, \dots, \omega_{i_n}\}, \text{ (say)}$$

then

$$P(A) = \sum_{j=1}^n P(\{\omega_{ij}\}). \quad (3.2)$$

Since $P(\{\omega_{ij}\})$ are assumed to be the same for all i , from (3.1) and (3.2) we have

$$P(\{\omega_{ij}\}) = \frac{1}{N}, \quad P(A) = \frac{n}{N}. \quad (3.3)$$

Thus, probability is a numerical function defined on the class of events associated with an experiment.

Example 3.1. Suppose a die with faces marked 1, 2, ..., 6 is thrown. Since any one of these 6 faces may turn up, assuming them to be equally likely, denoting these events by $\{\omega_1\}, \{\omega_2\}, \dots, \{\omega_6\}$, respectively, we have

$$\Omega = \{\omega_1, \omega_2, \dots, \omega_6\},$$

$$P(\{\omega_i\}) = 1/6, \quad (i = 1, 2, \dots, 6).$$

If A is the event that an even number turns up,

$$P(A) = P(\{\omega_2, \omega_4, \omega_6\}) = 3/6 = 1/2.$$

From (3.3), probability of an event depends only on the number of sample points contained in it. We can easily verify that this probability function defined on subsets of Ω , possesses the following properties:

$$\left. \begin{array}{ll} \text{(i)} & P(A + B) = P(A) + P(B), \\ \text{(ii)} & A \subset B \Rightarrow P(A) \leq P(B), \\ \text{(iii)} & P(\Omega) = 1, \\ \text{(iv)} & 0 \leq P(A) \leq 1. \end{array} \right\} \quad (3.4)$$

In order to establish these, we see that the additivity assumption is implied by (i), and is equivalent to it. It follows from (i) that

$$P\left(\sum_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i). \quad (3.5)$$

(ii) is a consequence of (i) and (iv), (iii) implies that P is a 'normed' function, while (iv) implies that it is non-negative. These properties will form the basis of the axiomatic definition of probability, to be given later.

This method of assigning probabilities to events was found useful when applied to the outcomes of the game of chance. The concept of randomness or randomization means in reality that we are dealing with phenomenon whose outcomes are equally likely. An important

limitation of this method is that it breaks down if the number of outcomes is infinite, countable or uncountable.

(c) STATISTICAL METHOD

Suppose A is a set of outcomes of an experiment. Suppose it is possible to repeat the experiment a large number of times. Denote by n_A , the number of occurrences of A in n repetitions. n_A is called the frequency of A , and n_A/n , the *relative frequency*. It is an observed fact that the relative frequencies stabilise to a certain value as n becomes large and this stable value will be called the probability of A , i.e., $P(A) = \lim (n_A/n)$, as $n \rightarrow \infty$.

If H denotes the occurrence of 'head' in a single throw of a coin, and if in a large number of throws, 50 per cent of throws result in heads, $P(H)$ will be $1/2$; if 40 per cent result in heads $P(H)$ will be $2/5$ and so on. According to classical definition, since there are only two possible outcomes, probability of events such as 'throwing a head', 'giving birth to a boy', etc., is always $1/2$. But, according to statistical definition, these events may have probability p , $0 \leq p \leq 1$.

The relative frequency of a sum of two disjoint events A , B is the sum of the relative frequencies of A and B . Since the limit of a sum $(n_A + n_B)/n$, is the sum of limits of n_A/n and n_B/n , properties (i)-(iv) of (3.4) hold true even in the present case.

Stability of relative frequencies for large values of n constitutes the basis of statistical theory. The present method of assigning probability based on this principle of *statistical regularity* is called the *statistical definition of probability*. This enables us to use the theory of probability in statistical analysis. It provides a link between the outside world and the mathematical theory of probability we study. It enables us to construct probability (or stochastic) models of natural phenomena and study them. If the model is good, conclusions derived from the models would be valid to an extent that the principle of statistical regularity is valid. This forms the basis of applied probability theory.

According to this 'definition', probability is a limiting concept. Hence it can never be determined in the real world. But this does not refrain us from using this method for assigning probabilities to various events, in much the same way as we are not deterred from using geometry in everyday life, even though the notions of line, point, etc., on which geometry is based, are limiting concepts. The 'definition' also assumes that the conditions of the experiment are reproducible. This assumption may not be valid, if 'time' is one of the variables involved. However, if past is any guide to the future and if by and large nature is stable, the above method of assigning probabilities to real-life situations should be satisfactory.

(d) SUBJECTIVE METHOD

In everyday life we hear or make statements such as 'Probably I shall miss the boat,' 'Probably Mr. X will be at home now.' Such statements can be made more precise by 'The chance of my missing the boat is 60%,' 'The chance that Mr. X will be at home now is 75%' etc. Here 0.6, 0.75, etc., measure one's belief in the occurrence of the event. This *subjective method* is another method of assigning probabilities to various events. Of course, one has to be careful to see that there is no 'inconsistency' in such an assignment.

When the experiment is not repeatable, this method may be adopted for assigning probabilities to events. Since different persons may assign different probabilities, one cannot arrive at objective conclusions using probabilities assigned by this subjective method.

(e) AXIOMATIC DEFINITION

Incorporating all the essential properties of the probabilities as explained above, Kolmogorov (1933) gave the following axiomatic definition of probability. Let $\mathcal{A}$ be the σ -field of events associated with the outcomes of a certain experiment and Ω be the sure event. *Probability* is a real-valued set function $P(\cdot)$ defined on $\mathcal{A}$ such that for any $A, A_1, A_2, \dots \in \mathcal{A}$,

$$\left. \begin{array}{ll} \text{(i)} & P(A) \geq 0, \text{ (non-negative);} \\ \text{(ii)} & P(\Omega) = 1, \text{ (normed);} \\ \text{(iii)} & \text{if } A = \sum_{i=1}^n A_i, P(A) = \sum_{i=1}^n P(A_i) \text{ (finite additivity);} \\ \text{(iv)} & \text{if } A = \sum_{i=1}^{\infty} A_i, P(A) = \sum_{i=1}^{\infty} P(A_i) \text{ (\sigma-additivity);} \end{array} \right\} \quad (3.6)$$

This definition incorporates the convenience of the subjective method, usefulness of statistical 'definition' and is satisfactory intuitively. Of course, (iv) includes (iii) as a special case. Hence (iii) may be omitted, when (iv) is assumed.

Some probability-theorists assume that probability is a function possessing the properties (i)–(iii) only. Many assume that it possesses the properties (i), (ii) and (iv), and *a fortiori* (iii). If $\mathcal{A}$ contains only a finite number of sets, then it will be sufficient if $P(\cdot)$ satisfies (i)–(iii) only. In general, if $\mathcal{A}$ is only a field of events, then $P(\cdot)$ defined on $\mathcal{A}$ need be only finitely additive. If $\mathcal{A}$ is a σ -field, then $P(\cdot)$ has to be σ -additive. We shall elaborate this idea in the next section. In practice, to verify (iii), we need prove it for $n = 2$ only.

In this book we shall assume that probabilities have been assigned to various events by some method and it satisfies the assumptions (i)–(iv). Such a function will be called a *probability function*. Thus, the triplet $(\Omega, \mathcal{A}, P)$, where Ω is the space of outcomes, $\mathcal{A}$ is the σ -field of events

associated with an experiment and P is a probability function defined on $\mathcal{A}$, is assumed to be given to us. This triplet is called a *probability space*.

If a function μ defined on $(\Omega, \mathcal{A})$ possesses properties (i) and (iv) it is called a *measure*. If $\mu(\Omega)$ is finite, μ is said to be a *finite measure*. Thus, probability is a normed or scaled measure. All the general properties of measures will be shared by probability function. It possesses some additional properties. Later on we shall also have occasion to refer to these general measures, finite or infinite in section 3.6. If $\mu(\Omega) = \infty$, but $\Omega = \sum A_i$ such that $\mu(A_i) < \infty$, then μ is called a *σ -finite measure*. Measures, such as length, defined on $(R, \mathcal{B})$ are σ -finite. The triplet, $(\Omega, \mathcal{A}, \mu)$ is called the *measure space*.

Probability function P on $\mathcal{A}$ is also called *probability distribution* on Ω . It tells us how the unit probability mass is distributed among different subsets $\mathcal{A}$ of Ω .

3.2 SOME SIMPLE PROPERTIES

Assuming (3.6) we can establish a number of properties of $P(\cdot)$. Suppose $\{A_i\}$ ($i = 1, 2, \dots, n$) forms a partition of Ω and suppose $A_i \in \mathcal{A}$ and $P(A_i)$ are given. If $A = \sum_{i=1}^k A_i$ (say), then $A \in \mathcal{A}$ and using (iii), $PA = \sum_{i=1}^k PA_i^*$.

Thus, $P(A)$ is defined for all sets $A \in \sigma\{A_i\}$. If n is infinite, using (iv), PA is defined for every $A \in \sigma\{A_i\}$. If Ω is finite, $\Omega = \{\omega_1, \omega_2, \dots, \omega_N\}$, say, then the singletons $\{\omega_i\}$ ($i = 1, 2, \dots, N$) form a partition of Ω . Every subset of Ω is a disjoint union of sets of this partition. Hence probability can be defined on all subsets of Ω , if we know the probabilities of these singletons, $P\{\omega_i\}$ ($i = 1, 2, \dots, N$). If $P\{\omega_i\} = 1/N$, for all i , then we are in the 'equally likely' case defined in Section 3.1. In this case if $A \subset \Omega$ and A contains n points, then $P(A) = n/N$. If $P\{\omega_i\} = P_i$ ($i = 1, 2, \dots, N$) and A contains $\omega_{i_1}, \omega_{i_2}, \dots, \omega_{i_n}$ (say), then as in (3.2),

$$P(A) = \sum_{j=1}^n P_{i_j},$$

which is the generalisation of the formula of the calculation of probability to the 'not equally likely' case. The question of defining probabilities for a general Ω will be taken up in Section 3.4.

Since $A + A^c = \Omega$, from (iii) $P(\Omega) = P(A) + P(A^c) = 1$. Hence

$$P(A^c) = 1 - P(A). \quad (3.7)$$

If $A \subset B$, $B = A + BA^c$ and hence from additivity $P(B) = P(A) + P(BA^c)$. But, from (i), $P(BA^c) \geq 0$. Therefore,

$$A \subset B \Rightarrow P(B) \geq P(A), \quad (3.8)$$

establishing monotonicity of P . Thus, we have

$$P(BA^c) = P(B - A) = P(B) - P(A), \quad (3.9)$$

*Parentheses are omitted when there is no confusion.

is the formula for the probability of proper difference of two events. From (3.7), since

$$\begin{aligned}\phi &= \Omega^c. \\ P(\phi) &= 0.\end{aligned}\tag{3.10}$$

From (i) and (3.8), for all $A \in \mathcal{A}$, we have the basic property

$$0 \leq P(A) \leq 1.\tag{3.11}$$

If $A_i \in \mathcal{A}$ and are disjoint then $B = \sum_{i=1}^{\infty} A_i$ belongs to $\mathcal{A}$ and

$$B_n = \sum_{i=1}^n A_i \subset B.$$

From (iii),

$$\sum_{i=1}^n P A_i = P(B_n) \leq P(B) \leq 1.$$

$P A_i \geq 0$, being less than unity in value $\sum_{i=1}^{\infty} P A_i$ is absolutely convergent and $P B_n \uparrow P B$. Then, condition (iv) gives us a unique definition of probability of events of the form $\sum_{i=1}^{\infty} A_i$. Moreover, convergence of the series implies that

$$P(A_n) \rightarrow 0 \text{ as } n \rightarrow \infty.\tag{3.12}$$

In order to obtain the probability of the union of two or more events we see that

$$A \cup B = A + A^c B, B = AB + A^c B.$$

$$P(A \cup B) = P(A) + P(A^c B),\tag{3.13}$$

$$P(B) = P(AB) + P(A^c B),\tag{3.14}$$

$$\text{and } P(A \cup B) = P(A) + P(B) - P(AB).\tag{3.15}$$

Since $P(AB) \geq 0$, we have

$$P(A \cup B) \leq P(A) + P(B) \text{ (sub-additivity).}\tag{3.16}$$

In general by induction we can prove (How?) that for arbitrary events A_i ,

$$P\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n P(A_i), \text{ (Boole's Inequality).}\tag{3.17}$$

From Lemma 1.1, an arbitrary union of events $A_1, A_2, \dots, A_n$ can be converted into union of disjoint events $B_1, B_2, \dots, B_n$. Hence

$$P\left(\bigcup_{i=1}^n A_i\right) = P\left(\sum_{i=1}^n B_i\right) = \sum_{i=1}^n P B_i,\tag{3.18}$$

and from Corollary 1.1,

$$P\left(\bigcup_{i=1}^n A_i\right) = P(A_1) + P(A_1^c A_2) + P(A_1^c A_2^c A_3) + \dots + P(A_1^c \dots A_{n-1}^c A_n).\tag{3.19}$$

Now

$$A_1^c A_2 = (\Omega - A_1) A_2 = A_2 - A_1 A_2;$$

and hence from (3.9)

$$P(A_1^c A_2) = P(A_2) - P(A_1 A_2),$$

and similarly

$$P(A_2^c A_3) = P(A_3) - P(A_2 A_3),$$

$$A_1^c A_2^c A_3 = (\Omega - A_1) A_2^c A_3 = A_2^c A_3 - A_1 (A_2^c A_3),$$

and hence from (3.9)

$$\begin{aligned} P(A_1^c A_2^c A_3) &= P(A_3) - P(A_2 A_3) - P(A_1 (A_3 - A_2 A_3)), \\ &= P(A_3) - P(A_2 A_3) - P(A_1 A_3) + P(A_1 A_2 A_3). \end{aligned}$$

$$\begin{aligned} P(A_1 \cup A_2 \cup A_3) &= P(A_1) + P(A_2) + P(A_3) - P(A_1 A_2) \\ &\quad - P(A_1 A_3) - P(A_2 A_3) + P(A_1 A_2 A_3). \end{aligned} \quad (3.20)$$

Now

$$P(A_2^c A_3^c A_4) = P(A_4) - P(A_3 A_4) - P(A_2 A_4) + P(A_2 A_3 A_4),$$

$$\begin{aligned} P(A_1^c A_2^c A_3^c A_4) &= P(A_4^c A_5^c A_4) - P(A_1 A_2^c A_3^c A_4), \\ &= P(A_4) - P(A_3 A_4) - P(A_2 A_4) + P(A_2 A_3 A_4) \\ &\quad - [P(A_1 A_4) - P(A_1 A_3 A_4) - P(A_1 A_2 A_4) + P(A_1 A_2 A_3 A_4)], \end{aligned}$$

and hence

$$\begin{aligned} &P(A_1 \cup A_2 \cup A_3 \cup A_4) \\ &= \sum_{i=1}^4 P(A_i) - \sum_{i<j} P(A_i A_j) + \sum_{i<j<k} P(A_i A_j A_k) - P(A_1 A_2 A_3 A_4). \end{aligned} \quad (3.21)$$

Continuing in this manner, in general we have

$$P\left(\bigcup_{i=1}^n A_i\right) = S_1 - S_2 + S_3 - \dots + (-1)^{n-1} S_n, \quad (3.22)$$

where

$$S_1 = \sum P(A_i), \quad S_2 = \sum_{i<j=1}^n P(A_i A_j), \quad S_3 = \sum_{i<j<k=1}^n P(A_i A_j A_k), \text{ etc.} \quad (3.23)$$

If $\{A_i\}$ is a partition of Ω , $B = \sum_i B A_i$ and hence

$$PB = \sum_i P(B A_i). \quad (3.24)$$

Here $\{A_i\}$ may be a finite or a countable partition.

A number of examples involving the calculation of probabilities of events are worked out in the standard books on probability theory. A few are given in the next section.

An event A such that $PA = 1$ is called an *almost sure* event. Then its complement, A^c , has probability zero. A^c is called a *null event*. If a certain phenomenon can happen with probability one, it is said to happen *almost surely* (a.s.). For example, while tossing a coin, if p is the probability of the occurrence of a head and $(1-p)$ is the probability

of the occurrence of a tail, the probability that the coin will fall on its edge (say) is zero. Hence the latter is a null event. If a certain property can hold for all ω except those belonging to a null set, it is said to hold almost surely or for *almost all* (a.a.) ω . In the case of general measures it is said to hold *almost everywhere* (a.e.).

3.3 DISCRETE PROBABILITY SPACE

(a) DEFINITION

In a sample space Ω , if the class of events $\mathcal{A}$ is generated by a countable partition of subsets of Ω then $(\Omega, \mathcal{A}, P)$ is called the *discrete probability space*. If $\mathcal{A} = \sigma\{A_i\}$, where $\{A_i\}$ is a partition of Ω , any event A is a countable union of $\{A_i\}$. Hence the probability of A can be determined knowing the probabilities p_i of A_i . If Ω is countable, A_i could be singletons $\{\omega_i\}$ ($i = 1, 2, \dots$). $\{p_1, p_2, \dots\}$ is called the probability distribution on Ω . Then $\mathcal{A}$ coincides with the power set.

If Ω is finite, and if $P\{\omega_i\}$ is the same for all $i = 1, 2, \dots, N$ (say), then this corresponds to the equally likely case with $(1/N, 1/N, \dots, 1/N)$ as the probability distribution. In the latter case, to determine the probability of any event A , it will be enough if we calculate n_A the number of ω 's in A and N , the number of ω 's in Ω . Here n_A is called the number of outcomes *favourable* to A , N the total number of outcomes and $P(A) = n_A/N$. Detailed study of such sample spaces is made by Feller (Vol. I, 1968). In this section we shall give a few examples of discrete, countable (finite or infinite) probability spaces.

(b) FINITE PROBABILITY SPACE (Equally likely outcomes)

An important tool for calculating the probabilities of events in the finite probability space is the theory of permutations and combinations. It is well known that the number of permutations of n different things taken r at a time is $(n)_r = n(n-1)\dots(n-r+1)$ and the number of combinations is $\binom{n}{r} = n! / [(n-r)! r!]$.

Example 3.2. (Binomial Distribution) Suppose n coins are tossed once or one coin is tossed n times. In both the cases, the space of outcomes Ω can be represented by the set of all n -tuples of the form $(HHTHTT\dots)$. Since each of the components may be either H or T , there are $N = 2^n$ such n -tuples. If all these have the same probability, viz.,

$$p_i = 2^{-n}, (i = 1, 2, \dots, N),$$

we can calculate the probabilities of events of the form

$A = (k \text{ heads and } n - k \text{ tails turn up}),$

$B = (\text{at least one head turns up}), \text{ etc.}$

The number of outcomes favourable to A are $\binom{n}{k}$ and those favourable to B are all n -tuples except the point $(TT \dots T)$. Hence

$$P(A) = \binom{n}{k} 2^{-n}, \quad P(B) = 1 - 2^{-n}.$$

Suppose we define the probability of the sample point $\omega_i = (HTH \dots)$ containing k , H 's and $(n - k)$, T 's to be

$$p_i = p^k q^{n-k} (q = (1 - p)), \quad (i = 1, 2, \dots, N). \quad (3.25)$$

We are no longer in the 'equally likely' case. Now

$P(B) = 1 - P\{TT \dots T\} = 1 - q^n$ and $P(A) = p^k q^{n-k} \binom{n}{k}$, since all the $\binom{n}{k}$ points favourable to A have the same probability $p^k q^{n-k}$. The probabilities of events such as {the no. of heads exceeds the no. of tails} can be calculated.

On Ω , suppose we define a r.v. X , denoting 'the number of heads turning up', X is a mapping from Ω to $\{0, 1, 2, \dots, n\} = \Omega'$. In the general case (3.25),

$$P[X = k] = \binom{n}{k} p^k q^{n-k} \quad (k = 0, 1, 2, \dots, n). \quad (3.26)$$

Suppose we associate with the point ' k ' of Ω' , the above probability (3.26). This set of probabilities on Ω' represents a 'distribution' of probability 'mass' on Ω' , the range space of X , called the *probability distribution of X* . Since (3.26) is the k th term in the expansion of the binomial $(q + p)^n$, it is called the *binomial distribution* and X is called the *binomial r.v.*

Example 3.3. (Hypergeometric Distribution) From a lot containing n items suppose r items are chosen. This may be done in $\binom{n}{r}$ ways. Suppose there are n_1 defective items in the lot. The number of defectives among the r items chosen may be $k (= 0, 1, 2, \dots, n_1)$. Assuming that all the $\binom{n}{r}$ outcomes are equally likely, the number of outcomes favourable to the event ' k defectives in the sample' is $\binom{n_1}{k} \binom{n_2}{r-k}$, where $n_2 = n - n_1$ is the number of non-defective items. This may be seen from the fact that k defectives can be chosen from n_1 defectives in $\binom{n_1}{k}$ ways and the remaining $(r - k)$ non-defectives can be chosen

from the n_2 non-defectives in $\binom{n_2}{r-k}$ ways. Hence

$$p_k = \frac{\binom{n_1}{k} \binom{n_2}{r-k}}{\binom{n}{r}}, k = 0, 1, 2, \dots, \min(n_1, r). \quad (3.27)$$

$\{p_k\}$ represents a probability distribution on the range space, $\Omega' = \{0, 1, \dots, \min(n_1, r)\}$ of X , representing the number of defectives observed in a sample of size r chosen from a lot of size n containing n_1 defectives. This distribution is called the *hypergeometric distribution* and X , *hypergeometric r.v.*

Example 3.4. (Multinomial Distribution) Suppose there are n cells and r balls each one of which is to be placed in one of these n cells. Total number of outcomes is $N = n^r$. If all these are equally likely, the probability $p(r_1, r_2, \dots, r_n)$ of the event that the n cells contain $r_1, r_2, \dots, r_n$ ($\sum r_i = r$) balls respectively can be obtained as

$$p(r_1, \dots, r_2, \dots, r_n) = \frac{r!}{r_1! \dots r_n!} \frac{1}{n^r} \quad (3.28)$$

by noting that the number of outcomes favourable to the event is the product of the number of combinations of r_1 balls out of a total of r balls in the first cell, r_2 balls out of the remaining $r - r_1$ balls in the second cell and r_n balls out of the remaining $r - r_1 - r_2$ balls in the third cell, and so on. But

$$\binom{r}{r_1} \binom{r-r_1}{r_2} \dots \binom{r-r_1-\dots-r_{n-2}}{r_{n-1}} = \frac{r!}{r_1! r_2! \dots r_n!} \quad (3.29)$$

Hence we get (3.28).

With every outcome of the experiment, we may associate an n -dimensional r.v. $\mathbf{X} = (X_1, X_2, \dots, X_n)$, where X_i denotes the number of balls in the i th cell ($i = 1, 2, \dots, n$). In the above probability space, (3.28) gives us the probabilities of the events,

$$[X_1 = r_1, X_2 = r_2, \dots, X_n = r_n] = [\mathbf{X} = \mathbf{r}],$$

for different vectors $\mathbf{r}$, such that $r_i \geq 0$, $\sum r_i = r$. The set of all such vectors $\mathbf{r}$ denotes the range space of $\mathbf{X}$. By assigning probabilities

$$P[\mathbf{X} = \mathbf{r}] = \frac{r!}{r_1! \dots r_n!} p_1^{r_1} \dots p_n^{r_n}, (\sum p_i = 1) \quad (3.30)$$

on this range space of $\mathbf{X}$, we get a distribution called the *multinomial distribution*. It can be seen that (3.30) is a term in the expansion of the multinomial $(p_1 + p_2 + \dots + p_n)^r$.

Example 3.5. (Classical occupancy problem). Suppose r balls are to be placed in n cells so that each ball may occupy any one of the cells. If the balls are distinguishable, the number of possible arrangements is n^r .

If all these are equally likely, the probability that a particular cell is empty is $(n-1)^r/n^r$. Similarly the probability that two pre assigned cells are empty is $(n-2)^r/n^r$. Let $p_m(r, n)$ be the probability that exactly m cells are empty. To calculate $p_m(r, n)$ we may proceed as follows:

$$\begin{aligned} p_0(r, n) &= 1 - P(\text{at least one of the cells is empty}), \\ &= 1 - P\left(\bigcup_{k=1}^n A_k\right) \end{aligned} \quad (3.31)$$

where A_k is the event that the cell marked k is empty. But from (3.22)

$$P\left(\bigcup_{k=1}^n A_k\right) = \sum_{k=1}^n P A_k - \sum_{j < k} P(A_j A_k) + \sum_{j < k < l} P(A_j A_k A_l) - \dots \quad (3.32)$$

Since $P(A_k) = [(n-1)/n]^r$, $P(A_j A_k) = [(n-2)/n]^r$, etc.,

$$p_0(r, n) = 1 - n\left(1 - \frac{1}{n}\right)^r + \binom{n}{2}\left(1 - \frac{2}{n}\right)^r - \dots + (-1)^k \binom{n}{k} \left(1 - \frac{k}{n}\right)^r + \dots \quad (3.33)$$

Then $n^r p_0(r, n)$ is the number of arrangements with no cell empty.

To determine $p_m(r, n)$ we note that any m of the n cells may be empty, which can happen in $\binom{n}{m}$ ways. Since each of the remaining $(n-m)$ cells is not empty, the number of such arrangements is $(n-m)^r p_0(r, n-m)$. Hence

$$\begin{aligned} p_m(r, n) &= \binom{n}{m} (n-m)^r p_0(r, n-m) n^{-r}, \\ &= \binom{n}{m} \sum_{k=0}^{n-m} (-1)^k \binom{n-m}{k} \left(1 - \frac{m+k}{n}\right)^r. \end{aligned} \quad (3.34)$$

(c) COUNTABLE PROBABILITY SPACE

Suppose $\Omega = \{\omega_1, \omega_2, \dots\}$ and $\mathcal{A}$ is the σ -field of all subsets of Ω . It is generated by all the singletons $\{\omega_i\}$. Suppose $P(\{\omega_i\})$ is specified for all i and $P(\{\omega_i\}) = p_i$. Evidently $p_i \geq 0$, and $\sum p_i = 1$. $P(A)$ is the sum of the probabilities of all ω contained in A . It is impossible that all the sample points are equally likely. Hence we have to view the probability distribution $\{p_i\}$ to be the distribution of unit probability mass among the sample points following a certain rule or law. This law is sometimes referred to as the *probability distribution* or *probability law*.

Example 3.6. Suppose $\Omega = \{0, 1, 2, \dots\}$ and

$$P(\{x\}) = (1-\theta)\theta^x, \quad (x=0, 1, 2, \dots), \quad (0 < \theta < 1). \quad (3.35)$$

Then $p_x = P(\{x\}) \geq 0$, $\sum_{x \in \Omega} p_x = 1$, and on every sub-set A of Ω

$$P(A) = \sum_{x \in A} p_x.$$

Thus, $(\Omega, \mathcal{A}, P)$, where $\mathcal{A}$ is the class of all subsets of Ω , is a probability space. This distribution is called a *geometric distribution*.

Example 3.7. Once again let $\Omega = \{0, 1, 2, \dots\}$ and $\mathcal{A}$ be the class of all subsets of Ω . If P specifies that

$$P(\{x\}) = \frac{e^{-\lambda} \lambda^x}{x!} (\lambda > 0), (x = 0, 1, 2, \dots), \quad (3.36)$$

then (3.36) specifies a *Poisson distribution*.

Let X be a mapping from Ω to $\{0, 1, 2, \dots\}$ such that

$$X^{-1}\{x\} = [X(\omega) = x] = A_x.$$

$\{A_x\}$ forms a partition of Ω , which generates $\mathcal{A}$. Let $P[X=x] = P(A_x) = p_x$, say. Then the probability distribution on $\mathcal{A}$ is specified by the probabilities of the events $[X=x]$. If $P[X=x] = (e^{-\lambda} \lambda^x)/x!$, as given by (3.36), then X is said to be a *Poisson r.v.* If $P[X=x] = (1-\theta)\theta^x$, as given by (3.35), then X is said to be a *geometric r.v.* If

$$P[X=x] = \binom{n+x-1}{n-1} p^n q^x, \quad (0 < p < 1, q = 1-p), \quad (3.37)$$

then X is called a *negative binomial r.v.*, and the corresponding probability distribution is called the *negative binomial distribution*.

To find the probability of the event, $[X = \text{odd number}]$, we note that,

$$P(X = \text{odd}) = \sum_{x=\text{odd}} p_x$$

and this probability equals $e^{-\lambda} \sum_{n=0}^{\infty} \frac{\lambda^{2n+1}}{(2n+1)!}$, for a Poisson random

variable and equals $(1-\theta) \sum_{n=0}^{\infty} \theta^{2n+1} = \frac{\theta}{1+\theta}$, for a geometric random variable. Similarly the probabilities of the events $[X \geq x]$, $[X > \theta]$, $[X = \text{even}]$, when the probability distribution is specified by (3.36), (3.37) or (3.35), can be calculated. (Obtain!).

A random variable (or vector) X taking only a countable number (finite or infinite) of distinct values $(x_0, x_1, x_2, \dots)$, say, is called a *discrete r.v.* The σ -field induced by X is generated by the countable partition $\{[X = x_i], i = 0, 1, 2, \dots\}$ and the probability distribution on this σ -field is completely specified, if $P[X = x_i]$ are given for all i . Binomial, Poisson, multinomial and geometric random variables are examples of discrete r.v.'s.

3.4 GENERAL PROBABILITY SPACE

(a) CONTINUITY PROPERTY

Let Ω be the space of all possible outcomes of an experiment under

consideration. The number of outcomes may be uncountable. Let $\mathcal{A}$ be the σ -field of events and P a probability function defined on $\mathcal{A}$ satisfying the conditions (i)–(iv) of (3.6).

LEMMA 3.1: In the presence of (iii), (iv) is equivalent to the following:

$$(iv)': B_n \in \mathcal{A}, B_n \uparrow B \Rightarrow P(B_n) \uparrow P(B) \text{ as } n \rightarrow \infty. \quad (3.38)$$

Proof: Let $A_n = B_n - B_{n-1}$ so that A_n 's are disjoint and

$$B_n = \sum_{i=1}^n A_i. \text{ Then } B_n \uparrow \sum_{i=1}^{\infty} A_i = B \text{ as } n \rightarrow \infty.$$

$$\text{But } P(B_n) = \sum_{i=1}^n P(A_i) \uparrow \sum_{i=1}^{\infty} P(A_i).$$

If (iv)' holds, $B_n \uparrow B \Rightarrow P(B_n) \rightarrow P(B)$ and hence taking

$$B_n = \sum_{i=1}^n A_i, \sum_{i=1}^{\infty} P(A_i) = P\left(\sum_{i=1}^{\infty} A_i\right)$$

and (iv) follows. Conversely, let (iv) hold. Since

$$P(B) = \sum_{i=1}^{\infty} P(A_i) = \lim_n \sum_{i=1}^n P(A_i) = \lim_n P(B_n),$$

(iv)' follows (*).

COROLLARY 3.1: If $A_n \in \mathcal{A}$, $A_n \downarrow A$, then

$$P(A_n) \downarrow P(A). \quad (3.39)$$

The Corollary follows because,

$$\begin{aligned} A_n \downarrow A &\Rightarrow A_n^c \uparrow A^c \Rightarrow P(A_n^c) \rightarrow P(A^c) \Rightarrow \{1 - P(A_n)\} \\ &\rightarrow \{1 - P(A)\} \Rightarrow P(A_n) \rightarrow P(A). \end{aligned}$$

The Lemma 3.1 and its corollary imply that probability is continuous from below and from above. In fact, from the following Corollary 3.2 we see that it is continuous.

COROLLARY 3.2: If $A_n \rightarrow A$, then $P(A_n) \rightarrow P(A)$. (3.40)

$$\text{Proof: } \bigcap_{k \geq n} A_k \subset A_n \subset \bigcup_{k \geq n} A_k \Rightarrow P\left(\bigcap_{k \geq n} A_k\right) \leq P(A_n) \leq P\left(\bigcup_{k \geq n} A_k\right) \quad (3.41)$$

$$\text{But } \bigcap_{k \geq n} A_k \uparrow \lim_{k \geq n} A_n \text{ and } \bigcup_{k \geq n} A_k \downarrow \lim_{k \geq n} A_n.$$

$$\text{Since } A_n \rightarrow A, \lim_{k \geq n} A_n = \overline{\lim}_{k \geq n} A_n = A.$$

$$\text{From Lemma 3.1, } P\left(\bigcap_{k \geq n} A_k\right) \uparrow P(A)$$

and from Corollary 3.1, $P\left(\bigcup_{k \geq n} A_k\right) \downarrow P(A)$. Hence from 3.41, $P(A_n) \rightarrow P(A)$. (*)

Thus, σ -additivity condition (iv) is sometimes referred to as the continuity condition of the probability function. The notions of continuity from below, continuity from above and continuity in general, are similar to the notions of continuity from the left, from the right and in the absolute sense for point functions.

Even though Lemma 3.1, Corollaries 3.1, 3.2 are proved for probability measures, results contained therein continue to hold for finite measures with obvious modifications (which?) in the proof. If the measure is σ -finite, it will be continuous from above but we cannot establish continuity from below.

A set function satisfying (iv) only, defined on a σ -field, coincides with the difference of two measures. Thus, by relaxing one or more of the conditions 3.6 we get various generalisations of probability measure (see Appendix B).

(b) EXTENSION OF PROBABILITY MEASURE

In the case of discrete probability space, we have seen that if we define probabilities of the events $A_i (i = 1, 2, \dots)$ forming a partition of Ω and which generate $\mathcal{A}$, we can obtain the probabilities of all the sets of $\mathcal{A}$. Now the question arises whether we can do the same if $\mathcal{A}$ is generated by an arbitrary class $\mathcal{F}$ of subsets of Ω , and $P(\cdot)$ is defined for every set of $\mathcal{F}$.

Evidently the sets of $\mathcal{A}$ are obtained by complementation and countable operations of unions (or intersections) on sets of $\mathcal{F}$. Axiom (iii) defines probability on all finite disjoint unions of sets of $\mathcal{F}$, if it is defined on sets of $\mathcal{F}$; given a countable (or finite) union of sets of $\mathcal{F}$ one can express it as the union of disjoint sets of $\mathcal{F}$, only if $\mathcal{F}$ is a field. Suppose we assume this to be the case. There may be more than one way of expressing a countable union as the union of disjoint sets. The question arises whether it is possible to get a unique value for $P(\bigcup_{i=1}^{\infty} A_i)$, ($A_i \in \mathcal{F}$) for $\bigcup_i A_i \in \sigma(\mathcal{F})$ but $\bigcup_i A_i \notin \mathcal{F}$, given $P(\cdot)$ for all sets of $\mathcal{F}$. The answer is in the affirmative on account of the Caratheodory extension theorem given below.

Suppose P is defined on a certain field $\mathcal{F}$. $\tilde{P}$, defined on $\sigma(\mathcal{F})$ is such that on sets of $\mathcal{F}$, $\tilde{P}$ coincides with P . Then $\tilde{P}$ is called an *extension* of P to $\sigma(\mathcal{F})$ and P is called the restriction of $\tilde{P}$ to $\mathcal{F}$. $\tilde{P}$ has a wider domain than P and the two coincide on the common domain. Proof of the following theorem can be found in any book on measure theory (see Halmos, P.R. (1958), Billingsley (1979) or Appendix A).

THEOREM 3.1. (Caratheodory Extension Theorem) *Probability defined on a field $\mathcal{F}$ can be extended uniquely to the minimal σ -field containing $\mathcal{F}$.*

An important measurable space is $(R, \mathcal{B})$, where $\mathcal{B}$ is the Borel field of subsets of the real line R . $\mathcal{B}$ is the minimal σ -field containing the class of all intervals of the form $(a, b]$, $a, b \in R$, say. But this latter class is closed under finite intersections, but it is not a field. So, it is not clear from the above Theorem 3.1, whether, probability function defined on all such intervals $(a, b]$, can be extended uniquely to all Borel sets. However, the answer is still in the affirmative from the following Lemma.

LEMMA 3.2: *Probability function defined on all intervals of the form $(a, b]$, defines uniquely an extension to the minimal field containing all the intervals.*

Proof: Let $\mathcal{B}_I$ be the class of all intervals of the form $(a, b]$, and $\mathcal{B}_0$ be the minimal field containing $\mathcal{B}_I$. It contains all finite sums of sets of $\mathcal{B}_I$ and complements of sets of $\mathcal{B}_I$, i.e., sets of the form $(-\infty, a] \cup (b, \infty)$.

Let $A \in \mathcal{B}_0$ and $A = \sum_{k=1}^m A_k$, $A_k \in \mathcal{B}_I$. Then $P(A) = \sum_k P A_k$, (by (iii)).

If A is also equal to $\sum_{j=1}^n A'_j$, $A'_j \in \mathcal{B}_I$, then $P(A) = \sum_j P A'_j$. We have to prove that $\sum_k P A_k = \sum_j P A'_j$.

Since P on $\mathcal{B}_I$ is additive, and

$$A'_j = A A'_j = \sum_k A_k A'_j,$$

$$A_k = A A_k = \sum_j A_k A'_j,$$

it follows that

$$\begin{aligned} \sum_j P(A'_j) &= \sum_j P\left(\sum_k (A_k A'_j)\right) = \sum_{j,k} P(A_k A'_j) \\ &= \sum_{k,j} P(A'_j A_k) = \sum_k P\left(\sum_j A'_j A_k\right) = \sum_k P(A_k). \end{aligned}$$

Hence we have the Lemma, since $P(A^c) = 1 - PA$, for all $A \in \mathcal{B}_I^*$.

From the Lemma, defining probabilities on all sets of the form $(a, b]$, one can extend it uniquely to $\mathcal{B}_0$, the minimal field containing such intervals and using Theorem 3.1, it may be extended further to all Borel sets.

In Lemma 3.2, we may take the class of all intervals of the form $(-\infty, a]$, $a \in R$, instead of the class of all intervals of the form $(a, b]$, or any class $\mathcal{C}$ which generates the Borel field. Defining probability on the sets of this class, we can extend it to all Borel sets. If $F(x) = P(-\infty, x]$, F is called the *probability distribution function* or the *distribution function* on R .

Example 3.8. Suppose $\Omega = \{\omega; 0 \leq \omega \leq 1\}$. Now Ω is an uncountable space. Suppose the σ -field of events is the Borel field restricted to subsets of $[0, 1]$. Thus, $\mathcal{A} = \{B \cap [0, 1]; B \in \mathcal{B}\}$, the smallest σ -field

An important measurable space is $(R, \mathcal{B})$, where $\mathcal{B}$ is the Borel field of subsets of the real line R . $\mathcal{B}$ is the minimal σ -field containing the class of all intervals of the form $(a, b]$, $a, b \in R$, say. But this latter class is closed under finite intersections, but it is not a field. So, it is not clear from the above Theorem 3.1, whether, probability function defined on all such intervals $(a, b]$, can be extended uniquely to all Borel sets. However, the answer is still in the affirmative from the following Lemma.

LEMMA 3.2: *Probability function defined on all intervals of the form $(a, b]$, defines uniquely an extension to the minimal field containing all the intervals.*

Proof: Let $\mathcal{B}_I$ be the class of all intervals of the form $(a, b]$, and $\mathcal{B}_0$ be the minimal field containing $\mathcal{B}_I$. It contains all finite sums of sets of $\mathcal{B}_I$ and complements of sets of $\mathcal{B}_I$, i.e., sets of the form $(-\infty, a] \cup (b, \infty)$.

Let $A \in \mathcal{B}_0$ and $A = \sum_{k=1}^m A_k$, $A_k \in \mathcal{B}_I$. Then $P(A) = \sum_k P(A_k)$, (by (iii)).

If A is also equal to $\sum_{j=1}^n A'_j$, $A'_j \in \mathcal{B}_I$, then $P(A) = \sum_j P(A'_j)$. We have to prove that $\sum_k P(A_k) = \sum_j P(A'_j)$.

Since P on $\mathcal{B}_I$ is additive, and

$$A'_j = AA'_j = \sum_k A_k A'_j,$$

$$A_k = AA_k = \sum_j A_k A'_j,$$

it follows that

$$\begin{aligned} \sum_j P(A'_j) &= \sum_j P\left(\sum_k (A_k A'_j)\right) = \sum_{j,k} P(A_k A'_j) \\ &= \sum_{k,j} P(A'_j A_k) = \sum_k P\left(\sum_j A'_j A_k\right) = \sum_k P(A_k). \end{aligned}$$

Hence we have the Lemma, since $P(A^c) = 1 - PA$, for all $A \in \mathcal{B}_I$.

From the Lemma, defining probabilities on all sets of the form $(a, b]$, one can extend it uniquely to $\mathcal{B}_0$, the minimal field containing such intervals and using Theorem 3.1, it may be extended further to all Borel sets.

In Lemma 3.2, we may take the class of all intervals of the form $(-\infty, a]$, $a \in R$, instead of the class of all intervals of the form $(a, b]$, or any class $\mathcal{C}$ which generates the Borel field. Defining probability on the sets of this class, we can extend it to all Borel sets. If $F(x) = P(-\infty, x]$, F is called the *probability distribution function* or the *distribution function* on R .

Example 3.8. Suppose $\Omega = \{\omega; 0 \leq \omega \leq 1\}$. Now Ω is an uncountable space. Suppose the σ -field of events is the Borel field restricted to subsets of $[0, 1]$. Thus, $\mathcal{A} = \{B \cap [0, 1]; B \in \mathcal{B}\}$, the smallest σ -field

generated by the class of subsets of the form $[0, x)$, $(0 < x \leq 1)$. We define

$$P[0, x) = x, (0 < x \leq 1), \quad (3.42)$$

and $P(\Omega) = 1$. This probability function has a unique extension to $\mathcal{A}$, which is called the *uniform* or the *rectangular distribution* on $[0, 1]$.

Now $P[0, 1/n) = 1/n$, and $[0, 1/n) \rightarrow \{0\}$ as $n \rightarrow \infty$. By the continuity property of P -function, $P[0, 1/n) \rightarrow P\{0\} = 0$. In general, we can prove that $P\{x\} = 0$, whatever be the singleton $\{x\}$. If A is a countable union of singletons, each one of which has probability zero, then $P(A) = 0$.

Let $\Omega = \{\omega: 0 \leq \omega \leq a\}$ and $\mathcal{A}$ be the Borel field restricted to subsets of $[0, a]$. Defining

$$P([0, x)) = x/a, (0 < x \leq a), \quad (3.43)$$

and $P(\Omega) = 1$, and P can be extended to $\mathcal{A}$ in a unique manner. P is called the *rectangular distribution* on $[0, a]$.

Example 3.9. Suppose $\Omega = (-\infty, \infty)$ and $\mathcal{A}$ is the Borel field. Define

$$F(x) = P(-\infty, x] = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt, \quad -\infty < x < \infty, \quad (3.44)$$

with $P(R) = 1$. This function P defined on intervals of the form $(-\infty, x]$, can be extended to all Borel sets in a unique manner. Hence $(R, \mathcal{B}, P)$ is a probability space. The probability distribution on R is called the *normal distribution*.

In (3.44), F is defined as the Riemann integral of a non-negative function $f(x)$, whose integral over R is equal to unity, $f(x)$ denotes the density of the probability mass at a point x and is hence called the *probability density function*. If the probability density function is

$$f(x) = \begin{cases} 1/a, & 0 \leq x \leq a, \\ 0, & \text{otherwise} \end{cases}$$

we get the *rectangular distribution* on $[0, a]$. If

$$f(x) = \begin{cases} (\Gamma_p)^{-1} e^{-x} x^{p-1}, & x \geq 0, \\ 0, & \text{otherwise,} \end{cases} \quad (3.45)$$

we have the *gamma distribution*. Taking different functions which are non-negative and whose integral over $(-\infty, \infty)$ or its subset is unity, we get different probability distributions.

Example 3.10. Let $\Omega = R^2 = \{(x, y) : -\infty < x, y < \infty\}$ and $\mathcal{A}$ be the Borel field of subsets of R^2 . Consider the rectangles of the form

$$\{(x, y) : -\infty < x \leq a, -\infty < y \leq b\}, a, b \in R.$$

Since these rectangles generate $\mathcal{B}_2$, and they are closed under finite intersections, defining probabilities on these rectangles, one can extend them to all Borel sets of R^2 . Let $f(x, y)$ be a non-negative function of (x, y) such that its integral over R^2 is unity. Defining

$$F(a, b) = P[-\infty < x \leq a, -\infty < y \leq b] = \int_{-\infty}^a \int_{-\infty}^b f(x, y) dy dx, \quad (3.46)$$

we get a probability distribution on R^2 defined by the probability density function $f(x, y)$.

If $\Omega = R^k$ and $\mathcal{B}_k$ is the Borel field in R^k , P may be defined by means of a function $f(x_1, x_2, \dots, x_k)$ with $f(x_1, \dots, x_k) \geq 0$ and

$$\int \dots \int_{R^k} f(x_1 \dots x_k) dx_1 \dots dx_k = 1, \quad (3.47)$$

such that

$$\begin{aligned} F(a_1, a_2, \dots, a_k) &= P[-\infty < x_1 \leq a_1, \dots, -\infty < x_k \leq a_k] \\ &= \int_{-\infty}^{a_1} \int_{-\infty}^{a_2} \dots \int_{-\infty}^{a_k} f(x_1 \dots x_k) dx_k \dots dx_1 \end{aligned} \quad (3.48)$$

$f(x_1 \dots x_k)$ is the probability density function on R^k .

If a probability density function exists for a certain distribution, then it is generally called an *absolutely continuous distribution* (or simply *continuous distribution*). We would like to emphasize here that we do not restrict ourselves to discrete or absolutely continuous distributions only. We also consider cases when the probability on $(R, \mathcal{B})$ is induced by a weighted sum of distributions, one of which may be discrete and the other, absolutely continuous.

3.5 INDUCED PROBABILITY SPACE

(a) PROBABILITY SPACE INDUCED BY X

Let X be a r.v. on $(\Omega, \mathcal{A}, P)$. For any $B \in \mathcal{B}$, we may define

$$\begin{aligned} P_X(B) &= P\{\omega : X(\omega) \in B\}, \\ &= P[X^{-1}(B)]. \end{aligned} \quad (3.49)$$

Then $P_X(B) \geq 0$, and

$$\begin{aligned} P_X(\sum_j B_j) &= P(X^{-1}(\sum_j B_j)), \\ &= P \sum_j X^{-1}(B_j), \\ &= \sum_j P(X^{-1}(B_j)), \\ &= \sum_j P_X(B_j), \end{aligned} \quad (3.50)$$

Since $P_X(R) = 1$, P_X on $(R, \mathcal{B})$ is a probability measure. Hence $(R, \mathcal{B}, P_X)$ is a new probability space, called the *space induced by X on its range space*. P_X is called the *probability distribution of X* .

In Example 3.2, we consider the space of all outcomes of n tosses of a coin. We define a r.v. X called the 'number of heads turning up',

taking the values $0, 1, 2, \dots, n$ with probabilities

$$P[X = k] = \binom{n}{k} p^k q^{n-k}, \quad (k = 0, 1, 2, \dots, n).$$

Denoting $\{0, 1, 2, \dots, n\}$ by Ω , the above defines probabilities of all subsets of Ω .

Moreover, if g is any Borel function of X ,

$$\begin{aligned} P[g(X(\omega)) \in B] &= P[X(\omega) \in g^{-1}(B)] \\ &= P_X(g^{-1}(B)). \end{aligned} \quad (3.51)$$

Since g is a Borel function $g^{-1}(B)$ is a Borel set. Hence the probability distribution of X determines uniquely the probability distribution of any Borel function of X . In the above example,

$$\begin{aligned} P[2X = k] &= P[X = k/2] = 0, \text{ if } k \text{ is odd,} \\ &= \binom{n}{k/2} p^{k/2} q^{n-k/2}, \text{ if } k \text{ is even.} \end{aligned}$$

Suppose

$$P[X \leq x] = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \exp(-t^2/2) dt, \quad \forall x \in R. \quad (3.52)$$

Then the probability measure P_X induced by X has the normal distribution on R . Then X is said to be a *normal r.v.*

In general, if the induced probability measure has a certain probability distribution, the r.v. itself is said to have that distribution. If P_X has the gamma distribution given by (3.45), X is called a *gamma r.v.* The probability density function of P_X is also called the probability density of the r.v. X . This is also true when X is a random vector, with more than one component.

(b) DISTRIBUTION OF BOREL FUNCTIONS OF RANDOM VARIABLES

Now

$$\begin{aligned} P[X^2 \leq k] &= P[-\sqrt{k} \leq X \leq +\sqrt{k}], \\ P[a \leq X^2 \leq b] &= P[(-\sqrt{b} \leq x \leq -\sqrt{a}) \cup (+\sqrt{a} \leq x \leq +\sqrt{b})] \end{aligned} \quad (3.53)$$

hence we can determine the distribution of X^2 . Similarly, we can determine the distributions of e^X , $\log X$ and other one-valued functions $Y = g(X)$. In fact, if g is increasing,

$$P[g(X) \leq y] = P[X \leq g^{-1}(y)]. \quad (3.54)$$

If $f(x)$ is the probability density function (p.d.f.) of X and if g^{-1} exists g , g^{-1} are differentiable, using the calculus of transformation of variables,

$$\begin{aligned} P[g(x_1) \leq g(X) \leq g(x_2)] &= \int_{x_1}^{x_2} f(x) dx \\ &= \int_{y_1}^{y_2} f(h(y)) h'(y) dy, \end{aligned} \quad (3.55)$$

where $y_1 = g(x_1)$, $y_2 = g(x_2)$, $g^{-1}(y) = x = h(y)$. Using the technique sometimes we can obtain the distribution function of $g(X)$.

Example 3.11. If $\log Y = X$ is normal with probability density function

$$\frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2}x^2 \right\}, \quad -\infty < x < \infty,$$

Y has *log-normal distribution* with probability density

$$\frac{1}{y\sqrt{2\pi}} \exp \left\{ -\frac{1}{2}(\log y)^2 \right\}, \quad (0 < y < \infty).$$

Example 3.12. If the probability density function of (X, Y) is

$$f(x, y) = \begin{cases} 1, & 0 \leq x, y \leq 1 \\ 0, & \text{otherwise,} \end{cases}$$

we can obtain the probability distribution of $(X + Y)$ as follows:

$$\begin{aligned} P[X + Y \leq u] &= \int \int_{\{x+y \leq u\}} f(x, y) dx dy, \\ &= \begin{cases} u^2/2, & 0 \leq u \leq 1, \\ 1 - (2-u)^2/2, & 1 \leq u \leq 2. \end{cases} \end{aligned}$$

We shall study more about probability density functions and distribution functions in the next chapter.

Usually in statistical work, one is given only the induced probability space. It is not necessary that one should have the knowledge of the fundamental probability space $(\Omega, \mathcal{A}, P)$, if one is interested in studying the properties of the given random variables and their Borel functions only. However, in order to study the calculus of r.v.'s, by introducing new random variables such as limits of r.v.'s, it is necessary to introduce the above concept of fundamental probability space.

3.6 OTHER MEASURES

(a) GENERALIZED PROBABILITY MEASURE

We have seen in section (3.1) that probability is a non-negative, σ -additive, normed set function on $(\Omega, \mathcal{A})$. Without the condition of normedness, if we have a non-negative and additive set function, it is called a measure, denoted by μ .

If $\mu(\Omega)$ is nonzero and finite, considering $\mu(\cdot)/\mu(\Omega)$ as a set function on $(\Omega, \mathcal{A})$, we see that it satisfies all the properties (i)–(iv) and hence may be viewed as a probability measure. Thus, normedness can be achieved by dividing the measure function by $\mu(\Omega)$. Hence finite measures may be viewed as generalisations of probability measure.

(b) CONDITIONAL PROBABILITY MEASURE

Let B be an arbitrary but fixed subset of Ω belonging to $\mathcal{A}$. Consider the class of events

$$B \cap \mathcal{A} = \{B \cap A : A \in \mathcal{A}\} = \mathcal{A}_B, \text{ (say).} \quad (3.56)$$

This class is a σ -field of subsets of B . Thus, $(B, \mathcal{A}_B)$ is a measurable space. Probability measure P defined on this space is not normed, since $P(B)$ may not be unity, in general. However, as in (a) it may be normed, by defining the measure P^* as

$$P^*(B \cap A) = \frac{P(B \cap A)}{P(B)}, \quad (3.57)$$

on $\mathcal{A}_B$, if $P(B) > 0$. $(B, \mathcal{A}_B, P^*)$ is a probability space. Domain of P^* is the class of events which are subsets of B . For a given B , we may view P^* to be a set function defined on elements of $\mathcal{A}$. Then it is called the *conditional probability measure*, given B , denoted by P_B or $P(\cdot | B)$. Thus,

$$P_B(A) = P^*(B \cap A) = P(BA)/P(B) = P(A | B). \quad (3.58)$$

Evidently P_B is a probability measure on $(\Omega, \mathcal{A})$, since

$$P_B(A) \geq 0, \quad P_B(\Omega) = 1,$$

and

$$\begin{aligned} P_B(\sum_j A_j) &= P^*(B \cap \sum_j A_j) = P(B \cap \sum_j A_j)/P(B), \\ &= \sum_j P(B \cap A_j)/P(B) = \sum_j P_B(A_j). \end{aligned} \quad (3.59)$$

Thus, $(\Omega, \mathcal{A}, P_B)$ is a probability space. $P_B(A)$ is called the *conditional probability of A given B* . $P_B(\cdot)$ is called the *conditional probability distribution on $(\Omega, \mathcal{A})$* .

Let X and Y be random variables and $(R^2, \mathcal{B}_2, P_{XY})$ be the probability space induced by (X, Y) . Consider any Borel set $A \in \mathcal{B}_2$ and the set

$$B = [X = x_0, -\infty < y < \infty] = [X = x_0].$$

Then $A \cap B = [X = x_0, (x, y) \in A]$.

$$P_X B = P[X = x_0], \quad P_{XY}(A \cap B) = P[X \in B, (X, Y) \in A].$$

Conditional probability of A given B is

$$P_B(A) = P[X = x_0, (X, Y) \in A] / P[X = x_0], \quad (3.60)$$

if $P[X = x_0] > 0$. It is also called the *conditional probability of A given $X = x_0$* . In general, if

$$B = [X \in B_1, -\infty < y < \infty] = [X \in B_1],$$

$$P_{B_1}(A) = P[X \in B_1, (X, Y) \in A] / P[X \in B_1], \quad (3.61)$$

where $P[X \in B_1] > 0$. $P_B(\cdot)$ is called the *conditional probability of A given that $X \in B_1$* and is denoted as $P((\cdot) | X \in B_1)$. Evidently, for

different B_i we get different conditional probabilities. We shall discuss "conditioning" in detail in Chapter 12.

(c) COUNTING MEASURE

Suppose Ω is the real line R and $\mathcal{A}$ is the class of all Borel sets in R . Define a set function μ such that

$$\mu(B) = \text{number of integers (positive, negative or zero) in } B.$$

Then

$$\mu(\phi) = 0, \mu(\sum_i B_i) = \sum_i \mu(B_i), B_i \in \mathcal{B}.$$

Thus, μ is a measure on R . It is called the *counting measure*. The measure associates unit mass with each one of the points $0, \pm 1, \pm 2, \dots$

But $\mu(R)$ is not finite. However

$$R = \sum_{n=-\infty}^{\infty} [n, n+1) \text{ and } \mu[n, n+1) = 1.$$

Hence μ is σ -finite on R . We may note that $\mu(B) = 0$ does not imply that $B = \phi$ (why?) and $\mu(\cdot)$ takes the values $0, 1, 2, \dots$ only.

Counting measure may be defined on $(R^k, \mathcal{B}_k)$ as follows. For $B_k \in \mathcal{B}_k$, define

$$\mu(B_k) = \text{number of points } (x_1, \dots, x_k) \text{ with integral (positive, negative or zero) coordinates lying in } B_k.$$

(d) LEBESGUE MEASURE

Consider the real line R and the class $\mathcal{J}$ of all bounded intervals of the form $(a, b]$ ($a < b$). On the elements of $\mathcal{J}$, we define a set function μ such that

$$\mu(a, b] = b - a,$$

$$\mu(\phi) = 0,$$

$$\mu(\sum_i E_i) = \sum_i \mu(E_i),$$

where $E_1, E_2, \dots$ are disjoint sets of $\mathcal{J}$. Thus, μ is a measure defined on $\mathcal{J}$. Since $\mathcal{J}$ is closed under finite intersections, using Lemma 3.2, μ can be extended uniquely to all Borel sets. This extended measure on $(R, \mathcal{B})$ is called the *Lebesgue measure*.

Since $R = \sum_{n=-\infty}^{\infty} (n, n+1]$ and $\mu(n, n+1] = 1$, μ is a σ -finite measure.

Since $\{x\} = \lim (x - 1/n, x]$ and $\mu(x - 1/n, x] = 1/n$, on account of continuity of μ ; $\mu(\{x\}) = 0$ for all singletons $\{x\}$. A set whose μ -measure is zero is called a μ -null set. By additivity, countable union of μ -null sets is μ -null. Thus, Lebesgue measure of a set containing a countable number of points of R is zero. Moreover

$$\begin{aligned}\mu(a, b] &= \mu(a, b) + \mu(\{b\}), \\ &= \mu(a, b).\end{aligned}$$

Even though all singletons are μ -null Borel sets, it is well known from measure theory (c.f. Halmos, P.R. (1958)) that there exist μ -null sets which are not Borel sets. Adding these μ -null sets to the class of all Borel sets, we get the class $\mathcal{L}$ of *Lebesgue-measurable* sets. If $L \in \mathcal{L}$, L is either a Borel set (B) or a union of a Borel set and a μ -null set, $(B \cup N)$, say, or a μ -null set (N). On $\mathcal{L}$ define a measure $\bar{\mu}$ such that

$$\bar{\mu}(B \cup N) = \mu(B). \quad (3.62)$$

Then $\bar{\mu}$ is an extension of μ and is called the *completion of μ* . It is also called the *completed Lebesgue measure* defined on $\mathcal{L}$.

In general, given a measure μ on a σ -field $\mathcal{A}$, we may get the completed measure $\bar{\mu}$ defined on $\mathcal{A}_\mu$ the class of all sets $A \cup N$, where $A \in \mathcal{A}$ and N is μ -null, as follows. Define $\bar{\mu}(A \cup N) = \mu(A)$. $\mathcal{A}_\mu$ is called the completion of $\mathcal{A}$ with respect to μ and $\bar{\mu}$ is called the *completed measure* on the σ -field (why?) $\mathcal{A}_\mu$. We may note that $\mathcal{A}_\mu$ depends on μ . Hence, for different measures we get different completions.

(c) LEBESGUE-STIELTJES MEASURE

Let g be any monotone non-decreasing point function on R , which is finite or having bounded variation. Without loss of generality, it may be taken to be continuous on the right (or left), by taking the value of the function at a jump point to be the limit of its values from the right (or left). Thus, $g(x) = g(x+) = \lim_{x_n \downarrow x} g(x_n)$ for all $x \in R$. On all sets of $\mathcal{J}$ defined in (d) define a measure μ as follows:

- (i) $\mu(a, b] = g(b) - g(a)$,
= increment of g in the interval $(a, b]$. (3.63)
- (ii) $\mu(\{a\}) = g(a) - g(a-)$
= magnitude of jump of g at a .
- (iii) $\mu(\sum E_i) = \sum \mu(E_i)$.

where $\{E_i\}$ is a class of disjoint set of $\mathcal{J}$. As in (d), since μ is non-negative and continuous on $\mathcal{J}$, it can be extended to all Borel sets and hence it is a measure on $(R, \mathcal{B})$. This measure is called the *Lebesgue-Stieltjes (L.S.) measure* induced by the function g . If $g(+\infty) - g(-\infty)$ is finite, this measure is finite, otherwise it is σ -finite.

The measure μ depends on the increments of g and not on its absolute value. If we increase or decrease g by an arbitrary constant, μ will not be affected. Thus, μ will not determine g uniquely. It will correspond to a class of functions with the same increments as that of $g, g(-\infty)$ being arbitrary. Further we may write

$$g(b) - g(a) = \int_{a+}^b dg, \quad (3.64)$$

where the integral is understood to be in Stieltjes sense. If g increases by jumps only, with magnitudes g_i at the points x_i ,

$$g(b) - g(a) = \sum_{i: x_i \in (a, b]} g_i.$$

If $g(x)$ is differentiable in $(a, b]$ with derivative $g'(x)$, then $g'(x) \geq 0$ and

$$g(b) - g(a) = \int_a^b g'(x) dx,$$

the integral being defined in the Riemann sense. Evidently, Lebesgue measure is a particular case of L.S. measure with $g'(x) = 1$.

If $g(+\infty) - g(-\infty) = 1$, then L.S. measure induced by g may be viewed as a probability measure P on $(R, \mathcal{B})$, with

$$P(a, b] = g(b) - g(a).$$

Evidently $P(R) = 1$ and P can be defined on all Borel sets by the extension theorem and Lemma 3.2. If $g'(x)$ exists, then

$$P(a, b] = \int_a^b g'(x) dx$$

implies that $g'(x)$ coincides with the probability density function defined in section (3.4). Of course we do not rule out the possibility that $g(x)$ might increase both continuously as well as by jumps. If $g(+\infty) - g(-\infty)$ is finite, dividing $g(x)$ by this quantity, we can get a probability measure corresponding to any finite L.S. measure. This result is sometimes referred to as the *correspondence theorem* (see § 4.4). It plays an important role in the study of distribution functions in Chapter 4 (see also section 4.4). It is more convenient to work with a point function such as $g(x)$, $x \in R$ than, with a set function such as $P(B)$, $B \in \mathcal{B}$, even though P determines $g(x)$ uniquely up to an additive constant (*utac*) only.

(f) SIGNED MEASURE

If a σ -additive set function ν defined on $\mathcal{A}$ can take positive and negative values, but assuming at most one of the values $\pm\infty$, satisfying $\nu(\phi) = 0$, it is called a *signed measure*. For example, if $\nu = \mu_1 - \mu_2$, where μ_1 and μ_2 are measures defined on $\mathcal{A}$, with at least one measure finite, then ν is a signed measure. In fact the converse statement is also true; viz., any signed measure may be viewed as the difference of two measures, of which at least one is finite. This is the content of the so called *Jordan-Hahn decomposition theorem* (See Appendix B).

COMPLEMENTS AND PROBLEMS

1. Show that

$$(i) \quad P(AB^c \cup BA^c) = PA + PB - 2P(AB),$$

$$(ii) P(AB) \leq PA \leq P(A \cup B) \leq PA + PB,$$

$$(iii) P(A_1 \cap \dots \cap A_n) \geq \sum_1^n PA_i - (n-1).$$

2. In example (3.2), calculate the probability of the events [no. of heads exceeds no. of tails by r], $r = 0, 1, \dots, n$.
3. If A_1, A_2, A_3 are such that

$$P(A_1 \cap A_2 \cap A_3) = P(A_1^c A_2^c A_3^c),$$

show that each is equal to

$$\frac{1}{2} [1 - PA_1 - PA_2 - PA_3 + PA_1 A_2 + PA_2 A_3 + PA_1 A_3].$$

4. Let $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$ and let $P\{\omega_1\} = 1/6, P\{\omega_2\} = 1/3, P\{\omega_3\} = 1/5, P\{\omega_4\} = 3/10$.

$$\text{Let } A_n = \begin{cases} \{\omega_1, \omega_2\}, & \text{if } n \text{ is odd,} \\ \{\omega_2, \omega_3\}, & \text{if } n \text{ is even.} \end{cases}$$

Find $P(\lim A_n), P(\overline{\lim A_n}), \lim PA_n, \overline{\lim PA_n}$.

5. Show that, if $\sum PA_n < \infty$, then $P(\lim A_n) = 0$.
6. Let Ω be a countably infinite set and let $\mathcal{A}$ consist of all subsets of Ω . Define $\mu(A) = 0$, if A is finite and $\mu(A) = \infty$ if A is infinite.
 - (i) Show that μ is finitely additive but not σ -additive
 - (ii) Show that Ω is the limit of an increasing sequence of sets A_n with $\mu(A_n) = 0$; but $\mu(\Omega) = \infty$. What can you say for $\{\mu(A_n^c)\}$?
7. If a set function assumes the value 1 at the empty set, show that it cannot be additive.
8. Let Ω be a countable space, $\mathcal{A}$ be a σ -field on Ω . If $\mathcal{S}(\Omega)$ is the power set of Ω , show that every singleton $\{\omega\} \subset \Omega$ is measurable iff $\mathcal{A} = \mathcal{S}(\Omega)$. In this case P can be defined on $(\Omega, \mathcal{A})$ if it is defined on every singleton $\{\omega_i\}$.
9. Show that the sum of two measures defined on the same σ -field is a measure. Show that the sum of a complete measure and any other measure defined on the same class is a complete measure.
10. If μ is a complete measure, any function f defined on a set E with $\mu(E) = 0$ is measurable.
11. f and g are said to be *equivalent* if they differ on a set of measure zero. Let μ be a complete measure. If f and g are equivalent on a set E and one of them is measurable, then the other is also measurable.
12. Let μ be a complete measure and $\{X_n\}$ be a sequence of measurable functions converging to X a.e. (μ). Show that X is also measurable. Is this result true if μ is not complete?
13. If $p_m(r, n)$ is the probability of finding exactly m cells empty when r balls are distributed randomly in n cells; obtain $\lim_{n, r} p_m(r, n)$ so that $\lambda = n e^{-(r/n)}$ remains bounded and m is fixed.

14. Show that binomial distribution converges to Poisson distribution in the sense that

$$P[X = k] = \binom{n}{k} p^k q^{n-k} \rightarrow e^{-\lambda} \lambda^k / k! = P[Y = k],$$

if as $n \rightarrow \infty$, $p \rightarrow 0$ such that $np = \lambda$, where X is binomial with index n and Y is Poisson with parameter λ .

15. Find the Lebesgue-Stieltjes measure generated by the following function $g(x)$.

$$\begin{aligned} g(x) &= 0, x < 1, \\ &= 1/2, 1 \leq x < 3, \\ &= 2/3, 3 \leq x < 7, \\ &= x - 6, 7 \leq x < 10, \\ &= 4, x \geq 10. \end{aligned}$$

16. A r.v. X has the following distribution function

$$\begin{aligned} P[X \leq x] &= 0, x < 0 \\ &= (1 - p) + p(1 - e^{-\lambda x}), 0 \leq x < T, \\ &= 1, x > T. \end{aligned}$$

Explain the induced probability space and express $P_X(B)$ in terms of p and λ .

17. Conditional probability of $[X_N = x]$ given $N = n$ is $1/(n+1)$ ($x = 0, 1, \dots, n$). If $P[N = n] = e^{-\lambda} \lambda^n / n!$, obtain the unconditional probability $P[X_N = x]$.
18. If X has p.d.f. e^{-x} , $x > 0$, find the probability densities of

$$(i) X^3; (ii) a + bX; (iii) e^{-X}; (iv) \log X.$$

19. If the p.d.f. of (X, Y) is $\theta^3 e^{-\theta(x+y)} = f(x, y)$, find the distribution of

$$\begin{aligned} (i) & X - Y, \\ (ii) & |X - Y|, \\ (iii) & X + Y, \\ (iv) & (X + Y)/X. \end{aligned}$$

(Hint: Make a transformation of (X, Y) to (U, V) with U as the above variable and integrate out V .)

20. (Completion) Let $\mathcal{N}$ be the class of null-sets of $(\Omega, \mathcal{A}, \mu)$. Show that the class $\mathcal{A}_\mu$ of sets of the form $A \cup N$, $A \in \mathcal{A}$, $N \in \mathcal{N}$, is the σ -field generated by $\mathcal{A} \cup \mathcal{N}$.
21. Consider the measure-space $(R, \mathcal{A}, \mu)$, where $\mathcal{A}$ is the power set of $(0, \pm 1, \pm 2, \dots)$ and μ is the counting measure. What is the completed σ -field $\mathcal{A}_\mu$? Does $\mathcal{A}_\mu$ coincide with the class of all Lebesgue-measurable sets?

22. If μ is a σ -finite measure on $(\Omega, \mathcal{A})$, show that there exists a finite measure μ^* such that $\mu(A) = 0$ iff $\mu^*(A) = 0$ for all $A \in \mathcal{A}$. (Hint: If $\Omega = \sum A_i$, such that $\mu(A_i) < \infty$, define $\mu^*(A) = \sum 2^{-i} \mu(AA_i)$).
23. If μ and ν are σ -finite measures on $(\Omega, \mathcal{A})$, show that $\mu + \nu$ is also σ -finite. In general, if $\mu_1, \mu_2, \dots, \mu_n$ are σ -finite, show that $\sum \mu_i$ is σ -finite. Can you have n infinite?

4. Distribution Functions

4.1 DISTRIBUTION FUNCTION OF A RANDOM VARIABLE

(a) DEFINING PROPERTIES

In statistical practice we generally take the relative frequency as an 'estimate' of the probability of an event. In n experiments if the value of X is observed to be less than or equal to x k times, k/n is an 'estimate' of

$$F_X(x) = P[X \leq x].$$

Hence the probability distribution of X is 'estimated'. If X is discrete and m is the number of times X is observed to have the value x , $P[X = x]$ is estimated by m/n (see article 4.4 also). Thus, distribution functions form the foundation of the theory of probability and statistics. Now we shall study some properties of the d.f.

Let X be a real r.v. on the probability space $(\Omega, \mathcal{A}, P)$. For $x \in R$, define

$$(i) \quad P[X \leq x] = F_X(x), \quad (4.1)$$

$$(ii) \quad P_X(a, b) = P[a < x \leq b], \\ = F_X(b) - F_X(a). \quad (b > a) \quad (4.2)$$

$F_X(x)$ is called the *distribution function* (d.f.) or *cumulative distribution function* of X . If $F_X(-\infty) = 0$, (i) and (ii) determine each other. In general, to a certain P_X there corresponds a family of distribution functions, differing by a constant. If we impose the restriction that the d.f. should vanish at $-\infty$, then P determines F uniquely.

Example 4.1. If $X = I(A)$ with $P(A) = p$,

$$F_X(x) = \begin{cases} 0, & x < 0, \\ 1 - p, & 0 \leq x < 1, \\ 1, & x \geq 1, \end{cases} \quad (4.3)$$

since X takes the values 0 and 1 only, with probabilities $(1 - p)$ and p respectively. $F_X(x)$ has jumps of magnitude $(1 - p)$ and p at 0 and 1 respectively.

Conversely, if $F(x)$ has jumps of magnitude p and $(1 - p)$ at a and b respectively, it is the d.f. of a r.v. taking the values a and b with

probabilities p and $(1 - p)$ respectively. If $F(x)$ has a jump of magnitude unity at $x = c$, it is the d.f. of a degenerate r.v. X , which is a constant equal to c .

Example 4.2. Let the probability space be $(R, \mathcal{B}, P)$ and X be the identity mapping of R to R , where P is the normal probability distribution defined in section (3.4). Then

$$F_X(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \exp(-t^2/2) dt. \quad (4.4)$$

In Examples 4.1 and 4.2, $F_X(x)$ is observed to be monotone non-decreasing. In example 4.1, $F_X(x)$ increases by jumps, while in Example 4.2, it increases continuously. Since $F_X(\cdot)$ represents the probability of a certain event, its maximum value is unity and it is non-negative. In statistics, d.f.'s play an important role as models describing physical phenomena. Some more d.f.'s are given in Complements and Problems.

THEOREM 4.1: *The distribution function F_X of a r.v. X is non-decreasing, continuous on the right with $F_X(-\infty) = 0$ and $F_X(+\infty) = 1$. Conversely, every function F , with the above properties is the d.f. of a r.v. on some probability space.*

Proof: For $x < x'$,

$$[X \leq x'] = [X \leq x] + [x < X \leq x']$$

and, therefore,

$$P[X \leq x'] = P[X \leq x] + P[x < X \leq x']. \quad (4.5)$$

Since

$$P[x < X \leq x'] \geq 0,$$

$$F_X(x') - F_X(x) \geq 0, \quad (4.6)$$

implying that $F_X(x)$ is monotone non-decreasing in x . Consider $\{x_n'\}$, $x_n' \downarrow x$. Since $[x < X \leq x_n'] \rightarrow \phi$ as $x_n' \downarrow x$, $F_X(x_n') - F_X(x) \rightarrow 0$ as $x_n' \downarrow x$. Since this is true for every sequence $\{x_n'\}$, F_X is continuous from the right (see Exercise 8).

Define $A_n = [X \leq n]$, $n = 1, 2, \dots$

Then $A_{n+1} \supseteq A_n$ and $A_n \uparrow [X < \infty] = X^{-1}(R) = \Omega$.

Then $P(A_n) \rightarrow P(\Omega) = 1$ as $n \rightarrow \infty$. Hence $F_X(+\infty) = 1$.

Defining $B_n = [X \leq -n]$, $n = 1, 2, \dots$, $B_n \downarrow \phi$ as $n \rightarrow \infty$. Therefore, $P(B_n) \rightarrow 0$ as $n \rightarrow \infty$ and $F_X(-\infty) = 0$.

To prove the converse, we have seen in section 3.6 that to a monotone non-decreasing function F on R with $F(-\infty) = 0$ and $F(+\infty) = 1$, there corresponds a Lebesgue-Stieltjes measure, which may be viewed as probability measure P on $(R, \mathcal{B})$. On an interval $(a, b]$,

$$P(a, b] = F(b) - F(a).$$

Taking $\Omega = R$, $\mathcal{A} = \mathcal{B}$ and P , defined above, F may be viewed as the d.f.

of the r.v. X , defined on R , such that $X(x) = x, x \in R$. This identity function is the required r.v. (*).

(b) ALTERNATE DEFINITION

We may note that if $P[X < x] = F'_x(x), F'_x(x)$, say, then

$$P[X < x'] - P[X < x] = P[x \leq X < x'] \quad (4.7)$$

tends to zero as $x \uparrow x'$ and F'_x is, therefore, continuous from the left. Some authors call F'_x the distribution function of X , which possesses all the properties of F_x , except that it is continuous from the left while F_x is continuous from the right. If at a point of discontinuity x of F , $F(x)$ is defined as $\{F(x+0) + F(x-0)\}/2$, say, where $F(x+0)$ is the limit of $F(x)$ from the right and $F(x-0)$, the limit from the left, then F will be continuous neither from the left nor from the right (Prove!) (see Figure 4.1). Sometimes we denote $F(x+0)$ by $F(x+)$ and $F(x-0)$ by $F(x-)$.

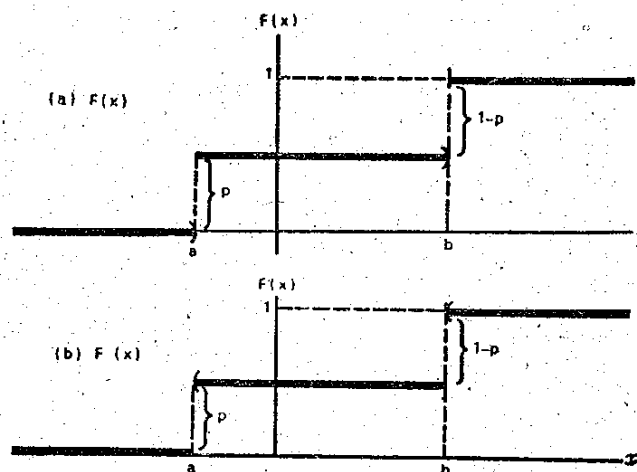


Fig. 4.1.

Since $P[x < X \leq x'] \rightarrow P[X = x']$ as $x \uparrow x'$, $F_x(x)$ will be continuous at x' if $P[X = x'] = 0$. In this case $F'_x(x)$ will also be continuous at x' . If $P[X = x'] > 0$, then there will be a jump of magnitude $P[X = x']$ at x' for F_x and F'_x . They coincide at all points of continuity of either function. At a point of discontinuity x' of F_x and F'_x , $F_x(x') = F_x(x'+0)$ while $F'_x(x') = F_x(x'-0)$. If X takes only a finite number of values $x_1, \dots, x_n$, say, then $F_x(x)$ has jumps at these points and F_x increases by jumps only. Such a function F_x which increases by jumps only is called a *step-function* (see Figure 4.1). The number of jumps may be finite or countable.

It may be mentioned here that since F is monotone, we may take

$$F(-\infty) = \lim_{x \rightarrow -\infty} F(x), \text{ and } F(+\infty) = \lim_{x \rightarrow +\infty} F(x).$$

If X is real-valued (not extended real-valued) r.v.,

$$F(+\infty) - F(-\infty) = P(X^{-1}(R)) = 1$$

and there is no confusion. If X can take the values $\pm \infty$ with positive probabilities, then

$$\begin{aligned} P(X^{-1}(\bar{R})) &= 1 = P([-\infty < X < \infty] \cup [X = +\infty] \cup [X = -\infty]) \\ &= P(X^{-1}(R)) + P[X = +\infty] + P[X = -\infty]. \end{aligned}$$

In this case we have to clarify whether

$$F(+\infty) = P(X^{-1}(R)) \text{ or } P(X^{-1}(\bar{R})),$$

$$F(-\infty) = P(\{X = -\infty\}) \text{ or } 0.$$

Suppose a random variable X induces the probability space $(R, \mathcal{B}, P)$, where P has geometric distribution (see Example 3.6). Then X is called a geometric random variable. Since X takes the values $0, 1, 2, \dots$ only, $F_X(x)$ has jumps at these points and is a step function. The magnitude of jump at x is $(1-\theta)\theta^x$, ($0 < \theta < 1$) (see Section 3.3). Similarly, the distribution function of Poisson r.v., binomial r.v., negative binomial r.v., are step-functions. These r.v.'s are defined similarly to the above geometric r.v. More generally, discrete r.v.'s induce discrete distributions. Distributions of normal, gamma or rectangular r.v.'s are continuous. A necessary and sufficient condition for a d.f. to be continuous at x is that $P[X = x] = 0$ ($x \in R$). (Why?) This does not imply that $F(x)$ is differentiable or that its derivative $f(x)$ exists.

It may be emphasized that the d.f. of a r.v. need not be either a step-function or continuous. It can increase continuously as well as by jumps. Such a distribution is given in Example (4.3).

The set of values of x for which $F(x)$ increases is called the *support* of F . If X is the r.v. having the d.f. F , and A is the support of F , $P[X \in A^c] = 0$. The set of non-negative integers from the support of Poisson distribution. The set of non-negative values is the support of gamma distribution.

4.2 DISCOMPOSITION OF D.F.'s

(a) MIXTURE OF D.F.'s

If $F(x)$ possesses the properties of Theorem 4.1, then we shall call $F(x)$ the d.f. of a r.v. If $F(x)$ is monotone non-decreasing right-continuous with $F(-\infty) = 0$ and $F(+\infty) = c$, a non-zero constant, then F will be called a *general distribution function* or simply a *distribution function*. Given a general d.f. F dividing it by $F(+\infty)$, one can get the d.f. of a r.v. As already stated in Chapter 3, general d.f. corresponds to a finite

Lebesgue-Stieltjes measure.

If $F_i(x)$ are d.f.'s of r.v.'s, and $\alpha_i \geq 0$, $\sum_{i=1}^n \alpha_i = 1$, then

$$G(x) = \sum_{i=1}^n \alpha_i F_i(x) \quad (4.8)$$

is also a d.f. of a r.v. It is called a *mixture* of d.f.'s $F_i(x)$, with weights $\alpha_i \geq 0$ ($i = 1, 2, \dots, n$). If $\sum \alpha_i \neq 1$, then $G(x)$ will be a general distribution function. It plays an important role in the study of Bayesian inference; $\{\alpha_i\}$ is called prior distribution. It also occurs in the study of conditional distributions.

(b) DECOMPOSITION OF D.F.'S

THEOREM 4.2: (*Jordan Decomposition Theorem*): Every d.f. F has a countable set of discontinuity points. Moreover $F = F_c + F_d$, where F_c is continuous and F_d is a step function. The decomposition is unique.

Proof: Consider the interval $(l, l+1]$, ($l = 0, \pm 1, \pm 2, \dots$). Let $x_1 < x_2 < \dots < x_n$ be the n points of discontinuity of $F(x)$ in $(l, l+1]$ with jumps of magnitude exceeding $1/m$. Since

$$\begin{aligned} F(l) &\leq F(x_1 - 0) < F(x_1) \leq F(x_2 - 0) < F(x_2) < \dots \\ &\leq F(x_n - 0) < F(x_n) \leq F(l+1), \text{ denoting } p(x_k) = F(x_k) - F(x_k - 0), \\ \sum_{k=1}^n p(x_k) &= \sum_{k=1}^n [F(x_k) - F(x_k - 0)] \leq F(l+1) - F(l). \end{aligned}$$

Thus, the number of discontinuities in $(l, l+1]$, of magnitude exceeding $1/m$ cannot be larger than $m[F(l+1) - F(l)]$. Hence as $m \rightarrow \infty$ we see that the number of discontinuities in $(l, l+1]$ is countable. Since the number of such intervals is countable, the first part of the Theorem is proved.

Define

$$F_d(x) = \sum_{x_k \leq x} p(x_k), \quad x \in R \quad (4.9)$$

$$F_c(x) = F(x) - F_d(x). \quad (4.10)$$

As in Theorem 4.1, F_d is non-decreasing, non-negative and continuous on the right. On account of 4.10, F_c is also non-negative and continuous on the right. It is also continuous on the left and non-decreasing, since for $x < x'$,

$$\begin{aligned} F_c(x') - F_c(x) &= F(x') - F(x) - \sum_{x < x_k \leq x'} p(x_k), \\ &= F(x' - 0) - F(x) - \sum_{x < x_k \leq x'} p(x_k), \end{aligned}$$

is non-negative and tends to zero as $x \uparrow x'$. This decomposition is unique, because, otherwise, if

$$\begin{aligned} F &= F_c + F_d = F_c' + F_d' \\ F_c - F_c' &= F_d' - F_d \end{aligned} \quad (4.11)$$

the left hand side of which is continuous, while the right hand side is a step-function. This is a contradiction, unless both sides vanish simultaneously. This implies that $F_c = F_c'$ and $F_d = F_d'$. (*)

Example 4.3. Suppose X denotes the life-time of an equipment which may fail immediately on installation with probability $1 - p$ or live up to the age x with probability $p(1 - e^{-\mu x})$. Then the d.f. of X is

$$\begin{aligned} F(x) &= (1 - p) + p(1 - e^{-\mu x}), \quad x \geq 0 \\ &= 0, \quad x < 0 \end{aligned} \quad (4.12)$$

Here $F(0-) = 0$, $F(0) = 1 - p$, $F(\infty) = 1$, $F(x)$ is a sum of a step-function with jump of magnitude $(1 - p)$ at '0' and of a continuous function. It is also a mixture of a d.f. with jump of magnitude unity at '0' and of an exponential d.f.

$$F(x) = 1 - e^{-\mu x}, \quad x \geq 0.$$

Theorem 4.2 implies that a general d.f. can be split up into two general d.f.'s in a unique manner such that the first one is continuous and the second one is a step-function. The continuous part can be further split as a differentiable part and singular part in a unique manner following Lebesgue decomposition theorem given in Appendix C. Norming appropriately we can view these three parts as d.f.'s of r.v.'s and the original d.f. which has been split up as the mixture of these three types of d.f.'s.

4.3 DISTRIBUTION FUNCTIONS OF VECTOR RANDOM VARIABLES

(a) BIVARIATE CASE

Let (X, Y) be a two-dimensional r.v. Its distribution function $F(x, y)$ is defined as

$$F(x, y) = P[X \leq x, Y \leq y] = P[(X \leq x] \cap [Y \leq y]]. \quad (4.13)$$

It is a point function with domain R^2 and range $[0, 1]$. Hence,

$$F(x, \infty) = P([X \leq x] \cap \Omega) = P[X \leq x] = F_X(x).$$

Similarly $F(\infty, y) = F_Y(y)$

$F(x, y)$ is called the *joint distribution function* of X and Y , while F_X and F_Y are called the *marginal distribution functions*, to emphasize that we are having the d.f. of a single variable under a bivariate set-up. Marginal

d.f.'s are projections of probability mass on single dimensions.

Now

$$\begin{aligned} P[x_1 < X \leq x_2, y_1 < Y \leq y_2] \\ = P\{[x_1 < X \leq x_2] \cap [y_1 < Y \leq y_2]\} \end{aligned}$$

But $[x_1 < X \leq x_2] = [X \leq x_2] - [X \leq x_1]$

and $[y_1 < Y \leq y_2] = [Y \leq y_2] - [Y \leq y_1]$.

Hence,

$$\begin{aligned} [x_1 < X \leq x_2] \cap [y_1 < Y \leq y_2] &= [X \leq x_2, Y \leq y_2] - [X \leq x_1, Y \leq y_2] \\ &\quad - [X \leq x_2, Y \leq y_1] + [X \leq x_1, Y \leq y_1], \end{aligned}$$

$$\text{and } P[x_1 < X \leq x_2, y_1 < Y \leq y_2] = F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) + F(x_1, y_1) \geq 0 \quad (4.14)$$

whenever $x_1 < x_2, y_1 < y_2$ (4.14) is a necessary condition for $F(x, y)$ to be a distribution function. Other necessary properties are

$$F(+\infty, +\infty) = 1. \quad (4.15)$$

$$\text{and } F(-\infty, y) = 0 = F(x, -\infty), \quad \forall x, y \in R \quad (4.16)$$

(Prove!). Conditions (4.14)–(4.16) will also be sufficient for $F(x, y)$ to be the d.f. of a two-dimensional vector r.v., since (4.14) determines the probability on all rectangles in R^2 which, by extension theorem 3.1 and Lemma 3.2 can be extended uniquely to all Borel sets of R^2 (Prove!).

If (X, Y) takes only countable values (x_i, y_j) , ($i, j = 1, 2, \dots$) with positive probability, and

$$p_{ij} = P[X = x_i, Y = y_j]$$

$$\text{then, } F(x, y) = \sum_{x_i \leq x, y_j \leq y} p_{ij}. \quad (4.17)$$

Thus, $F(x, y)$ has jumps at these points (x_i, y_j) . Evidently

$$\sum_i \sum_j p_{ij} = 1,$$

$$p_{i.} = \sum_j p_{ij} = P[X = x_i] \text{ and } p_{.j} = \sum_i p_{ij} = P[Y = y_j]$$

denote the marginal distributions of X and Y in this discrete case.

If $\partial^2 F / \partial x \partial y = f(x, y)$ exists at almost all points $(x, y) \in R^2$ then $f(x, y) \geq 0$ and it is called the *probability density function* of (X, Y) . Then,

$$P[X \leq x, Y \leq y] = \int_{-\infty}^x \int_{-\infty}^y f(u, v) \, dv \, du \quad (4.18)$$

$$\text{and } f_Y(y) = \int_{-\infty}^{\infty} f(u, y) \, du, \quad f_X(x) = \int_{-\infty}^{\infty} f(x, v) \, dv$$

are the probability density functions of Y and X respectively. Thus, if (X, Y) has a probability density function in R^2 , X and Y also have p.d.f. but the converse is not true in general (Why?).

Example 4.4. Let the joint probability density function of (X, Y) be

$$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left(\frac{x^2}{\sigma_1^2} - \frac{2\rho xy}{\sigma_1\sigma_2} + \frac{y^2}{\sigma_2^2}\right)\right) \quad (-\infty < x, y < +\infty).$$

Then (X, Y) is said to be bivariate normal r.v. We can verify that the marginal distributions of X and Y are also normal (Verify!).

(b) k -VARIATE CASE

Suppose $(X_1, \dots, X_k)$ is a k -dimensional vector r.v. on Ω to R^k . Then sets of the form

$$[-\infty < X_1 \leq x_1, \dots, -\infty < X_k \leq x_k] = \bigcap_{i=1}^k [-\infty < X_i \leq x_i] \quad (4.19)$$

are measurable. Probability of (4.19) is denoted by $F(x_1, \dots, x_k)$, and is called the *distribution function of $(X_1, \dots, X_k)$* . Since (4.19) will be empty if any one of the sets is ϕ and is Ω if all the sets in the RHS of (4.19) are equal to Ω , we have

$$0 = F(-\infty, x_2, \dots, x_k) = F(x_1, -\infty, x_3, \dots) = \dots = F(x_1, \dots, x_{k-1}, -\infty) \quad (4.20)$$

$$1 = F(+\infty, \dots, +\infty).$$

$$\left. \begin{aligned} F_{X_i}(x_i) &= F(+\infty, \dots, x_i, +\infty, \dots, +\infty), \\ F_{X_i X_j}(x_i, x_j) &= F(+\infty, \dots, +\infty, x_i, +\infty, \dots, +\infty, x_j, \dots, +\infty), \\ F_{(X_1, \dots, X_{k-1})}(x_1, \dots, x_{k-1}) &= F(x_1, \dots, x_{k-1}, +\infty). \end{aligned} \right\} \quad (4.21)$$

Equations in (4.21) represent the marginal distributions of

$$X_i, (X_i, X_j), (X_1, \dots, X_{k-1}).$$

These, together with the non-negativity of the following expression in (4.22), are necessary and sufficient for $F(x_1, \dots, x_k)$ to be the d.f. of a k -dimensional vector r.v. (Why?).

$$\begin{aligned} 0 \leq P\left\{\bigcap_{i=1}^k [a_i < X_i \leq b_i]\right\} &= P\left[\bigcap_{i=1}^k \{(-\infty < X_i \leq b_i) - (-\infty < X_i \leq a_i)\}\right] \\ &= \sum (-1)^{k-l} F(c_1, c_2, \dots, c_k), \end{aligned} \quad (4.22)$$

where c_i denotes either b_i or a_i , ($i = 1, 2, \dots, k$) l is the number of a_i 's and the summation is extended over all the 2^k such assignments of c_i 's as l varies from 0 to k .

Example 4.5. A simple k -dimensional distribution is given by the following. Consider the k -dimensional cuboid bounded by the surfaces $X_i = x_{i_0}$, $X_i = x_{i_1} (> x_{i_0})$ ($i = 1, 2, \dots, k$). Its volume is $\prod_{i=1}^k (x_{i_1} - x_{i_0})$. If the probability that $(X_1 \dots X_k)$ will fall inside a region is proportional to the volume common to this region and the cuboid,

$$F(x_1, \dots, x_k) = \frac{1}{\pi} \left(\frac{x_1 - x_{i_0}}{x_{i_1} - x_{i_0}} \right), (x_{i_0} \leq x_1 \leq x_{i_1}), \quad (4.23)$$

where $(x_1 - x_{i_0})/(x_{i_1} - x_{i_0})$ is to be taken as unity if $x_1 \geq x_{i_1}$ and zero if $x_1 \leq x_{i_0}$.

Multinomial distribution given in example 3.4 is an example of a discrete multivariate distribution.

4.4 CORRESPONDENCE THEOREM

(a) D.F. ON A DENSE SUBSET OF R

An arbitrary distribution function is defined as a non-decreasing finite function continuous from the right. Thus,

$$\begin{aligned} F(x-0) &= \lim_{x_n \uparrow x} F(x_n) = \sup_{x' < x} F(x'), \\ F(x+0) &= \lim_{x_n \downarrow x} F(x_n) = \inf_{x' > x} F(x') = F(x), \end{aligned}$$

exist and are bounded by $F(-\infty)$ and $F(+\infty)$.

A set D is said to be *dense* in R , if any point of R is either in D or is a limit point of D . For example, the set of all rationals are dense in the real line. Countable dense subset is called a separating subset. Rationals also form a separating subset of R .

LEMMA 4.1: Let F_D be a non-decreasing finite function defined on D , a dense subset of R . Define a function $F(x)$ on R by

$$\begin{aligned} F(x) &= \inf_{x_n > x} F_D(x_n), \quad x_n \in D, \quad x \in R \cap D^c, \\ &= F_D(x), \quad x \in D. \end{aligned} \quad (4.24)$$

Then $F(x)$ is a d.f.

Proof: We note that F is an extension to R of F_D defined on D . For every $x \in R$, there exists a sequence $\{x_n\} \subset D$ such that $x_n \downarrow x$, $x_n \geq x$, by definition of D . Since $\{F_D(x_n)\}$ is monotone non-increasing, bounded below, $F(x) = \lim F_D(x_n)$ exists, and is finite. Since F coincides with F_D on D and F_D is non-decreasing, F is also non-decreasing (Prove!). By (4.24) it is continuous on the right. Hence it is a d.f. It will be a d.f. of a r.v., if $F_D(-\infty) = F(-\infty) = 0$ and $F_D(+\infty) = F(+\infty) = 1$. (*)

It follows from the lemma that if two d.f.'s $F_1(x)$ and $F_2(x)$ coincide on a set D dense in R , they coincide everywhere. They have the same jump-points with jumps of the same size. If right continuity is not insisted upon, $F_1(x)$ and $F_2(x)$ may differ at these jump-points.

(b) CORRESPONDENCE THEOREM

In the previous chapter, we have seen that there is a close relation between probability measures on $(R, \mathcal{B})$ and the distribution function $F(x)$, while the probability measure is a set function, d.f. is a point function.

Hence it is easy to work with the latter. Moreover, usually in statistics, we are given the d.f. associated with a r.v., and not its probability distribution. Therefore, we would like to know whether there is a 1-1 correspondence between these two. The answer is in the affirmative and we state the following theorem without proof.

THEOREM 4.3: *Every probability measure P on $(R, \mathcal{B})$ determines uniquely a d.f. F , through the correspondence*

$$P(-\infty, x] = F(x). \quad (4.25)$$

Evidently, $F(-\infty) = 0$, $F(+\infty) = 1$. Conversely, every d.f. F of a r.v. determines uniquely a probability measure P on $(R, \mathcal{B})$, through (4.25).

A part of the proof of this theorem is contained in the discussion on Lebesgue-Stieltjes measure given in the previous chapter. Detailed proof can be had from any book on measure theory (cf. Halmos, 1958). On account of this theorem, in the case of $(R, \mathcal{B})$, we may use the terms, d.f. F and probability distribution P interchangeably. This is also true when Ω is a subset A of R and $\mathcal{A}$ is the class of Borel sets restricted to subsets of A . If the probability space $(R, \mathcal{B}, P)$ is that which is induced by a r.v. X , then

$$P(-\infty, x] = P[X \leq x] = F(x),$$

so that $F(x)$ is the d.f. of X .

(c) EMPIRICAL DISTRIBUTION FUNCTION

We have stated that if k values of X are less than or equal to x out of n observed values, then k/n is an 'estimate' of the d.f. of X . If $x_{(1)}, x_{(2)}, \dots, x_{(n)}$ are the n observed values of X , arranged in the increasing order of magnitude called 'order statistics', we define

$$F_n(x) = \begin{cases} 0, & x < x_{(1)} \\ k/n, & x_{(k)} \leq x < x_{(k+1)}, \quad k = 1, 2, \dots, n-1, \\ 1, & x \geq x_{(n)}. \end{cases}$$

Then $F_n(x)$ is called the *empirical d.f. of X* . It is a step-function with jumps of magnitude $1/n$ at each one of the observed values. If $F_n(x)$ is supposed to approximate $F(x)$, to which it converges as $n \rightarrow \infty$, $\sup |F_n(x) - F(x)|$ is a measure of this approximation. This measure has the properties of a 'distance' or a metric (why?) and is called *Kolmogorov-Smirnov distance* between F_n and F . $F(x)$ is a theoretical distribution function serving as a model, while $F_n(x)$ is the observed distribution function. Kolmogorov-Smirnov distance can be used to verify whether the theoretical model is satisfactory and agrees with the observed values.

COMPLEMENTS AND PROBLEMS

1. If a distribution function F possesses the properties $F(+\infty) = 1$ and $F(-\infty) = 0$, we call it the d.f. of a r.v. In statistics to suit a particular physical phenomena we would like to choose an appropriate d.f. among a certain family of d.f.'s. Families of d.f.'s are obtained by varying parameters underlying the specification of a certain d.f.

(i) Normal family:

$$F(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2} dx, \quad -\infty < \mu < \infty, \sigma > 0;$$

(ii) Exponential family:

$$F(x) = \begin{cases} \int_0^x \frac{1}{\theta} e^{-x/\theta} dx, & x \geq 0, \theta > 0, \\ 0, & x < 0; \end{cases}$$

(iii) Cauchy family:

$$F(x) = \frac{1}{\pi} \int_{-\infty}^x \frac{dx}{1 + (x - \mu)^2}, \quad -\infty < \mu < \infty;$$

(iv) Gamma family:

$$F(x) = \begin{cases} \int_0^x \frac{1}{p! \theta^p} e^{-x/\theta} x^{p-1} dx, & x \geq 0 \\ 0, & x < 0, p > 0, \theta > 0 \end{cases}$$

- (v) Discrete d.f.'s such as binomial, Poisson, geometric, negative binomial have parameters p, λ etc., varying which we get families of d.f.'s.
2. Define two d.f.'s as *equivalent* if they differ by a constant. Show that this is an equivalence relation (reflexive, transitive and symmetric). Note that there is a unique probability distribution corresponding to an equivalent class.
3. A function $F: R \rightarrow R$ is defined by
- $$F(x) = 0, \quad x \leq 0$$
- $$= \alpha + ke^{-(1/2)x^2}, \quad x \geq 0.$$
- Determine values of α and k so that $F(x)$ is a d.f. of a r.v.
4. If F is the d.f. of a r.v. X , and F is continuous, obtain the d.f. of
- (i) $3X^2 + 2$ (ii) $Y = F(X)$.
5. If X is standard normal r.v., show that
- $$P[a < X^2 < b] = 2P[-\sqrt{b} < X < -\sqrt{a}]$$
6. Let $F_1(x)$ and $F_2(x)$ be the d.f.'s of a binomial and normal r.v.'s. Sketch the graph of $\frac{1}{2}(F_1(x) + F_2(x))$.
7. A convolution of two d.f.'s $F_1(x)$ and $F_2(x)$ is given by F defined as

$$\begin{aligned}
 F(x) &= \int F_1(x-y) dF_2(y), \quad x \in R \\
 &= \int F_2(x-y) dF_1(y), \quad x \in R
 \end{aligned}$$

and denoted as $F = F_1 * F_2$.

Show that (i) F is a distribution function of a r.v. if F_1 and F_2 are.

(ii) If F_1 and F_2 vanish for $x < 0$, $F(x) = \int_0^x F_1(x-y) dF_2(y)$.

(iii) F is continuous iff either F_1 or F_2 is continuous.

8. Let $F(x, y) = \begin{cases} 0, & \text{if } x < 0, y < 0 \text{ or } x + y < 1 \\ 1, & \text{elsewhere.} \end{cases}$

Show that $F(x, y)$ is not a distribution function.

(Hint: $P[1/2 < X \leq 1, 1/2 < Y < 1] < 0$)

9. Show that

$$f(x, y) = \frac{1}{4} [1 + xy(x^2 - y^2)], \quad (|x| < 1, |y| < 1)$$

is the density function of a bivariate r.v. Obtain the marginal d.f.'s of X and Y .

10. Joint distribution of an infinite sequence of r.v.'s can be defined as follows. Define probability distribution on all cylinder sets with finite-dimensional rectangular base. If they are 'consistent' then this probability distribution can be extended uniquely to all Borel sets of $\mathcal{B}_\infty$ (see Loevé, 1962).

5. Expectation and Moments

5.1 DEFINITION OF EXPECTATION

(a) SIMPLE R.V.'S

In this chapter we shall assume that the probability space $(\Omega, \mathcal{A}, P)$ is given. All the random variables under consideration will be defined on this space.

Definition of expectation of r.v.'s will be built up in stages. First, we shall define the expectation of a simple r.v.; then that of a non-negative r.v., and finally we shall define that of an arbitrary r.v. We shall also study some important properties of expectations. We may mention here that the expectation with respect to a probability measure corresponds to the integral with respect to an arbitrary finite measure.

Let X be a simple r.v.,

$$X = \sum_{k=1}^n x_k I(A_k), \quad A_k \in \mathcal{A}, \quad \sum A_k = \Omega.$$

Its expectation, EX , is defined as

$$EX = \sum_k x_k P A_k = \sum_k x_k P[X = x_k]. \quad (5.1)$$

It is also called the *integral* of X with respect to P -measure and is denoted by

$$\int_{\Omega} X dP \quad \text{or} \quad \int X dP \quad \text{or} \quad \int X.$$

If $X = I(A)$, taking the value unity on A and 0 otherwise,

$$EX = PA; \quad E(cI_A) = cPA.$$

Expectation may be viewed as the weighted average of x_k 's, with PA_k as the weight of x_k . Thus, it is also the centre of gravity of the values of X , when the mass (or weight) associated with x_k is PA_k ($k = 1, 2, \dots, n$). If $PA_k = 1/n$, then $EX = (1/n) \sum x_k$, which is nothing but the usual arithmetic mean of x_k 's. The expectation gives us an idea of the central value around which x_k 's lie. It is a measure of *central tendency* and is used as a parameter of location of the probability distribution of X . As the centre of gravity, it is also the *first moment* or the mean of the probability distribution of X or simply, the mean of X .

The theory of expectations we develop runs parallel to the theory of Lebesgue-integral or Lebesgue-Stieltjes integral developed in books on

measure theory. Later on, when we have occasions to deal with integrals with respect to general measures, we shall assume the corresponding results proved with P replaced by μ . If μ is finite, the proof requires no change. If μ is σ -finite, proofs of some of the theorems may not be valid and hence may require some modifications.

Now we shall prove some important properties of expectations of simple r.v.'s.

(i) (Linearity) If X and Y are simple r.v.'s,

$$E(X \pm Y) = EX \pm EY. \quad (5.2)$$

More generally, $E \sum_{i=1}^n X_i = \sum_{i=1}^n EX_i$, n finite.

Let $X = \sum_j x_j I(A_j)$, $Y = \sum_k y_k I(B_k)$, where $\{A_j\}$ and $\{B_k\}$ are partitions of Ω . Then

$$X \pm Y = \sum_j \sum_k (x_j \pm y_k) I(A_j B_k)$$

and is, therefore, a simple function. Hence

$$\begin{aligned} E(X \pm Y) &= \sum_j \sum_k (x_j \pm y_k) P(A_j B_k), \\ &= \sum_j x_j \sum_k P(A_j B_k) \pm \sum_k y_k \sum_j P(A_j B_k), \\ &= \sum_j x_j P A_j \pm \sum_k y_k P B_k, \\ &= EX \pm EY, \end{aligned}$$

since

$$P(A_j) = P \sum_k (A_j B_k) = \sum_k P(A_j B_k)$$

and

$$P B_k = P \left(\sum_j A_j B_k \right) = \sum_j P(A_j B_k).$$

(ii) (Scale preserving) $E(cX) = cEX$, because,

$$cX = c \sum_k x_k I(A_k) = \sum_k (cx_k) I(A_k),$$

and

$$E(cX) = \sum_k cx_k P A_k = c \sum_k x_k P A_k = cEX.$$

Evidently, if $X(\omega) = c$ for all $\omega \in \Omega$, $X = cI(\Omega)$, so that $EX = c$.

(iii) (Non-negativity): If $X \geq 0$, then $EX \geq 0$. If $X = \sum_k x_k I(A_k) \geq 0$, then $x_k \geq 0$ for all k . Hence $EX = \sum_k x_k P A_k \geq 0$. Evidently $\sum_k x_k P A_k$ will be non-negative, even if $x_k < 0$ for some k for which $P A_k = 0$. Thus, for EX to be non-negative, it is sufficient that X is non-negative on those sets which have non-zero probability, i.e., $X \geq 0$ almost surely (a.s.). Here $X \geq 0$ a.s. means that $[\omega : X(\omega) \geq 0]$ has probability one or $[\omega : X < 0]$ has probability zero. If $X = \sum_k x_k I(A_k)$ and $X \geq 0$ a.s., then $x_k \geq 0$, whenever $P A_k > 0$ and $x_k < 0$ implies that $P A_k = 0$. Since

$$EX = \sum_k x_k P A_k = \sum_{x_k > 0} x_k P A_k + \sum_{x_k < 0} x_k P A_k, \quad (5.3)$$

and the first term of R.H.S. of (5.3) is non-negative and the second term is zero, $EX \geq 0$.

Hence we have (iii)' If $X \geq 0$ a.s., then $EX \geq 0$.

COROLLARY 5.1: $X \geq Y$ a.s. implies that $EX \geq EY$. (5.4)

Since $X \geq Y$ a.s., $X - Y \geq 0$ a.s. and hence from (iii) $E(X - Y) \geq 0$. By (i), $E(X - Y) = EX - EY$. Therefore, $X \geq Y$ a.s. implies that $EX \geq EY$. If $X = 0$ a.s. then $X \geq 0$ a.s. and $-X \geq 0$ a.s., and hence $EX \geq 0$ and $-EX \geq 0$, implying that $EX = 0$. (Give an alternate proof!). It is very easy to see that, it is not necessary that $X = 0$ a.s. in order that $EX = 0$ (How?).

The *indefinite integral* (or indefinite expectation) of X over $A \in \mathcal{A}$ is defined as

$$\varphi(A) = E(XI(A)) = \int_A X dP = E\left(\sum_k x_k I(AA_k)\right) = \sum_k x_k P(AA_k). \quad (5.5)$$

This is a function on $\mathcal{A}$, and hence it is a set function.

(b) NON-NEGATIVE R.V.'S

If $X \geq 0$, not necessarily simple, then we know from Corollary 2.4, that there exists a monotone increasing sequence of non-negative simple functions $\{X_n\}$ converging to X . But $X_n \geq X_{n-1} \geq 0$ implies that

$$EX_n \geq EX_{n-1} \geq 0.$$

Thus, $\{EX_n\}$ is also a monotone increasing sequence, converging to a limit, $\lim EX_n$, which may be finite (non-negative) or $+\infty$. For $X \geq 0$, we define its expectation $EX = \lim EX_n$. It is possible that there exist more than one sequence of simple functions converging to X and hence more than one value for the expectation of X . The following Lemma 5.1 proves that such an ambiguity will not arise; and $\lim EX_n$ will not depend on the sequence $\{X_n\}$ of simple functions we start with.

LEMMA 5.1: If two non-decreasing sequences of non-negative simple functions $\{X_n\}$ and $\{Y_n\}$ have the same limit X , then

$$\lim EX_n = \lim EY_n = EX. \quad (5.6)$$

Proof: (i) Suppose we are able to prove that $0 \leq X_n \uparrow X$ and $\lim X_n \geq Y$, $Y \geq 0$ and simple, implies that

$$\lim EX_n \geq EY. \quad (5.7)$$

Then, since

$$0 \leq Y_n \uparrow X \text{ and } 0 \leq X_n \uparrow X, \\ 0 \leq Y_p \leq \lim Y_n,$$

Y_p simple, implies that

$$\lim EX_n \geq EY_p, \forall p. \quad (5.8)$$

Interchanging X_n and Y_n in (5.8), we have

$$\lim EY_n \geq EX_p, \forall p. \quad (5.9)$$

From (5.8) and (5.9)

$$\lim_n EX_n \geq \lim_p EY_p \geq \lim_p EX_p, \quad (5.10)$$

implying $\lim EX_n = \lim EY_n$, and hence proving the lemma. It only remains to establish (5.7), which we shall do now.

(ii) *Case 1:* Suppose $m = \min_{\omega} Y(\omega) > 0$ and $M = \max_{\omega} Y(\omega) < \infty$. Note that, since Y is simple, $\min Y$ and $\max Y$ exist. Let $m > \varepsilon > 0$, where ε is arbitrary. Since $\lim X_n \geq Y$, as $n \rightarrow \infty$,

$$A_n = [X_n > Y - \varepsilon] \uparrow \Omega. \quad (5.11)$$

Hence, using the linearity property of the expectation,

$$\begin{aligned} \int X_n &= \int X_n I(A_n) + \int X_n I(A_n^c), \\ &\geq \int X_n I(A_n), \quad (\because X_n \geq 0), \\ &\geq \int (Y - \varepsilon) I(A_n), \quad (\because \text{on } A_n, X_n > Y - \varepsilon), \\ &= \int Y I(A_n) - \varepsilon \int I(A_n), \quad (\text{by linearity}), \\ &= \int Y - \int Y I(A_n^c) - \varepsilon P A_n, \\ &\geq EY - M P A_n^c - \varepsilon P A_n. \end{aligned} \quad (5.12)$$

As $n \rightarrow \infty$, from (5.11) $P A_n \uparrow 1$ and $P A_n^c \downarrow 0$. Since ε is arbitrary and small (5.7) is established.

(iii) *Case 2:* ($M = \infty$). If $M = \infty$, consider

$$Y_c = Y I([Y < \infty]) + c I([Y = \infty]) \leq Y. \quad (5.13)$$

Thus, $Y_c(\omega)$ is equal to $Y(\omega)$ if $Y(\omega)$ is finite and equals c if $Y(\omega) = \infty$. Then Y_c has a finite maximum and from Case 1, inequality (5.7) is valid for X_n and Y_c , for every c . Thus, for every c ,

$$\lim EX_n \geq EY_c = \int_{[Y < \infty]} Y + c P[Y = \infty]. \quad (5.14)$$

But, since Y is simple, as $c \uparrow \infty$, $EY_c \uparrow EY$. Thus, we have (5.7) from (5.14). We may note here that, if $P[Y = \infty] > 0$, $EY = \infty$, so that $\lim EX_n = \infty$. If Y is finite a.s., $EY_c = EY$ and (5.14) implies (5.7).

(iv) *Case 3:* ($m = 0$). If $m = 0$, consider $(Y + c)$ whose minimum value is $c > 0$. Replacing X_n by $X_n + c$, since $\lim (X_n + c) \geq Y + c$, by (iii),

$$\lim E(X_n + c) \geq E(Y + c).$$

This implies that

$$c + \lim EX_n \geq EY + c. \quad (5.15)$$

As $c \rightarrow 0$, we have (5.7). Proof of the Lemma is complete (*).

(c) ARBITRARY R.V.'S

With each r.v. X , we have associated two non-negative r.v.'s X^+ and X^- , defined by

$$\begin{aligned} X^+ &= XI([X \geq 0]), \\ X^- &= -XI([X < 0]). \end{aligned} \quad (5.16)$$

Since X^+ and X^- are non-negative, their expectations are well defined. For an arbitrary r.v., we define

$$EX = EX^+ - EX^-, \quad (5.17)$$

provided the right hand side of (5.17) is not of the form $+\infty - \infty$. Thus, EX exists if at least one of EX^+ or EX^- is finite. If both EX^+ and EX^- are infinite, EX does not exist. Evidently, (5.17) will be finite if and only if both EX^+ and EX^- are finite. If EX is finite, X is said to be *integrable*. X is integrable if and only if X^+ and X^- are integrable.

If X is an elementary r.v., corresponding to (5.17), we have

$$EX = \sum_{k: X_k \geq 0} x_k P A_k - \sum_{k: X_k < 0} x_k P A_k$$

provided at least one of the summations on the R.H.S. is finite.

Considering the probability space $(R, \mathcal{B}, P_x)$ induced by X , we may define EX as follows. If X takes the values $x_1, \dots, x_n, \dots$, with probabilities $p_1, p_2, \dots, p_n, \dots$ then from (5.1),

$$EX = \sum_k x_k P[X = x_k] = \sum_k x_k p_k, \quad (5.18)$$

where the summation may extend to infinity, if the RHS exists. If P_x is the L.S. measure induced by F_x with

$$F_x(x) = P_x(-\infty, x] = \int_{-\infty}^{x+} dF(x), \quad x \in R,$$

then

$$E(X) = \int_{-\infty}^{\infty} x dF(x)$$

is called the *Lebesgue-Stieltjes integral* of X on R . If F_x possesses a derivative $f(x)$, then

$$EX = + \int_0^{\infty} x f(x) dx - \int_{-\infty}^0 (-x) f(x) dx, \quad (5.19)$$

where the integral is the usual Riemann integral (see Example 5.1). In general, we may write

$$EX = \int_{-\infty}^{\infty} x dF(x), \quad (5.20)$$

where the integral represents the summation (5.18), if F_x is a step-function with jumps of magnitude p_k at x_k , ($k=1, 2, \dots$). It represents the

Riemann-integral (5.19), if F_X possesses a derivative for almost all $x \in R$, i.e. $F'_X(x)$ exists except perhaps for those x belonging to a Lebesgue-null set, (a set with Lebesgue-measure zero) and F_X has no jumps.

Example 5.1. Let X be a Cauchy r.v. with probability density function

$$f(x) = [\pi(1+x^2)]^{-1}, \quad -\infty < x < \infty.$$

Then
$$E(X^+) = \frac{1}{\pi} \int_0^\infty \frac{x \, dx}{1+x^2} = \infty = E(X^-) = \frac{1}{\pi} \int_{-\infty}^0 \frac{(-x) \, dx}{1+x^2}.$$

Hence EX does not exist. However, the Cauchy principal value

$$\frac{1}{\pi} \int_{-\infty}^\infty \frac{x \, dx}{1+x^2} = \lim_{A \rightarrow \infty} \int_{-A}^A \frac{x \, dx}{\pi(1+x^2)}$$

exists and is equal to zero.

This example shows that Lebesgue integral may not exist, while Cauchy principal value may exist.

Example 5.2. Let X take the values $\{0, \pm 1, \pm 2, \dots\}$

with $P[X=k] = P[X=-k] = (3/\pi^2)/k^2, k=1, 2, \dots$

$$P[X=0] = 0.$$

Hence X is an elementary r.v. But

$$EX^+ = \sum_{k=1}^\infty (3/\pi^2) 1/k = \infty = E(X^-).$$

Thus EX does not exist. If X takes the values $1, 2, \dots$ with $P[X=k] = (6/\pi^2) 1/k^2$ then $EX^+ = \infty, EX^- = 0$, so that EX exists, though infinite.

It is easy to verify that the expectation of a standard normal variate is zero and that of the binomial variate with

$$P[X=k] = \binom{n}{k} p^k q^{n-k}, \quad (k=0, 1, \dots, n),$$

is np (Derive!) (see Complements and Problems).

If we have an arbitrary measure space $(\Omega, \mathcal{A}, \mu)$, the integral of X with respect to measure μ is defined in the same way as the expectation. The treatment of this subject is available in any book on measure theory (cf. Halmos, 1958). For example, on $(R, \mathcal{B})$, we may consider counting measure, Lebesgue measure, Lebesgue-Stieltjes measure, rather than probability measure and define integrals with respect to these measures μ and denote them as $\int X \, d\mu$.

Let X , defined on R , take the value x_i , at i ($i=0, \pm 1, \pm 2, \dots$).

Then
$$\sum_{-\infty}^\infty x_i = \int_{-\infty}^\infty X \, d\nu,$$

where ν is the counting measure (Why?). Thus, any summation may be viewed as an integral with respect to the counting measure. Similarly

weighted sum can be viewed as L.S. integral with respect to a step-function, with magnitude of jumps corresponding to the weights. Thus, using the present definition of expectation we need not separate discrete r.v.'s and r.v.'s taking values in an interval. If F is a mixture of discrete and absolutely continuous distributions, then EX is of the form (see Sec. 12.1 also).

$$\sum x_i p_i + \int_{-\infty}^{\infty} x f(x) dx,$$

where p_i is the magnitude of jump of F at x_i ($i = 0, 1, 2, \dots$) and $f(x)$ is the derivative of the absolutely continuous (differentiable) part of F .

So far we defined the expectation of an arbitrary r.v. X . If $f(X)$ is a Borel function of X , then $f(X)$ is also a r.v., and its expectation is defined as

$$\begin{aligned} Ef(X) &= \int_{\Omega} f(x) dP, \\ &= \int_R f(x) dP_X. \end{aligned} \quad (5.21)$$

Since the σ -field induced by $f(X)$ is contained in the σ -field induced by X , hence $(R, \mathcal{B}, P_X)$ may be taken to be the basic probability space for calculating EX and $Ef(X)$. For example,

$$EX^r = \int_R x^r dF(x).$$

5.2 PROPERTIES OF EXPECTATION

(a) LINEARITY PROPERTY

In section (5.1) we have derived some properties of expectations of simple r.v.'s. Now we shall establish those and some more properties for arbitrary r.v.'s. All the r.v.'s will be assumed to be real valued and defined on $(\Omega, \mathcal{A}, P)$. They need not be discrete or be continuous only. They can be arbitrary. First, we note the following.

LEMMA 5.2: *If $X \geq 0$ and is integrable, X can be infinite at most on a set of probability measure zero.*

Proof: Otherwise, since

$$X = XI[X < \infty] + XI[X = \infty],$$

and $XI[X < \infty] = \lim X_n$, where X_n is simple and $X_n \uparrow XI[X < \infty]$,

we have $X = \lim (X_n + \infty \cdot I\{X = \infty\})$,

and $EX = \lim E(X_n + \infty \cdot I\{X = \infty\})$,

The latter is infinite if $P\{X = \infty\} > 0$. Since $EX < \infty$ implies that $EX^+ < \infty$ and $EX^- < \infty$, $X^+ < \infty$ a.s. and $X^- < \infty$ a.s. Hence $X < \infty$ a.s. (see also Problem 10). (*)

weighted sum can be viewed as L.S. integral with respect to a step-function, with magnitude of jumps corresponding to the weights. Thus, using the present definition of expectation we need not separate discrete r.v.'s and r.v.'s taking values in an interval. If F is a mixture of discrete and absolutely continuous distributions, then EX is of the form (see Sec. 12.1 also).

$$\sum x_i p_i + \int_{-\infty}^{\infty} x f(x) dx,$$

where p_i is the magnitude of jump of F at x_i ($i = 0, 1, 2, \dots$) and $f(x)$ is the derivative of the absolutely continuous (differentiable) part of F .

So far we defined the expectation of an arbitrary r.v. X . If $f(X)$ is a Borel function of X , then $f(X)$ is also a r.v., and its expectation is defined as

$$\begin{aligned} Ef(X) &= \int_{\Omega} f(x) dP, \\ &= \int_R f(x) dP_X. \end{aligned} \quad (5.21)$$

Since the σ -field induced by $f(X)$ is contained in the σ -field induced by X , hence $(R, \mathcal{B}, P_X)$ may be taken to be the basic probability space for calculating EX and $Ef(X)$. For example,

$$EX = \int_R x dF(x).$$

5.2 PROPERTIES OF EXPECTATION

(a) LINEARITY PROPERTY

In section (5.1) we have derived some properties of expectations of simple r.v.'s. Now we shall establish those and some more properties for arbitrary r.v.'s. All the r.v.'s will be assumed to be real valued and defined on $(\Omega, \mathcal{A}, P)$. They need not be discrete or be continuous only. They can be arbitrary. First, we note the following.

LEMMA 5.2: *If $X \geq 0$ and is integrable, X can be infinite at most on a set of probability measure zero.*

Proof: Otherwise, since

$$X = XI[X < \infty] + XI[X = \infty],$$

and $XI[X < \infty] = \lim X_n$, where X_n is simple and $X_n \uparrow XI[X < \infty]$,

we have $X = \lim (X_n + \infty \cdot I\{X = \infty\})$,

and $EX = \lim E(X_n + \infty \cdot I\{X = \infty\})$.

The latter is infinite if $P\{X = \infty\} > 0$. Since $EX < \infty$ implies that $EX^+ < \infty$ and $EX^- < \infty$, $X^+ < \infty$ a.s. and $X^- < \infty$ a.s. Hence $X < \infty$ a.s. (see also Problem 10). (*)

THEOREM 5.1: Suppose EX , EY and $EX + EY$ exist. Then

$$(i) \quad X \geq 0 \text{ a.s.} \Rightarrow EX \geq 0 \text{ (Non-negativity).} \quad (5.22)$$

$$(ii) \quad E(cX) = cEX, \quad (5.23)$$

$$(iii) \quad X = 0 \text{ a.s.} \Rightarrow EX = 0. \quad (5.24)$$

$$(iv) \quad E(X + Y) = EX + EY \text{ (Linearity)} \quad (5.25)$$

$$(v) \quad X \geq Y \text{ a.s.} \Rightarrow EX \geq EY \quad (5.26)$$

$$(vi) \quad X \text{ is integrable} \Leftrightarrow |X| \text{ is integrable,} \quad (5.27)$$

$$|X| < Y, Y \text{ integrable} \Rightarrow X \text{ is integrable.} \quad (5.28)$$

Proof: (i) $X \geq 0$ a.s. $\Rightarrow \exists \{X_n\}$, $X_n \geq 0$ a.s. and simple $\ni X_n \uparrow X$. Hence from 'non-negativity' property for simple random variables, $EX_n \geq 0$. Since X_n is monotone non-decreasing, EX_n is also monotone non-decreasing. Since $EX = \lim EX_n$, $EX \geq 0$. This proves (5.20).

(ii) Suppose $X \geq 0$. Then $\exists \{X_n\} \ni X_n$'s are simple and $0 \leq X_n \uparrow X$. Suppose $c > 0$. Then $0 \leq cX_n \uparrow cX$ and $E(cX_n) = cEX_n$ by linearity property (i) of expectations for simple functions. Hence $cEX_n \uparrow cEX$. But since $X \geq 0$,

$$\begin{aligned} E(cX) &= \lim E(cX_n), \text{ (by definition)} \\ &= \lim c(EX_n), \\ &= c \lim EX_n, \\ &= cEX. \end{aligned}$$

For arbitrary X and $c > 0$

$$\begin{aligned} cX &= cX^+ - cX^- \\ \Rightarrow E(cX) &= E(cX^+) - E(cX^-), \text{ (by definition)} \\ &= cEX^+ - cEX^-, \\ &\quad (\because cX^+ \text{ and } cX^- \text{ are non-negative}) \\ &= c(EX^+ - EX^-), \\ &= cEX. \end{aligned}$$

The case when $c < 0$ can be treated in a similar manner (Prove !). Thus, we have (5.23). From (5.22) and (5.23) we have that $E(-X) = -EX$ and that $X \leq 0$ a.s. implies that $EX \leq 0$.

$$\begin{aligned} (iii) \quad X = 0 \text{ a.s.} &\Rightarrow X \geq 0 \text{ a.s. and } -X \geq 0 \text{ a.s.,} \\ &\Rightarrow EX \geq 0 \text{ and } E(-X) = -EX \geq 0, \\ &\Rightarrow EX = 0. \end{aligned}$$

Hence we have (5.24). Converse of this is not true. However, $X \geq 0$ a.s. and $EX = 0 \Rightarrow X = 0$ a.s. (Why?).

(iv) In order to prove that $E(X + Y) = EX + EY$, first we shall prove it

THEOREM 5.1: Suppose EX , EY and $EX + EY$ exist. Then

$$(i) \quad X \geq 0 \text{ a.s.} \Rightarrow EX \geq 0 \text{ (Non-negativity).} \quad (5.22)$$

$$(ii) \quad E(cX) = cEX, \quad (5.23)$$

$$(iii) \quad X = 0 \text{ a.s.} \Rightarrow EX = 0. \quad (5.24)$$

$$(iv) \quad E(X + Y) = EX + EY \text{ (Linearity)} \quad (5.25)$$

$$(v) \quad X \geq Y \text{ a.s.} \Rightarrow EX \geq EY \quad (5.26)$$

$$(vi) \quad X \text{ is integrable} \Leftrightarrow |X| \text{ is integrable,} \quad (5.27)$$

$$|X| < Y, Y \text{ integrable} \Rightarrow X \text{ is integrable.} \quad (5.28)$$

Proof: (i) $X \geq 0$ a.s. $\Rightarrow \exists \{X_n\}$, $X_n \geq 0$ a.s. and simple $\ni X_n \uparrow X$. Hence from 'non-negativity' property for simple random variables, $EX_n \geq 0$. Since X_n is monotone non-decreasing, EX_n is also monotone non-decreasing. Since $EX = \lim EX_n$, $EX \geq 0$. This proves (5.20).

(ii) Suppose $X \geq 0$. Then $\exists \{X_n\} \ni X_n$'s are simple and $0 \leq X_n \uparrow X$. Suppose $c > 0$. Then $0 \leq cX_n \uparrow cX$ and $E(cX_n) = cEX_n$ by linearity property (i) of expectations for simple functions. Hence $cEX_n \uparrow cEX$. But since $X \geq 0$,

$$\begin{aligned} E(cX) &= \lim E(cX_n), \text{ (by definition)} \\ &= \lim c(EX_n), \\ &= c \lim EX_n, \\ &= cEX. \end{aligned}$$

For arbitrary X and $c > 0$

$$\begin{aligned} cX &= cX^+ - cX^- \\ \Rightarrow E(cX) &= E(cX^+) - E(cX^-), \text{ (by definition)} \\ &= cEX^+ - cEX^-, \\ &\quad (\because cX^+ \text{ and } cX^- \text{ are non-negative}) \\ &= c(EX^+ - EX^-), \\ &= cEX. \end{aligned}$$

The case when $c < 0$ can be treated in a similar manner (Prove !). Thus, we have (5.23). From (5.22) and (5.23) we have that $E(-X) = -EX$ and that $X \leq 0$ a.s. implies that $EX \leq 0$.

$$\begin{aligned} (iii) \quad X = 0 \text{ a.s.} &\Rightarrow X \geq 0 \text{ a.s. and } -X \geq 0 \text{ a.s.,} \\ &\Rightarrow EX \geq 0 \text{ and } E(-X) = -EX \geq 0, \\ &\Rightarrow EX = 0. \end{aligned}$$

Hence we have (5.24). Converse of this is not true. However, $X \geq 0$ a.s. and $EX = 0 \Rightarrow X = 0$ a.s. (Why?).

(iv) In order to prove that $E(X + Y) = EX + EY$, first we shall prove it

for the case when $X \geq 0$ and $Y \geq 0$. In this case there exist sequences of simple r.v.'s $\{X_n\}$, $\{Y_n\}$ and $\{X_n + Y_n\}$ such that

$$0 \leq X_n \uparrow X, 0 \leq Y_n \uparrow Y, 0 \leq X_n + Y_n \uparrow X + Y.$$

But

$$E(X_n + Y_n) = EX_n + EY_n, EX_n \uparrow EX, EY_n \uparrow EY, E(X_n + Y_n) \uparrow E(X + Y).$$

Hence we have (5.25) for non-negative r.v.'s

It follows that, if $A \in \mathcal{A}$, and $X \geq 0$ a.s., then

$$EX = E(XI(A)) + E(XI(A^c)), \quad (5.29)$$

Since $XI(A) \geq 0$ a.s. and $XI(A^c) \geq 0$ a.s., this implies that $EX < \infty$ if and only if $EXI(A) < \infty$ and $EXI(A^c) < \infty$, whatever be A . Moreover, if there are a finite number of non-negative r.v.'s $X_1 \dots X_k$, by induction we can prove (How?) that

$$E(X_1 + \dots + X_k) = \sum_{i=1}^k EX_i. \quad (5.30)$$

To prove (5.25) in the general case, let us partition Ω into $A_1, A_2, \dots, A_6$ as follows:

$$A_1 = [X \geq 0, Y < 0, X + Y \geq 0],$$

$$A_2 = [X < 0, Y \geq 0, X + Y \geq 0],$$

$$A_3 = [X \geq 0, Y \geq 0, X + Y \geq 0],$$

$$A_4 = [X < 0, Y \geq 0, X + Y < 0],$$

$$A_5 = [X \geq 0, Y < 0, X + Y < 0],$$

$$A_6 = [X < 0, Y < 0, X + Y < 0].$$

Since EX , EY and $EX + EY$ exist, except for the trivial case of all the three integrals infinite and of the same sign, we can assume that at least one of the r.v.'s, say Y , is integrable. Hence it is finite a.s. Hence $X + Y$ is defined a.s.

On A_1 , $X + Y \geq 0$ and $-Y > 0$. Hence additive property holds for non-negative functions $(X + Y)I(A_1)$ and $-YI(A_1)$. Thus,

$$E(X + Y)I(A_1) + E(-YI(A_1)) = E(XI(A_1))$$

and hence by adding $E(YI(A_1))$ which is finite,

$$E(X + Y)I(A_1) = E(XI(A_1)) + E(YI(A_1)). \quad (5.31)$$

On A_2 , $(X + Y) \geq 0$ and $-X > 0$. Hence as before

$$E(X + Y)I(A_2) - EXI(A_2) = E(YI(A_2)) < \infty.$$

This implies that $EXI(A_2) < \infty$. Adding this, we get

$$E(X + Y)I(A_2) = E(YI(A_2)) + E(XI(A_2))$$

On A_3 , $X \geq 0$, $Y \geq 0$ implies (by (iv))

$$E((X + Y)I(A_3)) = E(XI(A_3)) + E(YI(A_3)).$$

On A_4 , $-(X + Y) \geq 0$, $Y \geq 0$.

Hence $E(-(X + Y)I(A_4)) + E(YI(A_4)) = E(-XI(A_4))$

and subtracting $E(YI(A_4))$ and changing the sign throughout,

$$E(YI(A_4)) + E(XI(A_4)) = E((X + Y)I(A_4)).$$

Similarly, in general

$$E(XI(A_i)) + E(YI(A_i)) = E((X + Y)I(A_i)), (i = 1, 2, \dots, 6). \quad (5.32)$$

But by (5.30),

$$\begin{aligned} E\{X(I(A_1) + I(A_3) + I(A_5))\} &= EX^+ \\ &= E(XI(A_1)) + E(XI(A_3)) + E(XI(A_5)), \\ EX(I(A_2) + I(A_4) + I(A_6)) &= -EX^- \\ &= E(XI(A_2)) + E(XI(A_4)) + E(XI(A_6)), \end{aligned}$$

so that $\sum_{i=1}^6 E(XI(A_i)) = EX$.

Similarly $EY(I(A_2) + I(A_3) + I(A_4)) = E(Y^+)$,

$$-EY(I(A_1) + I(A_5) + I(A_6)) = EY^-,$$

and therefore, $\sum E(YI(A_i)) = EY$.

Similarly, $\sum E(X + Y)I(A_i) = E(X + Y)$.

From (5.32) this implies that in general

$$E(X + Y) = EX + EY,$$

establishing (5.25). From (5.23) and (5.25), it follows that

$$E(aX + bY) = aEX + bEY.$$

(v) $X \geq Y$ a.s. $\Rightarrow X - Y \geq 0$ a.s.

$$\Rightarrow E(X - Y) \geq 0, \quad (\text{by 5.22})$$

$$\Rightarrow EX - EY \geq 0, \quad (\text{by 5.25})$$

$$\Rightarrow EX \geq EY.$$

This proves (5.26). It follows that if $X = Y$ a.s., $EX = EY$.

(vi) If X is integrable, EX^+ and EX^- are finite. Then

$$E|X| = E(X^+ + X^-) = EX^+ + EX^- \quad (5.33)$$

is finite and hence $|X|$ is integrable. Conversely, if $|X|$ is integrable, since $|X| \geq X$, from (5.26), X is integrable.

If $|X| \leq Y$, from (5.26), $E|X| \leq EY$. If EY is finite then, $E|X|$ is finite implying that X is integrable. Thus, we have completed the proof of Theorem (5.1). (*)

COROLLARY 5.2; If $X_1, X_2, \dots, X_n$ are such that EX_i

and $\sum_{i=1}^n EX_i$ exist, then

$$E(\sum X_i) = \sum EX_i. \quad (5.34)$$

The Corollary can be proved by induction on n since it is true for $n = 2$ by (5.25). (Prove !)

COROLLARY 5.3: If X and Y are $\mathcal{A}$ -measurable functions such that

$$\int_A X = \int_A Y, \quad \forall A \in \mathcal{A}, \quad (5.35)$$

then $X = Y$ a.s.

Proof: If (5.35) holds good for all $A \in \mathcal{A}$, then

$$\int_A (X-Y)dP = 0, \quad \forall A \in \mathcal{A}.$$

Taking $A_\varepsilon = [(X-Y) \geq \varepsilon]$, ε arbitrary,

$$\int_{A_\varepsilon} (X-Y)dP \geq \varepsilon PA_\varepsilon > 0$$

a contradiction to (5.35), unless $PA_\varepsilon = 0$ for all $\varepsilon > 0$. Similarly, taking $A'_\varepsilon = [(X-Y) < -\varepsilon]$, we can prove that $PA'_\varepsilon = 0$ for all $\varepsilon > 0$. This implies $X-Y = 0$ a.s., establishing the Corollary. (*)

This Corollary is in the form of the converse of (5.24). It is useful in establishing several properties of 'conditioning' in Chapter 12.

(b) OTHER PROPERTIES

If $a \leq X \leq b$ a.s., then it follows from (5.26) that

$$a \leq EX \leq b.$$

Moreover, from (5.17) and (5.33),

$$|EX| \leq EX^+ + EX^- = E|X|. \quad (5.36)$$

If X takes only positive integral values with

$$P[X = i] = p_i, \quad (i = 1, 2, \dots),$$

then

$$\begin{aligned} EX &= \sum_{i=1}^{\infty} ip_i = (p_1 + p_2 + \dots + p_n + \dots) \\ &\quad + (p_2 + p_3 + \dots) + (p_3 + p_4 + \dots) + \dots \\ &= P[X \geq 1] + P[X \geq 2] + P[X \geq 3] + \dots \\ &= \sum_{n=1}^{\infty} P[X \geq n]. \end{aligned} \quad (5.37)$$

If $f(x)$ is the p.d.f. of X on R , and $X \geq 0$, in analogy with (5.37) we have

$$\begin{aligned}
 EX &= \int_0^{\infty} xf(x) dx, \\
 &= \int_0^{\infty} (1-F(x)) dx, \quad (5.38)
 \end{aligned}$$

which can be proved by integrating by parts assuming that

$$x(1-F(x)) \rightarrow 0 \text{ as } x \rightarrow \infty.$$

(5.38) depends only on $F(x)$ and can be defined even if $f(x)$ does not exist. Since X^+ and X^- are non-negative

$$\begin{aligned}
 EX^+ &= \int_0^{\infty} (1-F(x)) dx, \\
 EX^- &= \int_{-\infty}^0 F(x) dx.
 \end{aligned}$$

Hence we have the following expression for EX , which can be taken to be a definition of expectation of a r.v. (cf. Tucker, H.G. (1967))

$$EX = \int_0^{\infty} (1-F(x)) dx - \int_{-\infty}^0 F(x) dx, \quad (5.39)$$

provided it is not of the form $\infty - \infty$.

If (X, Y) are two r.v.'s, and $g(X, Y)$ is a Borel function of (X, Y) , then we have

$$E(g(X, Y)) = \int_{\Omega} g(X, Y) dP \quad (5.40)$$

If $(R^2, \mathcal{B}_2, P_{XY})$ is the probability space induced by (X, Y) , (5.40) reduces to

$$Eg(X, Y) = \int_{R^2} g(x, y) dP(X, Y) \quad (5.41)$$

If $F(x, y)$ is the joint d.f. of (X, Y) , then (5.41) becomes

$$\begin{aligned}
 Eg(X, Y) &= \int_{R^2} g(x, y) dF(x, y), \\
 &= \int_{R^2} g(x, y) f(x, y) dx \cdot dy \quad (5.42)
 \end{aligned}$$

if $f(x, y)$ is the p.d.f. of (X, Y) in R^2 . Similarly, one can define the expectation of $g(X_1, X_2, \dots, X_k)$.

(c) EXPECTATION OF COMPLEX R.V.

We know that Z is a complex r.v., if Z is of the form $X + iY$, where X and Y are r.v.'s and $i = \sqrt{-1}$. If

$$Z(\omega) = \rho(\omega) e^{i\theta(\omega)}, \quad \rho(\omega) \geq 0, \quad 0 \leq \theta(\omega) \leq 2\pi,$$

then ρ and θ are random variables. If $Z = X + iY$, we define

$$EZ = EX + iEY.$$

Z is said to be integrable iff X and Y are integrable.

If $Z = e^{iuX} = \cos uX + i \sin uX$; ($u \in R$) then

$$\varphi_X(u) = EZ = E \cos uX + iE \sin uX,$$

$$\text{i.e., } \varphi_X(u) = \int_R e^{iux} dF(x) = \int_R \cos ux dF(x) + i \int_R \sin ux dF(x).$$

$\varphi_X(u)$ is called the *characteristic function of X* , or the *characteristic function corresponding to $F(x)$* and plays an important role in studying the convergence properties of d.f.'s.

LEMMA 5.3: If Z is a complex r.v.

$$|EZ| \leq E|Z| \quad (5.43)$$

Proof: Let $Z = \rho e^{i\alpha}$, ρ, α , being random.

Let $EZ = re^{it}$. Then

$$\begin{aligned} |EZ| &= r = e^{-it} E(\rho e^{i\alpha}), \\ &= E\rho e^{i(\alpha-t)}, \\ &= E\rho \cos(\alpha-t) + iE\rho \sin(\alpha-t), \\ &= E(\rho \cos(\alpha-t)) \quad (\because \text{LHS is real}), \\ &\leq E\rho = E|Z|. \quad (*) \end{aligned}$$

LEMMA 5.4: If Z is complex, Z is integrable iff $|Z|$ is integrable.

Proof: From Lemma (5.3),

$$|EZ| \leq E|Z|.$$

This implies that $|Z|$ is integrable $\Rightarrow Z$ is integrable. Conversely, let Z be integrable. Then

$$\begin{aligned} |Z| &= \sqrt{X^2 + Y^2} \leq |X| + |Y| \\ &\Rightarrow E|Z| \leq E(|X| + |Y|). \end{aligned}$$

But Z is integrable implies that X and Y are integrable. Since this implies that $|X|$ and $|Y|$ are integrable, $E|Z| < \infty$, if Z is integrable. (*)

5.3 MOMENTS, INEQUALITIES

(a) MOMENTS

$E(X-a)^k$, ($k = 1, 2, \dots$) is called the *moment of k th order of X about a* or simply *k th moment of X about a* . EX^k is called the *k th moment about the origin or simply k th moment*. It may or may not exist. However moments of even order always exist, even though some of them may be infinite. $E|X-a|^r$ ($r > 0$) is called the *r th absolute moment of X* .

about a . $E|X|^r$ is called the r th absolute moment. Absolute moments always exist and are non-negative. They vanish iff the corresponding r.v. vanishes a.s.

$E(X - EX)^2$, the second moment about the expected value (assumed finite) is called the *variance* of X denoted by $\text{Var } X$. It measures the dispersion of different values of X around its mean value. In fact, $\text{Var } X$ is the least value of $E(X - a)^2$, minimum being taken over all values of a . Thus, $\text{Var } X$ always exists and vanishes iff X is a constant. $\sqrt{\text{Var } X}$ is called the *standard deviation* (s.d.) of X . Evidently

$$E(X - EX)^2 = EX^2 - (EX)^2 \geq 0.$$

(b) MOMENT GENERATING FUNCTION

If moments of all orders exist and are finite, then

$$M(\theta) = 1 + (EX)\theta + \frac{(EX^2)\theta^2}{2!} + \dots + \frac{(EX^n)\theta^n}{n!} + \dots \quad (5.44)$$

is called the *moment generating function* of X , provided the RHS of (5.44) is convergent at least for $|\theta| < \theta_0$, a positive number. In fact

$$\begin{aligned} M(\theta) &= \sum_{n=0}^{\infty} \frac{\theta^n}{n!} \int_{-\infty}^{\infty} x^n dF(x) = \int_{-\infty}^{\infty} \left(\sum_{n=0}^{\infty} \frac{\theta^n x^n}{n!} \right) dF(x) \\ &= \int_{-\infty}^{\infty} e^{\theta x} dF(x) = E(e^{\theta X}), \end{aligned} \quad (5.45)$$

since we may interchange integration and summation, because $e^{\theta X}$ and hence $M(\theta)$ is non-negative (see Fubini's theorem in Chapter 6).

Eq. (5.44) is the generating function (g.f.) of moments about the origin. If we are interested in moments about a point a , we replace X by $X - a$ in (5.45). By expanding $E(e^{\theta(X-a)})$ in power series, the coefficient of $\theta^n/n!$ is the n th moment of X about a . By differentiating $M(\theta)$ with respect to θ , n times and putting $\theta = 0$, we obtain

$$EX^n = \mu_n' = \left. \frac{d^n M(\theta)}{d\theta^n} \right|_{\theta=0}. \quad (5.46)$$

Example 5.3. If X is discrete, and has negative binomial distribution, with

$$\begin{aligned} P[X = x] &= \binom{r+x-1}{x} p^r q^x, \quad x = 0, 1, \dots, \\ M(\theta) &= \sum_{x=0}^{\infty} \binom{r+x-1}{x} p^r q^x e^{\theta x}, \\ &= \left(\frac{p}{1 - qe^{\theta}} \right)^r. \end{aligned} \quad (5.47)$$

Hence we can verify that $EX = rq/p$, $\text{Var } X = rq/p^2$.

Example 5.4. If X is continuous and has gamma distribution with

probability density function

$$f(x) = \begin{cases} \frac{\lambda}{\Gamma(p)} (\lambda x)^{p-1} e^{-\lambda x}, & x \geq 0, p > 0 \\ 0, & \text{otherwise,} \end{cases}$$

$$\begin{aligned} M(\theta) &= \int_0^\infty \frac{\lambda}{\Gamma(p)} (\lambda x)^{p-1} e^{-\lambda x + \theta x} dx, \\ &= (1 - \theta/\lambda)^{-p} \end{aligned} \quad (5.48)$$

Hence $EX = p/\lambda$, $\text{Var } X = p/\lambda^2$.

For distributions such as Cauchy, moment g.f.'s do not exist, since all moments are not finite. For nonnegative integer valued r.v.'s the probability g.f., $P(t)$ defined as $E(t^X)$, is as convenient as the moment g.f. to obtain moments and factorial moments (see chapter 7). A factorial moment of order k is defined as

$$\mu_{[k]} = E(X(X-1)\dots(X-k+1)). \quad (5.49)$$

Probability g.f. is obtained by putting t in the place e^θ in the moment g.f. Differentiating $P(t)$ k times and putting $t = 1$, we get $\mu_{[k]}$. For θ complex $M(\theta)$ is sometimes referred to as the Laplace transform of F . For t complex, $P(t)$ is called the Mellin transform of F . A more important transform, called the Fourier transform which is the same as $M(\theta)$, when θ is imaginary, will be discussed in Chapter 7. In the discrete case, Mellin transform corresponds to the probability generating function.

(c) INEQUALITIES

LEMMA 5.5: If $E|X|^r < \infty$, then $E|X|^{r'} < \infty$ for $0 < r' \leq r$ and EX^k exists and is finite for $k < r$, k an integer.

Proof: For $0 < r' < r$ and given ω

$$\begin{aligned} |X(\omega)|^{r'} &\leq 1, \text{ if } |X(\omega)| \leq 1, \\ &\leq |X(\omega)|^r, \text{ if } |X(\omega)| > 1. \end{aligned}$$

Hence for all ω ,

$$|X(\omega)|^{r'} \leq 1 + |X(\omega)|^r, \text{ for } 0 < r' < r. \quad (5.50)$$

Therefore, taking expectations of both sides,

$$E|X|^{r'} \leq 1 + E|X|^r \quad (5.51)$$

and

$$E|X|^r < \infty \text{ implies } E|X|^{r'} < \infty.$$

Since $|X(\omega)|^k = |X^k(\omega)|$ and X^k is integrable iff $|X^k|$ is integrable, Lemma 5.4 follows. (*)

The Lemma 5.5 implies that if a moment of a certain order is finite, then all moments of lower orders are also finite. Of course the converse of the Lemma is not true.

LEMMA 5.6: (*C_r-inequality*)

$$E |X + Y|^r \leq C_r E |X|^r + C_r E |Y|^r, \quad (5.52)$$

where

$$C_r = \begin{cases} 1 & \text{if } r \leq 1, \\ 2^{r-1} & \text{if } r \geq 1, \end{cases} \quad (5.53)$$

Proof: If $a > 0, b > 0$

$$\left(\frac{a}{a+b}\right)^r + \left(\frac{b}{a+b}\right)^r \geq 1, \text{ for } r \leq 1 \quad (5.54)$$

i.e.

$$a^r + b^r \geq (a+b)^r. \quad (5.55)$$

Hence for all ω and $r \leq 1$,

$$\begin{aligned} |X(\omega)|^r + |Y(\omega)|^r &\geq (|X(\omega)| + |Y(\omega)|)^r \\ &\geq |X(\omega) + Y(\omega)|^r. \end{aligned} \quad (5.56)$$

Consider the function

$$\phi(p) = p^r + (1-p)^r, \quad (r \geq 1), \quad (0 < p < 1).$$

 $\phi(p)$ has a minimum when $p = 1/2$. Thus, for $a > 0, b > 0$,

$$\left(\frac{a}{a+b}\right)^r + \left(\frac{b}{a+b}\right)^r \geq 2^{-(r-1)},$$

or

$$2^{r-1}(a^r + b^r) \geq (a+b)^r. \quad (5.57)$$

Thus, for all ω and all $r \geq 1$,

$$\begin{aligned} 2^{r-1}(|X(\omega)|^r + |Y(\omega)|^r) &\geq (|X(\omega)| + |Y(\omega)|)^r \\ &\geq |X(\omega) + Y(\omega)|^r. \end{aligned} \quad (5.58)$$

Taking expectations of (5.56) and (5.58), we obtain (5.52). Hence we have the *C_r-inequality*. (*)

This lemma implies that if the r th absolute moments of X and Y exist and are finite, so also is the r th absolute moment of $X + Y$. There are no restrictions on the distributions of X and Y . Thus, the class of r.v.'s whose s th absolute moment is finite is closed under addition.

LEMMA 5.7: (*Hölder's Inequality*)

$$E |XY| \leq E^{1/r} |X|^r E^{1/s} |Y|^s, \quad (5.59)$$

where $r > 1$ and $r^{-1} + s^{-1} = 1$.Here and elsewhere $E^{1/p} |X|^r$ denotes $\sqrt[p]{E |X|^r}$.*Proof:* Consider $\phi(p) = (p^r/r) + (p^{-s}/s)$, $0 < p < \infty$.Now $\phi(p)$ has a minimum when $p = 1$. Then $\phi(1) = 1$, putting

$$p = p_0 = a^{1/s}/b^{1/r} \quad (a > 0, b > 0),$$

$$\phi(p_0) = (br)^{-1} a^{r/s} + (as)^{-1} b^{s/r} \geq 1 = \phi(1),$$

$$\Rightarrow r^{-1} a^{r+s/s} + s^{-1} b^{r+s/r} \geq ab,$$

$$\Rightarrow r^{-1} a^r + s^{-1} b^s \geq ab,$$

since $(r + s)/rs = 1$. Putting

$$a = |X(\omega)| / (E|X|^r)^{1/r}, \quad b = |Y(\omega)| / (E|Y|^s)^{1/s},$$

and taking expectations and using the assumption that $(1/r) + (1/s) = 1$, we have Hölder's inequality. (*)

COROLLARY 5.4: Putting $r = s = 2$ in (5.59), we get

$$E|XY| \leq \sqrt{E|X|^2 E|Y|^2}. \quad (5.60)$$

(5.60) is called the *Schwarz' inequality*.

Since $E|XY| \geq |EXY|$, from (5.60) we have the inequality

$$|E(XY) - EXEY| \leq \sqrt{\text{Var } X \cdot \text{Var } Y} \quad (5.61)$$

i.e., $E[(X - EX)(Y - EY)]$, called the *covariance* of X and Y is less than the product of s.d.'s of X and Y . This implies that the *correlation coefficient* between X and Y , defined by

$$r_{X,Y} = \frac{\text{Cov}(X, Y)}{(\text{s.d. } X) \cdot (\text{s.d. } Y)},$$

has its absolute value always less than unity.

LEMMA 5.8: (*Minkowski Inequality*) For $r \geq 1$,

$$E^{1/r} |X + Y|^r \leq E^{1/r} |X|^r + E^{1/r} |Y|^r \quad (5.62)$$

Proof: $E|X + Y|^r = E(|X + Y| \cdot |X + Y|^{r-1})$

$$\leq E(|X| \cdot |X + Y|^{r-1}) + E(|Y| \cdot |X + Y|^{r-1}),$$

$$\leq E^{1/r} |X|^r E^{1-1/r} |X + Y|^{r-1} + E^{1/r} |Y|^r E^{1-1/r} |X + Y|^{r-1},$$

using Hölder's inequality with $s = r/(r - 1)$. Dividing by $E^{1-1/r} |X + Y|^{r-1}$, we have (5.62). (*)

(d) CONVEX FUNCTION (see Hardy, Littlewood and Polya, 1934)

Let $f(x)$ be a real valued Borel function defined on an open interval I (finite or infinite) $\subset R$. f is said to be *convex* (or *convex from below*) if for every pair of points x, x' of I ,

$$f\left(\frac{x + x'}{2}\right) \leq \frac{1}{2}f(x) + \frac{1}{2}f(x'). \quad (5.63)$$

If f is twice-differentiable on I , convexity is equivalent to $f''(x) \geq 0$ on I . A convex function is continuous (Why?). An alternative definition of a convex function is that, for every $x_0 \in I$, there corresponds a number $\lambda(x_0)$ such that for all $x \in I$,

$$\lambda(x_0)(x - x_0) \leq f(x) - f(x_0). \quad (5.64)$$

Here LHS may be interpreted as the tangent through x_0 , if it exists. (5.64) implies, that all the points of the curve $f(x)$ are above this tangent line. For example $|x|^r$ is a convex function of x .

LEMMA 5.9: (Jensen's Inequality) If f is convex and EX is finite, then

$$f(EX) \leq Ef(X). \quad (5.65)$$

Proof: Let X be a r.v. whose values lie in I and f be convex on I . Replacing x_0 by EX and x by X in (5.64) we have for all ω

$$\lambda(EX) (X(\omega) - EX) \leq f(X(\omega)) - f(EX). \quad (5.66)$$

Taking expectations, since the LHS vanishes, we have (5.65). (*)

If f is convex and monotone increasing and has an inverse f^{-1} , then, from (5.65),

$$EX \leq f^{-1}(Ef(X)). \quad (5.67)$$

For example, if $f(x) = x^r$, $r \geq 1$, $x \geq 0$, we have

$$EX \leq E^{1/r} |X|^r. \quad (5.68)$$

In fact (5.68) is a particular case of the inequality

$$E^{1/r} |X|^r \leq E^{1/p} |X|^p, \quad r < p \quad (5.69)$$

called Liapounov's inequality, which can be proved using Schwarz' inequality, (see Complements and Problems). This shows that $E^{1/r} |X|^r$ is a non-decreasing function of r .

(e) BASIC INEQUALITY

THEOREM 5.2 (Basic Inequality): Let X be an arbitrary r.v. and g on R be a non-negative Borel function. If g is even and is non-decreasing on $[0, \infty)$, then for every $a > 0$

$$\frac{Eg(X) - g(a)}{\text{a.s. sup } g(X)} \leq P[|X| \geq a] \leq \frac{Eg(X)}{g(a)}. \quad (5.70)$$

If g is non-decreasing on R and $P[|X| \geq a]$ is replaced by $P[X \geq a]$, (5.70) is true for every real a .

Here, a.s. sup $g(X)$ is the least among constants below which almost all values of $(X(\omega))$ lie.

Proof: Since g is a Borel function on R , it follows that $g(X)$ is a measurable function on Ω and is a r.v. Since g is non-negative, its integral exists. Hence

$$Eg(X) = \int_A g(X) + \int_{A^c} g(X) \quad (5.71)$$

where $A = [|X| \geq a]$. On A , since g is non-decreasing and even,

$$g(a) \leq g(X) \leq \text{a.s. sup } g(X). \quad (5.72)$$

Hence

$$g(a) \cdot PA \leq \int_A g(X) \leq (\text{a.s. sup } g(X)) \cdot PA \quad (5.73)$$

Since on A^c , $0 \leq g(X) \leq g(a)$,

$$0 \leq \int_{A^c} g(X) \leq g(a) PA^c \leq g(a). \quad (5.74)$$

Adding (5.73) and (5.74)

$$g(a) \cdot PA \leq Eg(X) \leq \text{a.s. sup } g(X) \cdot PA + g(a).$$

Hence we have (5.70)

$$\frac{Eg(X) - g(a)}{\text{a.s. sup } g(X)} \leq PA \leq \frac{Eg(X)}{g(a)}.$$

If g is non-negative and non-decreasing on R and if $A' = P[X \geq a]$, all the above statements are true for A' replacing A . (*)

COROLLARY 5.5:

$$\frac{E|X|^r - a^r}{\text{a.s. sup } |X|^r} \leq P(|X| \geq a) \leq \frac{E|X|^r}{a^r} \quad (5.75)$$

Proof follows by putting in (5.70)

$$g(x) = |x|^r, r > 0.$$

The right hand side of (5.75) is called the *Markov inequality*. For $r = 2$, Markov inequality reduces to *Chebyshev's inequality*.

The left hand side of (5.70) is non-trivial only if $g(X)$ is a bounded r.v. Therefore if $g(X)$ is unbounded, a.s. sup $g(X)$ is infinite and the left hand side of (5.70) vanishes.

Taking $g(X) = |x|^r / [1 + |x|^r]$ ($r > 0$), $g(x)$ is even, monotone increasing with a.s. sup $g(X) = 1$. Hence the basic inequality (5.70) becomes

$$E\left(\frac{|X|^r}{1 + |X|^r}\right) - \frac{a^r}{1 + a^r} \leq P[|Y| > a] \leq \frac{1 + a^r}{a^r} E\left(\frac{|X|^r}{1 + |X|^r}\right) \quad (5.76)$$

A direct proof of Markov inequality (or Chebyshev's inequality) can be given (How?). With $g(x) = |x|^r$, if x takes only the values 0 and a then the right hand part of (5.75) becomes an equality.

(f) INEQUALITIES OF MOMENTS OF COMPLEX R.V.'S

If X, Y and Z are complex r.v.'s, covariance of X and Y is defined as

$$\begin{aligned} C(X, Y) &= E(X - EX)(\overline{Y - EY}), \\ &= E(X - EX)(\overline{Y} - \overline{EY}), \quad (\because \overline{EY} = E\overline{Y}) \end{aligned}$$

where $\overline{Y}$ denotes the complex conjugate of Y , etc. $\text{Var } X = \sigma^2(X)$, is defined as

$$C(X, X) = E|X - EX|^2.$$

It may be noted that

$$\begin{aligned} C(\bar{Y}, \bar{X}) &= \overline{E(Y - EY)(\bar{X} - EX)} \\ &= \overline{E(Y - EY)(X - EX)}, \\ &= C(X, Y). \end{aligned} \quad (5.77)$$

$$C(aX + bY, Z) = aC(X, Z) + bC(Y, Z),$$

$$C(X, X) = E|X - EX|^2 = E(U^2 + V^2),$$

if $X - EX = U + iV$. Hence $C(X, X) = 0 \Leftrightarrow EU^2 = 0 = EV^2$.

i.e. $\Leftrightarrow U = 0$ a.s. and $V = 0$ a.s.

i.e. $\Leftrightarrow X - EX = 0$ a.s.

i.e. $\Leftrightarrow X = EX$ a.s.

This may be compared with the corresponding result for real r.v.'s.

Since for all real α

$$0 \leq E(\alpha|X| - |Y|)^2 = \alpha^2 E|X|^2 - 2\alpha E|XY| + E|Y|^2,$$

putting $\alpha = E|XY| / E|X|^2$, we have the analogue of Schwarz inequality for complex r.v.'s,

$$E^2|XY| \leq E|X|^2 E|Y|^2. \quad (5.78)$$

It follows that if $E|X|^2$ and $E|Y|^2$ are finite, then $E|XY|$ is finite.

COMPLEMENTS AND PROBLEMS

1. Obtain the mean and variance of the following discrete distributions:

(a) Binomial: $P[X = x] = \binom{n}{x} p^x q^{n-x} \quad (x = 0, 1, \dots, n);$

(b) Poisson: $P[X = x] = e^{-\lambda} \lambda^x / x! \quad (x = 0, 1, 2, \dots);$

(c) Negative binomial: $P[X = x] = \binom{n+x-1}{x} p^x q^n \quad (x = 0, 1, \dots);$

(d) Hypergeometric:
$$P[X = x] = \frac{\binom{r}{x} \binom{n-r}{d-x}}{\binom{n}{d}} \quad (x = 0, 1, \dots, \min(r, d)).$$

The above distributions are called standard distributions.

2. Obtain the mean and variance of the following continuous distributions with probability density function $f(x)$:

(a) Normal: $f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-(1/(2\sigma^2))(x-\mu)^2}, \quad -\infty < x < \infty$
 $-\infty < \mu < \infty, \sigma > 0;$

(b) Exponential:

$$f(x) = \frac{1}{\sigma} \exp(-(x-\mu)/\sigma), \quad x \geq \mu, \sigma > 0, -\infty < \mu < \infty \\ = 0, \text{ otherwise;}$$

(c) Laplace:

$$f(x) = \frac{1}{2\sigma} \exp(-|x-\mu|/\sigma), \quad -\infty < x < \infty; \\ -\infty < \mu < \infty, \sigma > 0.$$

(d) Triangular:

$$f(x) = 1-|x|, \quad |x| < 1 \\ = 0, \text{ otherwise;}$$

(e) Gamma:

$$f(x) = \frac{1}{\Gamma(p)\theta^p} \exp(-x/\theta)x^{p-1}, \quad x \geq 0 \\ = 0, \text{ otherwise;}$$

(f) Lognormal:

$$f(x) = \frac{1}{\sigma(x-a)\sqrt{2\pi}} \exp(-(\log(x-a)-m)^2/2\sigma^2), \quad x \geq a, \\ -\infty < m < \infty, \sigma > 0. \\ = 0, \quad x < a;$$

(g) Beta:

$$f(x) = \frac{1}{B(p, q)} x^{p-1}(1-x)^{q-1}, \quad p, q > 0 (0 \leq x \leq 1) \\ = 0, \text{ otherwise.}$$

3. Obtain the means, variances and covariances of the following distributions. Hence calculate the covariance matrix: ($\mathbf{X} = (X_1, \dots, X_k)$, $\mathbf{x} = (x_1, \dots, x_k)$, etc.)

(a) Multinomial:

$$P[\mathbf{X} = \mathbf{x}] = \frac{n!}{x_1! \dots x_k!} p_1^{x_1} \dots p_k^{x_k} (\sum p_i = 1), (\sum x_i = n);$$

(b) Negative multinomial:

$$\frac{(n-1)!}{x_1! \dots x_k!} p_1^{x_1} \dots p_k^{x_k} (\sum_{i=1}^k p_i = 1, \sum_{i=1}^k x_i = r, \text{ fixed}, r < n, 1 \leq k);$$

(c) Multinormal:

$$f(\mathbf{x}) = \frac{1}{|\Sigma|^{1/2} (2\pi)^{k/2}} \exp[-\frac{1}{2}(\mathbf{x}-\mu) \Sigma^{-1}(\mathbf{x}-\mu)].$$

4. Obtain the mean and the variance of the following distribution function.

$$\begin{aligned}
 F(x) &= 0, \quad x < 0, \\
 &= p + (1-p)(1-e^{-\lambda x}), \quad 0 < x < T, \\
 &= 1, \quad x \geq T.
 \end{aligned}$$

This is truncated modified exponential distribution, serving as a good model for life-time distributions. Note that the d.f. is a mixture of discrete and continuous distributions.

5. Let $\Omega = \{0, 1, 2, \dots\}$ and $\mathcal{A}$ is the power set Ω . For any $A \subset \Omega$, define $\mu(A)$ = no. of non-negative integers in A . Evaluate

$$\int_{\Omega} X d\mu, \text{ if } X(\omega) = 2^{-\omega}, \omega \in \Omega.$$

6. Let $\Omega = (-5, 5)$, $\mathcal{A}$ is the class of Borel sets in Ω and λ is the Lebesgue-measure. Taking $(\Omega, \mathcal{A}, \lambda)$, to be the measure space,

evaluate $\int_A X d\lambda$, where $A = [-1, 4)$, and

$$\begin{aligned}
 X(\omega) &= 1/2, \quad \omega \in (-5, 2), \\
 &= 1/3, \quad \omega = 2, \\
 &= 1, \quad \omega \in (2, 3], \\
 &= 0, \quad \omega \in (3, 5).
 \end{aligned}$$

7. Let $\Omega = \{-2, -1, 3, 7\}$, $A = \{-2, 3, 7\}$, $\mu(-2) = 2$, $\mu(-1) = 1$, $\mu(3) = 3$, $\mu(7) = 7$. If $X(\omega) = 1$, if $\omega = 3$, $\omega = 7$ = -1, if $\omega = -2$, $\omega = -1$,

evaluate $\int_A X d\mu$.

8. If X is a r.v. and $A_n = [|X| \geq n]$, $n = 0, 1, 2, \dots$. Show that X is integrable iff $\sum PA_n < \infty$. (See 10.22).

9. If X and Y are two positive r.v.'s and $Z = \sup(X, Y)$, then show that

$$\int_{\{Z > a\}} Z \leq \int_{\{X > a\}} X + \int_{\{Y > a\}} Y, \quad (a > 0).$$

10. Let $f(x) = (-1)^n n$, $1/(n+1) < x \leq 1/n$, $(n = 1, 2, \dots)$. Compute

$$\int_{(0,1)} f d\mu, \quad \int_{(1/(j+1), 1/j)} f d\mu \quad (j = 1, 2, \dots)$$

where μ is the Lebesgue measure. Can you say that

$$\int_{(0,1)} f d\mu = \sum_j \int_{(1/(j+1), 1/j)} f d\mu = \sum_j \int_{(1/(j+1), 1/j)} |f| d\mu?$$

Comment.

11. Let X be integrable on Ω . Then for every $\varepsilon > 0$, there exists a certain $\delta > 0$, such that

$$\left| \int_A X d\mu \right| < \varepsilon, \text{ for } A \subset \Omega \text{ and } \mu(A) < \delta.$$

12. The following is the generalisation of (5.39). Let X be a r.v. with d.f. $F(x)$ and $E|X|^k < \infty$. Then

$$EX^k = k \int_0^\infty x^{k-1} [(1 - F(x)) + (-1)^k F(-x)] dx, \quad (k = 1, 2, \dots).$$

If X is symmetric about zero, and $EX^{2k} < \infty$, then

$$E(X^{2k}) = 4k \int_0^\infty x^{2k-1} (1 - F(x)) dx, \quad (k = 1, 2, \dots).$$

13. If X has d.f. F and $E|X|^\alpha < \infty$, $\alpha > 0$, then

$$\lim_{x \rightarrow \infty} x^\alpha [1 - F(x) + F(-x)] = 0.$$

$$\left(\text{Hint: } \int_{-\infty}^\infty |X|^\alpha dF < \infty \Rightarrow \left(\int_{-\infty}^{-x_0} + \int_{x_0}^\infty \right) |X|^\alpha dF \rightarrow 0. \right)$$

14. If $F = F_1 * F_2$ (see problem 7 of chapter 4), and X, Y, Z have d.f.'s F, F_1 and F_2 , show that

$$EX = EY + EZ,$$

$$\text{Var } X = \text{Var } Y + \text{Var } Z,$$

$$EYZ = EY \cdot EZ.$$

15. Find the variance of the rectangular distribution on (a, b) . Hence obtain a lower bound to the probability of the interval

$$\{(3a - b)/2, (3b - a)/2\}.$$

16. Prove that

$$P[X \geq C] = e^{-Cr} E(e^{rX}), \quad r > 0.$$

17. (Bernstein) Let $M(r) = E(e^{r(X-\mu)})$, $\mu = EX$, $r > 0$. Then

$$P\left[X \geq \mu + \frac{t + \log M(r)}{r}\right] \leq e^{-t}, \quad t > 0.$$

Choosing r such that $[t + \log M(r)]/r$ is a minimum, makes the inequality sharp.

18. If $v_p = E|X|^p$, show that $\log v_r$ is convex in r and $(1/r) \log v_r$ is non-decreasing in r . Hence prove Liapounov's inequality,

$$\log v_{n+p} \leq \frac{q}{p+q} \log v_n + \frac{p}{p+q} \log v_{n+p+q} \quad (p > 0, q > 0).$$

$$\left(\text{Hint: Use Schwarz inequality to show that } \partial^2 \left(\frac{\log v(r)}{r} \right) / \partial r^2 \geq 0 \right)$$

19. Show that if X is bounded a.s.,

$$\lim_{r \rightarrow \infty} E^{1/r} |X|^r = \text{a.s. } \sup |X| = s.$$

$$\left(\text{Hint: } s \geq E^{1/r} |X|^r \geq E^{1/r} (|X|^r I(|X| > C)) \geq C P^{1/r} [|X| \geq C] \rightarrow s \text{ as } r \rightarrow \infty \text{ and } C \uparrow s \right)$$

20. Give examples of r.v.'s for which the following inequalities become equalities,

- (a) C_r — inequality,
- (b) Hölder's inequality,
- (c) Basic inequality.

21. If X has the d.f. given by

$$F(x) = \begin{cases} 0, & x < 0 \\ 1 - p \int_0^\infty e^{-x|y-y|} dy, & 0 \leq x < \infty, 0 < p \leq 1, \end{cases}$$

show that all the moments of X are finite, but the m.g.f. does not exist.

(Hint: $EX^k = p(k!)^2$, but $\sum_0^\infty (k!) \theta^k$ does not converge for any $\theta > 0$).

6. Convergence of Random Variables

6.1 CONVERGENCE IN PROBABILITY

(a) SET OF MUTUAL CONVERGENCE

A sequence of r.v.'s $\{X_n\}$ is said to converge to a r.v. X , if $\{X_n(\omega)\}$ converges to $X(\omega) < \infty$ as $n \rightarrow \infty$, for all $\omega \in \Omega$. In this case $\{X_n\}$ is said to converge to X everywhere. If $X_n(\omega)$ converges to $X(\omega)$ only for $\omega \in C$, then C is called the set of convergence of $\{X_n\}$. We know that $C \in \mathcal{A}$ and $\lim X_n$ is a r.v. (see section 2.3). $\{9\}$

Evidently, C is the set of all $\omega \in \Omega$, at which, whatever be $\epsilon > 0$, $|X_n(\omega) - X(\omega)| < \epsilon$, for all n greater than some $n = N_\epsilon(\omega)$ sufficiently large. Symbolically, for $r = n + m$, $m \geq 1$,

$$\begin{aligned} C &= [\omega : X_n(\omega) \rightarrow X(\omega)], \\ &= \bigcap_{\epsilon > 0} \bigcup_n \bigcap_m [\omega : |X_{n+m}(\omega) - X(\omega)| < \epsilon]. \end{aligned} \quad (6.1)$$

Equivalently, replacing 'for every $\epsilon > 0$ ' by 'for every $1/k$, $k = 1, 2, \dots$ ' we have

$$C = \bigcap_k \bigcup_n \bigcap_m [\omega : |X_{n+m}(\omega) - X(\omega)| < 1/k]. \quad (6.2)$$

Since C is now obtained from countable operations on measurable sets, C is measurable. This provides a direct proof that $C \in \mathcal{A}$.

It is well known that if $X_n(\omega) \rightarrow X(\omega)$, where $X(\omega)$ is finite, then $X_m(\omega) - X_n(\omega) \rightarrow 0$ as $m, n \rightarrow \infty$ and conversely. This is called the *Cauchy criterion* for convergence. Thus, $X_n \rightarrow X < \infty$, iff $X_n - X_m \rightarrow 0$ or equivalently $X_{n+m} - X_n \rightarrow 0$ as $n \rightarrow \infty$ uniformly in m . If a sequence of r.v.'s converges in the Cauchy sense, it is said to converge *mutually*. If C' is the set of *mutual convergence*

$$\begin{aligned} C' &= [\omega : X_{n+m}(\omega) - X_n(\omega) \rightarrow 0], \\ &= \bigcap_k \bigcup_n \bigcap_m [\omega : |X_{n+m}(\omega) - X_n(\omega)| < 1/k]. \end{aligned} \quad (6.3)$$

Since mutual convergence implies convergence, $C' \subset C$. Since X is finite on C , $C' = C$.

(b) DEFINITION

A sequence of r.v.'s $\{X_n\}$ is said to converge to X in probability, denoted

$$X_n \xrightarrow{P} X, \text{ if for every } \varepsilon > 0, \text{ as } n \rightarrow \infty, \\ P[|X_n - X| \geq \varepsilon] \rightarrow 0; \quad (6.4)$$

equivalently, if for every $\varepsilon > 0$, as $n \rightarrow \infty$

$$P[|X_n - X| < \varepsilon] \rightarrow 1.$$

This concept plays an important role in statistics. Consistency of estimators, weak laws of large numbers (to be discussed later) are instances of this concept. Heuristically, it means that the difference between X_n and X is likely to be small with large probability, for n large. It does not imply that $X_n(\omega) - X(\omega)$ will be small for every ω , even for n sufficiently large.

Two r.v.'s X and X' are said to be *equivalent*, if $X = X'$ a.s. The set of all ω , where X and X' are not equal is

$$[\omega : X(\omega) \neq X'(\omega)] = \bigcup_k [X - X' \geq 1/k]. \quad (6.5)$$

X and X' would be equivalent if $P[X \neq X'] = 0$ or equivalently if $P[|X - X'| \geq 1/k] = 0$ for all k .

LEMMA 6.1: $X_n \xrightarrow{P} X$ and $X_n \xrightarrow{P} X' \Rightarrow X$ and X' are equivalent.

This lemma shows that a sequence of r.v.'s cannot converge in probability to two essentially different r.v.'s. The proof follows from (6.5) and from the inequality

$$P[|X - X'| \geq 1/k] \leq P[|X - X_n| \geq 1/2k] \cup P[|X' - X_n| \geq 1/2k], \\ \leq P[|X - X_n| \geq 1/2k] + P[|X' - X_n| \geq 1/2k], \quad (6.6)$$

since the terms of the R.H.S. of (6.6) tend to zero as $n \rightarrow \infty$, whatever be k .

LEMMA 6.2: $X_n \xrightarrow{P} 0$ if $E|X_n|^r \rightarrow 0$. (6.7)

This Lemma provides us with a set of sufficient conditions for the convergence in probability. Replacing X_n by $(X_n - X)$ in (6.7), since

$$X_n \xrightarrow{P} X < \infty \text{ iff } X_n - X \xrightarrow{P} 0,$$

$E|X_n - X|^r \rightarrow 0$, implies that $X_n \xrightarrow{P} X$. The proof follows from Markov's inequality (5.75), since

$$P[|X_n| \geq \varepsilon] \leq E|X_n|^r / \varepsilon^r$$

and the R.H.S. tends to zero as $n \rightarrow \infty$. In particular, it follows that

$$X_n \xrightarrow{P} c, \text{ a constant, if } E(X_n - c)^2 \rightarrow 0.$$

(c) CRITERION FOR CONVERGENCE IN PROBABILITY

The following Lemma 6.3 gives us a set of necessary and sufficient

conditions for convergence in probability. It is based on the basic inequality (5.40).

LEMMA 6.3: $X_n \xrightarrow{P} 0$ iff $E\left(\frac{|X_n|}{1+|X_n|}\right) \rightarrow 0$, as $n \rightarrow \infty$. (6.8)

Proof: For any X , the r.v. $|X|/(1+|X|)$ is bounded by unity. Taking $g(x) = |x|/(1+|x|)$ in (5.70), since a.s. $\sup g(x) = 1$, for $\varepsilon > 0$

$$E\left(\frac{|X_n|}{1+|X_n|}\right) - \frac{\varepsilon}{1+\varepsilon} < P[|X_n| \geq \varepsilon] \leq E\left(\frac{|X_n|}{1+|X_n|}\right) \Big/ \left(\frac{\varepsilon}{1+\varepsilon}\right). \quad (6.9)$$

From the R.H.S. of the above inequality, we see that

$$E\left(\frac{|X_n|}{1+|X_n|}\right) \rightarrow 0 \Rightarrow P[|X_n| \geq \varepsilon] \rightarrow 0.$$

From the L.H.S. of the above inequality, we have

$$P[|X_n| \geq \varepsilon] \rightarrow 0 \Rightarrow E\left(\frac{|X_n|}{1+|X_n|}\right) - \frac{\varepsilon}{1+\varepsilon} \rightarrow 0.$$

But $\frac{|X_n|}{(1+|X_n|)}$ is non-negative r.v., so that $E\left(\frac{|X_n|}{(1+|X_n|)}\right) \geq 0$. Since ε is arbitrary, proof of the lemma is complete. (*)

LEMMA 6.4: $X_n \xrightarrow{P} X \Rightarrow X_n - X_m \xrightarrow{P} 0$ as $n, m \rightarrow \infty$. (6.10)

Proof: Since X is a real r.v. it is finite a.s. Leaving out this null set,

$$[|X_n - X_m| \geq \varepsilon] \subset [|X_n - X| \geq \varepsilon/2] \cup [|X_m - X| \geq \varepsilon/2]$$

and hence

$$P[|X_n - X_m| \geq \varepsilon] \leq P[|X_n - X| \geq \varepsilon/2] + P[|X_m - X| \geq \varepsilon/2].$$

The R.H.S. of this inequality tends to zero as $n, m \rightarrow \infty$. The Lemma follows. (*)

This lemma shows that convergence in probability implies mutual convergence in probability. Converse of this lemma is also true and will be proved in the next section. Now we shall discuss a few examples. It may be noted that the concept of convergence in probability is a particular case of convergence in a certain measure. In fact $\{f_n\}$ is said to converge to f in measure μ if $\mu[|f_n - f| > \varepsilon] \rightarrow 0$ as $n \rightarrow \infty$, denoted $f_n \xrightarrow{\mu} f$. Since probability measure is normed, it possesses some special properties.

Example 6.1. Let $\Omega = [0, 1]$ and the P -measure is the Lebesgue measure on $[0, 1]$. Let

$$X_n(\omega) = \begin{cases} 1, & \text{if } \omega \in [0, 1/n), \\ 0, & \text{otherwise.} \end{cases}$$

Thus, $X_n = I(A_n)$, where $A_n = [\omega: \omega \in [0, 1/n]]$. $P(A_n) = 1/n$ and as $n \rightarrow \infty$, $PA_n \rightarrow 0$. Hence, as $n \rightarrow \infty$,

$$P[|X_n| > \varepsilon] = P[I(A_n) = 1] = PA_n \rightarrow 0,$$

implying that $X_n \rightarrow 0$ in probability or in Lebesgue measure. In fact

$$\begin{aligned} \lim X_n(\omega) &= 1, \text{ if } \omega = 0, \\ &= 0, \text{ if } \omega \in (0, 1). \end{aligned}$$

Then $\lim X_n = I\{0\} = X$ and

$$P[|X_n - X| \geq \varepsilon] = (0, 1/n),$$

the probability measure of which tends to zero as $n \rightarrow \infty$.

In general, if $\{A_n\}$ is such that $PA_n \rightarrow 0$, then for any real r.v. X ,

$$X I(A_n) \xrightarrow{P} 0, \quad (6.11)$$

because $P[|X I(A_n)| \geq \varepsilon] \leq PA_n \rightarrow 0$, whatever be $\varepsilon > 0$.

(d) CLOSURE UNDER ARITHMETIC OPERATIONS

The following Theorem (6.1), shows that the operation of convergence in probability is closed under arithmetic operations in the same manner as ordinary convergence. It is sometimes attributed to Slutsky.

THEOREM 6.1: Let $X_n \xrightarrow{P} X$ and $Y_n \xrightarrow{P} Y$. Then

- (i) $a X_n \xrightarrow{P} aX$, (a , real),
- (ii) $X_n + Y_n \xrightarrow{P} X + Y$,
- (iii) $X_n Y_n \xrightarrow{P} XY$,
- (iv) $X_n/Y_n \xrightarrow{P} X/Y$, if $P[Y_n = 0] = 0$, $\forall n$, and $P[Y = 0] = 0$.

Proof: (i) If $a = 0$, the result is trivially true. Suppose $a \neq 0$. For every $\varepsilon > 0$,

$$\begin{aligned} P[|aX_n - aX| \geq \varepsilon] &= P[|a| |X_n - X| \geq \varepsilon], \\ &= P[|X_n - X| \geq \varepsilon/|a|] \rightarrow 0, \end{aligned}$$

as $n \rightarrow \infty$. Hence (i) is satisfied.

(ii) In order to establish (ii) we see that

$$\begin{aligned} P[|(X_n + Y_n) - (X + Y)| \geq \varepsilon] &= P[|(X_n - X) + (Y_n - Y)| \geq \varepsilon], \\ &\leq P[|X_n - X| + |Y_n - Y| \geq \varepsilon], \\ &\leq P[|X_n - X| \geq \varepsilon/2] + P[|Y_n - Y| \geq \varepsilon/2]. \end{aligned}$$

Since both the terms on the R.H.S. tend to zero as $n \rightarrow \infty$, we have (ii). From (i) and (ii), any linear function $\{aX_n + bY_n\}$ of $\{X_n\}$ and $\{Y_n\}$

converges to $aX + bY$ in probability, if $X_n \xrightarrow{P} X$ and $Y_n \xrightarrow{P} Y$.

$$(iii) \quad P[|X_n Y_n - XY| \geq \varepsilon] = P[|X_n Y_n - X_n Y + X_n Y - XY| \geq \varepsilon] \quad \text{why less than } \varepsilon/2 \\ \leq P[|X_n(Y_n - Y)| \geq \varepsilon/2] + P[|Y(X_n - X)| \geq \varepsilon/2]. \quad (6.13)$$

$$\text{But} \quad [|X_n(Y_n - Y)| \geq \varepsilon/2] = [|X_n(Y_n - Y)| \geq \varepsilon/2, |Y_n - Y| \geq \delta] \\ \cup [|X_n(Y_n - Y)| \geq \varepsilon/2, |Y_n - Y| < \delta], \\ \subset [|Y_n - Y| \geq \delta] \cup [|X_n| \geq \varepsilon/2\delta].$$

Hence

$$P[|X_n(Y_n - Y)| \geq \varepsilon/2] \leq P[|Y_n - Y| \geq \delta] + P[|X_n| \geq \varepsilon/2\delta]. \quad (6.14)$$

Since $\{X_n\}$ converges to X , which is bounded in probability, X_n 's are uniformly bounded in probability (see below for definition). Hence, the second term on the R.H.S. of (6.14) can be made less than $\eta/2$ by taking δ sufficiently small, so that $\varepsilon/2\delta$ is sufficiently large. The first term on the R.H.S. of (6.14) can be made less than $\eta/2$ by taking n sufficiently large. Thus, the first term on the R.H.S. of (6.13) can be made smaller than η . Similarly the second term on the R.H.S. of (6.13) can be made smaller than η . Since η is arbitrary we have (iii).

(iv) In order to prove (iv), we may first prove that

$$P[|1/Y_n - 1/Y| \geq \varepsilon] \rightarrow 0, \quad (6.15)$$

if $Y_n \xrightarrow{P} Y$, and $P[Y_n = 0] = P[Y = 0] = 0 \forall n$, along the lines of the proof of (iii) (How?) and then use (iii). (*)

From (6.7) and (6.12), if $X_n \xrightarrow{P} X$ and $E(Y_n - c)^2 \rightarrow 0$, then

$$X_n + Y_n \xrightarrow{P} X + c, \quad (6.16)$$

$$X_n Y_n \xrightarrow{P} cX,$$

where $c = EY$ is some constant. The following theorem is concerned with the convergence of a continuous function $f(X_n)$ of X_n as $X_n \xrightarrow{P} X$. It is due to Slutsky and Frechet.

THEOREM 6.2: If $f(x)$ is a continuous real valued function and $X_n \xrightarrow{P} X$,

$$\text{then} \quad f(X_n) \xrightarrow{P} f(X). \quad (6.17)$$

Proof: Since f is continuous and hence Borel, $f(X_n)$ and $f(X)$ are r.v.'s. Since X is a real r.v., it is bounded in probability i.e., for any $\varepsilon > 0$, there exists $a > 0$ such that

$$P[|X| > a] < \varepsilon.$$

Since $X_n \xrightarrow{P} X$, for any $\xi > 0$, there exists an n_0 such that for all $n > n_0$,

$$P[|X_n - X| < \xi] \geq 1 - \varepsilon.$$

Since $f(x)$ is uniformly continuous on $[-a-\xi, a+\xi]$, for every $\eta > 0$, there exists an ξ such that $|f(y) - f(x)| < \eta$, whenever $|y - x| < \xi$, for x and $y \in [-a-\xi, a+\xi]$. Therefore for $X_n(\omega), X(\omega)$ in $[-a-\xi, a+\xi]$, $|X_n(\omega) - X(\omega)| < \xi \Rightarrow |f(X_n(\omega)) - f(X(\omega))| < \eta$. The latter is implied by $|X_n(\omega) - X(\omega)| < \xi$ and $|X(\omega)| \leq a$. Hence

$$P[|X_n - X| < \xi, |X| \leq a] \leq P(|f(X_n) - f(X)| < \eta, |X| \leq a). \quad (6.18)$$

$$\begin{aligned} \text{But } P(|X_n - X| < \xi, |X| \leq a) + P(|X_n - X| < \xi, |X| > a) \\ = P(|X_n - X| < \xi) \geq 1 - \varepsilon. \end{aligned}$$

Therefore L.H.S. of (6.18) $\geq 1 - 2\varepsilon$ implying that for every $\eta > 0$, there exists n_0 , such that $n > n_0$,

$$P(|f(X_n) - f(X)| < \eta) > 1 - 2\varepsilon.$$

Since η is arbitrary, $f(X_n) \xrightarrow{P} f(X)$. (*)

We may note that Theorem (6.2) continues to hold even when X_n and X are random vectors.

6.2 CONVERGENCE ALMOST SURELY

(a) COMPARISON WITH CONVERGENCE IN PROBABILITY

X_n is said to converge to X *almost surely* (a.s.) or *almost certainly* or *strongly*, denoted $X_n \xrightarrow{\text{a.s.}} X$, if $X_n(\omega) \rightarrow X(\omega)$ for all ω , except those belonging to a null set N . Thus, symbolically,

$$X_n \xrightarrow{\text{a.s.}} X \text{ iff } X_n(\omega) \rightarrow X(\omega) < \infty \text{ for } \omega \in N^c,$$

where $PN = 0$. In this case the set of convergence of $\{X_n\}$ has probability unity. Thus, $\liminf X_n$ and $\limsup X_n$ are equivalent r.v.'s.

LEMMA 6.5: $X_n \xrightarrow{\text{a.s.}} X$ iff as $n \rightarrow \infty$

$$P\left(\bigcup_{k=n}^{\infty} [\omega : |X_k - X| \geq 1/r]\right) \rightarrow 0, \forall r, \text{ an integer.} \quad (6.19)$$

Proof: Now $X_n(\omega) \rightarrow X(\omega)$, if for arbitrary $r > 1$, there exists some $n_0(\omega, r)$ such that for all $k \geq n_0(\omega, r)$, $|X_k(\omega) - X(\omega)| < 1/r$. Moreover, $X_n \xrightarrow{\text{a.s.}} X$ implies that $P[X_n \rightarrow X] = 1$. Thus, from (6.2) and de Morgan rule,

$$P\left(\bigcup_r \bigcap_n \bigcup_{k=n}^{\infty} [\omega : |X_k(\omega) - X(\omega)| \geq 1/r]\right) = 0,$$

i.e. for each r ,

$$P\left(\bigcap_n \bigcup_{k=n}^{\infty} [\omega : |X_k(\omega) - X(\omega)| \geq 1/r]\right) = 0;$$

equivalently for each r ,

$$\lim_{n \rightarrow \infty} P\left(\bigcup_{k=n}^{\infty} [|X_k - X| \geq 1/r]\right) = 0.$$

Hence we have the Lemma. (*)

Here (6.19) may be replaced by the complementary condition

$$P\left(\bigcap_{k=n}^{\infty} [|X_k - X| < 1/r]\right) \rightarrow 1.$$

COROLLARY 6.1: $X_n \xrightarrow{\text{a.s.}} X \Rightarrow X_n \xrightarrow{P} X$. (6.20)

This Corollary follows immediately, since for $\varepsilon > 1/r$,

$$\bigcup_{k=n}^{\infty} [|X_k - X| \geq 1/r] \supset [|X_n - X| \geq \varepsilon].$$

It follows from the Corollary and the Lemma 6.1 that $X_n \xrightarrow{\text{a.s.}} X$ and $X_n \xrightarrow{\text{a.s.}} X'$, implies that $X = X'$ a.s. In general, converse of Corollary (6.1) is not true. However, a partial converse is given by the following.

THEOREM 6.3: If $X_n \xrightarrow{P} X$, then there exists a subsequence $\{X_{n_k}\}$ of $\{X_n\}$ which converges a.s. to X .

Before proving the Theorem 6.3, we prove the following Lemma 6.6.

LEMMA 6.6: A sequence of r.v.'s converges a.s. to a r.v. iff the sequence converges mutually, a.s.

Proof: We may note here that the statements ' $X_m(\omega) - X_n(\omega) \rightarrow 0$ as $m, n \rightarrow \infty$ ' and ' $X_{n+m}(\omega) - X_n(\omega) \rightarrow 0$ as $n \rightarrow \infty$ and then $m \rightarrow \infty$ ' are equivalent. Hence

$$X_n \xrightarrow{\text{a.s.}} X \Rightarrow X_n(\omega) \rightarrow X(\omega),$$

for all $\omega \in N^c$, where $PN = 0$. Since X is a real r.v., $X(\omega) < \infty$ for almost all ω and hence the exceptional set may be clubbed with N . Thus,

$$\begin{aligned} X_n \xrightarrow{\text{a.s.}} X &\Rightarrow X_n(\omega) - X_m(\omega) \rightarrow 0 \text{ as } m, n \rightarrow \infty, (\omega \in N^c), \\ &\Rightarrow X_n - X_m \xrightarrow{\text{a.s.}} 0. \end{aligned} \quad (6.21)$$

Conversely, if as $n \rightarrow \infty$, $X_{n+m} - X_n \xrightarrow{\text{a.s.}} 0$ for every m , i.e., $X_{n+m}(\omega) - X_n(\omega) \rightarrow 0$, for $\omega \in N_m^c$, then, $\{X_{n+m}(\omega)\}$ converges to a finite limit, $X(\omega)$, say. Hence

$$X_n(\omega) \rightarrow X(\omega) \text{ for } \omega \in \bigcap_{m=1}^{\infty} N_m^c = N^c \text{ say,}$$

where $N = \bigcup_{m=1}^{\infty} N_m$,

is a countable union of null sets and is hence a null set. Thus we have the Lemma. (*)

On account of the Lemmas 6.5 and 6.6, the criterion for a.s. convergence of a sequence $\{X_n\}$ of r.v.'s can be written as

$$P\left(\bigcup_{m=1}^{\infty} [|X_{n+m} - X_n| \geq \varepsilon]\right) \rightarrow 0, \quad (6.22)$$

as $n \rightarrow \infty$, whatever be $\varepsilon > 0$.

Proof (of Theorem 6.3): From Lemma (6.4)

$$X_n \xrightarrow{P} X \Rightarrow X_n - X_m \xrightarrow{P} 0, \text{ i.e. } X_{n+m} - X_n \xrightarrow{P} 0, \text{ as } m, n \rightarrow \infty.$$

For every integer k , $\exists$ an integer $n(k)$ such that for $n \geq n(k)$ and all m

$$P[|X_{n+m} - X_n| \geq 1/2^k] < 1/2^k. \quad (6.23)$$

Let $n_1 = n(1)$, $n_2 = \max(n_1 + 1, n(2))$, $n_k = \max(n_{k-1} + 1, n(k))$, ($k = 2, 3, \dots$) so that $n_1 < n_2 < \dots \rightarrow \infty$ and (6.23) is satisfied for $n > n_k$. Let $\{X'_k\} = \{X_{n_k}\}$ be a subsequence of $\{X_n\}$. We shall show that $\{X'_k\}$ is the required subsequence converging almost surely to a certain X' , (say). For this, from (6.22) we have to show that, as $n \rightarrow \infty$

$$P[\bigcup_m |X'_{n+m} - X'_n| \geq \varepsilon] \rightarrow 0.$$

Let $A_k = [|X'_{k+1} - X'_k| \geq 1/2^k]$. Then from (6.23) $PA_k < 2^{-k}$. Let $B_n = \bigcup_{k \geq n} A_k$, then $PB_n < \sum_{k \geq n} 2^{-k} = 2^{-n+1}$. For n large enough so that $2^{-n+1} < \varepsilon$, for $\omega \in B_n^c = \bigcap_{k \geq n} A_k^c$ and for all m ,

$$\begin{aligned} |X'_{n+m}(\omega) - X'_n(\omega)| &\leq \sum_{k=n}^{n+m-1} |X'_{k+1}(\omega) - X'_k(\omega)|, \\ &\leq \sum_{k \geq n} |X'_{k+1}(\omega) - X'_k(\omega)| < 2^{-n+1} < \varepsilon. \end{aligned}$$

Hence $|X'_{n+m}(\omega) - X'_n(\omega)| \geq \varepsilon \Rightarrow \omega \in B_n$ and therefore,

$$P(\bigcup_m [|X'_{n+m} - X'_n| \geq \varepsilon]) \leq PB_n < 2^{-n+1} \rightarrow 0,$$

as $n \rightarrow \infty$. Thus, by Lemma 6.6, $\{X'_n\}$ converges a.s. to X' , say.

Hence by (6.20) $X'_n \xrightarrow{P} X'$. We have to prove that $X' = X$. This follows from the fact that $\{X'_n\}$, being a subsequence of $\{X_n\}$, converges to the same limit X in probability. By Lemma 6.1, $X = X'$ a.s. (*)

(b) COMPLETENESS OF THE SPACE OF ALL R.V.'S

A certain topological space, in which a certain mode of convergence is defined, is said to be complete, if mutual convergence in that mode is equivalent to ordinary convergence in that mode. Consider the space $\mathcal{X}$ of all a.s. finite r.v.'s. Let $\{X_1, X_2, \dots, X_n, \dots\}$ be a sequence of elements of $\mathcal{X}$. Let the mode of convergence defined on $\mathcal{X}$ be

convergence a.s. Since $X_n \xrightarrow{\text{a.s.}} X \Leftrightarrow X_n - X_m \xrightarrow{\text{a.s.}} 0$, a.s. convergence

and a.s. mutual convergence are equivalent. Thus $\mathcal{X}$ is a.s. complete under 'convergence a.s.' Theorem 6.4 given below implies that it is complete under convergence in probability also.

THEOREM 6.4: $X_m - X_n \xrightarrow{P} 0 \Leftrightarrow X_n \xrightarrow{P} X$, where X is some r.v.

Proof: By Lemma 6.4, $X_n \xrightarrow{P} X \Rightarrow X_{n+m} - X_n \xrightarrow{P} 0$.

To prove the converse, let $X_{n+m} - X_n \xrightarrow{P} 0$. Then by Theorem 6.3, there exists a subsequence $\{X(n_k)\}$ which converges a.s. mutually. Hence from Lemma 6.6, $X(n_k) \longrightarrow$ to some X a.s. Thus, for every $\varepsilon > 0$,

$$P[|X - X_n| \geq \varepsilon] \leq P[|X - X(n_k)| \geq \varepsilon/2] + P[|X(n_k) - X_n| \geq \varepsilon/2].$$

R.H.S. tends to zero as $n_k \rightarrow \infty$ and hence $X_n \xrightarrow{P} X$. (*)

Example 6.2. Let us consider the double sequence $\{X_{nk}\}$ of r.v.'s on $\Omega = [0, 1]$ with Lebesgue measure as P , defined as follows

$$X_{nk}(\omega) = 1, \text{ for } \omega \in [(k-1)/n, k/n],$$

$$\begin{pmatrix} k = 1, 2, \dots, n; \\ n = 1, 2, \dots \end{pmatrix}$$

$$= 0, \text{ otherwise.}$$

$X_{11}, X_{21}, X_{22}, X_{31}, X_{32}, X_{33}, \dots$ being the sequence. Taking $X \equiv 0$, we shall investigate the convergence of $\{X_{nk}\}$ to X . Now

$$P[\omega; |X_{nk}(\omega) - X(\omega)| > \varepsilon] = P[X_{nk} = 1] = 1/n \rightarrow 0.$$

Hence $\{X_{nk}\} \rightarrow 0$ in probability. Consider

$$\bigcup_{n=m}^{\infty} \left(\bigcup_{k=1}^n [X_{nk} = 1] \right) = [X_{m1} = 1] \cup \dots \cup [X_{mm} = 1] \cup \dots$$

$$= [0, 1].$$

Hence $P[\bigcup_{n=m}^{\infty} \bigcup_{k=1}^n |X_{nk} - X| > \varepsilon] = 1$, for all m . Thus, $\{X_{nk}\}$ does not converge a.s. Note that the subsequence $\{X_{n1}\}$ of $\{X_{nk}\}$ converges a.s. (Prove!).

In Theorem 6.2, we can replace convergence in probability by convergence a.s., by replacing the set $[|X_n - X| < \xi]$ by

$$\bigcap_{k=n}^{\infty} [|X_k - X| < \xi]$$

and noting that

$$\bigcap_{k=n}^{\infty} [|X_k - X| < \zeta, |X| \leq a] \subset \bigcap_{k=n}^{\infty} [|f(X_k) - f(X)| < \eta, |X| \leq a].$$

Evidently, Theorem 6.1 is also valid for convergence a.s. (Prove!).