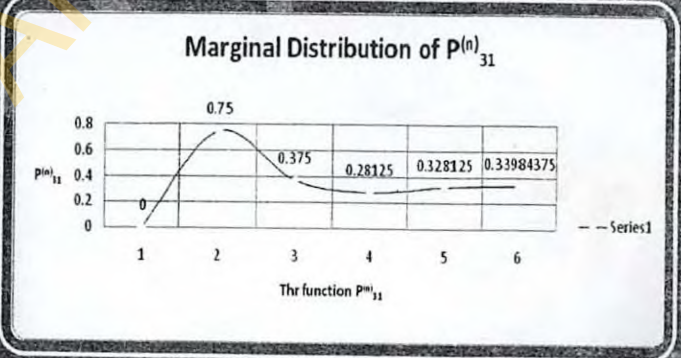
INTRODUCTION TO PROBABILITY AND STOCHASTIC PROCESSES

WITH APPLICATIONS



$$\bigcup_{i=1}^{\infty}(A_i) = \sum_{i=1}^{\infty} P(A_i) \quad i = 1, 2, 3 \ldots$$

and $A_i \cap A_j = \phi$

Shittu, Olanrewaju I.
Otekunrin, Oluwaseun O.
Udomboso, Christopher G.
Adepoju, Kazeem A.

INTRODUCTION TO PROBABILITY AND STOCHASTIC PROCESSES

FP

ABOUT THE BOOK

Probability theory has a very long history dating back to the seventeenth century. It is a well-established branch of mathematics that has applications in every area of human discipline and daily experiences.

This is an introductory textbook dealing with probability and stochastic processes. It is designed for undergraduate and postgraduate students in Statistics, Mathematics, the Physical and Social Sciences, Engineering and Computer Science. It presents a thorough treatment of probability and stochastic ideas and methods necessary for a firm understanding of the subject. The text can be used in a variety of course lengths, levels, and areas of emphasis.

The material is divided into three parts. The first part covers basic probability topics for undergraduate students. The second part covers advanced probability topics that are of interest to postgraduate students while the third part deals with topics in stochastic processes that are taught both at undergraduate and postgraduate levels.
Very little statistical background is assumed in order to obtain full benefits from the use of the text. Also, numerous examples and practice questions are included to aid understanding of all the subject areas covered by the book.

The publication of this book is a demonstration of our commitment to the provision of relevant and current materials for Statistics students in higher institutions of learning.

FASCO PUBLISHERS
ISBN 978-978-52890-0-8

INTRODUCTION TO PROBABILITY AND STOCHASTIC PROCESSES (WITH APPLICATIONS)

First Published: May, 2014

Fasco Publishers, 67 Gbadebo Street, Mokola, Ibadan.

Tel: 08032934309, 08051036056

ISBN 978 – 978 – 52890 – 0 – 8

Printed in Ibadan by Fasco Printing Works

FOREWARD

This impressive book by Shittu O.I., Otekunrin O. A., Udomboso C. G., and Adepoju K A. (of the Department of Statistics, University of Ibadan, Nigeria) encompasses the essence of probability, and stochastic processes under a common shade. The authors are to be commended for their lucid presentation as well as their broad coverage of the subject matter.

A special feature of the book is that the exercises and the test form an integrated pattern. These exercises are designed to encourage the student to reread the text, practice them and become thoroughly familiar with the techniques described. This will help in impressing on the student the methods and logic of establishing the techniques.

Student involved in statistics-oriented discipline, and professional statistician in a wide variety of fields will find this book a highly useful volume for study, and application. This is a scholarly undertaking by the four authors, and I have a full appreciation for the job nicely done.

G.N. Amahia (Ph.D)
Professor of Statistics
University of Ibadan
Ibadan.

PREFACE

Probability theory has a very long history dating back to the seventeenth century. It is a well-established branch of mathematics that has applications in every area of human discipline and daily experiences.

This is an introductory textbook dealing with probability and stochastic processes. It is designed for undergraduate and postgraduate students in Statistics, Mathematics, the physical and social sciences, engineering and computer science. It presents a thorough treatment of probability and stochastic ideas and methods necessary for a firm understanding of the subject. The text can be usedin a variety of course lengths, levels, and areas of emphasis.

The material is divided into three parts. The first part covers basic probability topics for undergraduate students. The second part covers advanced probability topics that are of interest to postgraduate students while the third part deals with topics in stochastic processes that are taught both at undergraduate and postgraduate levels.
Very little statistical background is assumed in order to obtain fullbenefits from the use of the text. Also, numerous examples and practice questions are included to aid understanding of all the subject areas covered by the book.

The publication of this book is a demonstration of our commitment to the provision of relevant and current materials for Statistics students in higher institutions of learning of the authors.

This text which cannot be said to be exhaustive was developed from the years of learning and teaching of probability and stochastic processes. While we claim responsibility for some errors that could have been made inadvertently in this first edition, we welcome comments and objective criticisms from the users of this book.

TABLE OF CONTENTS

PART ONE

CHAPTER 1
THE MATHEMATICS OF CHOICE

1.0 Introduction

Many real life situations requires enumerating the number of possible ways of taking a number of decisions out of many available ones, or the number of ways an event can occur, number of possible outcomes of an experiment. All of the above require the act of either counting, choosing, arranging or a combination of the above. Therefore, it is just apt to introduce the reader first, to some basic principles of counting.

1.1 Fundamental Principle of Counting

If one experiment can result in n possible outcomes an experiment can result in k possible outcomes, then nk is the total number of possible outcomes from the two experiments.

Consider a finite sequence of decisions. Suppose the number of choices for each individual decision is independent of decisions made previously in the sequence. Then, the number of ways to make the whole sequence of decisions is the product of these numbers of choices i.e. $n!$

Example 1: The number of four-letter words that can formed by rearranging the letters in the word PLAN is 4! =24.

PLAN	PLNA	PALN	PANL	PNLA	PNAL
LPAN	LPNA	LAPN	LANP	LNPA	LNAP
APLN	APNL	ALPN	ALNP	ANPL	ANLP
NPLA	NPAL	NLPA	NLAP	NAPL	NALP

1.1.2 The Second Counting Principle (The Principle of Inclusion and Exclusion)

If a set is the disjoint union of two (or more) subsets, then the number of elements in the set is the sum of the numbers of elements in the subsets. i.e.

$n(A \cup B) = n(A) + n(B)$ implying that $|A \cup B| = |A| + |B|$ if A and B are disjoint.

Theorem 1: $|A \cup B| < |A| + |B|$ if A and B are not disjoint.

This is because $|A| + |B|$ counts every element of $A \cap B$ twice. Let us illustrate this with the following example.

Example 2: If $A = (2, 3, 4, 5, 6)$. $|A| = 5$ and $B = (3, 4, 5, 6, 7)$. $|B| = 5$

then. $|A| + |B| = 10$

$A \cup B = 2, 3, 4, 5, 6, 7$

$|A \cup B| = 6$

Since A and B are not disjoint, $|A \cup B| < |A| + |B|$

Compensating for this double counting yields the formula

$|A \cup B| = |A| + |B| - |A \cap B|$............eqn.(1)

From our example, $A \cap B = 3, 4, 5, 6$

$|A \cap B| = 4$

$|A \cup B| = 5 + 5 - 4$

$= 6$

Thus proving equation (1)

Theorem 2: $|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|$ for three sets A, B and C.

Proof:

We know from equation (1) above that $|A \cup B| = |A| + |B| - |A \cap B|$

Then, for 3 sets, $|A \cup B \cup C| = |A \cup [B \cup C]|$

$$= |A| + |B \cup C| - |A \cap [B \cup C]|$$

Applying equation (1) to $|B \cup C|$ gives

$|A \cup B \cup C| = |A| + [|B| + |C| - |B \cap C|] - |A \cap [B \cup C]|$.........eqn (2)

Because $A \cap [B \cup C] = (A \cap B) \cup (A \cap C)$, we can apply equation(1) again to obtain

$|A \cap [B \cup C]| = |A \cap B| + |A \cap C| - |A \cap B \cap C|$.........eqn (3)

Finally, a combination of equations (2) and (3) yields

$|A \cup B \cup C| = [|A| + |B| + |C|] - [|A \cap B| + |A \cap C| + |B \cap C|] + |A \cap B \cap C|$.........eqn (4)

Thus proving theorem 2.

From this derivation, we notice that an element of $A \cap B \cap C$ is counted 7 times in equation(4), the first 3 times with a plus sign, then 3 times with a minus sign and then once more with a plus sign.

Example 1.1: If $A = \{1,2,3,4\} B = \{3,4,5,6\}$ $C = \{2,4,6,7\}$ then

$$A \cup B \cup C = \{1,2,3,4,5,6,7\}$$

$|A \cup B \cup C| = 7$……………….(a)

$|A| = 4$

$|B| = 4$

$|C| = 4$

$|A| + |B| + |C| = 12$

$$A \cap B = 3,4, A \cap C = 2,4, B \cap C = 4,6$$

In this example, $|A \cap B| = |A \cap C| = |B \cap C| = 2$ so that

$|A \cap B| + |A \cap C| + |B \cap C| = 6$ and

$$A \cap B \cap C = 4, |A \cap B \cap C| = 1$$

Therefore, $|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|$

$= 12 - 6 + 1 = 7$……………… (b)

Thus, (a) = (b) thus establishing theorem 2.

Generally, the Principle of Inclusion and Exclusion (PIE) states that:

If $A_1, A_2, \ldots, A_n$ are finite sets, the cardinality of their union

$$|A_1 \cup A_2 \cup \ldots \cup A_n| =$$

$$\sum_{i=1}^{n}|A_i| - \sum_{1\le i<j\le n}|A_i \cap A_j| + \sum_{1\le i<j<k\le n}|A_i \cap A_j \cap A_k| - \cdots + (-1)^{n+1}|\cap A_i|$$

Proof:

On the left is the number of elements in the union of n sets. On the right, we first count elements in each of the sets separately and add them up. If the sets A_i are not disjoint, the elements that belong to at least two of the sets A_i, or the intersections $A_i \cap A_j$, are counted more than once. We wish to consider every such intersection, but each only once. Since $A_i \cap A_j = A_j \cap A_i$, we should consider only pairs (A_i, A_j) with $i < j$.

When we subtract the sum of the number of elements in such pairwise intersections, some elements may have been subtracted more than once. Those are the elements that belong to at least three of the sets A_i. We add the sum of the elements of intersections of the sets taken three at a time. (Note: the condition $i < j < k$ ensures that every intersection is counted only once)

The process continues with sums being alternately added or subtracted until we come to the last term which is the intersection of all sets A_i thus proving the theorem.

Let $S = A_1 \cup A_2 \cup \ldots \cup A_n$ and $A_i^{\,c} = S \backslash A_i$ then the PIE principle can also be expressed as

$|A_1^{\,c} \cap \ldots \cap A_n^{\,c}| =$

$$|S| - \sum_{i=1}^{n}|A_i| + \sum_{1\le i<j\le n}|A_i \cap A_j| - \sum_{1\le i<j<k\le n}|A_i \cap A_j \cap A_k| + \cdots - (-1)^{n+1}|\cap A_i|$$

Example 1.2: Let A be the subset of the first 700 hundred numbers $S = \{1,2,\ldots,700\}$ that are divisible by7. Find the number of elements in S that are not divisible by 7.

Solution:

$$A = \{7,14,21,28,35,42,49,\ldots\}$$

$$|A| = 100$$

$$|A^c| = |S| - |A|$$

$$= 700 - 100$$

$$= 600$$

Example 1.3: Find the number of integers from 1 to 1000 that are not divisible by 5, 6 and 8

Solution: Let A_1, A_2, A_3 be the subset consisting of those integers that are divisible by 5, 6 and 8. The number we are interested in is

$|A_1^{\,|} \cap A_2^{\,|} \cap A_3^{\,|}| = 1000 - |A_1| - |A_2| - |A_3| + |A_1 \cap A_2| + |A_1 \cap A_3| + |A_2 \cap A_3| - |A_1 \cap A_2 \cap A_3|$

$$|A_1| = \left\lfloor \frac{1000}{5} \right\rfloor = 200 \quad |A_2| = \left\lfloor \frac{1000}{6} \right\rfloor = 166 \quad |A_3| = \left\lfloor \frac{100}{8} \right\rfloor = 125$$

Note: The results for $|A_1|, |A_2|$ and $|A_3|$ were achieved by using the round down. notation $\lfloor\ \rfloor$ which involves the dropping of the fractional part.

To compute the number in a 2 and 3 – set interaction, we use the least common multiple (LCM) i.e.

$$|A_1 \cap A_2| = \left\lfloor \frac{1000}{30} \right\rfloor = 33$$

$$|A_1 \cap A_3| = \left\lfloor \frac{1000}{40} \right\rfloor = 25$$

$$|A_2 \cap A_3| = \left\lfloor \frac{1000}{24} \right\rfloor = 41$$

$$|A_1 \cap A_2 \cap A_3| = \left\lfloor \frac{1000}{12} \right\rfloor = 8$$

Thus, $|A_1^1 \cap A_2^1 \cap A_3^1| = 1000 - 200 - 166 - 125 + 33 + 25 + 41 - 8 = 600$

Important remarks:

Definition 1: The number of ordered selections of r elements chosen from an n-element set is $P(n, r)$.

$$\begin{aligned} P(n,r) &= n(n-1)(n-2)\ldots.(n-[r-1]) \\ &= n(n-1)(n-2)\ldots.(n-r+1) \\ &= \frac{n!}{(n-r)!} \\ &= r!\,C(n,r) \end{aligned}$$

Example 1.4: Suppose 6 members of Adeola's School Parent Teachers Association meet to select a 3-member delegation to represent the association at a statewide convention. If the laws stipulate that each delegation be comprised of a delegate, a first alternate and a second alternate. How many ways can the 6 members comply from among themselves?

Solution: $P(6,3) = 120$ ways or $3!\,C(6,3) = 120$ ways

Definition 2: The number of ways of making a sequence of r decisions each of which has n choices is n^r if order matters and the selection is with replacement.

Definition 3: The number of different ways to choose r things from n things with replacement if order does not matter is $C(r + n - 1,\ r)$

Example 1.5: How many different three letter "words" can be produced using the "alphabet" ALEXY?

Solution: Since there are no restrictions on the number of times a letter can be used, $5^3 = 125$ such words can be formed.

Example 1.6: At a fundraising luncheon, each of 50 patrons is given a numbered ticket, while its duplicate is placed in a bowl from which prize-winning numbers will be drawn.

i) If the prizes are #50, #100, and #150, how many outcomes are possible assuming that winning tickets are not returned to the bowl.

ii) If the prizes are the same, say, #70 each for the 3 prizes, how many outcomes are possible assuming that winning tickets are not returned to the bowl?

iii) If the winning tickets are returned to the bowl, how many outcomes are possible when the prizes are as in (i) and (ii) respectively?

Solution:

i) $P(50,3) = 117600$ different outcomes are possible

ii) Since the prizes are the same, then order is not important implying that there are $C(50,3) = 19600$ different possible outcomes.

iii) a) Different prizes, with replacement, order matters: $50^3 = 125000$

b) Same prizes, with replacement, order does not matter:

$C(3 + 50 - 1, 3) = 22100$

Note: In choosing with replacement, elements may be chosen more than once. If order does not matter, then we are only concerned with the multiplicity with which each element is chosen.

Table 1.1 summarizes the four scenarios that we have considered.

Table 1.1

	Order Matters	Order does not matter
Without replacement	$P(n,r)$	$C(n,r)$
With replacement	n^r	$C(r+n-1,r)$

1.3 Permutation: The ordered arrangement of n distinct items taking all or r of them at a time is called permutation. The items are usually assumed to be arranged on a line without replacement such that if two of the r objects are interchanged, it results into different permutation (arrangement).

The number of permutation of n items taking r at a time is denoted

$$n_{P_r} = \frac{n!}{(n-r)!}$$

This is the same as the number of ways, in which r spaces can be fill taking n different items at a time.

The first place is filled in n way, the second $(n-1)$ ways … and r place if filled in $(n+r+1)$ ways. This r places is filled in $(a-1)(n-2) \ldots (n-1r+1)!$ ways.

$$\therefore n_{P_r} = n(n-1)(n-2)\ldots(n-r-1)$$

The number of permutations of n distinct items taking all at a time is

$$n(n-1)(n-2)\ldots 3.2.1. = n! \text{ ways}$$

The symbol $n!$ is called m factorial and we define $0! = 1$.

Example 1.7: Evaluate 5_{P_s}

Solution:

$$5_{P_s} = \frac{5!}{(5-3)!}$$

$$= \frac{5 \times 4 \times 3! \times 2!}{2!}$$

$$= 5 \times 4 \times 3$$

$$= 60 \; ways$$

Example 1.4: If $12_{P_r} = 17160$, find r.

Solution:

$$13_{P_r} = 13(12)(11)\ldots(12-r+1) = 13\,(12)(11)(10)$$

$$= 13 - r \qquad \therefore r = 3$$

Example 1.8: How many different words of three letters can be formed with letters A, B, C, D, E and F no letter is repeated?

Solution: The first letter can be arranged in 6 ways

The second letter can be arranged in 5 ways

The third letter can be arranged in 4 ways.

Total number of arrangement is 6 x 5 x 4 x 3.

Alternatively

$$6_{P_3} = \frac{6!}{(6-3)!} = 6 \times 5 \times 4 = 120 \text{ ways.}$$

(A) **Permutation of n things, not all of which are distinct.**

The number of permutations of n things taking all at a time where p of them are alike of one kind, q are alike of another kind and r alike of the third kind is

$$\frac{n!}{p!q!r!}$$

Example 1.9: In how many ways can the letters of the word STATISTICS be arranged.

Solution:

T occurs 3 times

I occurs 2 times

s occurs 3 times.

So the number of possible arrangement is

$$\frac{10!}{3!2!3!} = 50400 \text{ ways.}$$

(B) **When certain things always or never occur:**

(i) **s item will always occur:** Given n items to arrange taking r at a time out of which S of them will always occur, keep aside the S items and arrange the remaining $(n-s)$ items taking $(r-s)$ at a time.

The S items can be arranged taking S at a time in r_{P_s} ways.

The total number of permutations is $n - S_{P_{r-s}} X \, r_{P_s}$.

(ii) **Never occur:** Leave out the S items and find the number of permutation of $(n-s)$ items taking r at a time, i.e.

$$n - S_{P_r} = \frac{(n-s)!}{(n-s-r)!}$$

Example 1.10: A committee of 7 representative of a class consists of class captain and his deputy. On a visit to the Head-teacher there are four seats. How many ways can the committee be seated it:

(i) there is no restriction

(ii) the class captain and his deputy must sit.

(iii) one of the students committed a crime and can not sit down even if there are enough seats, and

(iv) determine the probability of the event in (ii) and (iii) above.

Solution:

(i) When there is no restriction

$$n = 7, r = 4$$

$$7P_4 = \frac{7!}{(7-4)!} = 7 \times 6 \times 5 \times 4 = 840 \text{ ways}$$

(ii) keep aside the class captain and his deputy:

$$4_{P_2}(n-2)P_{r-2} = 5P_2 X\, 4P_2 = \frac{5!}{(5-3)!}\, x\, \frac{4!}{(4-2)!}$$

$$= 5\, x\, 4\, x\, 3\, x\, 12$$

$$= 12\, x\, 60$$

$$= 720 \text{ ways}$$

(iii) Leave out the criminal then we have

$$P_4 = P_4 = \frac{6!}{(6-4)!}$$

$$= 6\, x\, 5\, x\, 4\, x\, 3$$

$$= 360 \text{ ways}$$

(C) **Permutation when two things are not to occur together:**

Procedure

(a) Find permutation without restriction

(b) Find permutation when two things occur together.

(c) The difference between (a) and (b) gives the number of arrangement when two things do not occur together.

Example 1.11: In how many ways can 10 different books be arranged on a shelf if two particular books are not to stand together?

Solution:

If the two books are to stand together, regard the two books as one, then the number of arrangement is $219F_9 = 2\, x\, 9! = 72560$ ways. Number of arrangement without restriction is $P_{10} = 10! = 3628800$ ways so the permutation when the two books are not to stand together is

$$10! - 2\, x\, 9!$$

$$= 3628800 - 725760$$

$$= 2903040 \text{ ways}$$

Example 1.12: Letters of the word "ARRANGE' are to be arranged. Find the probability if:

(i) two r's do not occur together

(ii) if the two R's and two A's do occur together

Solution:

(i) Without restriction, number of arrangement is $\frac{7!}{2!2!} = 1260$ ways. When two R's occur together is $\frac{6!}{2!} = 360$ way when two R's do not occur together is $1260 - 360 = 900$ ways.

$$\text{P (two R's not occur together)} = \frac{900}{1260} = 0.714$$

(ii) If two R's and two A's do occur together we have (A, A) (R, R) N G. E i.e.,

$$P_5 = 5! = 120 \text{ ways.}$$

(D) **When the number of items not occurring together is more than two**

Some kind of logic would have to be applied here. It is better illustrated with an example.

Example 1.13: In how many ways can 5 blue cars and 4 red cards be arranged in a straight car park two red cars are not to stand together.

Solution: First, the first 5 cards are positioned as indicated below

X B X B X B X B X B X

The blue cars can be arranged in 5! ways. Now there are 6 vacant positions (marked X). The remaining 4 red cars can be arranged in $P_4 = 360$ ways. The required number of ways of parking 5 blue cars and 4 red cars is $5!\, X\, P_4$

$= 120 \ x \ 360$

$= 43200$ ways

(E) **When Items are repeated:**

The number of permutation of n different items taking r at a time, when each item may occur an number of times is n^r.

Example: 1.14: A die is rolled 4 times what is the sample space.

Solution:

A die has six faces, hence may occur in 6 ways.

The sample space is

$$6^4 = 1296$$

(F) **Formation of numbers with digits:**

The idea of permutation can be applied in the formation of numbers with digits. This is particularly useful in a raffle draw. Let us illustrate with a simple case.

Example 1.15: Suppose the five digits 1, 2, 3, 4, 5 are given. To find the total number of numbers which can be formed under different conditions.

(a) Without restriction $= P_5 = 5! = 120$ ways.

(b) Suppose 5 always occur in the tenth place. Now the tenth place is fixed, then the remaining four places can be fitted with four digits as $P_4 = 4! = 24$ ways. i.e.

1 2 3 4 5   2 1 3 5 4
1 2 4 5 3   2 1 4 5 3

1 3 2 5 4   2 3 1 5 4   x 2 = 24 ways
1 3 4 5 2   2 3 4 5 1
1 4 3 5 2   2 4 1 5 3

1 4 2 5 3   2 4 3 5 1

(c) Suppose we have to form a number divisible by 2. Then the unit's place must be occupied by 2 or 4 which can be arranged in 2 ways.

The remaining 4 digits can be fitted in

$$P_4 = 4! = 6 \text{ ways}$$

So, the total number of numbers divisible by $2 = 24 \ x \ 2 = 48$.

(d) Suppose we have to form numbers which begin with 1 and <u>end</u> with 3. Here the first and the last places are fixed.

Then, the remaining 3 digits can be filled in.

$$P_3 = 3! = 6 \text{ ways}$$

i.e.

1 2 4 5 3
1 2 5 4 3
1 4 2 5 3
1 4 5 2 3 = 6 ways
1 5 2 4 3
1 5 4 2 3

(e) Suppose we have to form a number where 1 or 3 is in the beginning <u>or</u> the end. Then the two digits can be arranged among themselves in 2! ways. Hence total number of arrangement will be $P_3 X \ 2 = 12$ ways.

(f) Suppose we have to form numbers greater than 30,000. Here there should be 3 or 4 or 5 in ten thousand's place which can be filled in 3 ways.

The remaining 4 digits filled in 4! ways.

Therefore, we have, i.e.

3 1 2 4 5
3 2 1 4 5 etc

i.e., total number of numbers

$$3 \ X \ P_4$$

$$= 3 \ X \ 24 = 72$$

Example 1.16: How many numbers can be formed with digits 1, 2, 4, 0, 5 when any is not repeated in any number?

Solution: There are 5 digits in all including zero. The number of single digit numbers is P_1. The number of two digit number is P_2. Out of this, some have zero in the tenth place and so reduces to one digit number. Hence the number of two digit numbers is $P_2 - P_1$. Similarly, the nutrber of three digit number is $P_3 - P_2$.

The total number of numbers is

$$P_1 + (P_2 - P_1) + (P_3 - P_2) + (P_4 - P_3) + (P_5 - P_4)$$

$$4 \quad + \quad 16 \quad + \quad 48 \quad + \quad 96 \quad + \quad 96$$

260 numbers.

Example 1.17:

(i) Find the sum of all the numbers that can be formed with digits 1, 3, 4, 7, 5, 9 taking all at a time.

(ii) Find the probability of having a number with 3 in the tenth place.

Solution:

(i) We need to consider when each digit occupy a particular place. The number of permutation when 1 is in the unit place is $P_5 = 5! = 120$. The number of permutation when any of the given numbers occupy the unit place is also $5! = 120$ ways. Hence we can sum all the numbers in the unit place a $120(29)x\ 1 = 3480\ x\ 1$

Similarly the sum of numbers in the 10th place is also

$$120(1+3+4+5+7+9) = 2480\ x\ 10$$
$$= 34800$$

In the same manner, the sum of all the numbers is

$$3480\ (100{,}000 + 10{,}000 + 1{,}000 + 100 + 10 + 1)$$
$$= 3430\ (111111) = 386666280$$

(ii) The number of numbers taking all at a time without restriction is

$$P_6 = 6! = 720$$

The number of numbers when 3 occupy the tenth place is $P_5 = 120$

$$\text{Pr (a number 3 in the tenth place)} = \frac{120}{720} = 0.1667$$

(G) **Formation of words with letters:**

This is similar to what we illustrated in *Formation of numbers with digits*.

Example 1.18: Suppose the letters of the word STAPLER is given to form words.

(a) If there is no restriction, the number of words is

$$P_7 = 7! = 5040 \text{ words.}$$

(b) Suppose all words to be formed begins with *S*. The remaining 6 places can be filled in 6! = 720.

(c) Suppose all words to be formed begins with *S* or ends with *E*. The two positions can be filled in $P_2 = 2$ ways. The other 6 digits can be filled in $P_6 = 6! = 70$ ways.

Hence total number of words is 2 x 120 = 240 words.

(d) If all words formed must begin with *S* and end with *E*. The two places are now fixed. Then the remaining 5 places can be filled in 5! = 120 ways. Hence, 120 words are formed.

(e) Suppose two vowels *A* and *E* are to stand together. Regard *A* and E as one

a, E, STPLR

STPLR can be arranged among themselves in 6! = 120 ways.

The two vowels can be arranged in 2 ways.

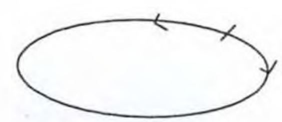
Hence the total number of words is 2 x 120 = 240 words.

(f) If three particular letter are to occupy the even places. The first letter can be filled in 3 ways, the second in 3 ways and the third in 1 way, a total of 6 ways. Then. the remaining 4 letters can be filled in 4! = 24 ways. Hence, the total number of words is 6 x 24 = 144

(H) **Ordered:**

Arrangement of items round a circle:

Things can be arranged round a circle in (i) clockwise and (ii) anti- clockwise direction.



Example 1.19: In how many ways can 7 people sit round a circular dinning table

$$= \frac{1}{2}(7-1)!$$
$$= 360 \text{ ways}$$

(i) The number of arrangements when the direction (clockwise or anticlockwise) is specified is $(n-1)!$. This is because one of the items can be used as a starting point.

(ii) When the direction of arrangement is not specified is $\frac{1}{2}(n-1)!$ ways.

Example 2.17: How many ways can 20 different beads be arranged to form a necklace?

$$= \frac{1}{2}(n-1)!$$
$$= \frac{1}{2}(19!) \text{ ways}$$

Example 1.20: A round table conference is to be held by 10 persons such that 2 particular person may wish to sit together.

Solution: Regard the 2 people as one. We now have 9 persons. The two persons can be arranged in 2! ways. The 9 persons can be arranged in $(9-1)!$ ways. The total number of arrangement is

$$8! \times 2! = 80640 \text{ ways}$$

1.4 Combination

The number of arrangement or 'selection' of n different items taking some or all of the number of things at a time irrespective of the order is referred to as combination.

The number of combination n things taking r at a time is denoted by

$$\binom{n}{r} or\ C_r = \frac{n!}{(n-r)!r!}$$

Most of the problems on selection without replacement can be solved using combination approach.

Example 1.21: In how many ways can a committee of 5 be selected from amongst 6 boys and 7 girls; if the committee must consist of (i) 2 boys and 3 girls, (ii) at most 3 boys?

Solution: There are a total of 13 persons.

(i) The total number of combination is 2 boys can be selected from 6 boys in $\binom{6}{2}$ ways.

3 girls can be selected from 7 girls in $\binom{7}{3}$ ways.

Total number of combination is

$$\binom{6}{2}\binom{7}{3} = 15 \times 35 = 525 \text{ ways}$$

(ii) There could be 0, 1, 2 and maximum of 3 boys. Hence the total number of combination is

$$\binom{6}{0}\binom{7}{5} + \binom{6}{1}\binom{7}{4} + \binom{6}{2}\binom{7}{3} + \binom{6}{3}\binom{7}{4}$$
$$= 21 + 210 + 525 + 420$$
$$= 1176 \text{ ways}$$

Example 1.22: A box contains 20 balls all of which are of the same size. 15 of them are Red and 5 Black balls. 4 balls are selected at random from the box, find the probability of having:

(i) exactly 2 black balls.

(ii) at least 1 red ball

Solution:

(i) The first thing to do is to find the combination of any 4 balls out of 20 (i.e. sample space) $\binom{20}{4}$.

Number of ways of choosing 2 black from 5 is $\binom{5}{2}$.

Number of ways of choosing the remaining 2 from 15 red balls is $\binom{15}{2}$.

Number of outcomes of favour of the event is $\binom{5}{2}\binom{15}{2}$

$$P(2\ black\ and\ 2\ red\ balls) = \frac{\binom{5}{2}\binom{15}{2}}{\binom{20}{4}} = 0.217$$

(ii) The probability of having at least 1 red ball is

$$= \frac{\binom{15}{1}\binom{5}{3} + \binom{15}{2}\binom{15}{2} + \binom{15}{3}\binom{5}{1} + \binom{15}{4}\binom{5}{0}}{\binom{20}{40}}$$

$$= \frac{75 + 1050 + 2275 + 1365}{4845}$$

$$= 0.983$$

(A) **Combination when a particular thing must be included or not included**

(i) The number of ways of choosing r things out of n in which k particular thing always occur is $\binom{n-k}{r-k}$

(ii) The number of ways of choosing r things out of n which k particular thing never occur is $\binom{n-k}{r}$

Example 1.23: 15 players were invited for a crucial football match. In how many ways can 11 players be chosen if

(i) the skipper must be included

(ii) a particular player is injured and must not be included.

(iii) player A must be included and player B must not be included.

Solution:

(i) If the skipper is selected first, we have 14 players left to select the remaining 10 players.

The required number is $\binom{14}{10} = 1001$ ways.

(ii) Remove the injured player, now select 11 from the remaining 14 players.

The required number is $\binom{14}{11} = 364$ ways.

(iii) If we remove B and select player A.

Then required number is $\binom{13}{10} = 286$ ways.

Example 1.24: A certain examination consists of 12 questions divided into two parts of 6 questions each. How many ways can a student choose any 8 questions if he must attempt exactly 5 questions from the first part?

Solution: From the first part, questions are selected in $\binom{6}{5} = 6$ ways.

In the second part, 3 questions are selected $\binom{6}{3} = 20$ ways.

The required number is $\binom{6}{5}\binom{6}{3} = 120$ ways.

(B) **When all items are alike and each of them may be disposed off in 2 ways:**

In this situation, the item may be included or rejected. The total number of ways of disposing all things is $2 \, x \, 2 \, x \ldots x \, n$ times $= 2^n$. This include a case where all the items are rejected.

Hence, the total number of ways in which one or more things are included is $2^n - 1$.

This is equivalent to $\binom{n}{n} + \binom{n}{n-1} + \cdots + \binom{n}{1}$

i.e. $\binom{n}{n} + \binom{n}{n-1} + \cdots + \binom{n}{1} = 2^n - 1$

Example 1.25: In how many ways can a student solve one or more questions out of 8 in a paper?

Solution: The student may either solve a question or leave it (i.e. 2 ways). The total number of ways of solving one or two or all the questions is

$$\binom{8}{1} + \binom{8}{2} + \cdots + \binom{8}{8} = 2^n - 1$$

$$= 256 - 1$$

$$= 255 \text{ ways}$$

Note:

If it must include a case where none of the questions is solved, then the required number is

$$\binom{8}{0} + \binom{8}{1} + \binom{8}{2} + \cdots + \binom{8}{8} = 2^8$$

$$= 256 \text{ ways}$$

Example 1.26: How many different products can be formed with the letters a, b, c, d, e and f.

Solution: The number of ways in which one or more of the six letter

$$= 2^6 = 1$$

But this includes a single letter which is not a product. Hence, the number of products i.e. $2^6 - 6 - 1 = 57$.

(C) **When some items are alike and each of them can be disposed in a way:**

Given $n = [p + q + r + s + \cdots]$ items out of which p, q, r, s of them are alike and

p can be chosen in $(p + 1)$ ways

q can be chosen in $(q + 1)$ ways

r can be chosen in $(r + 1)$ ways.

then the total number of combinations is $(p + 1)(q + 1)(r + 1)(s + 1) - 1$ ways.

Example 1.27: How many factors has 2160?

Solution: The factors of 2160 are i.e.

$$2160 = 16 \, x \, 27 \, x \, 5$$
$$= 2^4 x \, 3^3 x \, 5^1$$

But

2^4 can be formed in 5 ways.

3^3 can be formed in 4 ways.

5^1 can be formed in 2 ways.

Hence the total number of factors are $5 \, x \, 4 \, x \, 2 = 40$.

(D) **When Sharing (Dividing) n items into different groups:**

A number of items can shared among a group of people equally or in given proportion.

(i) If $n = p + q + r$ and $p = q = r$.

Then the number of ways of sharing n things equally is $\frac{n!}{(p!)^3}$

(ii) If $n = p + q + r$ and $p \neq q \neq r$, then the number of ways of sharing n things proportionally is $\frac{n!}{p!q!r!}$

Example 1.28:

(a) In how many ways can a deck of 52 cards be shared among 4 players equally?

Solution: $\frac{52!}{(13!)^4} = 5.36 \, x \, 10^{28}$

(b) If the group of 13 cards are to be arranged, in how many ways can this be done?

Solution: $\frac{52!}{(13!)^4} = 1.28 \, x \, 10^{30}$

Example 1.29: How many ways can 18 books be divided?

(i) equally or

(ii) in ratio 1:2:3

Solution:

(i) 18 books can be divided into 3 groups of 6 each. Then the required number is $\frac{18!}{(6!)^3} = 17{,}153{,}138$ ways

(ii) To divide 18 books in ratio 1:2:3 each group would consist of 3, 6, 9 respectively.

Hence the required number is $\frac{18!}{3!6!9!} = 4{,}084{,}080$ ways.

(E) **Permutation and Combination Occurring Simultaneously**

Some problems require the application of the permutation and combination approaches simultaneously. We shall give a theory which may be proved.

Theorem: If there are m different things of one kid, n different thins of the 2nd kind and k different things of the 3rd kind. The number of permutation which can be formed containing r of the first, s of the second and j of the third is

$$\binom{m}{r} x \binom{n}{s} x \binom{k}{j} x \, (r * s + j)!$$

Example 1.30: How many ways can 5 boys and 4 girls selected from among 12 boys and 9 girls be arranged on a bench?

Solution: 5 boys are selected from 12 in $\binom{12}{5}$ ways.

4 girls are selected from 9 in $\binom{9}{4}$ ways.

but the 9 people can be arranged among themselves in $ap_9 = 9!$ ways

The required number is

$$\binom{12}{5}\binom{9}{4}\, 9! = 3.62 \times 10^{10}$$

Example 1.31:

(a) How many words each containing 2 vowels and 3.consonants can be formed with 5 vowels and 8 consonants?

(b) How many words can be formed if

(i) 'a' must be included

(ii) the words must contain at least two consonants?

Solution:

(a) 2 vowels can be chosen from 5 in $\binom{5}{2}$

3 consonants can be chosen from 8 in $\binom{8}{3}$

the 5 letters can be arranged among themselves in 5! ways.

The required number is

$$\binom{5}{2}\binom{8}{3}\, 5! = 560 \times 120$$

(b) 'a' is a vowel = 67200 ways.

(i) if 'a' must be included, we need one more vowel. The required number is

$$\binom{5}{1}\binom{8}{3}\, 5! = 33600 \text{ ways}$$

(ii) If the word must contain at least 2 consonant, then it could contain 2 or more consonants.

The required number is

$$\binom{5}{3}\binom{8}{2}\, 5! + \binom{5}{2}\binom{8}{2}\, 5!$$

$$= 33600 + 67200$$

$$= 100800 \text{ ways}$$

(F) **Combination with repetition**

Sometimes we are interested in the number of combinations of items when each of the items may be repeated. Given n items, the number of combinations taking r at a time then repetitions are allowed is denoted by nHr where

$$nHr = \binom{n+r-1}{r} = \frac{(n+r-1)!}{(n+r-1)!r!}$$

$$= \frac{(n+r-1)(n+r-2)\dots(n+r-r-1)\,(n-1)n}{r!}$$

$$= \frac{(n+r-1)(n+r-2)\dots n}{r!}$$

Example 1.32: How many combinations of 4 digit numbers can be formed from the digits 2, 4, 5, 7, 8, 9 if the digits may be repeated at least once?

Solution: There are 6 digits, to take any 4 at a time, the required number is

$$6H_r = \left(\frac{6+4-1}{4!}\right) = \frac{9!}{4!5!}$$

$$= 126$$

Example 1.33: In an experiment, 2 dice are rolled once. Find the total number of outcomes if

(i) they are distinct

(ii) they are of distinguishable

Solution

On a single die there are 1, 2, 3, 4, 5, 6 (6 numbers)

(i) If they are distinct, the total number of outcomes is $6^2 = 36$

(ii) If they are not distinguishable, then any number on the die may be repeated. Hence the required total number of outcomes is

$$\left(\frac{6+2-1}{2}\right) = \frac{7!}{2!5!} = 21.$$

Multinomial Coefficients

This is a generalized version of basic counting principle.

Consider a set of n-distinct items to be divided into r distinct groups of sizes $n_1, n_2, n_3, \dots, n_r$ where $\sum_{i=1}^{r} n_i = n$.

The number of possible choices for the first group is $\binom{n}{n_1}$, second group is $\binom{n-n_1}{n_2}$; third group is $\binom{n-n_1-n_2}{n_3}$……

The total possible division for $n_1, n_2, n_3, \ldots, n_r$ is therefore

$$\frac{n!}{n!(n-n_1)!} \cdot \frac{(n-n_1)!}{(n-n_1-n_2)!n_2!} \cdots \cdots \frac{(n-n_1-n_2-\cdots-n_{r-1})!}{0!n_r!}$$

$$= \binom{n}{n_1}\binom{n-n_1}{n_2}\cdots\binom{n-n_1-n_2-\cdots-n_{r-1}}{n_r}$$

$$= \frac{n!}{n_1!\; n_2!\, n_3! \ldots n_r!}$$

Example: There are 12 Super Falcons to be divided into two teams of 6 girls each. How many different divisions is possible.

Solution: There are $\frac{12!}{5!5!} = 33{,}264$ divisions.

Exercises:

1. A U.I. football team plays 8 games in succession, winning 3, losing 3 and ending 2 in a tie. Show that the number of ways this can happen is $\binom{8}{3}\binom{5}{3} = \frac{8!}{3!3!2!}$
2. Find n and r such that the following equation is true $\binom{13}{5} + 2\binom{13}{6} + \binom{13}{7} = \binom{n}{r}$

1.5 Stirling Numbers of the Second Kind

Definition 4: Let S be a set. A partition of S is an ordered collection of pairwise, disjoint, nonempty subsets of S whose union is all of S. The subsets of a partition are called blocks.

For $S = A_1 \cup A_2 \cup A_3 \cup \ldots \cup A_k$ to be a partition of S:

i. $A_i \cap A_j = \emptyset$ whenever $i \neq j$

ii. $A_j \neq \emptyset$, $1 \leq j \leq k$

Two partitions are equal if and only if they have the same blocks.

For instance, $\{1\} \cup \{2,3\}$, $\{1\} \cup \{3,2\}$, $\{2,3\} \cup \{1\}$ and $\{3,2\} \cup \{1\}$ are 4 different looking ways of writing the same two-block partition of $S = \{1,2,3\}$

The other partition of $S = \{1,2,3\}$ are

$\{1\} \cup \{2\} \cup \{3\}$- 3 blocks

$\{1,2\} \cup \{3\}$- 2 blocks

$\{1,3\} \cup \{2\}$- 2 blocks

$\{1,2,3\}$- 1 block

Thus, S has a total of 5 different partitions made up of:

One of $\{1\} \cup \{2,3\}$, $\{1\} \cup \{3,2\}$, $\{2,3\} \cup \{1\}$ and $\{3,2\} \cup \{1\}$

$\{1\} \cup \{2\} \cup \{3\}$

$\{1,2\} \cup \{3\}$

$\{1,3\} \cup \{2\}$

$\{1,2,3\}$

Definition 5: The number partitions of $\{1,2,3,\ldots m\}$ into n blocks is denoted by $S(m,n)$ and this is known as the Stirling number of the second kind.

Note: $S(m,n) = 0$ if $n < 1$ or $n > m$.

Also, $S(m,1) = 1 = S(m,m)$. This is because there is just one way to partition $\{1,2,3,\ldots m\}$ into a single block and $\{1\} \cup \{2\} \cup \{3\} \cup \ldots \cup \{m\}$ is the unique unordered way of expressing $\{1,2,3,\ldots m\}$ as the disjoint union of m nonempty subsets.

1.5.1 Stirling's Identity

Stirling's Identity: For any two positive integers m and r,

$$r!\,S(m,r) = \sum_{t=1}^{r} (-1)^{r+t} C(r,t)\, t^m$$

Therefore $S(m,r) = \frac{1}{r!}\sum_{t=1}^{r}(-1)^{r+t}C(r,t)\,t^m$

Example 1.34: $S(4,1) = C(1,1)1^4 = 1$

$$S(4,2) = \frac{1}{2}[-C(2,1)1^4 + C(2,2)2^4]$$

$$= \frac{1}{2}[-2 + 16] = 7$$

$$S(4,3) = \frac{1}{6}[C(3,1)1^4 - C(3,2)2^4 + C(3,3)3^4]$$

$$= \frac{1}{6}[3 - 48 + 81] = 6$$

$$S(4,4) = \frac{1}{24}[-C(4,1)1^4 + C(4,2)2^4 - C(4,3)3^4 + C(4,4)4^4]$$

$$= \frac{1}{24}[-4 + 96 - 324 + 256] = 1$$

1.5.2 Application of Stirling's number of the second kind to distribution of objects into urns

We are interested in the question "In how many different ways can m balls be distributed among n urns?" We are going to answer this question by considering whether the balls and urns are labelled or not and whether a particular urn can be left empty?

We will consider 4 variations:

Variation 1: In how many ways can m labelled balls be distributed among n unlabelled urns if no urn is left empty? This is the same as "In how many ways can the set $\{1, 2, 3, \ldots m\}$ be partitioned into n blocks. This is $S(m, n)$.

Example 1.35: In how many ways can 4 labelled balls be distributed among 2 unlabelled urns if no urn is left empty?

Solution: $S(4, 2) = 7$ that is if the balls are labelled $1, 2, 3, 4$ then the 7 possibilities are

$\{1\}\&\{2,3,4\}$
$\{2\}\&\{1,3,4\}$
$\{3\}\&\{1,2,4\}$
$\{4\}\&\{1,2,3\}$
$\{1,2\}\&\{3,4\}$
$\{1,3\}\&\{2,4\}$
$\{1,4\}\&\{2,3\}$

Because the urns are unlabelled,
$\{2\}\&\{1,3,4\} = \{1,3,4\}\&\{2\}$ etc.

Variation 2: In how many ways can m labelled balls be distributed among n unlabelled urns?

Solution: This is $S(m,1) + S(m,2) + \cdots + S(m,n)$. This is the same as finding the number of ways in which $\{1,2,\ldots,m\}$ can be partitioned into n or fewer blocks since it is no longer a requirement that no urn be left empty.

Example 7: The number of ways to distribute four labelled balls among two unlabelled urns is $S(4,1) + S(4,2) = 1 + 7 = 8$ i. e.

$S(4,1) = \{1,2,3,4\}\&\{\ \ \}$

$S(4,2) = \{1\}\&\{2,3,4\}, \{2\}\&\{1,3,4\},$
$\{3\}\&\{1,2,4\}, \{4\}\&\{1,2,3\}, \{1,2\}\&\{3,4\}, \{1,3\}\&\{2,4\}, \{1,4\}\&\{2,3\}$

Variation 3: In how many ways can m labelled balls be distributed among n labelled urns? This is n^m.

Example 1.36: Five labelled balls can be distributed among 3 labelled urns in $3^5 = 243$ ways.

Variation 4: In how many ways can m labelled balls be distributed among n labelled urns if no urn is left empty? This is $n!\,S(m,n)$.

There are $S(m,n)$ ways to distribute m labelled balls among n unlabelled urns using variation 1. After the distribution of the balls, there are $n!$ ways to label the urns. By the fundamental principle of counting, the answer is $n!\,S(m,n)$.

Example 9: In how many ways can 5 labelled balls be distributed among 3 labelled urns if no urn is left empty?

Solution: $3!\,S(5,3)$

Example: Suppose that a secretary prepares 5 letters and 5 envelopes to send to 5 different people. If the letters were randomly stuffed into the envelopes, a match occurs if a letter is inserted in the proper envelope.

(i) In how many ways can the letters be stuffed into the envelopes so that no letter falls into the proper envelope?

(ii) What is the probability that none of the letters is placed in the right envelope?

(iii) What is the probability that at least one of the letters is placed in the right envelope?

(iv) What is the probability that exactly 3 of the letters were placed in the right envelope?

Solution:

(i) The total number of derangements for the 5 letters is

$$D_5 = 5!\left[1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \frac{1}{5!}\right]$$

$= 120[1 - 1 + 0.5 + 0.1667 + 0.0417 + 0.00833]$

$= 120(0.71673)$

=86.008 ways

(ii) Probability that none of the letters is placed in the right envelope is given as

$$\frac{D_5}{5!} = 1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \frac{1}{5!}$$

$= 0.716$

(iii) The probability that at least one of theletters is placed in the right envelope is

1 - Prob [None of the letters is placed in the right envelope]

$= 1 - (0.716)$

$= 0.2833$

(iv) The probability that exactly 3 of the letters were placed in the right envelope is given by

$$\frac{\binom{N}{K}(N-K)!\left[1-1+\frac{1}{2!}-\frac{1}{3!}+\ldots+\frac{(-1)^{N-K}}{(N-K)!}\right]}{N!}$$

$$= \frac{\binom{5}{3}(5-3)!\left[1-1+\frac{1}{2!}\right]}{5!}$$

$= 0.083$

1.6 Allocation and Matching Problems

Introduction

Matching and allocation are some of the classic problems in probability theory. This problems dated back to the early 18th century has many variations. There are many ways to describe the problem. One such description is the example of matching letters with envelops. Suppose there are n letters with n matching envelops (assume that each

letter has only one matching envelop). Then it is possible to determine the probability that the secretary stuffs the letters randomly into right envelops.

1.6.1 Derangements

Definition 1: A derangement of (1, 2,….n) is a permutation $i_1, i_2, \ldots, i_n$ of (1, 2, …,n) such that $i_1 \neq 1, i_2 \neq 2, \ldots, i_n \neq n$.

Thus, a derangement of (1. 2,…,n) is a permutation $i_1, i_2, \ldots, i_n$ of (1, 2,…,n) in which no integer is in its natural position: $i_1 \neq 1, i_2 \neq 2, \ldots, i_n \neq n$.

Denote by D_n the number of derangement of (1, 2, …., n)

Consider the following example for illustration:

Example 1: At a party, 10 gentlemen check their hats. In how many ways can their hats be returned so that no gentleman gets the hat with which he arrived?

This problem consists of an n-element set X in which each element has a specified location. We are required/asked to find the number of permutations of the set X in which no element is in its specified location.

Here, the set X is the set of 10 hats and the specified location of a hat is (the head of) the gentlemen to which it belongs.

Let us take X to be the set {1, 2,……,10} in which the location of each of the integers is that specified by its position in the sequence 1, 2,…….10.

Theorem: For n ≥ 1, $D_n = n!\left[1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \ldots\ldots + (-1)^n\frac{1}{n!}\right]$

Proof: Let S be the set of all n! permutations of (1, 2, …..,n). For j = 1, 2,…, n, let p_j be the property that in a permutation, j is in its natural position. Thus, the permutation $i_1, i_2, \ldots, i_n$ of (1, 2, ….,n) has property p_j provided $i_j = j$. A permutation of (1, 2,…n) is a derangement if and only if it has none of the properties $p_1, p_2, \ldots, p_n$.

Let A_j denote the set of permutations of (1, 2,….n) with property p_j, (j = 1, 2, ….., n). The derangements of (1,2,…,n) are those permutations in $A_1^c \cap A_2^c \cap \ldots\ldots \cap A_n^c$.

Thus, $D_n = \left|A_1^c \cap A_2^c \cap \ldots\ldots \cap A_n^c\right|$

The PIE is used to evaluate D_n as follows:

The permutation in A_1 are of the form 1, $i_2 \ldots i_n$, where $i_2 \ldots i_n$ is a permutation of (2,.......n). Thus, $|A_1| = (n-1)!$ And more generally for $|A_j| = (n-1)!$ for j = 1, 2, ...n.

The permutations in $A_1 \cap A_2$ are of the form 1, 2, $i_3 \ldots i_n$ where $i_3 \ldots\ldots i_n$ is a permutation of (3, ...n). Thus, $|A_1 \cap A_2| = (n-2)!$

Generally, $|A_i \cap A_j| = (n-2)!$ for any 2 combinations (i, j) of (1, 2,....,n).

For any integer k, with $1 \le k \le n$, the permutations in $A_1 \cap A_2 \cap \ldots \cap A_k$ are of the form 1, 2, ...,k, $i_{k+1} \ldots i_n$, where $i_{k+1} \ldots, i_n$ is a permutation of (k+1,.....,n). Thus, $|A_1 \cap A_2 \cap \ldots \cap A_k| = (n-k)!$.

Generally, $|A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}| = (n-k)!$ for any k-combination $(i_1, i_2, \ldots, i_k)$ of (1, 2,....n):

Since there are $\binom{n}{k}$ k – combinations of (1, 2, ...n), applying the inclusion-exclusion principle, we obtain:

$$D_n = n! - \binom{n}{1}(n-1)! + \binom{n}{2}(n-2)! - \binom{n}{3}(n-3)! + \ldots\ldots + (-1)^n\binom{n}{n}(n-n)!$$

$$= n! - \frac{n!}{1!} + \frac{n!}{2!} - \frac{n!}{3!} + \ldots\ldots + (-1)^n \frac{n!}{n!}$$

$$= n!\left[1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \ldots\ldots + (-1)^n \frac{1}{n!}\right]$$

Thus, from example 1 above,

$$D_{10} = 10!\left[1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \frac{1}{5!} + \frac{1}{6!} - \frac{1}{7!} + \frac{1}{8!} - \frac{1}{9!} + \frac{1}{10!}\right]$$

You should be able to supply the final answer for D_{10}

Note: (i) The series expansion for $e^{-1} = 1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} \ldots\ldots$

(ii) $\frac{Dn}{n!}$ is the ratio of the number of derangement of (1, 2,....n) to the total number of permutations of (1, 2,....n).

Thus, $\frac{Dn}{n!}$ is the probability that it is a derangement if we select a permutation of (1, 2,....n) at random.

1.6.2 The Matching Problem

Suppose that an absent minded secretary prepares n letters and envelopes to send to n different people. If the letters were randomly stuffed into the envelopes, a match occurs if a letter is inserted in the proper envelope.

Example 2: Suppose that each of N men in a room throws his shirt into the centre of the room. The shirts are first mixed up and then each man randomly selects a shirt.

(1) What is the probability that none of the men selects his own shirt?

(2) What is the probability that at least one of the men selects his own shirt?

(3) What is the probability that exactly k of the men select their own shirt?

Solution:

1. From our discussion on derangement, the probability that none of the men selects his own shirt is

$$\frac{D_N}{N!} = 1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \ldots\ldots + (-1)^N \frac{1}{N!}$$

2. The probability that at least one of the men selects his own shirt is

1 – Prob [None selects his own shirt]

$$= 1 - \left[1 - 1 + \frac{1}{2!} - \frac{1}{3!} + \ldots\ldots + \frac{(-1)^N}{N!}\right]$$

$$= 1 - 1 + 1 - \frac{1}{2!} + \frac{1}{3!} \ldots\ldots\ldots - \frac{(-1)^N}{N!}$$

$$= 1 - \frac{1}{2!} + \frac{1}{3!} \ldots\ldots\ldots\ldots - \frac{(-1)^N}{N!}$$

3. The probability that exactly k of the men select their own shirt is as follows: First fix attention on a particular set of k men. The number of ways in which this and only this k men can select their own shirt is equal to the number of ways in which the other N-k men can select among their shirts in such a way that none of them selects his own shirt.

The probability that none of the N-K men, (selecting among their shirts), selects his own shirt is $1 - 1 + \frac{1}{2!} - \frac{1}{3!} + ... + \frac{(-1)^{N-K}}{(N-K)!}$

It follows that the number of ways in which the set of men selecting their own shirts corresponds to the set of k men under consideration is

$$(N-K)!\left[1 - 1 + \frac{1}{2!} - \frac{1}{3!} + ... + \frac{(-1)^{N-K}}{(N-K)!}\right]$$

Also, as there are $\binom{N}{K}$ possible selections of a group of K men, it follows that there are

$$\binom{N}{K}(N-K)!\left[1-1+\frac{1}{2!}-\frac{1}{3!}+...+\frac{(-1)^{N-K}}{(N-K)!}\right]$$

ways in which exactly K of the men select their own shirts.

The probability required is thus

$$\frac{\binom{N}{K}(N-K)!\left[1-1+\frac{1}{2!}-\frac{1}{3!}+...+\frac{(-1)^{N-K}}{(N-K)!}\right]}{N!}$$

$$\frac{1-1+\frac{1}{2!}-\frac{1}{3!}+......+\frac{(-1)^{N-K}}{(N-K)!}}{K!}$$

This result is approximately $\frac{e^{-1}}{K!}$, for large N, k = 0,1, ...

Example 3:

Suppose there are a group of six men and six women. They are to be paired in groups of 2 for the purpose of determining roommates.

(i) What is the probability that both groups will have the same number of male and female.

(ii) What is the probability that there are no male and female as roommates?

Solution:

1. 6 men, 6 women divided into 2 groups

(i) two groups of 6persons each

$$\frac{12!}{2!6!6!} \Big/ \frac{12!}{6!6!}$$

$$= \frac{14.4375}{924}$$

$$= 0.0156$$

(ii)

All males and all females $\frac{6!}{2^3 3!3!} \times \frac{6!}{2^3 3!3!} \Big/ \frac{12!}{2^6 3!3!}$

$$= \frac{\left(\frac{6!}{2^3 3!3!}\right)^2}{\frac{12!}{2^6 6!6!}} = \frac{(2.5)^2}{14.4375} = \frac{6.25}{14.43} = 0.4329$$

Example 4:

(a) State the principle of inclusion and exclusion.

(b) Suppose 15% of apple and 10 consignments were toxic. If the consignment consists of 60% apple and 40% mango, what is the probability that a fruit selected at random is toxic?

Solutions:

(b) 15% of apple are toxic, 10% of mangoes are toxic

Consignment: 60% apple, 40% mango

Let F represent fruit; A: apple, M: mango

Let T represent toxic fruit

(i) $P(T) = P(A|T)P(A) + P(M|T)P(M)$

$= 0.15(0.6) + 0.10(0.4)$

$= 0.09 + 0.04$

$= 0.13$

(ii) $P(A|T) = \frac{P(A|T)P(A)}{P(T)}$

$= \frac{0.09}{0.13}$

$= 0.0117$

Example 5:

3.(a) Give the Stirling's identity.

(b)(i) In how many ways can 10 labelled balls be distributed among 7 labelled urns

(ii) What is the probability if the urns are unlabeled and non of them is left empty.

Solution:

(a). Stirling Identity

$$S(m,r) = \frac{1}{r!}\sum_{t=1}^{r}(-1)^{r+t}\binom{r}{t}t^m$$

where m and r are positive integers

b(i) m = 10 labelled balls

n = 7labelled balls

Number of ways is $n^m = 7^{10}$

$= 282,475,249$ways

(uses the principle of inclusion and exclusion)

b(ii) $S(10,7) =$

$\frac{1}{7!}\left[\binom{7}{1}1^{10}(-1)^8 + \binom{7}{2}2^{10}(-1)^9 + \binom{7}{3}3^{10}(-1)^{10} + \binom{7}{4}4^{10}(-1)^{11} + \binom{7}{5}5^{10}(-1)^{12} + \binom{7}{6}6^{10}(-1)^{13} + \binom{7}{7}7^{10}(-1)^{14}\right]$

$= \frac{29635200}{7!}$

$= 5880$ways

Therefore, $Pr[S(10,7)] = \frac{S(10,7)}{n^m}$

$= \frac{5880}{282475247}$

$= 7.369 \times 10^{-14}$

Example 6:

Suppose that each of the 10 men in a room throws in their cap into the center of the room to be picked by 10 ladies in the annual marriage fixing ceremony. What is the probability that

(i) No lady picks the cap of the man of her choice.

(ii) At least one lady picks the cap of the man of her choice.

(iii) Exactly 7 ladies could not pick the cap of men of their choice.

Solution:

10men and 10 ladies

$\frac{D_n}{n!} = \left[1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \cdots + \frac{1}{10!}\right]$

(i) $Pr\ (No\ lady\ picked\ a\ cap) = [1 - 1 + 0.5 - 0.1667 + 0.0417 - 0.0083 + 0.0014 - 0.0002 + 0.000 - 0.000 + 0.000]$

$= 0.3679$

(ii) $Pr\ (at\ least\ one\ lady\ picked\ a\ cap) = 1 - Pr\ (No\ lady\ picked\ a\ cap)$

$= 1 - 0.3679$

$= 0.6321$

(iii) $n - k where\ n = 10, k = 7$

$10 - 7 = 3$

$$P_{(k)} = \binom{n}{k}(n-k)!\left[1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \cdots (-1)^{n-k}\frac{1}{(n-k)!}\right]$$

$\frac{D_n}{7!} = \frac{\left[1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!}\right]}{7!}$

$= \frac{1 - 1 + 0.5 - 0.1667}{7!}$

$= \frac{0.333}{5040}$

$= 0.00006$

$\equiv 6.61 \times 10^{-5}$

Therefore Pr(exactly 7 ladies could not pick the cap of men of their choice) is $1 - \frac{D_n}{7!}$

i.e$1 - 6.66 \times 10^{-5} = 0.9993$

Practice Questions

(1) Show that $\binom{n}{r} = \binom{n}{n-r}$

(2) If $C_{n-4} = 15$; find n.

(3) An examination question is divided into three sections A, B, C with 3, 4 and 5 question respectively. A student is required to answer t questions each from. Sections A and B and 3 from Section C. In how many ways can he write the examination?

(4) In how many ways can he solve one or more question in Section C.

(5) If the paper is one of the professional examination papers where candidates are required to attempt as many questions as possible, find the total number of ways a candidate can write the examination if must attempt at least one question?

(6) In how many ways can a person purchase two or more items out of 5?

(7) A nursery school pupil learning simple arithmetic is given 5 counters with digits 2, 1, 3, 0, 4, 5 to form numbers. Find the probability that the pupil is about to form a

(a(i)) 3 digit number

(ii) a number greater than 400,000

(b) Using all the digits except 0, how many numbers can be formed and what is their sum?

(8) How many ways can the letters of the sentence "Daddy did a deadly deed" be formed?

(9) A boy found a keylock for which the combination was unknown, but correct combination is a four digit number d_1, d_2, d_3, d_4, where $d_i, i = 1, 2, 3, 4$ is selected from 1, 2, 3, 4, 5, 6, 7, 8. How many different lock combinations are possible results in such keylock?

(10) Ten children are to be grouped into two clubs in such a way that five will belong to each club. If in watch club a secretary and a president is to selected, in how many ways can this be done?

(11) A shelf contains Chemistry, Mathematics and Economic text books. In how many ways can S books be selected?

(12) Show that:

a. $nP(n-1, r) = P(n, r+1)$

b. $P(n+1, r) = rP(n, r-1) + P(n, r)$

13. In how many ways can four elements be chosen from a ten-element set:

a. with replacement if order matters?

b. with replacement if order does not matter?

c. without replacement if order does not matter?

d. without replacement if order matters?

3. In how many ways can six balls be distributed among four urns if :

a. the urns are labelled but the balls are not?

b. the balls are labelled but the urns are not?

c. both balls and urns are labelled?

d. neither balls nor urns are labelled?

14. Show that $D_5 = 44$

15. Seven gentlemen check their hats at a party. How many different ways can their hats be returned so that:

a) no gentleman receives his own hat?

b) at least one gentleman receives his own hat?

c) at least two gentlemen receive their own hat?

CHAPTER 2
ELEMENTS OF PROBABILITY

2.1 Introduction

The definition of probability is as varied as the values of any random variable. Its definition depends on the extent to level one is knowledgeable of the use and power of probability concept.

Probability can be defined as a measure of uncertainty concerning a phenomenon. It can also be defined as a real value that measures the degree of belief one has in the occurrence of a specified event. Probability is also described as the study of random phenomena. Most phenomena studies in the Physical Science, Biological Sciences, Engineering and even Social Sciences are looked at not only from deterministic but also from a random point of view. Therefore the theory of probability has as its central feature, the concept of a repeatable random experiment, the outcome of which is uncertain.

To the Statistician, probability remains the vehicle that enables him use information in the sample to make inferences or describe a population from which the sample was obtained. Thus the study of probability prepares a strong background for reliable statistical inference. No wonder Professor Sir John Kingman remarked in a review Lecture in 1984 on the 150th anniversary of founding of the Royal Statistical Society that "the theory of Probability lies at the root of all statistical theory".

2.2 Definition of Terms and Concepts

Before we define probability as a concept, it is necessary to review the definition of some probability terms that shall be employed in our discussions.

(a) **A Trial:** Is any process or an act which generate a number of outcome which can not be predicted a priori. A trial usually results into only one of the possible outcomes e.g., A toss of a coin once, will lead to either a Had (H) or a tail (T) turning up. The selection of a card from a deck of well shuffled cards result in one of the cards being drawn.

(b) **A Random Experiment:** Is any operation which when repeated generates a number of outcomes which cannot be predetermined. e.g. A toss of two coins at a time; draw of two cards from a deck one after the other; a random selection of a ball from a box and examine the colour.

(c) **An outcome:** This is a possible result of a trial or an experiment. In a toss of two coins, an outcome could be any one of HH, HT, TH, TT. The possible outcomes in a throw of a die are, 1, 2, 3, 4, 5, 6.

(d) **Sample Space:** Is the totaliy of all possible outcomes of an experiment. It is a set of all finite or countably infinite number of elementary outcomes $e_1, e_2, \dots, e_n$ It is usually represented by $S = \{e_1, e_2, \dots, e_n\}$

The sample space in a toss of a coin and a die is represented by

H	1H	2H	3H	4H	5H	6H
T	1T	2T	3T	4T	5T	6T
	1	2	3	4	5	6

i.e. $S = \{1H, 2H, 3H, 4H, 5H, 1T, 2T, 3T, 4T, 5T, 6T\}$

The sample space when a die is thrown twice is

$S = \{11, 1, 2, 1, 3, 1, 4, 1, 5, 1, 6, 1, 2, 22, \dots, 66\}$

(e) **An Event:** Is a subset of a sample space.

It consists of one or more possible outcomes of an experiment. It is usually denoted by capital letters $A, B, C, D, \dots$. It should be noted that a subset in a given set could consist of all the possible outcomes or none of the outcomes of the given set.

e.g. When a die is tossed once, we define. Set

$A = \{set\ of\ even\ number\} = \{2, 4, 6\}$

$B = \{set\ of\ prime\ number\} = \{1, 3, 5\}$

$C = \{set\ of\ number\ greater\ than\ 7\} = \{\phi\}$

(f) **Mutually exclusive events:** Two events A and B are said to be mutually exclusive, if the occurrence of A prevents the occurrence of B. This implies that the two events can not occur together i.e. $A \cap B =$ e.g. the occurrence of H prevent the occurrence of T in a toss of a coin.

(g) **Mutually Exhaustive Events:** Events $A_1, A_2, A_3, A_4, \dots, A_n$ are said to be mutually exhaustive if they constitute the sample space. i.e.

$$\sum_{i=1}^{n} A_i = S.$$

However, some events could be both mutually exclusive and exhaustive. This implies that they are disjointed and yet their sum is equal to the sample space. This would be illustrated later in (1.8). It should be noted that the last two probability terms are associated with one experiment only.

(h) **Independent Events:** Two events A and B are said to be independent if the occurrence of A does not affect B. This implies that the two events can occur together. e.g. the event of an event number and a Tail in a throw of a coin and a die at once.

(i) **Sure/Certain Event:** The sample space S is the only sure event. The probability of a certain event E is one $(P(E) = 1)$

(j) **Impossible Event:** This is the complement of the sure event. It is an empty set $\emptyset$.

2.3 The Approaches to the definition of Probability

The three conceptual approaches to the definition of probability (1) the classical approach, (2) the relative frequency approach and (3) the axiomatic approach, (4) subjective approach. These three concepts are explained as follows:

(a) Classical or 'a priori' Approach

If there are $n$number of exhaustive, mutually exclusive and equally likely cases of an event and suppose that n_A of them are favourable to the happenings of an event A under the given set of conditions, then $(A) = \frac{n_A}{n}$. An example is the toss of a die once. The six possible outcomes are 1,2,3,4,5,6. The probability of occurrence of a 2 is $\frac{1}{6}$. The probability is 'a priori', that is it can be determined before carrying out the experiment.

This method assumes that the elementary outcomes of an experiment are equally likely. It defines the probability of an elementary event e_i as 1 divided by the total

number of outcomes for an experiment. There is no requirement that the experiment be performed before the probability is determined, i.e.

$$P(A) \frac{Number\ of\ outcomes\ in\ favour\ of\ A}{Total\ number\ of\ outcomes\ for\ experiment} = \frac{n_A}{N}$$

Where N is the total number of possible outcomes

ThusProbability is a measure of likelihood that a specific event will occur.

Example 2.3.1: Find the probability of obtaining any number in a simple thrown of a die.

Solution: The experiment has six outcomes 1, 2, 3, 4, 5, 6.

$$P(a\ number) \frac{1}{Total\ number\ of\ outcomes} = \frac{1}{6}$$

Example 2.3.2: Find the probability of obtaining an event number in one roll of a die.

Solution: Let A be the event of an even number,

$$A = \{2, 4, 6\};\ n(A) = 3$$

$$S = \{1, 2, 3, 4, 5, 6\};\ n(S) = 6$$

$$P(A) \frac{Number\ of\ outcomes\ included\ in\ A}{Total\ number\ of\ outcomes} = \frac{3}{6} = 0.5$$

This approach to the definition of probability only holds for finite sample space where elementary events are equally likely. However this assumption is not always true in the real life as all events are not equally likely. After all we are not equally endowed.

(b) **Frequency or 'aposteriori' probabilityApproach:** This method defines probability as an idealization of the proportion of times that a certain event will occur in repeated trials of an experiment under the same condition. Thus, in an experiment is repeated N times and $n(A)$, is the number of times that A occur, then the relative frequency is

$$\frac{n(A)}{N}$$

But relative frequencies are not probabilities but approximate probabilities. If the experiment is repeated indefinitely, the relative frequency will approach the actual or theoretical probability.

$$\therefore P(A) = \lim_{n \to \infty} \frac{n(A)}{N}$$

However, there is a requirement that the experiment be performed before the probability is determined. Hence, the probability is determined aposteriori. It should he noted that some events in real life cannot be repeated before the probability is determined. Even if it can be determined the limit may not converge.

Example 2.3: Fifty of the 800 cars that enters the University of Ibadan on a graduation day are found to be Jeep. Assuming different cars comes into the campus randomly, what is the probability that the next car is a Jeep?

Solution: Let *N* be the total number of cars and n be the total number of Lexus. Then

$$N=800,\ n=50$$

Using the relative frequency concept of probability, the probability that the next car being a Lexus is

$$P(Lexus) = \frac{n}{N} = \frac{50}{800} = 0.0625$$

(c) **Subjective Probability:** is the probability assigned to an event based on subjective judgement, experience, information and believe. Such probabilities assigned arbitrarily are usually influenced by the biases and experience of the person assigning it.

For instance the probability of the following events are subjective:

1. The probability that Jude, who is taking statistics in the second semester will score seven points in the course.
2. The probability that a particular Football Club win the maiden match with another club.
3. The probability that Ade will win the case he has filed against his landlord.

Since subjective probabilities is based on the individual's own judgement, it is rarely used in practice as it lacks the theoretical backing.

(d) **Axiomatic or theoretical Approach:** To circumvent the difficulties posed by the earlier approaches to the definition of probability and based on the study of random of random phenomena, researchers have developed a mathematical expression of certain aspects of the real world. The probability of a certain part of the

real world occurring at random is then determined satisfying certain properties (called axioms).

2.4 Probability of an event

If A is an event from an experiment E with sample space *S*, the real valued function *P(A)*is called the probability of A which satisfy the following axioms:

(1) $0 \le P(A) \le 1$ for every event A

(2) $P(S) = 1$

(3) $P(A_1 \cup A_2 \cup \ldots) = P(A_1) + P(A_2) + \ldots$

$$= \sum_{i=1}^{\infty} P(A_1)$$

for every finite or infinite sequence of disjoint event $A_1, A_2 \ldots$

2.5 Consequences of Probability Axioms

Theorem I

(a) If *A* is a given event and A^C is the compliment of *A*, then $P(A^C) = 1 - P(A)$.

Proof: $A \cup A^C = S$

$P(A + A^C) = P(S) = 1$ by axiom (2)

$\therefore P(A) + P(A^C) = 1$ A and A^C are mutually exclusive

$= P(A^C) = 1 - P(A)$.

(b) **Theorem II:**

Given that $\phi \subset S$, then $P(A) = 0$

Proof:

$S \cup \phi = S.$

$P(S \cup \emptyset) = S = 1$ by axiom (2)

$P(S) + P(\phi) = 1$ since $P(S) = 1$

$1 + P(\phi) = 1$

$= P(\phi) = 0.$

2.6 Rules of Probability

Theorem 1: Let S be a sample space and $P(.)$ be a probability function on S; then the probability that the event A does not happen is $1 - P(A)$ i.e. $P(A') = 1 - P(A)$.

Proof:

From definition. $A \cap A' = \emptyset$; $A \cup A' = S$

$$P(A \cup A') = P(S)$$
$$P(A \cup A') = P(S) = 1$$
$$P(A \cup A') = P(A) + P(A') = 1$$
$$P(A') = 1 - P(A)$$

Theorem 2: Let S be a sample space with probability function $P(.)$; then $0 \leq P(A) \leq 1$ for any event A in S.

Proof:

By property (1). $P(A) \geq 0$

We need to show that $P(A) \leq 1$

From theorem (1). $P(A) + P(A') = 1$

But $P(A') \geq 0$

So. $P(A) = 1 - P(A') \leq 1$

Theorem 3: Let S be a sample space with a probability function $P(.)$. If $\emptyset$ is the impossible event, then $P(\emptyset) = 0$.

Proof: Observe that $\emptyset = S'$

From property (3), we get $P(S \cup S') = P(S) + P(S')$

$$P(S) + P(\emptyset)$$

But $S \cup S' = S$ and $P(S) = 1$

Therefore $P(\emptyset) = 0$

Theorem 4: If A_1 and A_2 are subsets of S such that $A_1 \subset A_2$, then $P(A_1) \leq P(A_2)$.

Theorem 5: Commutative laws:

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

Theorem 6: Associative laws:

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

Theorem 7: Distributive laws:

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

$$(A')' = A$$
$$A' = S \setminus A$$

Thus

$$A \cap S = A$$
$$A \cup S = S$$
$$A \cap \emptyset = \emptyset$$
$$A \cup \emptyset = A$$

Also

$$A \cap A' = \emptyset$$
$$A \cup A' = S$$
$$A \cap A = A$$
$$A \cup A = A$$

Theorem 11: De Morgan's laws:

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

Theorem 12:

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

Theorem 13:

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

If A and B are disjoint, that is $P(A \cap B) = \emptyset$,

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

Theorem 14:

$P(\emptyset) = 0$

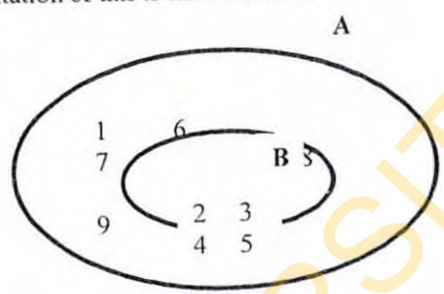
Theorem 15: Multiplicative law of Probability

If there are two events A and B, probabilities of their happening being $P(A)$ and $P(B)$ respectively, then the probability $P(AB)$ of the simultaneous occurrence of the events A and B is equal to the probability of A multiplied by the conditional probability of B(i. e. the probability of B when A has occurred) or the probability of B multiplied by the conditional probability of A i.e.$P(AB) = P(A)P(B/A) = P(B)P(A/B$

2.7 Venn Diagrams

A set is a collection of objects, which can be distinguished from each other. The objects comprising the set are called the elements of the set and they may be finite or infinite in number.

Venn diagrams are diagrammatical representation of sets. For instance, consider the set $A = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$, suppose that A has a subset $B = \{2, 3, 4, 5\}$. The diagrammatic representation of this is shown below.

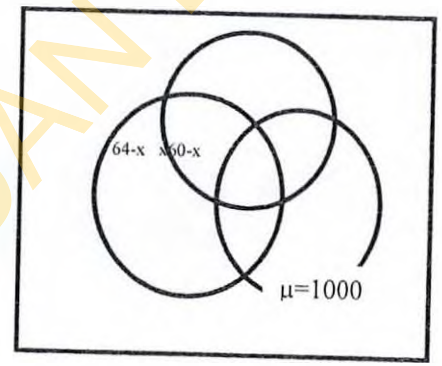


Solving Problems using Venn diagrams

Example 1: In a sample of 1000 foodstuff stores taken at an Ibadan market, the following facts emerged:

200 of them stock rice, 240 stock beans, 250 stock gaari, 64 stock both beans and rice, 97 stock both rice and gaari, while 60 stock beans and gaari. If 430 do not stock rice, do not stock beans and do not stock gaari, how many of the stores stock rice, beans and gaari?

Solution:



Let: R represent rice stores

B represents beans stores

G represents Gaari stores

Let x represents those that stock all the 3 food items

Those that stock gaari alone are $250 - [(97 - x) + x + (60 - x)] = 93 + x$

Those that stock beans alone are $240 - [(60 - x) + (x) + (64 - x)] = 116 + x$

Those that stock rice alone are $200 - [(64 - x) + x + (97 - x)] = 39 + x$

430 did not stock any of the food items

Therefore, $1000 = (39 + x) + (93 + x) + (116 + x) + x + (64 - x) + (60 - x) + (97 - x) + 430$

And $x = 1000 - 899 = 101$

Therefore 101 stores stock rice, beans and gaari.

2.8 The Principle of Inclusion and Exclusion

2.8.1 The Second Counting Principle

If a set is the disjoint union of two (or more) subsets, then the number of elements in the set is the sum of the numbers of elements in the subsets. i.e.

$n(A \cup B) = n(A) + n(B)$ implying that $|A \cup B| = |A| + |B|$ if A and B are disjoint.

Theorem 1: $|A \cup B| < |A| + |B|$ if A and B are not disjoint.

This is because $|A| + |B|$ counts every element of $A \cap B$ twice. Let us illustrate this with the following example.

Example 2: If $A = (2, 3, 4, 5, 6)$, $|A| = 5$ and $B = (3, 4, 5, 6, 7)$, $|B| = 5$

then, $|A| + |B| = 10$

$A \cup B = 2, 3, 4, 5, 6, 7$

$|A \cup B| = 6$

Since A and B are not disjoint, $|A \cup B| < |A| + |B|$

Compensating for this double counting yields the formula

$|A \cup B| = |A| + |B| - |A \cap B|$............eqn.(1)

From our example, $A \cap B = 3, 4, 5, 6$

$|A \cap B| = 4$

$|A \cup B| = 5 + 5 - 4$

$= 6$

thus proving equation (1)

Theorem 2: $|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|$ for three sets A, B and C.

Proof:

We know from equation (1) above that $|A \cup B| = |A| + |B| - |A \cap B|$

Then, for 3 sets, $|A \cup B \cup C| = |A \cup [B \cup C]|$

$$= |A| + |B \cup C| - |A \cap [B \cup C]|$$

Applying equation (1) to $|B \cup C|$ gives

$|A \cup B \cup C| = |A| + [|B| + |C| - |B \cap C|] - |A \cap [B \cup C]|$.........eqn (2)

Because $A \cap [B \cup C] = (A \cap B) \cup (A \cap C)$, we can apply equation(1) again to obtain

$|A \cap [B \cup C]| = |A \cap B| + |A \cap C| - |A \cap B \cap C|$.........eqn (3)

Finally, a combination of equations (2) and (3) yields

$|A \cup B \cup C| = [|A| + |B| + |C|] - [|A \cap B| + |A \cap C| + |B \cap C|] + |A \cap B \cap C|$.........eqn (4)

Thus proving theorem 2.

From this derivation, we notice that an element of $A \cap B \cap C$ is counted 7 times in equation(4), the first 3 times with a plus sign, then 3 times with a minus sign and then once more with a plus sign.

Example 3: If $A = \{1, 2, 3, 4\}$ $B = \{3, 4, 5, 6\}$ $C = \{2, 4, 6, 7\}$ then

$$A \cup B \cup C = \{1, 2, 3, 4, 5, 6, 7\}$$

$|A \cup B \cup C| = 7$....................(a)

$|A| = 4$

$|B| = 4$

$|C| = 4$

$|A| + |B| + |C| = 12$

$$A \cap B = 3, 4, A \cap C = 2, 4, B \cap C = 4, 6$$

In this example, $|A \cap B| = |A \cap C| = |B \cap C| = 2$ so that

$|A \cap B| + |A \cap C| + |B \cap C| = 6$ and

$$A \cap B \cap C = 4, |A \cap B \cap C| = 1$$

Therefore, $|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|$

$= 12 - 6 + 1 = 7$.................. (b)

Thus, (a) = (b) thus establishing theorem 2.

Generally, the Principle of Inclusion and Exclusion (PIE) states that:

If $A_1, A_2, \ldots, A_n$ are finite sets, the cardinality of their union

$$|A_1 \cup A_2 \cup \ldots \cup A_n| = \sum_{i=1}^{n} |A_i| - \sum_{1 \le i < j \le n} |A_i \cap A_j| + \sum_{1 \le i < j < k \le n} |A_i \cap A_j \cap A_k| - \cdots + (-1)^{n+1} |\cap A_i|$$

Proof:

On the left is the number of elements in the union of n sets. On the right, we first count elements in each of the sets separately and add them up. If the sets A_i are not disjoint, the elements that belong to at least two of the sets A_i, or the intersections $A_i \cap A_j$, are counted more than once. We wish to consider every such intersection, but each only once. Since $A_i \cap A_j = A_j \cap A_i$, we should consider only pairs (A_i, A_j) with $i < j$.

When we subtract the sum of the number of elements in such pairwise intersections, some elements may have been subtracted more than once. Those are the elements that belong to at least three of the sets A_i. We add the sum of the elements of intersections of the sets taken three at a time. (Note: the condition $i < j < k$ ensures that every intersection is counted only once)

The process continues with sums being alternately added or subtracted until we come to the last term which is the intersection of all sets A_i thus proving the theorem.

Let $S = A_1 \cup A_2 \cup \ldots \cup A_n$ and $A_i^{\,c} = S \backslash A_i$ then the PIE principle can also be expressed as

$|A_1^{\,c} \cap \ldots \cap A_n^{\,c}| =$

$$|S| - \sum_{i=1}^{n} |A_i| + \sum_{1 \le i < j \le n} |A_i \cap A_j| - \sum_{1 \le i < j < k \le n} |A_i \cap A_j \cap A_k| + \cdots$$
$$- (-1)^{n+1} |\cap A_i|$$

Example 4: Let A be the subset of the first 700 hundred numbers $S = \{1, 2, \ldots, 700\}$ that are divisible by 7. Find the number of elements in S that are not divisible by 7.

Solution:

$$A = \{7, 14, 21, 28, 35, 42, 49, \ldots\}$$
$$|A| = 100$$
$$|A^c| = |S| - |A|$$
$$= 700 - 100$$
$$= 600$$

Example 5: Find the number of integers from 1 to 1000 that are not divisible by 5, 6 and 8

Solution: Let A_1, A_2, A_3 be the subset consisting of those integers that are divisible by 5, 6 and 8. The number we are interested in is

$|A_1' \cap A_2' \cap A_3'| = 1000 - |A_1| - |A_2| - |A_3| + |A_1 \cap A_2| + |A_1 \cap A_3| + |A_2 \cap A_3| - |A_1 \cap A_2 \cap A_3|$

$$|A_1| = \left\lfloor \frac{1000}{5} \right\rfloor = 200 \quad |A_2| = \left\lfloor \frac{1000}{6} \right\rfloor = 166 \quad |A_3| = \left\lfloor \frac{100}{8} \right\rfloor = 125$$

Note: The results for $|A_1|$, $|A_2|$ and $|A_3|$ were achieved by using the round down, notation $\lfloor \ \rfloor$ which involves the dropping of the fractional part.

To compute the number in a 2 and 3 – set interaction, we use the least common multiple (LCM) i.e.

$$|A_1 \cap A_2| = \left\lfloor \frac{1000}{30} \right\rfloor = 33$$

$$|A_1 \cap A_3| = \left\lfloor \frac{1000}{40} \right\rfloor = 25$$

$$|A_2 \cap A_3| = \left\lfloor \frac{1000}{24} \right\rfloor = 41$$

$$|A_1 \cap A_2 \cap A_3| = \left\lfloor \frac{1000}{12} \right\rfloor = 8$$

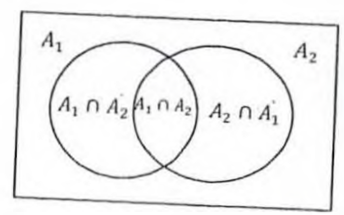
Thus, $|A_1' \cap A_2' \cap A_3'| = 1000 - 200 - 166 - 125 + 33 + 25 + 41 - 8 = 600$ If A_1 and A_2 are any two events of an experiment with sample space S, then we have the addition rule

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$$

Proof:

In a Venn diagram

Fig. 1.1



$P(A_1 \cup A_2) = P(A_1) \cup P(A_2) = 1$

$P(A_1 \cup A_2) = P(A_1) + P(A_2 \cap A_1^C)$

but $P(A_2 \cap A_1^C) = P(A_2) - P(A_1 \cap A_2)$

$\therefore P(A_1 \cup A_2) = P(A_1) + P(A_2) P(A_1 \cap A_2) Addition\ rule$

$\therefore P(A_1 \cup A_2) = P(A_1) + P(A_2) P(A_1 \cap A_2) Addition\ rule$

However, if A_1 and A_2 have no point in common, that is when A_1 and A_2 are mutually exclusive

$$P(A_1 \cap A_2) = 0 \text{ since } A_1 \cap A_2 = \emptyset$$

We have $P = (A_1 \cup A_2) = P(A_1) + P(A_2)$ Special Addition rule $P = (A_1 \cup A_2) = P(A_1) + P(A_2)$ Special Addition rule

Using the same procedure fort any three events *A*, *B* and *C*.

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

Example: A coin is rolled three times, what is the probability of getting (i) 1 head, (ii) 2 heads, (iii) at least 2 heads.

Solution: Let *H* and *T* represent Head and Tail respectively.

Let the sample space be defined as

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

(i) $P(1 \text{ head}) = \{HTT, THT, TTH\} = \frac{3}{8}\}$

(ii) $P(2 \text{ head}) = \{HHT, THH, HTH\} = \frac{3}{8}\}$

(iii) $P(\text{at least 2 head}) = P(2 \text{ heads}) + P(3 \text{ heads})$

$= \frac{3}{8} + \frac{1}{8} = \frac{4}{8} = 0.5$

Note: The events of 2 heads and 3 heads are mutually exclusive.

Examples: A bag contains 8 black balls; 3 red balls, 4 green balls and 5 yellow balls all of which are of the same size. If a ball is drawn at random from the bag, what is the probability that the ball is (i) black, (ii) either yellow or green (iii) not black, (iv) neither black nor green, (v) black and yellow?

Solution: Let *B*, *R*, *G* and *Y* represent the event of black, red, green and yellow ball respectively. Total number of balls = 20.

(i) $P(B) = \frac{n(B)}{n(S)} = \frac{8}{20} = 0.4$

(ii) $P(Y \cup G) = P(Y) + P(G)$

$\frac{5}{20} + \frac{4}{20} = \frac{9}{20} = 0.5$

(since only one ball is drawn $P(Y \cap G) = 0$)

(iii) $P(B^{'}) = 1 - P(B) = 1 - \frac{8}{20} = 0.6$

(iv) $P(B \cup G)^c = 1 - P(B \cup G)$

$= 1 - [P(B) + P(B)]$

$= 1 - \left[\frac{8}{20} + \frac{4}{20}\right]$

$= \frac{8}{20}$

$= 0.4$

Alternatively,

P(neither Black nor Green) = P(Yellow or Red)

$= P(Y) + P(B)$

$= \frac{5}{20} + \frac{3}{20}$

$= \frac{8}{20} = 0.4$

(v) $P(B \cap Y) = 0$ see note in (ii) above.

Example: A survey of 500 students taking one or more courses in Algebra, Physics and Statistics during one semester revealed the following numbers of students in indicated subject:

Algebra 186	Algebra and Physics 83
Physics 295	Physics and Statistics 217
Statistics 329	Algebra and Statistics 63

A student is selected at random what is the probability that he takes

(i) all the three subjects

(ii) Statistics but not Physics

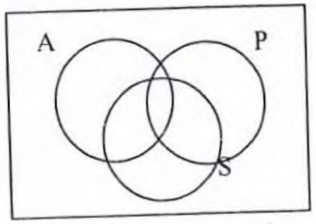
(iii) Statistics but not Physics and Algebra

(iv) Statistics, Algebra but not Physics

(v) Algebra or Physics

Solution: Let A, P and S denotes the event of a student taking Algebra, Physics and Statistics respectively.

Presenting the information in a Venn diagram we have



$$n(A \cap A \cap B^C) = n(A \cap S) - n(A \cap P \cap S) = 10$$
$$n(P \cap S \cap B^C) = n(P \cap S) - n(A \cap P \cap S) = 164$$
$$n(A \cap P \cap S^C) = n(A \cap P) - n(A \cap P \cap S) = 30$$

Using the addition rule, we can find the number of students that takes all the three subjects.

$$n(A \cup P \cup S) = n(A) + n(P) + n(A \cap P) - n(A \cap S) + n(A \cap P \cap S)$$
$$500 = 186 + 329 - 83 - 217 - 63 + n(A \cap P \cap S)$$
$$\therefore n(A \cap P \cap S) = 53$$
$$\therefore P(All\ three\ subjects) = \frac{53}{500} = 0.106$$

(ii) P(Statistics but not Physics)

$$\equiv P(S \cap P^C)$$
$$= P(S) - P(S \cap P)$$
$$= \frac{329}{500} - \frac{217}{500}$$
$$= \frac{112}{500} = 0.224$$

(iii) P (Statistics but not Physics and Algebra)

$$\equiv P(S) - P(A \cap P)$$
$$= P(S) - P(A \cap P) - P(S \cap P) + P(A \cap P \cap S)$$
$$= \frac{329}{500} - \frac{83}{500} - \frac{217}{500} + \frac{53}{500}$$
$$= \frac{82}{500}$$
$$= 0.164$$

(iv) P (Statistics, Algebra but not Physics)

$$\equiv P(S) - P(S \cap P^c)$$
$$= P(S) - [P(S \cap P) - P(A \cap P \cap S)]$$
$$= \frac{329}{500} - \frac{217}{500} + \frac{53}{500}$$
$$= \frac{165}{500}$$
$$= 0.33$$

(v) P (Algebra or Physics)

$$\equiv P(A \cap P)$$

i.e. $P(\cup P) = P(A) + P(P) - P(A \cup P)$

$$= \frac{186}{500} + \frac{295}{500} - \frac{83}{500}$$
$$= \frac{398}{500}$$
$$= 0.796$$

2.9 Conditional Probability and Independence

If A and B are any two events, the conditional probability of A given B is the probability that even A will occur given that event B has already occurred.

This is equivalent to the probability of events A and B (occurring simultaneously) divided by probability of event B.

i.e. $P(A/B) = \frac{P(A \cap B)}{P(B)}$ provided $P(B) \neq 0$

$$= P(A \cap B) = P(B)P\ (A/B) = P(A)P(B).$$

In general

$$P(A_1 \cap A_3 \cap \ldots A_n) = P(A_1)P(A_1/A_2)P(A_3/A_1 \cap A_2) \ldots P\ (A \cap)/(A_1 \ldots A_n)$$

Let A_1, A_2, A_2 denote the 1st, 2nd and 3rd cards

$$P(A_1 \cap A_2 \cap A_3) = P(A_1).P(A_2/A_1).P(A_3/A_1 \cap A_2)$$
$$= \frac{4}{52} x \frac{3}{51} x \frac{2}{50}$$
$$= \frac{24}{132600}$$
$$= 0.00018$$

Example: A bag contains 10 while balls and 15 black balls. Two balls are drawn in succession (a) with replacement (b) without replacement. What is the probability that

(i) the first ball is black and the second white

(ii) both are black

(iii) both are of the same colour

(iv) both are of different colours

(v) the second is black given that the first is white.

Solution: Let *B* and *W* denote black and white balls respectively.

(a) with replacement

(i) $P(B \cap W) = P(B).P(W)$

$$= \frac{15}{25} x \frac{10}{25} = 0.24$$

(ii) $P(B_1 \cap B_2) = P(B)\, x\, P(B)$

$$= \left(\frac{15}{25}\right) = 0.36$$

(iii) P (both black or both white) = $P(B_1 \cap B_2) + P(W_1 \cap W_2)$

$$= \left(\frac{15}{25}\right) + \left(\frac{10}{25}\right)$$

$$= 0.36 + 0.16$$

$$= 0.52$$

(iv) P (both are of different colours) = $P(B \cap W) + P(W \cap B)$

$$= \left[\frac{15}{25} x \frac{10}{25}\right] + \left[\frac{10}{25} x \frac{15}{25}\right]$$

$$= 2(0.24)$$

$$= 0.40$$

(v) $P(B/W) - \frac{P(B \cap W)}{P(W)} = \frac{0.24}{0.4}$

$$= 0.6$$

From the last result, we could see that the two events are independent, hence,

$$P(B/W) = P(W) = 0.6.$$

because the drawing is with replacement.

(b) without replacement

(i) $P(B \cap W) = P(B).P(W/B)$

$$= \frac{15}{25} x \frac{10}{25} = 0.25$$

(ii) $P(B_1 \cap B_2) = P(B_1).\ P(B_2/B_1)$

$$= \frac{15}{25} x \frac{14}{24} = 0.35$$

(iii) P (both black or both white) = $P(B_1)P(B_2/B_1) + P(W_1)P(W_2/W_1)$

$$= \frac{15}{25} x \frac{14}{24} + \frac{10}{25} x \frac{9}{24}$$

$$= 0.35 + 0.15$$

$$= 0.50$$

(iv) P (both are of different colours) = $P(B)P(W/B) + P(W)P(B/W)$

$$= \frac{15}{25} x \frac{10}{25} + \frac{10}{24} x \frac{15}{24}$$

$$= 0.25 + 0.25$$

$$= 0.50$$

(v) $P(B/W) = \frac{P(B \cap W)}{P(W)}$

$$= \frac{15}{25} x \frac{10}{24} / \frac{10}{25}$$

$$= \frac{0.25}{0.4}$$

$$= 0.625$$

2.10 Statistical Independence

Two events *A* and *B* are said to be independent if the probability that *B* occurs is not influenced by whether *A* has occurred or not.

i.e. $P(B) = P(B/A)$

Hence events *A* and *B* are independent if

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

Three events are said to be mutually independent if

(i) They are pairwise independent, i.e.

$$P(A \cap B) = P(A).P(B); P\ (A \cap C) = P(A) \cot P(C);$$

$P(B \cap C) = P(A).\ P(B).P(C)$ and

(ii) $P(A \cap B \cap C) = P(A).\ P(B).P(C)$

It should be noted that mutually exclusive events are not independent as the occurrence of one rules out the possibility of the other, i.e.

$$P(A/B) = P(B/A) = 0.$$

Example: What is the chance of getting two sixes in two rollings of a single die?

Solution:

P (six in 1st die) $= \frac{1}{6}$ P (six in 2nd die) $= \frac{1}{6}$

Since the two events are independent

$$\text{P (six in 1st and 2nd die)} = \frac{1}{6}.\frac{1}{6} = \frac{1}{36}$$

Example: *A* and *B* plays 12 games of Ayo (Yoruba traditional game). *A* wins 6 and *B* wins 4 and two are drawn. They agree to play three games more. Find the probability that:

(i) A wins all the three games

(ii) Two games end in a tie

(iii) *A* and *B* wins alternately

(iv) *B* wins at least one game.

Solution: Let *A* and *B* represent the event of *A* and *B* winning the game and *D* winning the game and *D* denote the event of a tie.

$$P(A) = \frac{6}{12} = \frac{1}{2}$$

$$P(B) = \frac{4}{12} = \frac{1}{3}$$

$$P(D) = \frac{2}{12} = \frac{1}{6}$$

(i) $\text{P(A wins all three)} = \frac{1}{2}x\frac{1}{2}x\frac{1}{2} = \frac{1}{8}$

(ii) $\text{P (2 games and in ties)} = \text{P(D. D. D)}^C + \text{P}(\text{D}^C.\text{D. D}) + \text{P}(\text{D. D}^C.\text{D})$

$$= \left(\frac{1}{6}x\frac{1}{6}x\frac{5}{6}\right) + \left(\frac{5}{6}x\frac{1}{6}x\frac{1}{6}\right) + \left(\frac{1}{6}x\frac{5}{6}x\frac{1}{6}\right)$$

$$= \frac{5}{72}$$

(iii) If A and E – B wins alternately in two mutually exclusive ways.

$$= P(ABA) + P(B.A.B)$$

$$= \left(\frac{1}{2}x\frac{1}{3}x\frac{1}{2}\right) + \left(\frac{1}{3}x\frac{1}{2}x\frac{1}{3}\right)$$

$$= \frac{5}{36}$$

(iv) P(B wins at least one game = 1 – P(no game)

$$= 1 - P(B_1'B_2'B_3')$$

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

$$= \frac{19}{27}$$

Example: An unbiased die is rolled *n* times

(i) Determine the probability that at least one six is observed in the *n* trials.

Calculate the value of *n* if the probability is to be approximately $\frac{1}{2}$

Solution:

$$\text{P(a six in a throw)} = \frac{1}{6}$$

$$\text{P(no six in a throw)} = \frac{5}{6}$$

(i) $\text{P(at least 1 six in } n \text{ trials)} = 1 - \text{P(no six in } n \text{ trials)}$

$$= 1 - \left(\frac{5}{6}\right)$$

(ii) If the probability is $\frac{1}{2}$; then

$$\frac{1}{2} = 1 - \left(\frac{5}{6}\right)$$

$$\Rightarrow \left(\frac{5}{6}\right) = \frac{1}{2}$$

$$n \log\left(\frac{5}{6}\right) = \log\left(\frac{1}{2}\right)$$

$$n = \frac{\log(1/2)}{\log(5/6)}$$

$$n = 4$$

Example: Determine the probability for each of the following events.

(a) A king or an ace or jack of clubs or queen of diamond appears in a single card from a well shuffled ordinary deck of cards.

(b) The sum of 8 appears in a single toss of. a pair of fair dice.

(c) A 7 or 11 comes up in a single toss of a pair of dice

Solution:

(a) P(King) $=\frac{4}{52}$; P(an ace) $=\frac{4}{52}$

P (Jack of club) $=\frac{1}{52}=\frac{4}{52}\cdot\frac{1}{4}$

P (Queen of diamond) $=\frac{1}{62}$

P (a kind, an ace, J. of club or Q. of diamond)

$$\left(\frac{4}{52}+\frac{4}{52}+\frac{1}{52}+\frac{1}{52}\right)=\frac{5}{26}$$

(b)

Dice	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

P(sum = 8) $=\frac{5}{36}$

(c) P(7) $=\frac{6}{36}$; P(11) $=\frac{2}{36}$

$$P(7 \text{ or } 11)=\frac{6}{36}+\frac{2}{36}=\frac{2}{9}$$

Example: A pair of fair coins is tossed once. Let A be the event of head on the first coin and B the event of head on the second coin first coin and B the event of head on the second coin while C is the event of exactly o head is events A, B and C mutually independent?

Solution:

$S=\{HH,HT,TH,TT\}$

$A=\{HH,HT\}, B=\{HH,TH\}$

$C=\{HT,TH\}$

$A\cap B=\{HH\}, A\cap C=\{HT\}, B\cap C=\{TH\}, A\cap B\cap c=\emptyset$

$\therefore P(A)=P(B)=P(C)=\frac{2}{4}=0.5$

$P(A\cap B)=P(A).P(B)=\frac{1}{4}; P(B\cap C)=P(B).P(C)=\frac{1}{4}$

$P(A\cap C)=P(A).P(C)=\frac{1}{4}; P(A\cap B\cap C)\neq P(A).P(B).P(C)$

Hence events A, B and C are not mutually exclusive.

Example: An urn contains 'P' white and 'q' black balls and the second contains 'C' white and 'd' black balls. A ball is drawn at random from the first and put into the second. Then a ball is drawn from the second urn. Find the probability that the ball is white.

Solution: This is a conditional probability.

Total number of ball in the 1st Urn is $(P+q)$

Total, number of ball in the 2nd Urn is $(c+d)$

Total number of ball in the 2nd after the first draw is $c+d+1$

P (white ball in the 2nd urn)

$=P(W)P(W/B+P(W)P)(W/W)$

$=\frac{c}{c+d+1}\left(\frac{p}{p+d}\right)+\frac{c}{c+d+1}\left(\frac{q}{p+q}\right)$

$=\frac{c(p+q)}{(c+d+1)(p+q)}$

$=\frac{c}{c+d+1}$

CHAPTER 3
CONDITIONAL PROBABILITY AND BAYES' THEOREM

3.1 Conditional Probability

Supposed A and B are any two events such that A is the prior event and B is the posterior event. There is the possibility that there are points of intersection between the two events such that the occurrence of one is conditioned on the other. Thus we give the following definition.

Definition 1: Let A and B be two events in the sample space S with given probability space [S, A, B, P(.)] where P(.) is a real valued function, the conditional probability of event A given the event B has occurred denoted by P[A/B], is defined by

$$P(A/B) = \frac{P(A\cap B)}{P(B)}, P(B)>0,$$ this implies that

$$P(A\cap B) = P(A/B).P(B)$$

Also $P(B/A) = \frac{P(A\cap B)}{P(A)}, P(A)>0$ which also implies $P(A\cap B) = P(B/A).P(A)$

Example 1: Two students are chosen at random from a class consisting of 18 boys and 12 girls. What is the probability that the two students selected are:
(a) both boys (b) both girls (c) of the same sex (d) a boy and a girl.

Solution: Let B_1 to be the event that the first student selected is a boy.
Let B_2 be the event that the second student selected is a boy.

Let $B_1 \cap B_2$ denote the event that the two students selected are both boys.

(i) $P(B_1 \cap B_2) = P(B_1).P(B_2/B_1)$ where

$$P(B_1) = \frac{18}{30} = 3/5$$

$$P(B_2/B_1) = 17/29$$

Therefore, $P(B_1 \cap B_2) = \frac{3}{5} \times \frac{17}{29}$

$$= \frac{51}{145}$$

(ii) Let $G_1 \cap G_2$ denote the event that the two students selected are both girls.

$$PG_1 \cap G_2 = P(G_1)\ P(G_2/G_1)$$

$$= \frac{12}{30} \times \frac{11}{29}$$

$$= \frac{132}{870} = \frac{22}{145}$$

(iii) $B_1B_2 \cup G_1G_2$ is the event that both students selected are of the same sex.

$$P(B_1B_2 \cup G_1G_2) = P(B_1B_2) + P(G_1G_2)$$

Since $B_1 \cap B_2$ *and* G_1G_2 are mutually exclusive

$$\therefore P(B_1B_2 \cup G_1G_2) = \frac{51}{145} + \frac{22}{145}$$

$$= \frac{73}{145}$$

(iv) $B_1G_2 \cup G_1B_2$ is the event that the two students selected are a boy and a girl.

$$(B_1G_2 \cup G_1B_2) = P(B_1G_2) + P(G_1B_2)$$

$$= P(B_1).P(G_2/B_1) + P(G_1).P(B_2/G_1)$$

$$= \frac{18}{30} \times \frac{12}{29} + \frac{12}{30} \times \frac{18}{29}$$

$$= \frac{3}{5} \times 12/29 + \frac{2}{5} \times 18/29 = 72/145$$

Example 2: A boy has 10 identical marbles in a container consisting of 6 red and 4 blue marbles. He draws two marbles at random one after the other from the container without replacement. Find the probability that:

(a) the first draw is red while the second is blue
(b) both draws are of the same colour
(c) both draws are of different colours.

Solution:

(a) Let R_1 be the event that the first draw is red
Let B_2 be the event that the second draw is blue.

The event $R_1 \cap B_2$ is the event that the first draw is red while the second draw is blue.

$$P(R_1 \cap B_2) = P(R_1) . P(B_2 / R_1) \text{ where}$$

$$P(R_1) = 6/10$$

$$P(B_2 / R_1) = 4/9$$

$$\therefore P(R_1 \cap B_2) = \frac{6}{10} x \frac{4}{9}$$

$$= \frac{4}{15}$$

(b) Let R_1 be the event that the first draw is red.
Let R_2 be the event that the second draw is red.
Let B_1 be the event that the first draw is blue
Let B_2 be the event that the second draw is blue.

Therefore

$P(R_1R_2 \cup B_1B_2) = P(R_1R_2) + P(B_1B_2)$ since R_1R_2 and B_1B_2 are mutually excusive.

$$P(R_1R_2) = P(R_1)\ P(R_2 / R_1)$$

$$= \frac{6}{10} x \frac{5}{9}$$

$$= 1/3$$

$$P(B_1B_2) = P(B_1)\ x\ P(B_2 / B_1)$$

$$= \frac{4}{10} x \frac{3}{9}$$

$$= \frac{12}{15}$$

Therefore, $P(R_1R_2 \cap B_1B_2) = \frac{1}{3} + \frac{2}{15} = \frac{7}{15}$

3.2 Independence

Recall that P [A/B] $= \dfrac{P[A \cap B]}{P(B)} \quad P(B) > 0$

Definition 2: Two events A and B are said to be stochastically or statistically independent if and only if any one of the following conditions is satisfied:

(i) P(A∩B) = P(A)P(B)

(ii) P(A/B) = P(A) if P(B) > 0

(iii) P(B/A) = P(B) if P(A) >0

It is easily shown that (i) implies (ii), (ii) implies (iii) and (iii) implies (i). See Post-test (2).

Therefore, $P(A \cap B) = P(A / B)P(B) = P(B / A)P(A)$ if P(A) and P(B) are non-zero. This implies that one of the events is independent of the other. In fact,

$$P[A / B] = \frac{P(A \cap B)}{P(B)} = \frac{P(B / A)P(A)}{P(B)} = \frac{P(B)P(A)}{P(B)} = P(A)$$

So, if P(A), P(B) > 0 and one of the events is independent of the other, then the second event is also independent of the first. Thus, independence is a symmetric relation.

Remark: Two mutually exclusive events A and B are independent if and only if P(A) P(B) = 0 which is true if and only if either P(A) or P(B) = 0

Also, if P(A) ≠ 0 and P(B) ≠ 0, then A and B independent implies that they are not mutually exclusive.

Definition 2: Events $A_1, A_2, \ldots, A_n$ from A in the probability space $[S, A, P(.)]$ are said to be completely independent if and only if

(i) $P(A_i \cap A_j) = P(A_i)P(A_j)$ for $i \neq j$

(ii) $P(A_i \cap A_j \cap A_k) = P(A_i)P(A_j)P(A_k)$ for $i \neq j, j \neq k, i \neq k$

(iii) $P\left[\bigcap_{i=1}^{n} A_i\right] = \prod_{i=1}^{n} P(A_i)$

Note: (i) These events are said to be pairwise independent if

$$P(A_i \cap A_j) = P(A_i)P(A_j) \text{ for all } i \neq j$$

(ii) Pairwise independence does not imply independence

(iii) A and B mutually exclusive implies that they are not independent.

Example 3: Suppose two dice are tossed. Let A denote the event of an odd total, B, the event of an ace on the first die, and C the event of a total of seven.

(i) Are A and B independent?

(ii) Are A and C independent?

(iii) Are B and C independent?

Solution:

$P[A/B] = \frac{1}{2} = P(A)$

$P[A/C] = 1 \neq P[A] = \frac{1}{2}$

$P[C/B] = \frac{1}{6} = P(C) = \frac{1}{6}$

So, A and B are independent

A is not independent of C

B and C are independent

Example 4: Let A_1 denote the event of an odd face on the first die,

Let A_2 denote the event of an odd face on the second die,

Let A_3 denote the event of an odd total in the random experiment consisting of two dice. Then,

$P(A_1)P(A_2) = \frac{1}{2} x \frac{1}{2} = P(A_1 \cap A_2)$

$P(A_1)P(A_3) = \frac{1}{2} x \frac{1}{2} = P[A_3/A_1]P(A_1) = P(A_1 \cap A_3)$

$P(A_2 \cap A_3) = \frac{1}{4} = P(A_2)\ P(A_3)$

Therefore, A_1, A_2 and A_3 are pairwise independent.

But $P(A_1 \cap A_2 \cap A_3) = 0 \neq \frac{1}{8} = P(A_1)P(A_2)P(A_3)$

So, A_1, A_2 and A_3 are not independent.

3.3 Bayes Theorem

Given that $P(A/B) = \dfrac{P(A \cap B)}{P(B)}$

$$P(B/A) = \frac{P(B \cap A)}{P(A)}$$

This implies that $P(A \cap B) = P(B \cap A) = P(B/A)P(A)$

Therefore $P(A/B) = \dfrac{P(B/A)P(A)}{P(B)}$

The above is known as Bayes theorem.

3.4 Total Probability Rule and Baye's Theorem

If there are two or more events where one is the prior and the other in the posterior event, it is often desirable to determine the probability that a particular event has occurred given that the other event has previously occurred. Even though this kind of problem can be solve by merely applying the addition and multiplication rule, much compact procedure has been developed called the Baye's theorem.

Baye's Theorem

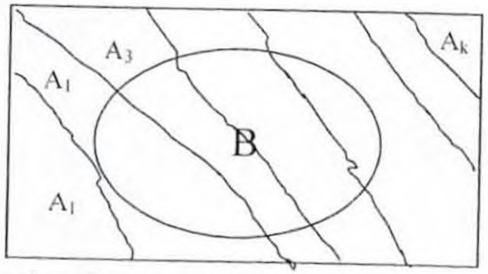
Let a sample space S of an experiment be partitioned into n mutually exclusive and exhaustive events $A_1, A_2, \ldots, A_n$. Let B be an arbitrary event that occurred after the experiment been performed. Such that $P(A_i) \neq 0, i = 1, 2, \ldots, n$ then,

$$P(B) = \sum_{i=1}^{n} P(A_i)(B/A_i)$$

and

$$P(A_i/B) = \frac{P(A_i)P(B/A_i)}{P(B)}$$

Proof: Let the events A_i and B be depicted as in Fig. 1.3



By definition of conditional probability, we have

$$P(B/A_i) = \frac{P(A_i \cap B)}{P(A_i)}$$

$$P(A_i \cap B) = P(A_i)P(B/A_i)$$

We know that

$$P(A_i/B) = \frac{P(A_i \cap B)}{P(B)}$$

Such that $P(A_i \cap B) = P(B)P(A_i/B)$

But total probability is

$$P(B) = P(A_i \cap B) + P(A_2 \cap B) + P(A_2 \cap B) + \cdots + P(A_n \cap B)$$

Using (1) in (3) we have

$$P(B) = P(A_1)P(B/A_1) + P(A_2)P(B/A_2) + \cdots + P(A_n)(B/A_n)$$

$$= \sum_{i=1}^{n} P(A_i)(B/A_i)$$

Using (3) in (2) we have the Bayes' formula defined as:

$$P(A_i/B) = \frac{P(A_i)P(B/A_i)}{\sum_i^n 1\, P(A_i)P\,(B/A_i)}$$

Example 1: The contents of 3 identical baskets B_i(i = 1, 2, 3) are:

B_1: 4 apples and 1 orange

B_2: 1 apple and 4 oranges

B_3: 2 apples and 3 oranges

A basket is selected at random and from it, a fruit is picked. The fruit picked turns out to be an apple on inspection. What is the probability that it come from the first basket

Solution:

Let E be the event of picking an apple.

Using the table below:

State of Nature	$P(B_i)$	$P(E/B_i)$	$P(B_i)P(E/B_i)$	$P(B_i/E)$
B_1 (4A. 10)	$\frac{1}{3}$	$\frac{4}{5}$	$\frac{4}{15}$	$\frac{4}{7}$
B_2 (1A. 40)	$\frac{1}{3}$	$\frac{1}{5}$	$\frac{1}{15}$	$\frac{1}{7}$
B_3 (2A. 30)	$\frac{1}{3}$	$\frac{2}{5}$	$\frac{2}{15}$	$\frac{2}{7}$
Total	1	-	$\frac{7}{15}$	1

The required probability

$$P(B_1/E) = \frac{4}{7}$$

i.e. $$P(B_1/E) = \frac{P(E/B_1)P(B_1)}{\sum_{i=1}^{3} P(B_i)P(E/B_i)}$$

$$= \frac{\frac{4}{5}\cdot\frac{1}{3}}{\frac{7}{15}} = \frac{4}{7}$$

Example 2: In a certain town, there are only two brands of hamburgers available, Brand A and Brand B. It is known that people who eat Brand A hamburger have a 30% probability of suffering stomach pain and those who eat Brand B hamburger have a 25% probability of suffering stomach pain. Twice as many people eat Brand B compared to Brand A hamburgers. However, no one eats both varieties. Supposing one day, you meet someone suffering from stomach pain who has just eaten a hamburger what is the probability that they have eaten Brand A and what is the probability that they have eaten, Brand B?

Solution: Let S denote people who have just eaten a hamburger

Let A denote people who have eaten Brand A hamburger

Let B denote people who have eaten Brand B hamburger

Let C denote people who are suffering stomach pains

We are given that

$$P(A) = \frac{1}{3}$$

$P(B) = \frac{2}{3}$

$P(C/A) = 0.3$

$P(C/B) = 0.25$

$S = A \cup B$

As those who have stomach pain have either eaten Brand A or B. then $A \cap B = \phi$

$P(C) = P(C \cap S) = P(C \cap A) + P(C \cap B)$

$= P(C/A)P(A) + P(C/B)P(B)$

$0.3 \times \frac{1}{3} + 0.25 \times \frac{2}{3}$

$= \frac{8}{30}$

Then $P(A/C) = \frac{P(C/A)P(A)}{P(C)}$

$= \frac{0.3 X (\frac{1}{3})}{\frac{8}{30}} = \frac{3}{8}$

And $P(B/C) = \frac{P(C/B)P(B)}{P(C)} = \frac{0.25x(\frac{2}{3})}{\frac{8}{30}}$

$= \frac{5}{8}$

So, if someone has stomach pain, the probability that they have eaten Brand A hamburger is $\frac{3}{8}$ and the probability that they have eaten Brand B is $\frac{5}{8}$.

Example 4: Suppose 15% of apple and 10 consignments were toxic. If the consignment consists of 60% apple and 40% mango, what is the probability that a fruit selected at random is toxic?

Solution: Let B be the event of toxic fruit and. A_1, A_2 be events of selected fruit being an apply and a mango respectively.

$$P(A_1) = \frac{60}{100} = 0.6; \quad P(A_2) = \frac{40}{100} = 0.4$$

$$P(B/A_1) = \frac{15}{100} = 0.15; \quad P(B/A_2) = \frac{10}{100} = 0.1$$

$$P(B) = P(A_1)P(B/A_1) + P(A_2)P(B/A_2)$$

$$= (0.6 \, x \, 0.15) + (0.4 \, x \, 0.1) = 0.13$$

Example 5: Every Saturday a fisherman goes to the river, the sea and a lake to catch fishes with probabilities $\frac{1}{4}$; $\frac{1}{2}$ and $\frac{1}{4}$ and respectively. If he goes to the sea, there is an

80% chance of catching fish, the corresponding figures for the river and the lake are 40% and 60% respectively.

(a) Find the probability that he catches fish on a given Saturday.

(b) What is the probability that he catches fish an at least three of the fire consecutive Saturdays?

(c) If on a particular Saturday, he comes home without catching anything, where is it most likely he has been?

(d) His friend, who is also a fisherman, chooses among the three locations with equal probabilities. Find the probability that the two fishermen will meet at least once in the next three weekends? (Any assumptions made should be clearly stated).

Solution: Let S, R and L denote the event that he goes to the sea, the river and the lake respectively and F denote the event that he catches fish.

$$P(S) = \frac{1}{2}; P(F/S) = \frac{4}{5}$$

$$P(R) = \frac{1}{4}; P(F/R) = \frac{2}{5}$$

$$P(L) = \frac{1}{4}; P(F/L) = \frac{3}{5}$$

(a) Using the idea of total probability,

$$P(F) = P(S)P(F/S) + P(R)P(F/R) + P(L)P(F/L)$$

$$= \frac{1}{2} x \frac{4}{5} + \frac{1}{4} x \frac{2}{5} + \frac{1}{4} x \frac{3}{5}$$

$$= \frac{13}{20} = 0.65$$

(b) Let the number of Saturdays on which he catches fish be a random variable X with $B\left(5, \frac{13}{20}\right)$.

$$P(X \geq 3) = P(X = 3) + P(X = 4) + P(X = 5)$$

$$= \binom{5}{3} (0.65)^3 (6.35)^2 + \binom{5}{3} (0.65)^4 (0.35)^1 + \binom{5}{5} (0.65)^5 (0.35)^0$$

$$= 0.3364 + 0.3124 + 0.116$$

$$= 0.765$$

Here we need to calculate the probability that he goes to each of the locations without catching fish.

$$P(S/F^1) = \frac{P(S\cap F^1)}{P(F^1)}$$

$$= \frac{P(S)P(F^1/S)}{P(F^1)} = \frac{\frac{1}{2}x\frac{1}{5}}{\frac{7}{20}} = \frac{2}{7} = 0.286$$

Similarly,

$$P(R/F^1) = \frac{P(R)P(F^1/R)}{P(F^1)} = \frac{\frac{1}{4}x\frac{3}{5}}{\frac{7}{20}} = \frac{3}{7} = 0.429$$

$$P(L/F^1) = \frac{P(L)P(F^1/R)}{P(F^1)} = \frac{\frac{1}{4}x\frac{2}{5}}{\frac{7}{20}} = \frac{2}{7} = 0.286$$

So it is most likely that he has been to the river.

(d) Let S_1, S_2 denote the event that the first and second fisherman goes to the sea respectively, and define R_1, R_2, L_1, L_2 similarly.

The probability that they meet on a given Saturday (assuming independence) is

$$P(S_1 \cap S_2) + P(R_1 \cap R_2) + P(L_1 \cap L_2)$$
$$= \frac{1}{2}x\frac{1}{3} + \frac{1}{4}x\frac{1}{3} + \frac{1}{4}x\frac{1}{3}$$
$$= \frac{1}{3} = 0.33$$

Probability that they fail to meet on a Saturday is

$$\left(1 - \frac{1}{3}\right) = \frac{2}{3} = 0.666$$

The probability that they fail to meet on three consecutive Saturdays is

$$\left(1 - \frac{1}{3}\right) = \frac{8}{27} = 0.296$$

The probability that they meet at least once in three weekends is

$= 1 - \text{P(failed to meet)}$

$= 1 - 0.296$

$= 0.703$

Example 6: Suppose 15% of apple and 10 consignments were toxic. If the consignment consists of 60% apple and 40% mango, what is the probability that a fruit selected at random is toxic?

Solution: Let B be the event of toxic fruit and A_1, A_2be events of selectedfruit being an apply and a mango respectively.

$$P(A_1) = \frac{60}{100} = 0.6;\quad P(A_2) = \frac{40}{100} = 0.4$$

$$P(B/A_1) = \frac{15}{100} = 0.15;\quad P(B/A_2) = \frac{10}{100} = 0.1$$

$$P(B) = P(A_1)P(B/A_1) + P(A_2)P(B/A_2)$$
$$= (0.6\ x\ 0.15) + (0.4\ x\ 0.1) = 0.13$$

Example 7: Every Saturday a fisherman goes to the river, the sea and a lake to catch fishes with probabilities $\frac{1}{4}, \frac{1}{2}$ *and* $\frac{1}{4}$ respectively. If he goes to the sea, there is an 80% chance of catching fish, the corresponding figures for the river and the lake are 40% and 60% respectively.

(a) Find the probability that he catches fish on a given Saturday.

(b) What is the probability that he catches fish an at least three of the fire consecutive Saturdays?

(c) If on a particular Saturday, he comes home without catching anything, where is it most likely he has been?

(d) His friend, who is also a fisherman, chooses among the three locations with equal probabilities. Find the probability that the two fishermen will meet at least once in the next three weekends? (Any assumptions made should be clearly stated).

Solution: Let S, R and L denote the event that he goes to the sea, the river and the lake respectively and F denote the event that he catches fish.

$$P(S) = \frac{1}{2}; P(F/S) = \frac{4}{5}$$
$$P(R) = \frac{1}{4}; P(F/R) = \frac{2}{5}$$
$$P(L) = \frac{1}{4}; P(F/L) = \frac{3}{5}$$

(a) Using the idea of total probability

$$P(F) = P(S)P(F/S) + P(R)P(F/R) + P(L)P(F/L)$$
$$= \frac{1}{2}x\frac{4}{5}x\frac{1}{4}x\frac{2}{5}x\frac{1}{4}x\frac{3}{5}$$
$$= \frac{13}{20} = 0.65$$

(b) Let the number of Saturdays on which he catches fish be a random variable X with $B\left(5,\frac{13}{20}\right)$

$$P(X \geq 3) = P(X = 3) + P(X = 4) + P(X = 5) = \binom{5}{3}(0.65)^3(6.35)^2 + \binom{5}{3}(0.65)^4(0.35)^1 + \binom{5}{5}(0.65)^5(0.35)^0$$

$$= 0.3364 + 0.3124 + 0.116$$

$$= 0.765$$

Here we need to calculate the probability that he goes to each of the locations without catching fish

$$P(S/F') = \frac{P(S \cap F')}{P(F')}$$

$$= \frac{P(S)P(F'/S)}{P(F')} = \frac{\frac{1}{2}x\frac{1}{5}}{\frac{7}{20}} = \frac{2}{7} = 0.286$$

Similarly,

$$P(R/F') = \frac{P(R)P(F'/R)}{P(F')} = \frac{\frac{1}{4}x\frac{3}{5}}{\frac{7}{20}} = \frac{3}{7} = 0.429$$

$$P(L/F') = \frac{P(L)P(P'/R)}{P(F')} = \frac{\frac{1}{4}x\frac{2}{5}}{\frac{7}{20}} = \frac{2}{7} = 0.286$$

So it is mostly likely that he has been to the river.

(d) Let S_1, S_2 denote the event that the first and second fishermen goes to the sea respectively, and define R_1, R_2, L_1, L_2similarly.

.the probability that they meet on a given Saturday (assuming independence) is

$$P(S_1 \cap S_2) + P(R_1 \cap R_2) + P(L_1 \cap L_2)$$

$$= \frac{1}{2}x\frac{1}{3}x\frac{1}{4}x\frac{1}{3}x\frac{1}{4}x\frac{1}{3}$$

$$= \frac{1}{3} = 0.33$$

Probability that they fail to meet on a Saturday is

$$\left(1 - \frac{1}{3}\right) = \frac{2}{3} = 0.666$$

The probability that they fail to meet on three consecutive Saturdays is

$$\left(1 - \frac{1}{3}\right)^3 = \frac{8}{27} = 0.296$$

The probability that they meet at least once in three weekends is

$$= 1 - P\,(failed\ to\ meet)$$

$$= 1 - 0.296$$

$$= 0.703$$

Practice Questions

1. If A_1, A_2, and A_3 be any three events, prove that

$$P\,(A_1 + A_2 + A_3) = \sum_{i=1}^{3} P(A_1) - \sum_{i=j} P(A_i A_j)\,P\,(A_1 + A_2 + A_3)$$

It is important to note that addition theorem can be validly applied only when the mutually exclusive events belong to the same set.

2. A newspaper vendor sells three papers: the Times, the Punch and the Commet. 70 customers bought the Times, 60 the Punch and 50 the Commet on a particular day. 17 bought Times and the Punch and 15 the Punch and the Commet and 16 the Commet and the Time while 3 customers bought all three papers. Every customer bought at least one type of paper. Using Venn diagram or otherwise; find:

(i) how many customers patronized the newsagent on that particular day?

(ii) how many customers bought a single paper?

(iii) how many customers bought Times but not Commet?

(iv) how many customers bought the Punch or Commet, but not the Times?

3. A random sample of 60 candidates who sat for Part I and II of an examination in 1984 is taken. The table below shows the number of candidates who passed or failed each part of the examination.

		Part I		
Part II	Pass	Pass	Fail	Total
	Fail		20	35
Total		24		60

i) copy and complete the table

ii) if a candidate is chosen at random from the sample, use the table to find the probability that the candidate:

a) passed part II

b) passed parts I and II

c) passed part II but failed part I.

iii) if a candidate is chosen at random from the subgroup of those who failed Part I, find the probability that the candidate passed Part II.

4. Given that:

(i) $P(A \cap B) = P(A)P(B)$

(ii) $P(A/B) = P(A)$ if $P(B) > 0$

(iii) $P(B/A) = P(B)$ if $P(A) > 0$

Show that (i) implies (ii), (ii) implies (iii) and (iii) implies (i)

5. Consider the experiment of tossing 2coins. Let the sample space S = {(H,H), (H,T), (T,H), (T,H)} and assume that each point is equally likely. Find:

i) the probability of two heads given a head on the first coin

ii) the probability of two heads given at least one head.

6. Given that two dice are tossed. What is the probability that their sum will be 6 given that one face shows 2?

7. A certain brand of compact disc (CD) player has an unreliable integrated circuit $[IC]$, which fails to function on 1% of the models as soon as the player is connected. On 20% of these occasions, the light displays fail and the buttons fail to respond, so that it appears exactly the same as if the power connection is faulty. No other component failure causes that symptom. However, 2% of people who buy the CD player fail to fit the plug correctly, in such a way that they also experience a complete loss of power. A customer rings the supplier of the CD players saying that the light displays and buttons are not functioning on the CD. What is the probability that the fault is due to the IC failing as opposed to the poorly fitted plug?

8. An electronic has 3 components and the failure of any one of them may or may not cause the device to shut off automatically. Furthermore, these failures are the only possible causes for a shut-off and the probability that two of the components will fail simultaneously is negligible. At any time, component B_1 will fail with probability 0.1, component B_2 will fail with probability 0.3 and

component B_3 will fail with probability 0.6. Also, if component B_1 fails, the device will shut off with probability 0.2; if component B_2 fails, the device will shut off with probability 0.5, if component B_3 fails, the device will shut off with probability 0.1. The device suddenly shuts off, what is the probability that the shut off was caused by the failure of component B_1.

9. Stores X, Y, Z sell brands A, B and C of men's shirts. A customer buys 50% of his shirts at X, 20% at Y and 30% at Z. Store X sells 25% brand A, 40% brand B and 25% brand C. Store Y sells 40% brand A, and 20% brand B and 30% brand C. Store Z sells 20%

CHAPTER4

FUNDAMENTALS OF PROBABILITY FUNCTIONS

4.1 Introduction

A random variable X is a real valued function that assigns values to every elementary outcomes of an experiment. Let E be an experiment, with elementary outcomes $e_1, e_2, e_3, e_4, \dots \dots$ in the sample space S, then$S = \{e_1, e_2, e_3, e_4, \dots \dots\}$.

A .random variable X can take values 1,2,3,4, ...for finite or countable infinite elementary event.

An event may consist of one or more elementary events, for example:

$A = \{e_1, e_3, e_{k+1}: e_i \epsilon S\}$

$B = \{\phi\}$ a null set

$C = \{e_1\}$ a singleton

$D = \{e_1, e_3\}$ a doubleton

Independent events: Two events A and B are independent if the occurrence of A has no influence on the occurrence of B and vise versa.

i.e $P(A \cap B) = P(A).P(B)$

Independent Random Variables

The random variable X and Y are said to be independent if for any two set of real numbers if for all A and B.

$$P\{X \leq a, y \leq b\} = P(X \leq a,)P(Y \leq b)$$

$$P(A \cap B = \quad P(A).\ P(B)$$

4.2 Probability Density Function (pdf)

Suppose X is a random variable and ∃ a function $f_{(x)}$ such that

(i) $f_{(x)} \geq 0$

(ii) $f_{(x)}$ has at most a finite number of discontinuity in every finite interval on the real line

(iii) $\int_{-\infty}^{\infty} f_{(x)}\, dx = 1$

(iv) For every interval $[a, b]$

$$P\{a \leq X \leq b\} = \int_a^b f_{(x)}\, dx$$

Then X is said to be a continuous random variable with pdf$f_{(x)}$

However, f (i) and (ii) above holds and

(iii) $\sum_{i=\infty}^{\infty} f_{(xi)} = 1, \quad and$

(iv) for all $i, i = 1, a+1, \dots, b \quad s.t.$

$$P(a \leq X \leq b) = \sum_{i=1}^{b} f_{(xi)}$$

Then X is said to be discrete random variable with probability mass function (pmf) $f_{(xi)}$

Note:

$$\frac{d}{dx}\int_a^x f_{(x)}\, dx = f_{(x)}; \quad F_{(x)} = \int_a^x f_{(x)}\, dx$$

Where $f_{(x)}$ is the pdfof the random variable X and $F_{(x)}$ is the distribution function, then

$$F_{(t)}^{-1} = f_{(t)} \text{and}$$

$$\frac{d}{dx}\left[\int_a^x f_{(t)} dt\right] = \frac{d}{dx}\left(\int_a^x F_{(t)}^{1} dt\right)$$

$$\frac{d}{dx}\left[F_{(t)}\right]_a^x = \frac{d}{dx}\left[F_{(x)} - F_{(a)}\right] = f_{(t)}$$

Consider a continuous random variable X defined on an interval [0, a]. Let x be a point on [0, a] i.e. a value of x.

$$\therefore P_{(>a)} = Pr\{x_0 < X \leq x_0 + x_a\}$$

It follows that

$$P_{(2x_a)} = Pr\{x_0 < X < x_0 + 2x_a\}$$

$$= Pr\{x_0 < X \leq x_0 + x\} + Pr\{x_0 + x < X \leq x_0 + 2x_a\}$$

$$= P_{(x)} + P_{(x)}$$

$$= 2P_{(x)}$$

It follows that

$$P_{(nx)} = nP_{(x)}$$

If $(0 < x < a)$ and we consider $P_{(x)}$ to be contiunuious at $x = 0$, then it is

$$\lim_{x \to 0} P_{(x)} = P_{(0)} = 0$$

It follows from the above that

$$\Pr(x = x_0) = 0 \ for any x_0.$$

Thus for a continuous random variables we define a probability density function (pdf) $f_{(x)}$ such that

$$Pr\{a < X \leq b\} = \int_a^b f_{(x)} dx$$

For all real values a and b

Equation (3) can be rewritten as

$$Pr\{a < X < a + h\} = hf_{(x)} + 0(h)$$

Or

$$Pr\{a < X < x + dx\} = f_{(x)} dx$$

From the above, we can deduce the following

$(i) f_{(x)} \leq 0$

$(ii) Pr\{a < X \leq b\} = \int_a^b d_{(x)} dx$

$(iii) \int_{-\prime}^{\prime} f_{(x)} dx = 1 \ = Pr\{-\infty < X \leq -\infty\}$

$(iv) \ 0 \leq f_{(x)} \leq 1$

In term of the joint distribution function, the distribution of X and Y is

$$F_{(a,b)} = F_x(a) F_y(b) \quad \forall a, b.$$

Example: Suppose that $n + m$ independent trials have a common probability of success P. If X is the number of success in the first n trials and Y, the number of success in the final n trials. Show that X and Y are independent.

Solution

$$P(X = x, \ Y = y) = \binom{n}{x} p^x q^{n-x} \binom{m}{y} P^y (1 - P)^{m-y} \quad 0 \leq X \leq n$$

$$= P(X = x) \ P(Y = y) \qquad\qquad 0 \leq y \leq m$$

X and Y will be dependent if Z is the number of successes in the $n + m$ trials i.e. $Z = X + Y$

Example

If X and Y are independent binomial random variable with respective parameters (n, p) and (m, p). Calculate the distribution of X + Y

Solution

Let

$$P(X + Y = K) = \sum_{i=0}^{n} P(X = i, Y = K - i)$$

$$= \sum_{i=0}^{n} P(X = i) \ P(Y = K - i)$$

$$= \sum_{i=0}^{n} \binom{n}{i} p^i q^{n-i} \binom{m}{k = i} p^{k-i} q^{m-k+i}$$

$$= p^k q^{m+n-k} \sum_{i=1}^{n} \binom{n}{i} \binom{m}{k - i}$$

$$= \binom{n + m}{k} p^k q^{m+n-k}$$

where

$$\binom{n + m}{k} = \sum_{i=0}^{n} \binom{n}{i} \binom{m}{k - i}$$

and $\binom{r}{j} = 0 \ when \ j > r$

4.3 Distribution Function

Distribution function forms the foundation of the theory of probability and statistics.

If the value of X observed in n–experiment is less than or equal to x k-times, then

$$F_x(x) = P(x \le x) = \frac{k}{n}$$

If X is discrete and m is the number of times X is observed in n trial, then

$$f_{(x)} = P(X = x) = \frac{m}{n}$$

$\therefore f_x(x)$ is the (cumulative) distribution function

$f_{(x)}$ is the probability density function.

Let X be a real random variable on the probability space (Ω, A, P). For $x \in \Re$, we define

(i) $P(X \le x) = F_x(x)$

(ii) $P_x(a,b) = P(a < x \le b)$

$$= f_X(b) - f_x(a) \cdot (b > a)$$

Example:

Let X have the distribution function

$$F_x(x) = \begin{cases} 0; & \text{if} \quad x < -1 \\ 1-\rho; & \text{if} \quad -1 \le x < 0 \\ 1-\rho+\frac{1}{2}\rho; & \text{if} \quad 0 \le x \le 2 \\ 1 \cdot & \text{if} \quad x > 2 \end{cases}$$

Find

(i) $\rho(X = -1)$

(ii) $\rho(X = 0)$

(iii) $\rho(X \ge 1)$

Solution

(i) $\rho(X = -1) = 1 - \rho$; $\exists$ *a jump discontinuity at* $x = 1$

(ii) $\rho(X = 0) = 0$; *F is contains at* $x = 0$

(iii) $\rho(X \ge 1)$ $F_{(0+)} - F_{(0-)} = 0$

$$= F_{(1)} - F_{(0)}$$

$$= 1 - \rho + \frac{1}{2}\rho - (1-\rho) = \frac{1}{2}\rho$$

Theorem

The distribution function $F_x(x)$ of a random variable is non-degreasing, continuous on the right with $F_x(-\infty) = 0$ and $F_x(-\infty) = 1$. Conversely every function F, with the above properties is the different from a random variable on some probability space.

Proof: For $x < x'$

$$[X \le x'] = [X \le x] + [x < X \le x']$$

$$\therefore \rho[X \le x'] = \rho[X \le x] + \rho[x < X \le x']$$

$$Since\ \rho[x < X \le x'] \ge 0$$

$$F_x(x') - F_x(x) \ge 0$$

This implies that $F_x(x)$ is monotone non-decreasing in x

Consider $\{x'_n\}$; $x'_n \downarrow x$

Since $[x < X \le d_n] \to \phi$ *as* $x'_n \downarrow x$

$$F_x(x'_n) - F_x(x) \to 0 \ \ as\ x'_n \downarrow x$$

Since this is true for every sequence $\{x'_n\}$ then $F_x(x)$ is continuous from the right.

For a continuous random variable X. the c.d.f is defined as

$$F_{(x)} = Pr(X \le x) = \int_{-\infty}^{x} f_{(t)}dt$$

If X assumes a value between and b

$$Pr\{a < X \le\} = P(X \le b) - P(X \le a)$$

$$= F_{(b)} - F_{(a)}$$

$$= \int_a^b f_{(x)} dx$$

From (5) we obtain

$$f_{(x)} = \frac{dF_{(x)}}{dx}$$

From (4) we can also define

$$Pr\{X > x\} = Pr\{x < X < +\infty\}$$
$$= Pr\{-\infty < X < +\infty\} - Pr\{-\infty < X \leq x\}$$
$$= \int_{-\infty}^{\infty} f_{(x)} dx - \int_{-\infty}^{\infty} f_{(x)} dx$$
$$1 - F_{(x)}$$

This is often referred to as the survivor function

$$S_{(x)} = 1 - F_{(x)}$$

Hazard function is a related quantity defined by $H_{(x)} = \frac{f_{(x)}}{1-F_{(x)}} = \frac{f_{(x)}}{S_{(x)}}$

For a discrete random variable X, the equivalence of pdf is probability mass function (pmf) defined as

$$P_{(x)} = P(X = x_i)$$

$$F_{(t)} = \sum_{t=1}^{t} P(X = x_t) = \sum_{t=1}^{t} Pr(X = x_t)$$

Example:

Let X be the number of success in single trial of an experiment with constant probability P. When the trial is repeated n times then

$$P(X = x) = \binom{n}{x} p^x q^{n-x} \quad where\ q = 1 - P$$

$$\sum_{i=1}^{n} P(X = x_i) = \sum_{i=1}^{t} \binom{n}{x} p^{x_i} q^{n-x_i}$$
$$=(P + q)^n = 1$$

Recall

$$E(X) = np, \quad Var\ (X) = npq = \sigma^2$$

$$F_{(t)} = \sum_{i=0}^{t} P(X = x_i)$$

Example:

Let the probability space by $(\Re, \beta, \rho)$ and X be the identity mapping of $\Re$ *to* R, where ρ is the normal probability distribution. Then

$$F_x(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t}{2}} dt$$

It can be noted that F_x increases continuously. Since $F_x(\cdot)$ represented the cumulative probability at an event, its maximum value is unity and non-negative.

$$i.e.\ F(-\infty) = 0 \quad = \lim_{x \to -\infty} F_{(x)}$$
$$F(+\infty) = 1 \quad = \lim_{x \to \infty} F_{(x)}$$

$\lim F_{(x)} \Rightarrow$ lim*it from the left*

$\lim F_{(x)} \Rightarrow$ lim*it from the left*

4.3.1 Distribution Function for Discrete Random Variables

Let us define the distribution function for the discrete random variable as

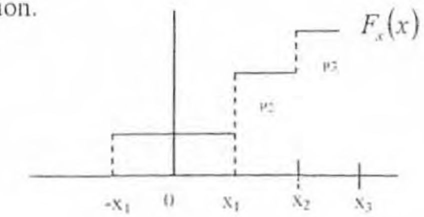
$$F_x^*(x) = P(X < x),\ then,$$
$$P(x \leq X < x^1) = P(X < x^1) - P(X < x)$$

which tends to zero as $x \uparrow x^1$ and F^* is $\therefore$ continuous from the left.

$F(x^+) \equiv F_x(x + 0)$ is the limit from the right

$F(x^-) \equiv F_x(x - 0)$ is the limit from the left

It is known that $F_x(x)$ for discrete random variable increases by jumps, and is called the step-function.



Example:

Consider a random variable X with distribution function given by

$$F_X(x) = \begin{cases} 0; & x < 0 \\ \frac{1}{4}; & 0 \le x < 1 \\ \frac{1}{4}; & 1 \le x < 2 \\ \frac{1}{3}; & 2 \le x < 3 \\ 1; & x \ge 3 \end{cases}$$

(i) Sketch the distribution function and hence or otherwise

(ii) *Calculate* $\Pr\{X = \frac{1}{2}\}$

(iii) *Calculate* $\Pr\{X = 1\}$

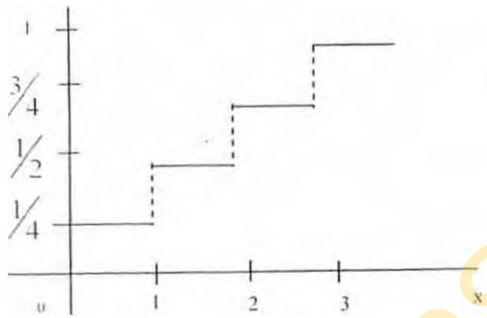
(iv) *Calculate* $\Pr\{X = 2\}$

(v) Calculate the conditional probability that X is greater than 2, given that X is greater than

(vi) $\Pr\{2 < X < 3\}$; (vii) $\Pr\{1 \le X < 2\}$

(viii) $\Pr\{0 \le X < 1\}$ (ix) $\Pr\{X \ge 2\}$; (x) $\Pr\{X \ge 3\}$

Solution



(ii) $\Pr\left(X = \frac{1}{2}\right) = \Pr\left(X^+ = \frac{1}{2}\right) - \Pr\left(X < \frac{1}{2}\right) = \frac{1}{4} - \frac{1}{4} = 0$

(iii) $\Pr(X = 1) = \Pr(X^+ - 1) - \Pr(X < 1) = \frac{1}{4} \cdot \frac{1}{4} = 0$

(iv) $\Pr(X = 2) = \Pr(X' \le 2) - P(X < 2) = \left(\frac{1}{2} + \frac{1}{3}\right) - \left(\frac{1}{4} + \frac{1}{4}\right)$

$= \frac{5}{6} - \frac{1}{2} =$

(v) $P(X > 2)/(X > 1) = \frac{P(X > 2)}{P(X > 1)} = \frac{1 - F(2)}{1 - F(1)} = \frac{\frac{1}{2}}{\frac{3}{4}} = \frac{4}{9}$

(vi) $F(3) - F(2) = 1 - \frac{2}{3} = \frac{1}{3}$

(vi) $F(2^-) - F(1^+) = \frac{1}{4} - \frac{1}{4} = 0$

(vii) $F(1^-) - F(0^+) = \frac{1}{4} - 0 = \frac{1}{4}$

(ix) $1 - F(2^-) = 1 - \frac{1}{4} = \frac{3}{4}$

(x) $1 - F(3^-) = 1 - 1 = 0.$

4.4 Jointly Distributed random variables

If the occurrence of event X that affects event Y we require the concept of conditional probability.

The conditional probability distribution function of X given Y for discrete random variable is given by: $P(X/Y) = \frac{P(XY)}{P(Y)}$; $P_{(y)} > 0$

$P(Y/X) = \frac{P(XY)}{P(X)}$; $P_{(x)} > 0$

While for continuous random variable: $f(X/Y) = \int \frac{f(x,y)dx\,dy}{f_Y(Y)}$

$f(Y/X) = \int \frac{f(x,y)\,dx\,dy}{f_X(X)}$

Definition: Let (Ω, ϵ, P)be a probability spaceandlet B be an event with $P(A) > 0$.Then the conditional probability of B given A is defined by $P(B/A) = \frac{P(AB)}{P(A)};\ P(A) > 0$

But $P(AB) = P(B/A)\,P(A) = P(A/B)\,P(B)$

Recall the Baye's theorem

$$P(B_k/A) = \frac{P(A/B_k)P(B_k)}{\sum_{i=1}^{n} P(A/B_i P(B_i)}$$

Two random variable. X and Y are jointly and continuously distributed if there exist a function $f_{(x,y)}$defined for all real x and y and a two dimensional plane C such that:

$$P\{(x,y)\epsilon\, C\} = \int x, y \int \epsilon\, C\, f_{(x,y)} dx\, dy$$

$$\{P\{X = x_i, y = y_i\} = \rho_{ij} \geq 0$$

$$\Sigma\Sigma\rho_{ij} = 1$$

The function $f_{(x,y)}$is called the joint of X and Y. Satisfying the following conditions

$$(i) f_{(x,y)} \geq= 1, \quad \vee\, x, y\, \epsilon\, C$$

$$(ii) \sum_{x}\sum_{y} f_{(x,y)} = 1, \qquad for\ X, Y\ discrete$$

$$\int x \int y\ f_{(x,y)} = 1, \quad for\ X, Y\ continues$$

For discrete random variable.

The joint distribution function of X and Y is given by:

$$F_{(X,Y)} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{(X,Y)} dx\ dy$$

$$= \sum_{i=1}^{x}\sum_{j=0}^{y} f_{(x,y)} = \sum_{1}^{x}\sum_{j}^{y} \rho_{ij}$$

and the marginal distribution d.f for X. is defined as

$$f_x(x) = \int_{-\infty}^{\infty} f_{(x,y)} dy \qquad X, Y, continues$$

$$P_x(x) = \sum_{j} P\{X = X_i\} = \sum \rho_{ij} \quad X, Y\ discrete$$

the m.d.f. for random variable Y is

$$f_y(y) = \int_{-\infty}^{\infty} f_{(x,y)} dx \qquad X, Y, continues$$

$$P_y(y) = \Sigma_j \rho\{Y = y_i\}\, X, Y\ discrete$$

Example: The joint d.f. at X and Y is given by $f_{(x,y)} = \begin{cases} 2e^{-x}e^{-2y} \\ 0 \quad 0 < x < \infty \end{cases}$

Compute (i) $P\{X < 1, y < {}^1/_2\}$

$(ii)\ P\{x < y\}$

$(iii) P(X < a)$

Example 2:

Given $f_{(x,y)} = \begin{cases} 2(x + y - 3xy^2) \\ 0\ elsewhere \quad 0 < y < 1 \end{cases};\ 0 < x < 1$

Find $(i)\ Pr\{0 < X < {}^3/_4\} (iv)\ P(X/Y < {}^1/_2)$

$(ii)\ P\{{}^1/_{10} < y < {}^3/_4\} (v)\ Pr(X < {}^3/_4 / Y < {}^1/_2)$

(iii) $P(x, y)$

4.5.1 Conditional Distribution of Jointly Distributed Random Variables

$$P(X/Y) = \frac{P(XY)}{P(y)}; \quad P(y) > 0$$

$$f(x/y) = \frac{\int f_{(x,y)}\ dx\ dy}{\int f_{(y)} dy}$$

$$f(y/x) = \frac{\int f_{(x,y)}\ dx\ dy}{\int f_{(x)} dx}$$

Exercise:

If X and Y are independent Poission random variable with respective parameters λ_1 and λ_2. Compute the distribution of X + Y.

Solution

Let $P(x+y=n) \equiv \Pr(X=k, Y=n-k)$

for $0 < k < n$

and disjoint events

$(X=k, Y=n-k)$

Exercise

Given the following probability distribution function

$X/_Y$	1	2	3	4
1	$^1/_{24}$	$^1/_{16}$	$^1/_{48}$	$^1/_8$
2	$^1/_{12}$	$^1/_8$	$^1/_{24}$	$^1/_4$
3	$^5/_{24}$	$^5/_{16}$	$^5/_{48}$	$^5/_8$
	$^1/_3$	$^1/_2$	$^1/_6$	1

Find $\rho_{(XY)}$

4.6 Independence of Functions of Random Variables

Two random variable's X and Y are said to be stochastically independent iff:

$$f_{(x,y)} = f_{1(x)} f_{2(x)}; \quad -\infty < x < \infty$$
$$-\infty < y < \infty$$

where $f_{(x,y)}$ is the joint density function of X and Y and $f_{1(x)}$ and $f_{2(y)}$ are the marginal pdf of X and Y respectively.

Theorem: Two random variables are stochastically independent if and only if the joint p.d.f can be written as a product of a negative function of x alone and a with negative function of y alone.

Where $f(x)$ is the pdf and random variable X and $f(x)$ is the distribution function.

Proof:

Let $F^{-1}(t) - P_{(t)}$ *then*

$$\frac{d}{dx}\left[\int_a^x f(t)\right]_a^x \quad \frac{d}{dx}\left(\int_a^x F_{(x)} - F_{(a)}\right) = f(x)$$

Two random variable's X and Y are said to be stochastically independent different:

$$f(x,y) = f_1(x) f_2(y); \quad -\infty < x < \infty$$
$$-\infty < y < \infty$$

Where $f(x,y)$ is the joint density function of X and Y and $f_1(x)$ and $f_2(y)$ are the marginal p.d.f of X and Y respectively.

Theorem: Two random variables are stochastically independent if and only if the joint p.d.f can be written as product of a non-negative function of x alone and a non-negative function of y alone.

Proof:

If $f_{(x,y)} = g(x)\,h(y)$ where $g(x)$ *and* $h(y)$ are non-negative function of x and y alone respectively, then the marginal pdf at X is given by

$$f_1(x) = \int_{-\infty}^{\infty} f_{(x,y)}\,dy$$
$$= \int_{-\infty}^{\infty} g(x)h(y)\,dy, \text{ where } g(x) \text{ is a function of } x \text{ alone}$$
$$\therefore f_1(x) = g(x)\int_{-\infty}^{\infty} h(y)\,dy$$
$$= c_1\, g(x)$$

Similarly, the marginal p.d.f of Y is given by

$$f_2(y) = \int_{-\infty}^{\infty} f_{(x,y)}\,dx$$
$$= \int_{-\infty}^{\infty} g(x)h(y)\,dx, \text{ where } h(y) \text{ is a function of } y \text{ alone}$$
$$\therefore f_2(y) = h(y)\int_{-\infty}^{\infty} g(y)\,dx$$
$$= c_2\, h(y)$$

But since $f_{(x,y)}$ is the joint pdf of X and

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{(x,y)}dy\,dx = 1, \text{ by hypothesis}$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{(x,y)}dx\,dy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)\,h(y)\,dx\,dy$$

Applying Fubini's theorem to the finite integer we have:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{(x,y)}\,dx\,dy = \left(\int_{-\infty}^{\infty} g(x)dx\right)\left(\int_{-\infty}^{\infty} h(y)dy\right)$$

Letting $C_1 * C_2 = 1$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{(x,y)} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} C_1C_2\,g(x)h(y) = 1$$

$$\therefore f_{(x,y)} = C_1g(x)\cdot C_2h(y)$$

$$= f_1(x)\cdot f_2(y)$$

which implies that X and Y are stochastically independent.

Fubini's Theorem: (1) A necessary and sufficient condition that a measurable subset A of $\Omega_1\ X\,\Omega_2$ has measure zero is that almost every $w_1 - section$ (or almost even $w_2 - section$ has $\mu_1 - measure$ (or $\mu_2 - measure$), zero.

$$If\ A = A_1\ X\ A_2,\ \lambda(A) = \int \mu_2(Aw_1)d\mu_1(w_1)$$

$$= \int \mu_1(Aw_1)d\mu_2(w_2)$$

$$= \mu(A_1)\,\mu(A_2)$$

Fubini's theorem gives condition under which it is possible to compute double integral using iterated integrals. It allows the order of integration to be changed in iterated integrals.

Theorem

Suppose A and B are complete measure spaces. Suppose $f_{(x,y)}$ *is* $A\,X\,B$ measurable if

$$\int_{AXB} |f(x,y)|\,d(x,y) < \infty$$

Then $\int_{AXB} f(x,y)d(x,y) = \int_A\left(\int_B f(x,y)dy\right)dx = \int_B\left(\int_A f(x,y)dx\right)dy$

The last two integrals being iterated integrals w.r.t. two measures respectively and the first then integral w.r.t. product of two measure.

OR *If* $f(x,y) = g(x)\,h(y)$ *for some function* g *and* h

then $\int_A g(x)dx \int_B h(y)dy = \int_{AXB} f(x,y)d(x,y)$

Where λ is a unique-infinite measure $A_1\ X\ A_2$

Fubini Theorem (2): If h is a non-negative function on $\Omega_1\ X\ \Omega_2$ then

$$\int h\,d\lambda = \int\int h\,d\mu_1\,d\mu_2$$

$$= \int\int h\,d\mu_2\,d\mu_1$$

The above reduces to Theorem (2) above in the case of indicator function of rectangles.

Lemma

Let X and Y be stochastically independent random variable the pdf of $Z = X + Y$ is given by $g(Z) = \int_{-\infty}^{\infty} f(x)h(y) = \int_{-\infty}^{\infty} f(x)h(Z - x)\,dx$

Where $f(x)$ is the p.d.f of X and

$h(y)$ is the p.d.f of Y.

Proof:

Let X have the p.d.f $f(x)$ and y has p.d.f $h(y)$.

pdf of $Z = P(\{Z \le Z\}) = P(X + Y \le Z)$ the joint pdf of X and Y is $f(x)\,h(y)$ since X and Y are stochastically independent.

$$\therefore G(Z) = \int_{-\infty}^{\infty}\int_{-\infty}^{Z-x} f(x)\,h(y)\,dx\,dy$$

$$= \int_{-\infty}^{\infty} f(x)\left[\int_{-\infty}^{Z-x} h(y)\,dy\right]dx$$

Since $\le G(Z) \le 1$, by Fubini's theorem

$$G(Z)=\int_{-\infty}^{\infty} f(x)H(Z-x)\,dx$$

$$\therefore g(x)=\frac{dG(Z)}{dZ}$$

$$=\int_{-\infty}^{\infty} f(x)\cdot\left[\frac{dG(Z)}{dZ}\cdot H(Z-x)\right]dx$$

By continuing of distribution function

$$g(Z)=\int_{-\infty}^{x} f(x)\,h(Z-x)\,dx$$

Example 1:

Let X and Y be stochastically independent random variables. each having the poission distribution with parameter λ. Find the distribution of $Z=X+Y$?

Solution

$$f(x)=\frac{e^{-\lambda}\lambda^{x}}{x!};\qquad h(y)=\frac{e^{-\lambda}\lambda^{y}}{y!}$$

$$g(Z)=\sum_{x=0}^{\infty} f(x)\,h(y)$$

$$=\sum_{x=0}^{\infty} f(x)\,h(Z-x);\qquad Since\ y=Z-x$$

$$=\sum_{x=0}^{\infty}\frac{e^{-2\lambda}\lambda^{x}\lambda^{Z-x}}{x!(Z-x)!}$$

Applying Binominal expansion, we have

$$g(Z)=\frac{e^{-2\lambda}}{Z!}\sum_{x=0}^{\infty}\frac{Z!\lambda^{x}\lambda^{Z-x}}{x!(Z-x)!}$$

$$=\frac{e^{-2\lambda}}{Z!}(\lambda+\lambda)^{Z}$$

$$=\frac{e^{-2\lambda}}{Z!}(2\lambda)^{Z}\qquad \sim P(2\lambda)$$

Exercise

Find $g(Z)$ if X and Y are independent with parameters λ_1 *and* λ_2 respectively show that the random variable $Y_1=X_1+X_2$ and $Y_2=\dfrac{X_1}{X_1+X_2}$ are stochastically independent from an exponential distribution.

Solution: Since X_1 and X_2 are independent, the joint p.d.f of x_1 *and* x_2, $\mu(x_1,x_2)$ is given

$$\mu(x_1,x_2)=\begin{cases} f(x_1)\,f(x_2) & 0<x_1,x_2<0 \\ 0 & elsewhere\end{cases}$$

$$=\begin{cases} e^{-(x_1+x_2)} & 0<x_1<\infty, 0<x_2<\infty \\ 0 & elsewhere\end{cases}$$

Nw $\quad y_1=\mu(x_1x_2)=x_1+x_2$

$$y_2=\mu(x_1x_2)=\frac{x_1}{x_1+x_2}$$

which defines a mapping $(1-1)$ transformation from the space

$$A=\{(x_1+x_2);0<x_1<\infty,\ 0<x_2<\infty\}$$

unto the space

$$B=\{(y_1+y_2);0<y_1<\infty,\ 0<y_2<1\}$$

The inverse transformations are given by

$$x_1=y_1\,y_2;\qquad x_2=y_1\,y_2=y_1(1-y_2)$$

$$\frac{\partial x_1}{\partial y_1}=y_2;\qquad \frac{\partial x_1}{\partial y_2}=y_1$$

$$\frac{\partial x_2}{\partial y_1}=1-y_1;\qquad \frac{\partial x_2}{\partial y_2}=-y_1$$

The Jacobian of the transformation

$$J = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix} = \begin{vmatrix} y_2 & y_1 \\ 1-y_2 & -y_1 \end{vmatrix}$$

$$= -y_2 y_1 - (y_1 - y_1 y_2)$$

$$= |-y_1| = y_1$$

$\therefore J \neq 0$, since y_1 is not dentically zero

$$\therefore g(y_1\ y_2) = \begin{cases} \theta(y_1\ y_2)|J| & o < y_1 < \infty, 0 < y_2 < 1 \\ 0 & elsewhere \end{cases}$$

$$g(y_1\ y_2) = \begin{cases} e^{-y_1} \cdot y_2 & o < y_1 < \infty, 0 < y_2 < 1 \\ 0 & elsewhere \end{cases}$$

$$g(y_1\ y_2) = \begin{cases} y_1 e^{-y_1} & o < y_1 < \infty, 0 < y_2 < 1 \\ 0 & elsewhere \end{cases}$$

Exercise:

Find the marginal pdf of y_1 and marginal p.d.f of y_2, hence or otherwise show that y_1 and y_2 are stochastically independent.

$$a(y_1) = y_1 e^{-y_1} \quad h(y_2) = 1, \qquad 0 < y =< 1$$

It should noted that

Y_1 has the gamma pdf with parameter $\alpha = 2, \beta = 1$

Y_2 has the uniform distribution over $(0,1)$

X_1 *and* X_2 have exponential distribution with parameter 1.

Definition (for more than two variable)

Let $X_1, X_2, ..., X_n$ n mutually stochastically independent random variable s. each of which has the same p.d.f $f(x)$ which may or may not be known. Then

$$f_{(x_1, x_2, ..., x_n)} = f_{(x_1)} f_{(x_2)} ... f_{(x_n)}$$

By stochastic independence, since the marginal pdfs

$$f_i(x_i) = f(x_i), i = 1, 2, ..., n$$

The random variables are said to be a random sample of size n from a distribution which has pdf $f(x)$.

Exercise

Let X_1 *and* X_2 be two stochastically independent random variables with p.d.f

$$\frac{1}{\Gamma(\alpha)} x_1^{\alpha-1} e^{-x_1}, a < x_1 < \infty$$

$$\frac{1}{\Gamma(\beta)} x_2^{\beta-1} e^{-x_2}, a < x_2 < \infty \qquad respectively.$$

Where $\Gamma(\cdot)$ is the gamma function.

Define $Y = \frac{X_1}{X_1 - X_2}$. By defining a suitable Y_2, a function of X_1 *and* X_2.

Calculate:

(a) the joint pdf of Y_1 *and* Y_2 and

hence (b) the marginal pdf of Y_1.

4.7 Functions of Random Variables

Suppose X is a characteristic of interest, the p.d.f $f_X(X)$ may refer to the pdf of a given population. Another characteristic Y (which may be a function of X may be of interest Therefore, there is need to obtain the distribution of the later variable.

Thus, given the pdf or c.d.f. of therandom variableX, the pdf or c.d.f. of anotherrandom variable Y may be obtained as a function of X.

There are two given major technique to achieve this. They are CDF technique and the Transformation technique.

4.7.1 The CDF Technique

Given the CDFofX $(F_X(X)$with some function of interest (say) $Y = g(x)$ is of interest.

The idea is to express the CDF of Y in terms of the distribution of X. Define set $A_y = \{X/g(x) \leq y)$ It follows that $[\{Y \leq y\}]$ $and\ X \in A_y]$

i.e. $F_y(y) = Pr(g(x) < y)$

In the continuous case

$$F_y(y) = \int_{x_1}^{x_2} f_X(x)\, dx$$

$$= F_x(x_2) - f_x(x_1)$$

and p.d.f of $f_y(y) = {}^{d}/_{dy} F_y(y)$

Example 1:

Given $F_x(x) = 1 - e^{-2x}, \quad 0 < x < \infty$

Find the pdf of $Y = e^x$

Solution

$$F_y(y) = P(Y \leq y)$$

$$= P[e^x \leq y]$$

$$= P[X \leq In\, y]$$

$$= F_x(In\, y)$$

but $f_x(x) = \frac{dy}{dx} f_x(x)$

$$= \frac{d}{dx}(1 - e^{-2x})$$

$$= 2e^{-2x}$$

and $F_y(y) = \int_0^{Iny} 2e^{-2x} dx$

$$F_y(y) = e^{-2x}\Big|_0^{Iny}$$

$$= e^{Iny^{-2}} + 1$$

$$= 1 - y^{-2}, \quad 1 < y < \infty$$

$$f_y(y) = \frac{dy}{dx} F_y(y)$$

$$= {}^{d}$$

Example 2:

Let $f_x(x) = 2x \qquad 0 < x < 1$

and $y = 3x + 1$

find the distribution of $g(y)$

Solution

$$y = 3x + y \Rightarrow X = \frac{y-1}{3}$$

$$F_y(y) = P(Y \leq y)$$

$$= P[3x + 1 \leq y]$$

$$= P\left[X \leq \frac{y-1}{3}\right]$$

$$= \int_0^{(y-1)/3} 2x\ dx$$

$$= x^2\Big|_0^{(y-1)/3}$$

$$= \left(\frac{y-1}{2}\right) - 0$$

$$F_y(y) = \frac{y-1}{9}$$

$$f_{(y)} = \frac{d}{dx}\left(\frac{y-1}{9}\right)$$

$$= \frac{2(y-1)}{9}; \qquad 1 < y < 4$$

4.7.2 Transformation Method

Let X be a continuous random variable with pdf $f_{(x)} > 0$ for $a < X < b\ and\ y = g(x)$. If there is a one-to-one transformation from $A = \{x/f_y(x) > 0\}\ on\ to\ B = \{Y/f_y(y) > 0\}$ with inverse transformation.

$X = w(y)$ if the derivative ${}^{d}/_{dy}$ w(y) exist, then

$$f_y(y) = f_x w(y) \left|\frac{d}{dy}\right| \qquad y \in B$$

Where $\left|\frac{d}{dy}\right|$ is the Jacobian of the transformation.

Y could be monotone increasing or decreasing $f_x(y) = H_x(y)\left|\frac{d}{dy}\right|$

Example 3: Using the last example

$$f_{(x)} = 2x \quad 0| < X < 1, \quad and\ y = 3x$$

$$X = \frac{y-1}{3}$$

$$\left|\frac{dx}{dy}\right| = \left|\frac{1}{3}\right|$$

$$g(y) = 2\left(\frac{y-1}{3}\right) * \frac{1}{3}$$

$$= \frac{2}{9}(y-1), \qquad 1 < y < 4$$

Example 3:

Given $f_{(x)} = \begin{cases} 2xe^{-x^2} & 0 < x < \infty \\ 0 & elsewhere \end{cases}$

determine the pdf of $y = X^2$

Solution

$$f_{(x)} = 2xe^{-x^2}$$

$$y = X^2 \Rightarrow X = y^{1/2}$$

$$\frac{dy}{dx} = \frac{1}{2}y^{-1/2}$$

$$g(y) = f_x(y)\left|\frac{d}{dy}\right|$$

$$= 2\,y^{1/2}e^{-y} * {}^{1}/_{2}\,y^{-1/2}$$

$$= e^{-y} \qquad 0 < y < \infty$$

4.7.3 Transformation that are not one-to-one

If $g(x)$ is not one-to-one over $A = \{x/f_x(x) > 0\}$; then thee is no unique solution to equation $y = g(x)$. It is usually possible to partition A into disjoint subsets $A_1, A_2, A_3 \ldots$ such that $\mu(x)$ is one-to-one over each A_j

$$f_y(y) = \sum_j f_x(w_j(y))$$

i.e. $f_y(y) = \sum_j f_x(x_j)$ where the sum is over x_j such that $\mu(x_j) = y$

Example

Let $f_{(x)} = \frac{4}{31}({}^{1}/_{2}),\ x = -2, -1, 0, 1, 2,$

Find the distribution of $Y = |X|$

Solution

$$f_y(0) = \frac{4}{31} \qquad i.e. \frac{4}{31}({}^{1}/_{2}),$$

$$f_y(1) = f_x(-1) + f_x(1) = \frac{8}{31} + \frac{2}{31} =$$

$$f_y(2) = f_x(-2) + f_x(2) = \frac{17}{31} \qquad \text{or}$$

$$f_{(y)} = \begin{cases} \frac{4}{31} & y = 0 \\ \frac{4}{31}\left[\left(\frac{1}{2}\right)^{-y} + \left(\frac{1}{2}\right)^{y}\right], & y = 1,\ 2 \end{cases}$$

Exercise

1. Let X have a Poisson distribution with p.d.f $f_{(x)} = \frac{e^{-\lambda}\lambda^x}{x!}$
   $x = 0, 1, 2, \ldots$
   Let $Y = 4X$, derive the pdf of Y.
2. A random variable X has pdf
   $f_{(x)} = 1 \qquad 0 < X < 1$
   Find the pdf of $Y = -2ln\,X$
3. If the random variable $X \sim N(0, 1)$, find the pdf of $Y = X^2$
4. Use the transformation method to solve the problem in example 1.

$$f_{(x)} = \begin{cases} 4X^3 & 0 < X < 1 \\ 0 & \text{elsewhere} \end{cases}$$

Use the CDF technique to derive the pdf of

(i) $Y = X^4$, (ii) $w = e^x$ (iii) $Z = \ln X$ (iv) $\mu = (X - 0.5)^2$

In the above example

$$P(x = 2/y = 3) = \frac{P(2,3)}{P(Y=3)}$$

$$= \frac{{}^{1}/_{24}}{{}^{1}/_{6}}$$

CHAPTER 5

SOME DISCRETE PROBABILITY DISTRIBUTIONS

5.0 Introduction

In this chapter, we will be studying some discrete probability distributions with a view to obtaining their men and variances.

5.1. Bernoulli Random Variable

A random variable X, that assumes only the value 0 or 1 is known as a Bernoulli random variable. The values 0, or 1 can be interpreted as events of failure and success respectively in an experiment usually referred to as ***Bernoulli trial.***

Definition 1: A random variable X is defined to have a Bernoulli distribution if the discrete density function of X is given by

$$f(x) = \begin{cases} p^x(1-p)^{1-x} \text{for } x = 0 \text{ or } 1 \\ 0 \quad otherwise \end{cases} = p^x(1-p)^{1-x} I_{\{0,\ 1\}}(x)$$

Where the p satisfies $0 \leq p \leq 1$. $1-p$ is usually denoted by q

Theorem 1: If X has a Bernoulli distribution, then

$E(X) = p$, $Var(X) = pq$

Proof:

$$E(X) = 0.q + 1.p = p$$
$$Var(X) = E(X^2) - (E[X])^2$$
$$= 0^2.q + 1^2.p - p^2 = pq$$

Bernoulli distribution is a special type of discrete distribution sometimes referred to as Indictor function. This implies that for a given arbitrary probability space$[S, A, P(.)]$, let A belong to A , define the random variable $X$to be the indicator function of A; that is $\chi(w) = I_A(w)$; then X has a Bernoulli distribution with parameter $p = P[X = 1] = P[A]$.

5.2 Binomial Distribution

In Bernoulli distribution, there is just one trial that can result in either success or failure. But, in Binomial distribution, we have repeated and independent trials of an experiment with two outcomes resulting in either success or failure, yes or no etc.

The probability of exactly x successes in n repeated trials is given by:

$$f(x) = \begin{cases} \binom{n}{x} p^x q^{n-x}; & x = 0, 1, 2, \ldots\ldots\ldots, n \\ 0 & elsewhere \end{cases}$$

where p is the probability of success

q = 1-p is the probability of failure

x is the number of successes in repeated trials.

f(x) is the probability density function (p.d.f).

5.2.1 Properties of Binomial distribution

(i) It has n independent trials

(ii) It has constant probability of success p and probability of failure $q = 1 - p$.

(iii) There is assigned probability to non-occurrence of events.

(iv) Each trial can result in one of only two possible outcomes called success or failure.

5.2.2 Mean and Variance of a Binomial Distribution

$$f(x) = \binom{n}{x} p^x q^{n-x}; \quad x = 0, 1, 2, \ldots\ldots\ldots, n$$

(i) Mean:

$$E(X) = \sum_{x=0}^{n} x\, f(x)$$

$$= \sum_{x=0}^{n} x \binom{n}{x} p^x q^{n-x}$$

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

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

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

$$= np \sum_{x=1}^{n-1} \frac{(n-1)!}{(n-x)!(x-1)!} p^{x-1} q^{n-x}$$

Let s = $x - 1$, x = s + 1

$$= np \sum_{s=0}^{n-1} \frac{(n-1)!}{(n-s-1)!s!} p^s q^{n-s-1}$$

$$= np \sum_{s=0}^{n-1} \binom{n-1}{s} p^s q^{n-s-1}$$

$$= np\,(p+q)^{n-1} = np$$

(ii) Variance:

$Var(X) = E(X^2) - [E(X)]^2$

$E[X^2] = E[X(X-1)] + E(X)$

$$E[X(X-1)] = \sum x(x-1)\, f(x)$$

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

$$= \sum x(x-1) \frac{n!}{(n-x)x!} p^x q^{n=x}$$

$$= \sum_{x=0}^{n} x(x-1) \frac{n(n-1)(n-2)!}{(n-x)!x(x-1)(x-2)!} p^2 p^{x=2} q^{n-x}$$

$$= n(n-1)p^2 \sum_{x=2}^{n-2} \frac{(n-2)!}{(n-x)!(x-2)!} p^{x=2} q^{n-x}$$

Let s = x-2   x = s + 2

$$= n(n-1)\,p^2 \sum_{s=0}^{n-2} \frac{(n-2)!}{(n-s-2)!s!} p^s q^{n-s-2}$$

$$= n(n-1)\,p^2 \sum_{s=0}^{n-2} \binom{n-2}{s} p^s q^{n-s-2}$$

$$E[X(X-1)] = n(n-1)p^2$$

$$\therefore E(X^2) = E[X(X-1)] + E(X)$$

$$= n(n-1)p^2 + np$$

$$\therefore V(X) = E(X^2) - [E(X)]^2$$

$$= n(n-1)\,p^2 + np - n^2p^2$$

$$= n^2p^2 - np^2 + np - n^2p^2$$

$$= np - np^2$$

$$= np\,(1-p)$$

$$= npq$$

Remark: The binomial distribution reduces to the Bernoulli distribution when $n = 1$

Example 1:

It is known that screw produced by a certain company will be defective with probability 0.02 independently of each other. The company sells the screws in packages of 10 and offer a money back guarantee that at most 1 of 10 screws is defective. What proportion of packages sold must the company replace?

Solution

Let X be the number of defective screws thus n = 1=, p = 0.02

Pr (at most one defective) $\equiv 1 - P(X=0) - P(X=1)$

$$P(X > 1) = 1 - P(X \le 1)$$

$$= 1 - \binom{10}{0}(0.2)^0(0.8)^{10} - \binom{10}{1}(0.2)^1(0.8)^9$$

What is the final answer?

Example 2:

A communication system consist of n components each of which will, independently function with probability p. The total system will be able to operate effectively it at least one-half of its components function.

For what value of p is the 7-components system were likely to operate more effectively than a 5 components system.

Solution

A 7-component system will be effective

$$If\ P(E_7 > 3) = P(E=4) + P(E=5) + (P(E=6) + P(E=7).$$

$$= 1 - P(E \le 3) = 1 - P(E=0) - P(E=1) - P(E=2) - P(E=3)$$

$$= \binom{7}{4} P^4 q^3 \binom{7}{5} + p^5 q^5 + \binom{7}{6} P^6 q^1 + P^7$$

A 5- component will be effective if

$$P(E_5 > 2) = P(E=3) + P(E=4) + P(E=5)$$

$$= \binom{5}{3} P^3 q^2 + \binom{5}{4} P^4 q^1 + P^5$$

The 7-component will be better if

$$P(E_7 > 3) > P(E_5 > 2);\ \text{for } q = 1 - p.$$

Complete this

Try for 5 and 3.

Example 3:

For what value of K will $P(X=K) / P(X=K-1)$ be greater or less than 1 if X is a b (n, p) and 0<P<1.

Solution

$$\frac{P(X=K)}{P(X=K-1)} = \frac{\binom{n}{k} P(1-P)^{n-k}}{\binom{n}{k-1} P^{k-1}(1-P)^{n-k+1}}$$

$$= \frac{(n-k-1)P}{k(1-P)}$$

$$\therefore P(X=k) \ge P(X=k-1) \quad \text{iff}$$

$$(n-k+1)P \ge k(1-P)$$

$$\text{i.e. } K \le (n+1)P$$

This implies that for the binomial distribution b (n, p), as *k* goes from 0 to n, P (x=k) first increases monotonically and then decreases monotonically, reaching its largest value when k is the maximum.

5.3 Poisson Distribution

When n becomes large and p is fairly small, the use of the binomial distribution in calculating the various probabilities becomes cumbersome. To overcome this problem, we use another probability function which approximates the binomial distribution. This probability function is known as the Poisson probability function which we shall be considering in this lecture.

A random variable closely related to the binomial random variable is one whose possible values 0, 1, 2, 3,..... represent the number of occurrences of some outcomes not in a given number of trials but in a given period of time or region of space. This variable is called the Poisson variable.

5.4 Properties of a Poisson Experiment

A Poisson experiment is a statistical experiment that has the following properties:

1. The experiment results in outcomes that can be classified as success or failures.
2. The average number of success(λ) that occur in a specified region is known.
3. The probability that a success will occur is proportional to the size of the region.
4. The probability that a success will occur in an extremely small region is virtually zero.

Note:

(The specified region may take many forms e.g. length, an area, a period of time, volume etc)

A Poisson random variable is the number of successes that result from a Poisson experiment.

The probability distribution of a Poisson random variable is called a Poisson distribution.

Given the mean number of successes λthat occur in a specified region, the probability density function (pdf) of Poisson distribution is given by

$$P(x:\lambda) = \frac{e^{-\lambda}(\lambda^x)}{x!}$$

where x is the actual number of successes that result from the experiment.

λ = np (n is the total number of observation in the experiment and p is the probability of success).

Note that mean λ and variance are equal i.e. λ= mean = variance. Also λ is the parameter of the distribution, with e= 2. 71828

Some examples of random variables that obey the Poisson probability law are:

1. The number of customers entering a post office on a given day
2. The number of misprints on a page (or a group of pages) of a book.
3. The number of packages of instant noodles sold in a particular store on a given day.

Identities:

$$e^{\lambda} = \sum_{x=0}^{\infty} \frac{\lambda^x}{x!}$$

$$= 1+\lambda+\frac{\lambda^2}{2!}+\frac{\lambda^3}{3!}+....$$

Using the result, we have

$$\sum_{x=0}^{\infty} P_{(x)} = \sum_{x=0}^{\infty} \frac{e^{-\lambda}\lambda^x}{x!}$$

$$= e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!}$$

$$= e^{-\lambda} \quad e^{\lambda} = 1$$

5.5 Mean and Variance of a Poisson Distribution

(i) Mean

$$E(X) = \sum_{x=0}^{\infty} xP(x)$$

$$= \sum_{x=0}^{\infty} x \frac{e^{-\lambda}\lambda^x}{x!} \qquad \lambda^x = \lambda^{x-1}.\lambda$$

$$= \sum_{x=0}^{\infty} x \frac{e^{-\lambda}\lambda^x}{x!}$$

$$= \sum_{x=1}^{\infty} x\lambda \frac{\lambda^{x-1}e^{-\lambda}}{x(x-1)!}$$

$$= \lambda \sum_{x=1}^{\infty} \frac{\lambda^{x-1}e^{-\lambda}}{(x-1)!}$$

Let s = x – 1

$$= \lambda \sum_{s=0}^{\infty} \frac{e^{-\lambda}\lambda^s}{s!}$$

$$= \lambda$$

(ii) Variance

$$Var(X) = E(X^2) - [E(x)]^2$$

$$E(x^2) = E[x(x-1)] + E(x)$$

$$E[x(x-1)] = \sum_{x=0}^{\infty} x(x-1) \frac{e^{-\lambda}\lambda^x}{x!}$$

$$= \lambda^2 \sum_{x=2}^{\infty} \frac{e^{-\lambda}\lambda^{x-2}}{(x-2)!} \qquad \lambda^x = \lambda^{x-2}\lambda^2$$

Let s = x – 2

$$= \lambda^2 \sum_{s=0}^{\infty} \frac{e^{-\lambda}\lambda^s}{s!}$$

$$= \lambda^2$$

$$\therefore Var(X) = \lambda^2 + \lambda - [\lambda]^2$$

$$= \lambda^2 + \lambda - \lambda^2$$

$$= \lambda$$

Example 1:

The average number of homes sold by Assurance Homes Company is 2 homes per day. What is the probability that exactly 3 homes will be sold tomorrow?

Solution: λ = 2 since 2 homes are sold per day on the average

x = 3, e = 2.71828

$$\therefore P(x; \lambda = np) = \frac{e^{-\lambda}\lambda^x}{x!}$$

$$= \frac{2.71828^{-2}\,(2)^3}{3!}$$

$$P(3; 2) = 0.180$$

Thus the probability of selling 3 homes is 0.180

5.6 The Poisson Distribution as an Approximation to the Binomial Distribution

Let n and p be the parameters of a binomial distribution.

Therefore mean $\lambda = np$

Variance $\sigma^2 = np(1-p)$

If $n \to \infty$ and $p \to 0$ simultaneously, in such a way that $\lambda = np$ is fixed, then we can say that $p = \lambda/n$ where λ is a fixed value.

Then as n increases, the binomial probabilities.

$P(x) = \binom{n}{x} p^x (1-p)^{n-x}$, $x = 0, 1, 2, \ldots$ Get closer and closer to the Poisson probabilities.

Proof: Given that

$$P(x) = \frac{e^{-\lambda}\lambda^x}{x!} \text{ where } \lambda = np, x = 0, 1, 2, \ldots$$

$$P(x: n, p) = p(x: n, \lambda/n)$$

$$= \binom{n}{x}\left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x} \quad x = 0, 1, 2, \ldots n$$

$$= \frac{n(n-1)(n-2)\ldots(n-x+1)(n-x)!}{x!(n-x)!} \frac{\lambda^x}{n^x} \left(1 - \frac{\lambda}{n}\right)^{n-x}$$

$$= \frac{n(n-1)(n-2)\ldots(n-x+1)}{x!\,n^x} \lambda^x \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-x}$$

$$= \frac{n(n-1)(n-2)\ldots\ldots(n-x+1)}{n^x} \frac{\lambda^x}{x!} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-x}$$

$$\text{but } \lim_{n \to \infty} \left(1 - \frac{\lambda}{n}\right)^n \equiv e^{-\lambda}$$

$$\lim_{n \to \infty} \left(1 - \frac{\lambda}{n}\right)^{-x} \equiv 1$$

$$\frac{n(n-1)(n-2)\ldots(n-x+1)}{n^x} \equiv 1$$

Therefore, we have $\frac{e^{-\lambda}\lambda^x}{x!}$, $x = 0, 1, 2\ldots$

Thus, the binomial pdf approaches the Poisson as n increases and p tends to zero.

5.7 Hypergeometric Distribution

Consider a lot consisting of $m + n$ items of which m of them are defective and the remaining n of them are non-defective. A sample of r items is drawn randomly without replacement. Let x denote the number of defective items that is observed in the sample. The random variable x is the hypergeometric random variable with parameters $m + n$ and m. Then, the number of ways selecting x defective items from m defective items is $\binom{m}{x}$; the number of ways of selecting $r - x$ non-defective items from n non-defective items is $\binom{n}{r-x}$. Therefore, total number of ways of selecting r items with x defective and $r - x$ non defective items is. $\binom{m}{x}\binom{n}{r-x}$

Finally, the number of ways one can select r different items from a collection of $m + n$ different items is $\binom{m+n}{r}$. Thus, the probability of observing x defective items in a sample of r items (probability density function) is

$$\frac{\binom{m}{x}\binom{n}{r-x}}{\binom{m+n}{r}} \text{ for } x = 0, 1, 2, \ldots, r, x \le m \text{ and } r - x \le n$$

Example 1:

In a lottery, a player selects 6 different numbers from 1,2,…,44 by buying a ticket for 1 naira (N1.00). Later in the week, the winning numbers will be drawn randomly by a device. If the player matches all six winning numbers, then he or she will win the jackpot of the week. If the player matches 4 or 5 numbers, he or she will receive a lesser cash prize. If a player buys one ticket, what are the chances of matching. (a) all 6 numbers (b) 4 numbers.

Solution: Let x denote the number of winning numbers in the ticket. If we regard winning numbers as defective, then x is a hypergeometric random variable with $m+n = 44, m = 6$ and $n = 38$.

(a) $$P(X=6) = \frac{\binom{6}{6}\binom{38}{0}}{\binom{44}{6}} = \frac{1}{\binom{44}{6}} = \frac{6!38!}{44!} = \frac{1}{7059052}$$

(b) $$P(X=4) = \frac{\binom{6}{4}\binom{38}{2}}{\binom{44}{6}} = 0.0014938$$

Example 2:

As part of a health survey, a researcher decides to investigate prevalence of cholera in 8 sub-urban areas out of a city's 28 sub-urban areas. If 6 of the sub-urban areas have a very high prevalence rate, what is the probability that none of them will be included in the researcher's sample?

Solution:

Recall that we have $f(x) = \frac{\binom{m}{x}\binom{n}{r-x}}{\binom{m+n}{r}}$ as the p. d. f for hypergeometric distribution

Here, $x = 0, n = 22, m + n = 28$ and $m = 6$

Then, we have $$\frac{\binom{6}{0}\binom{28-6}{8-0}}{\binom{28}{8}}$$

Complete this

5.8 Mean and Variance of Hypergeometric Distribution

(i) **Mean:**

$$E(x) = \sum_{x=0}^{r} xf(x)$$

$$= \sum_{x=0}^{r} x \frac{{}^{m}C_{x}\,{}^{n}C_{r-x}}{{}^{m+n}C_{r}}$$

$$= \sum_{x=0}^{r} x \frac{\frac{m!}{(m-x)!x!}\,{}^{n}C_{r-x}}{{}^{m+n}C_{r}}$$

$$= \sum_{x=1}^{r} \frac{\frac{x\,m(m-1)!}{(m-x)!x(x-1)!}\,{}^{n}C_{r-x}}{{}^{m+n}C_{r}}$$

$$= m\sum_{x=1}^{r} \frac{\frac{(m-1)!}{(m-x)!(x-1)!}\,{}^{n}C_{r-x}}{{}^{m+n}C_{r}}$$

let $s = x - 1, x = s + 1$

this implies $$\frac{m\sum_{s=0}^{r-1} \frac{(m-1)!}{(m-s-1)!s!}\,{}^{n}C_{r-s-1}}{{}^{m+n}C_{r}}$$

$$= \frac{m}{^{m+n}C_r} \cdot {^{m-1}C_s} \; {^{n}C_{r-s-1}}$$

$$= \frac{m}{^{m+n}C_r} \cdot {^{m+n-1}C_{r-1}}$$

$$= \frac{m}{\frac{m+n!}{(m+n-r)!r!}} \quad x \quad \frac{(m+n-1)!}{[(m+n-1)-(r-1)]!\,(r-1)!}$$

$$= \frac{m}{\frac{m+n!}{(m+n-r)!r!}} \quad x \quad \frac{(m+n-1)!}{(m+n-r)!(r-1)!}$$

Simplifying gives

$$E(x) \quad = \quad \frac{mr}{m+n}$$

(ii) Variance:

$$E(x^2) \quad = \quad E[x(x-1)+x]$$

$$= \quad E[x(x-1)] \; + \; E(x)$$

$$E[x(x-1)] \quad = \sum x(x-1)f(x)$$

$$= \sum_{x=0}^{r} x(x-1) \; \frac{^{m}C_x \; ^{n}C_{r-x}}{^{m+n}C_r}$$

Continuing gives

$$= \sum_{x=2}^{r} x(x-1) \frac{\frac{m(m-1)(m-2)!}{(m-x)!x(x-1)(x-2)!} \; ^{n}C_{r-x}}{^{m+n}C_r}$$

$$= \; m(m-1) \sum_{x=2}^{r} \frac{\frac{(m-2)!}{(m-x)!(x-2)!} \; ^{n}C_{r-x}}{^{m+n}C_r}$$

$$= \frac{m(m-1)}{^{m+n}C_r} \sum_{x=2}^{r} \frac{(m-2)!}{(m-x)!(x-2)!} \; ^{n}C_{r-x}$$

let s = x-2 . x = s + 2

$$= \frac{m(m-1)}{^{m+n}C_r} \sum_{s=0}^{r-2} \frac{(m-2)!}{(m-s-2)!s!} \; ^{n}C_{r-s-2}$$

$$= \frac{m(m-1)}{^{m+n}C_r} \; x \; ^{m-2}C_s \quad ^{n}C_{r-s-2}$$

$$= \frac{m(m-1)}{^{m+n}C_r} \; x \; ^{m+n-2}C_{r-2}$$

$$= \frac{m(m-1)}{\frac{m+n!}{(m+n-r)!r!}} \quad x \quad \frac{(m+n-2)!}{[(m+n-2)-(r-2)]!(r-2)!}$$

$$= \frac{m(m-1)(m+n-r)!r!}{(m+n)!} \quad x \cdot \frac{(m+n-2)!}{(m+n-r)!(r-2)!}$$

$$= \frac{m(m-1)r(r-1)(r-2)!}{(m+n)(m+n-1)(m+n-2)!} \quad \frac{(m+n-2)!}{(m+n-2)!}$$

$$= \frac{rm(m-1)(r-1)}{(m+n)(m+n-1)}$$

$$E(x^2) = E[x(x-1)] \; + \; E(x)$$

$$= \frac{m(m-1)r(r-1)}{(m+n)(m+n-1)} \; + \; \frac{rm}{m+n}$$

Therefore, $V(x) = E(x^2) - [E(x)]^2$

$$= \frac{m(m-1)r(r-1)}{(m+n)(m+n-1)} + \frac{rm}{m+n} - \left(\frac{rm}{m+n}\right)^2$$

Simplify this last expression to obtain $\frac{rm}{m+n}\left(1 - \frac{m}{m+n}\right)\left(\frac{m+n-r}{m+n-1}\right)$ (Post-Test 2)

Note: If the sampling was with replacement, r and $p = \frac{m}{m+n}$ would be the appropriate binomial parameter and its respective variance would be $r\frac{m}{m+n}\left(1 - \frac{m}{m+n}\right)$.

The binomial variance is slightly greater than the hypergeometric variance because of the factor $\left(\frac{m+n-r}{m+n-1}\right)$ in the hypergeometric variance.

As $m+n$ becomes very large compared to r, the hypergeometric distribution tends to the binomial distribution.

5.9 Binomial Distribution as an approximation to the Hypergeometric Distribution

Suppose the p.d.f of a hypergeometric distribution is given by

$$f(x) = \frac{\binom{m}{x}\binom{n}{r-x}}{\binom{m+n}{r}}$$

then, we have the following theorem.

Theorem:

Let m, n $\to\infty$ and suppose that

$$\frac{m}{m+n} = P_{m,n} \to P,\ 0 < P < 1$$

then $$\frac{\binom{m}{x}\binom{n}{r-x}}{\binom{m+n}{r}} \to \binom{r}{x} P^x q^{n-x},\ x = 0,1,2,\ldots\ldots,r$$

Proof:

We have

$$\frac{\binom{m}{x}\binom{n}{r-x}}{\binom{m+n}{r}} = \frac{\frac{m!}{(m-x)!x!}\ \frac{n!}{[n-(r-x)]!(r-x)!}}{\frac{(m+n)!}{(m+n-r)!r!}}$$

$$= \frac{m!}{(m-x)!x!}\ \frac{n!}{(n-r+x)!(r-x)!}\ x\ \frac{(m+n-r)!r!}{(m+n)!}$$

$$= \binom{r}{x}\frac{m!n!\,(m+n-r)!}{(m-x)!(n-r+x)!(m+n)!}$$

$$= \binom{r}{x}\frac{m(m-1)\ldots\ldots(m-x+1)(m-x)!\ n(n-1)\ldots(n-r+x+1)(n-r+x)!}{(m-x)!(n-r+x)!(m+n)\ldots\ldots[(m+n)-(r-1)]}$$

$$= \frac{\binom{r}{x}m(m-1)\ldots\ldots[m-(x-1)]\ n(n-1)\ldots\ldots[n-(r-x-1)]}{(m+n)\ldots\ldots[(m+n)-(r-1)]}$$

Divide through by m+n

$$\binom{r}{x}\left[\frac{\left(\frac{m}{m+n}\right)\left(\frac{m}{m+n}-\frac{1}{m+n}\right)\ldots\ldots\left(\frac{m}{m+n}-\frac{x-1}{m+n}\right)\times\left(\frac{n}{m+n}\right)\left(\frac{n}{m+n}-\frac{1}{m+n}\right)\ldots\left(\frac{n}{m+n}-\frac{r-x-1}{m+n}\right)}{\left(\frac{m+n}{m+n}\right)\ldots\ldots\left(\frac{m+n}{m+n}-\frac{r-1}{m+n}\right)}\right]$$

$$\Rightarrow \binom{r}{x}\left(\frac{m}{m+n}\right)\left(\frac{m}{m+n}-\frac{1}{m+n}\right)\ldots\left(\frac{m}{m+n}-\frac{x-1}{m+n}\right)\times$$

$$\left(\frac{n}{m+n}\right)\left(\frac{n}{m+n}-\frac{1}{m+n}\right)\ldots\left(\frac{n}{m+n}-\frac{r-x}{m}\right)\times\frac{1}{1\ldots\left[1-\frac{(r-1)}{(m+n)}\right]}$$

Since $\frac{m}{m+n} \Rightarrow P,\ hence\ \frac{n}{m+n} \Rightarrow 1$

Therefore

$$\lim_{m+n\to\infty}\binom{r}{x}(p)\left(p-\frac{1}{m+n}\right)\cdots\left(p-\frac{x-1}{m+n}\right)\times(q)\left(q-\frac{1}{m+n}\right)\cdots\left(q-\frac{r-x-1}{m+n}\right)\times\frac{1}{1\cdots\left(1-\frac{r-1}{m+n}\right)}$$

$$=\binom{r}{x}p^x q^{r-x}$$

This result implies that we can approximate the probabilities $\frac{\binom{m}{x}\binom{n}{r-x}}{\binom{m+n}{r}}$ by $\binom{r}{x}p^x q^{r-x}$ by setting $p=\frac{m}{m+n}$ provided m, n are large. This is true for all x = 0, 1, 2, …., r.

If m, n, are large, approximate the hypergeometric distribution by an appropriate binomial distribution. If the need arises, we may also go a step further in approximating the binomial distribution by the appropriate Poisson distribution.

5.10 Negative Binomial and Geometric Distributions

Negative binomial and Geometric distributions are two families of discrete distributions that are very important in Statistics. The Geometric distribution is so named because the values of the Geometric density are the terms of a geometric series while the Negative binomial distribution is sometimes also referred to as the Pascal's distribution.

5.11 Negative Binomial Distribution

Consider a succession of Bernoulli trials, let $P(r)$ denote the probability that exactly $r+k$ $(k>0)$, trials are needed to produce k successes. This will so happen when the last trial, that is, $(r+k)$th trials is a success with probability p and the previous $(r+k-1)$ trials must have $(k-1)$ successes with probability $r+k-1_{C_{k-1}}p^{k-1}q^r$, where $q=1-p$

$$P(r) = prob\ of\ (k-1)\ successes\ in\ (x+k-1) trials$$
$$\times\ prob\ of\ (x+k)th\ success$$
$$= r+k-1_{C_{k-1}}p^{k-1}q^r.p$$
$$= r+k-1_{C_{k-1}}P^k q^r \qquad r=0,1,2 \ldots\ldots\ldots\ldots \textbf{eqn. (1)}$$
$$=\frac{P^k(k+r-1)(k+r-2)\ldots.[k+r+1-(r+1)]}{r!}q^r$$
$$=\frac{P^k(k+r-1)(k+r-2)\ldots.(k+1)k}{r!}q^r$$
$$=P^k(-1)^r\frac{(-k)(-k-1)\ldots.(-k-r+1)}{r!}q^r$$
$$=P^k(-1)^r{-k_{C_r}}q^r$$
$$=-k_{C_r}P^k(-q)^r \ldots\ldots\ldots\ldots \textbf{eqn. (2)}$$

Note that:

(i) $r+k-1_{C_{k-1}}P^k q^r, \qquad r=0,1,2,\ldots\ldots$

$$= r+k-1_{C_r}P^k q^r, \qquad r=0,1,2,\ldots\ldots$$

(ii) $\sum_0^\infty P(r) = P^k\sum_{r=0}^\infty -k_{C_r}(-q)^r$

$$=P^k[1-q]^{-k}$$
$$=p^k p^{-k}=1$$

Equations (1) and (2) for $k\geq 0$ are known as negative binomial distribution.

5.11.1 Mean and Variance of the Negative Binomial Distribution

(i) Mean

Recall that the moment generating function (MGF) of a random variable X,

$$M(t)=E(e^{tx}),$$

using the moment generating function approach, therefore, from equation (1), the MGF of R is

$$M(t)=E(e^{tr})=\sum_{r=0}^\infty e^{tr}\binom{r+k-1}{r}P^k q^r,$$

But $(1-x)^{-n}=\sum_{j=0}^\infty\binom{-n}{j}(-x)^j=\sum_{j=0}^\infty\binom{n+j-1}{j}x^j$ for $-1<x<1$

Therefore

$M(t)=\sum_{r=0}^{\infty} e^{tr}\binom{-k}{r} p^k(-q)^r,$

$=\sum_{r=0}^{\infty}\binom{-k}{r} p^k(-qe^t)^r$

$$= p^k(1-qe^t)^{-k}$$

Now, $M'(t) = k\,qe^t p^k(1-qe^t)^{-k-1}$

$$E(R) = M'(t)|t = 0 = \frac{kq}{p}$$

$$E(R)^2 = M''(t)$$

$$= k\,qe^t p^k(1-qe^t)^{-k-1} + (k+1)qe^t p^k(1-qe^t)^{-k-2}kqe^t$$

Complete the solution using $V(R) = E(R)^2 - (E(R))^2$ (see Post –test 4)

5.12 Geometric Distribution

If in equation (1), we put $k = 1$, we have

$$r + k - 1_{C_{k-1}} P^k q^r$$

$$= r_{C_0} P \quad q^r$$

$$= q^r p,\ r = 0,1,2,\ldots$$

and $q = 1 - p$, we have geometric distribution.

The following describes the Geometric distribution.

Consider a sequence of Bernoulli trials with probability p of success. This sequence is observed until the first success occurs. Let R denote the number of failures before this first success. For instance, if the sequence starts with F representing failure and S success, with F, F, F, S ……. then R=3. i.e. this distribution describes the event of first success after nth independent trials with probability ρ, $0 < \rho < 1$.

Moreover, the probability of such a sequence is P $[R = 3] = (q)(q)(q)(p) = q^3 p = (1-p)^3 p$.

Generally, the p.d.f, f(r) = $P[R = r]$ of R is given by

$$f(r) = (1-p)^r p, \qquad r = 0,1,2 \ldots\ldots$$

$$f(r) = q^r p, \qquad r = 0,1,2,\ldots\ldots\ldots \qquad (1)$$

Some authors define the geometric distribution by assuming 1 (instead of 0) is the smallest mass point. The p.d.f then has the form

$$f(r) = \begin{cases} p(1-p)^{r-1}, & r = 1,2,3\ldots\ldots\ldots\ldots \\ 0, & elsewhere \end{cases} \qquad (2)$$

5.12.1 Mean and Variance of a Geometric Distribution

Consider equation (2)

(i) Mean

$$E(R) = \sum_{r=1}^{\infty} r\, p(1-p)^{r-1}$$

let $1 - p = q$

$$\Rightarrow E(R) = \sum_{r=1}^{\infty} r\, p(q)^{r-1}$$

$$= \sum_{r=1}^{\infty} p\frac{d}{dq}(q)^r$$

$$= p\frac{d}{dq}\sum_{r=1}^{\infty}(q)^r$$

$$= p\frac{d}{dq}(q + q^2 + q^3 + \cdots)$$

But $(q + q^2 + q^3 + \cdots) = q(1 + q + q^2 + q^3 + \cdots)$

$$= q\left(\frac{1}{1-q}\right)$$

Therefore, $E(R) = p\frac{d}{dq}\left(\frac{q}{1-q}\right)$

$$= p\left(\frac{(1-q)(1) - q(-1)}{(1-q)^2}\right)$$

$$= p\left(\frac{(1-q+q)}{(1-q)^2}\right)$$

$$= p\left(\frac{(1)}{(p)^2}\right)$$

$$E(R) = \frac{1}{p}$$

(ii) Variance

$$E(R^2) = \sum_{r=1}^{\infty} r^2\, p(1-p)^{r-1}$$
$$= \sum_{r=1}^{\infty} r^2\, p(q)^{r-1}$$
$$= p\sum_{r=1}^{\infty} \frac{d}{dq}(rq^r)$$
$$= p\frac{d}{dq}\sum_{r=1}^{\infty}(rq^r)$$
$$= p\frac{d}{dq}(q + 2q^2 + 3q^3 + \cdots)$$
$$= p\frac{d}{dq}q(1 + 2q + 3q^2 + 4q^3 + \cdots)$$

Recall that $1 + 2x + 3x^2 + 4x^3 + \cdots = \frac{1}{(1-x)^2}$

Therefore, we have $p\frac{d}{dq}q\left(\frac{1}{(1-q)^2}\right)$

$$= p\left[\frac{(1-q)^2(1) + 2q(1-q)}{(1-q)^4}\right]$$
$$= p\left[\frac{[(1-q)][(1-q+2q)]}{(1-q)^4}\right]$$
$$= p\left[\frac{1+q}{(1-q)^3}\right]$$
$$= p\left[\frac{1+q}{(1-q)^3}\right]$$
$$= \left[\frac{1+q}{(p)^2}\right]$$

$E(R^2) = \left[\frac{2-p}{(p)^2}\right]$ since $q = 1 - p$

Therefore, $V(R) = E(R)^2 - \left(E(R)\right)^2$

$$= \left[\frac{2-p}{(p)^2}\right] - \left(\frac{1}{p}\right)^2$$
$$= \frac{1-p}{p^2}$$

Thus, the mean and variance of this form geometric distribution are $\frac{1}{p}$ and $\frac{1-p}{p^2}$ respectively.

Example 1: A fair die is cast on successive independent trials until second six is observed. What is the probability of observing exactly 10 non-sixes before the second six is cast.

Solution: This is a negative binomial distribution problem. So,

$\binom{r+k-1}{k-1} p^k (1-p)^r \; r = 0, 1, 2 \ldots\ldots$

Therefore, we have $\binom{10+2-1}{1}\left(\frac{1}{6}\right)^2 \left(5/6\right)^{10} = 0.049$

Example 2:

Team A plays team B in a seven game with series. That is the series is over when either of the teams wins four games. For each game, $\rho(A \text{ wins}) = 0.6$ and the games are assumed to be independent. What is the probability that the series will end in exactly six games.

Solution:

The game will end is either A or B wins the game series.

$$\rho(\text{game ends}) = \rho(\text{A wins series in 6 games}) + \rho(\text{B wins series in 6 games})$$
$$= \binom{5}{3}(0.6)^4(0.4)^2 + \binom{5}{3}(0.4)^6(0.4)^2$$
$$= 0.207 + 0.092$$
$$= 0.299$$

Note: that

$$\rho(\text{A wins series in 6 games}) = \rho[\text{A looses 2 games before 4 wins}]$$
$$= \rho(Y = 2)$$
$$= \binom{5}{3}(0.6)^4(0.4)^2$$
$$= 0.207$$

Example 3:

In a sequence of independent rolls of a fair die;

i. What is the probability that the first four is observed in the sixth trial.

Solution: This is geometric distribution problem

$P(R = 5) = \left(\frac{5}{6}\right)^5\left(\frac{1}{6}\right) = 0.067$ where R denotes the number of non-fours before the occurrence of the first four.

i. What is the probability that at least six trials are required to observe a four.

Solution: $P[R \geq 5] = 1 - P[R \leq 4]$

$$= 1 - [P[R=0] + P[R=1] + P[R=2]] + P[R=3 + P[R=4]]$$

$$P[R=0] = \left(\frac{5}{6}\right)^0\left(\frac{1}{6}\right) = \frac{1}{6}$$

$$P[R=1] = \left(\frac{5}{6}\right)^1\left(\frac{1}{6}\right) = \frac{5}{36}$$

$$P[R=2] = \left(\frac{5}{6}\right)^2\left(\frac{1}{6}\right) = \frac{25}{216}$$

$$P[R=3] = \left(\frac{5}{6}\right)^3\left(\frac{1}{6}\right) = \frac{125}{1296}$$

$$P[R=4] = \left(\frac{5}{6}\right)^4\left(\frac{1}{6}\right) = \frac{625}{7776}$$

$$= 1 - \frac{4651}{7776}$$

Complete the solution

5.13 Multinomial Distribution

We know from binomial distribution that each trial of a binomial experiment can result in two and only two possible outcomes. In the multinomial experiment, however, each trial can have two or more possible outcomes. So, a binomial experiment is a special case of a multinomial experiment.

A multinomial experiment is a statistical experiment that has the following properties:

- The experiment consists of n repeated trials

- Each trial has a discrete number of possible outcomes
- The probability that a particular outcome will occur is constant for any given trial
- The trials are independent

A multinomial distribution is the probability distribution of outcomes from a multinomial experiment.

Definition: Suppose a multinomial experiment consists of n trials, and each trial can result in any of k possible outcomes $E_1, E_2, E_3, \ldots, E_k$. Suppose, also, that each possible outcome can occur with probabilities $p_1, p_2, p_3, \ldots, p_k$. Then, the probability p that E_1 occurs n_1 times, E_2 occurs n_2 times,......., and E_k occurs n_k times is

$p = \left[\frac{n!}{(n_1!n_2!\ldots n_k!)}\right][p_1^{n_1}\, p_2^{n_2} \ldots\ldots p_k^{n_k}]$ where $n = n_1 + n_2 + n_3 + \cdots + n_k$

Example 1:

A bowl consists of 2 red marbles, 3 green marbles and 5 blue marbles. 4 marbles are randomly selected from the bowl with replacement. What is the probability of selecting 2 green marbles and 2 blue marbles?

Solution:

The experiment consists of 4 trials, so $n = 4$.

The 4 trials produce 0 red marbles, 2 green marbles and 2 blue marbles; so

$n_{red} = 0$, $n_{green} = 2$, $n_{blue} = 2$

On any particular trial, the probability of drawing a red, green or blue marble is 0.2, 0.3 and 0.5 respectively.

Using the multinomial formula, we have

$$p = \left[\frac{n!}{(n_1!\,n_2!\ldots\ldots n_k!)}\right][p_1^{n_1}\, p_2^{n_2} \ldots\ldots p_k^{n_k}]$$

$$\left[\frac{4!}{(0!\,2!\,2!)}\right][(0.2)^0(0.3)^2(0.5)^2]$$

Therefore $p = 0.135$.

Example 2:

Suppose a card is drawn randomly from an ordinary deck of playing cards and then put back in the deck. This exercise is repeated five times. What is the probability of drawing 1 spade, 1 heart, 1 diamond and 2 clubs?

Solution:

The experiment consists of 5 trials, n=5

The 5 trials produce 1 spade, 1 heart, 1 diamond and 2 clubs; so $n_1 = 1, n_2 = 1, n_3 = 1, n_4 = 2$

On any particular trial, the probability of drawing a spade, heart, diamond or club is 0.25, 0.25, 0.25 and 0.25 respectively. Thus, $p_1 = 0.25,\ p_2 = 0.25,\ p_3 = 0.25, p_4 = 0.25$

Using the multinomial formula, we have

$$p = \left[\frac{n!}{(n_1!\, n_2! \,.....\, n_k!)}\right] [p_1{}^{n_1}\, p_2{}^{n_2} \,.....\, p_k{}^{n_k}]$$

$$\left[\frac{5!}{(1!\,1!\,1!\,2!)}\right] [(0.25)^1(0.25)^1(0.25)^1(0.25)^2]$$

$$p = 0.05859$$

Practice Questions

1. Suppose that a fair die is rolled 9 times. Find the probability that 1 appears 3 times, 2 and 3 twice each, 4 and 5 once each.
2. In a city on a particular night, television channels 4, 3 and 1 have the following audiences: channel 4 has 25 percent of the viewing audience, channel 3 has 20 percent of the viewing audience and channel 1 has 50 percent of the viewing audience. Find the probability that among ten television viewers randomly chosen in that city on that particular night, 4 will be watching channel 4, 3 will be watching channel 3 and 1 will be watching channel 1.

CHAPTER 6

SOME CONTINUOUS PROBABILITY DISTRIBUTIONS

6.0 Introduction

Having studied some discrete probability distributions in the last chapter, this chapter now deals with the study of some commonly used continuous probability distributions.

6.1 Normal Distribution

A random variable X is said to have come from the normal distribution if its probability density function (pdf) $f(x)$is define as:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, -\infty, x < \infty$$

With $\mu > 0$ and $\sigma^2 > 0$

The mean and variance of the normal distribution can be obtained as follows:

$$E(x^2) = \int_{-\infty}^{\infty} x^r f(x) dx$$

$$= \int_{-\infty}^{\infty} x^r \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x^r\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

Let $Z = \frac{x-\mu}{\sigma}, \frac{dZ}{dx} = \frac{1}{\sigma}$

$$x = \mu + \delta Z$$

$$E(X^r) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (\mu + \sigma Z)^r\, e^{-\frac{1}{2}Z^2} \sigma dZ$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (\mu + \sigma Z)^r\, e^{-\frac{Z^2}{2}} dZ$$

When r = 1

$$E(X) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\mu + \sigma Z)\, e^{-\frac{Z^2}{2}} dZ$$

$$= \frac{1}{\sqrt{2\pi}}\left[\mu \int_{-\infty}^{\infty} e^{-\frac{z^2}{2}} + \sigma \int_{-\infty}^{\infty} Ze^{-\frac{z^2}{2}} dZ\right]$$

$$= \mu \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} + \sigma \int_{-\infty}^{\infty} Z \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dZ$$

Recall that $\frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$ is a standardized normal distribution with 0 and variance 1.

Therefore

$E(X) = \mu(1) + \sigma(0)$

$Since \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} = 1$ and therefore

$\therefore E(X) = \mu$

$$E(Z) = \int_{-\infty}^{\infty} Z \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dZ$$

To obtain the variance, set r to 2 in equation (1) and use

$Var(X) = E(X^2) - [E(X)]^2$, we proceed as follows

$$E(X^2) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\mu + \sigma Z)^2 e^{-\frac{z^2}{2}} dZ$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\mu^2 + 2\mu\delta Z + \sigma^2 Z^2) e^{-\frac{z^2}{2}} dZ$$

$$= \mu \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz + 2\mu\delta \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\delta}} Ze^{-\frac{z^2}{2}} dZ + \sigma^2 \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} Z^2 e^{-\frac{z^2}{2}} dZ$$

$= \mu^2(1) + 2\mu\sigma(0) + \sigma^2(1)$

$E(X) = \mu^2 + \sigma^2$

Therefore

$$Var(X) = (\mu^2 + \sigma^2) - \mu^2$$
$$= \sigma^2$$

6.2 Exponential Distribution

The exponential distribution (also known as negative exponential distribution) is the probability distribution that describes the time between events in a Poisson process, i.e. a process in which events occur continuously and independently at a constant average rate. It is the continuous analogue of the geometric distribution, and it has the

key property of being memoryless. In addition to being used for the analysis of Poisson processes, it is found in various other contexts.

The exponential distribution is not the same as the class of exponential families of distributions, which is a large class of probability distributions that includes the exponential distribution as the baseline distribution

A random variable X is said to have an exponential distribution if is probability density function is defined as

$$f(x) = \lambda e^{-\lambda x}, \lambda > 0$$

Its corresponding moment about the origin is derived using

$$\mu_r^1 = E(X^r) = \int_{-\infty}^{\infty} x^r f(x) dx$$
$$= \int_{-\infty}^{\infty} x^r \lambda e^{-\lambda x} dx$$
$$= \lambda \int_{-\infty}^{\infty} x^r e^{-\lambda x} dx$$

$Let y = \lambda x, \quad \frac{dy}{dx} = ë$

$dx = \frac{dy}{\lambda}$ $and x = \frac{y}{\lambda}$

$$\mu_r^1 = \lambda \int_0^{\infty} \left(\frac{y}{\lambda}\right)^r e^{-y} \frac{dy}{\lambda}$$

$$= \frac{\lambda}{\lambda^{r+1}} \int_0^{<} y^r e^{-y} dy$$

$$= \frac{\Gamma r+1}{\lambda^r} since \Gamma\alpha = \int_0^{\infty} y^{\alpha-1} e^{-y} dy$$

Therefore, the r^{th} moment about origin of an exponential; distribution is

$$\mu_r^1 = \frac{\Gamma(r+1)}{\lambda^r}$$

The first four moments can be demanded as follows

When $r = 1$, we have the mean

$$\mu_r^1 = \mu = \frac{\Gamma 2}{\lambda} = \frac{1_1^1}{\lambda} = \frac{1}{\lambda}$$

$Since \Gamma\alpha = (á - 1)_1^1$

When $r = 2$

$$\mu_2^1 = \frac{\Gamma 3}{\lambda^2} = \frac{(3-1)_1^1}{\lambda^2} = \frac{2}{\lambda^2}$$

From which variance can be obtained as follows

$$Var(X) = E(X^2) - [E(X)]^2$$

$$= \mu_2^1 - (\mu_1^1)^2$$

$$= \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2$$

$$\frac{1}{\lambda^2}$$

When $r = 3$

$$\mu_3^1 = \frac{\Gamma 4}{\lambda^3} = \frac{(4-1)_1^1}{\lambda^3} = \frac{6}{\lambda^3}$$

and similarly with $r = 4$

$$\mu_4^1 = \frac{\Gamma 5}{\lambda^4} = \frac{(5-1)_1^1}{\lambda^4} = \frac{2\lambda}{\lambda^4}$$

6.3 Gamma Distribution

The gamma distribution is a two-parameter family of continuous probability distributions. The common exponential distribution and chi-squared distribution are special cases of the gamma distribution.

In each of these three forms, both parameters are positive real numbers.

The parameterization with k and θ appears to be more common in econometrics and certain other applied fields, where e.g. the gamma distribution is frequently used to model waiting times. For instance, in life testing, the waiting time until death is a random variable that is frequently modeled with a gamma distribution.

A continuous random variable X is said to have a Gamma distribution if its probability density function is defined as follow

$$f(x) = \frac{e^{-\frac{x}{\beta}} x^{\alpha-1}}{\Gamma\alpha\, \beta^{\alpha}}, x > 0, \alpha > 0, \beta > 0$$

6.3.1 Moments of Gamma Distribution

$$E(X^2) = \int_0^{\infty} \frac{x^r e^{-\frac{x}{\beta}} x^{\alpha-1}}{\Gamma\alpha\, \beta^{\alpha}} \; dx$$

$$= \frac{1}{\Gamma\alpha\, \beta^{\alpha}} \int_0^{\infty} x^{r+\alpha-1} e^{-\frac{x}{\beta}} \; dx$$

If we let $y = \frac{x}{\beta} \Longrightarrow \; dx = \beta dy$

$$E(X^r) = \frac{1}{\Gamma\alpha\, \hat{a}^{\alpha}} \int_0^{\infty} (\beta y)^{r+\alpha-1} e^{-y} \beta dy$$

$$E(X^r) = \frac{1}{\Gamma\alpha\, \beta^{\alpha}} \int_0^{\infty} \beta^{r+\alpha-1} y^{r+\alpha-1} e^{-y} \beta dy$$

$$= \frac{1}{\Gamma\alpha} \int_0^{\infty} y^{r+\alpha-1} e^{-y} dy$$

Recall from Gamma function that

$$\Gamma\alpha = \int_0^{\infty} e^{-x} x^{-1} \; dx, \;\; then$$

$E(X^r) = \frac{\beta^r}{\Gamma \acute{a}} \Gamma(r + \alpha)$ This gives the *rth* moment about the origin from which the first four moments can be derived.

When $r = 1$, we have

$$E(X) = \frac{\beta}{\Gamma\alpha} \Gamma(r + \alpha)$$

$$E(X) = \frac{\beta}{\Gamma\alpha} \alpha\Gamma\alpha$$

$$= \alpha\beta$$

When $r = 2$

$$E(X^2) = \frac{\beta^2}{\Gamma\alpha} \Gamma 2 + \alpha$$

$$= \frac{\beta^2(1+\alpha)\Gamma(1+\alpha)}{\Gamma\alpha}$$

$$= \frac{\beta^2(1+\alpha)\alpha\Gamma\alpha}{\Gamma\alpha}$$

$$= \alpha(1+\alpha)\beta^2$$

Therefore, we obtain the variance of X using the fact that

$$Var(X) = E(X^2) - [E(X)]^2$$

$$= \alpha(1+\alpha)\beta^2 - (\alpha\beta)^2$$

$$= \alpha\beta^2 + \alpha^2\beta^2 - \alpha^2\beta^2$$

$Var(X) = \alpha\beta^2$

$When\, r = 3$

$E(X^3) = \frac{\beta^3}{\Gamma\alpha}\Gamma(3+\alpha)$ and finally

$with\ r = 4$, we have

$$E(X^4) = \frac{\beta^4}{\Gamma\alpha}\Gamma(4+\alpha)$$

6.3.2 Moment Generation Function of Gamma Distribution

The moment generating function of a random variable X distributed as Gamma i.e. $X \sim GA(\alpha\beta)$ is derived as follows:

$$M_x(t) = E(e^{tx}) = \int_{-\infty}^{\infty} e^{tx} f(x)\ dx$$

$for\ -h < t < h$

$$M_x(t) = \frac{\int_0^{\infty} e^{tx} e^{-\frac{x}{\beta}} x^{\alpha-1}}{\Gamma\alpha\, \beta^{\alpha}}\ dx$$

$$= \frac{1}{\Gamma\alpha\, \beta^{\alpha}} \frac{\int_0^{\infty} e^{tx} e^{-\frac{x}{\beta}} x^{\alpha-1}}{\Gamma\alpha\, \beta^{\alpha}}\ dx$$

$$= \frac{1}{\Gamma\alpha\, \beta^{\alpha}} \int_0^{\infty} e^{x\left(t-\frac{1}{\beta}\right)} x^{\alpha-1}\ dx$$

$$= \frac{1}{\Gamma\alpha\, \beta^{\alpha}} \int_0^{\infty} e^{-x\left(\frac{1}{\beta}-t\right)} x^{\alpha-1}\ dx$$

$$= \frac{1}{\Gamma\alpha\, \beta^{\alpha}} \int_0^{\infty} e^{-x\left(\frac{1-\beta t}{\beta}\right)} x^{\alpha-1}\ dx$$

If we let $y = x\frac{(1-\beta t)}{\beta} \Rightarrow x = \frac{\beta y}{1-\beta t}$

$\frac{dy}{dx} = \frac{1-\beta t}{\beta} \Rightarrow dx = \frac{\beta dy}{1-\beta t}$

$$\therefore M_x(t) = \frac{1}{\Gamma\alpha\, \beta^{\alpha}} \int_0^{\infty} e^{-y} \left(\frac{\beta y}{1-\beta t}\right)^{\alpha-1} \frac{\beta dy}{1-\beta t}$$

$$= \frac{1}{\Gamma\alpha\, \hat{a}^{\alpha}} (1-\beta t)^{\alpha} \int_0^{\infty} e^{y} y^{\alpha-1}\ dy$$

$Since\ \Gamma x = \int_0^{\infty} e^{y} y^{\alpha-1}\ dy$ as before.

Then,

$M_x(t) = \frac{\Gamma\alpha}{\Gamma\alpha\,(1-\beta t)^{\alpha}}$

$M_x(t) = (1-\beta t)^{-\alpha}$

Differentiating the above and setting t to zero, we obtain the first four moments about the origin as follows

$M_x^1(t) = \alpha\beta(1-\beta t)^{-\alpha-1}$

$E(X) = M_x^1(0) = \alpha\beta$

$M_x^{11} = -\alpha\beta^2(-\alpha-1)(1-\beta t)^{-\alpha-2}$

$E(X^2) = M_x^{11}(0) = \alpha^2\beta^2 + \alpha\beta^2$

$Var(X) = E(X^2) - [E(X)]^2$

$= \alpha^2\beta^2 + \alpha\beta^2 - (\alpha\beta)^2$

$= \alpha\beta^2$

The characteristics function, the second characteristic function and the cumulate generating function can be obtained respectively as

$\phi_x(t) = (1-\beta i\, t)^{-\alpha},\ \Phi_x(t) = -\alpha \log(1-\beta i\, t)$ and

$K_x(t) = -\alpha \log(1-\beta t)$

6.3.3 Maximum Likelihood Estimation of parameter of the Gamma Distribution

Let $X_1, X_2, \ldots, X_n$ be a random sample of size n taking from Gamma distribution, the likelihood function is

$$L = \frac{e^{-\frac{\Sigma x}{\beta}} \prod_{i=1}^{n} x^{\alpha-1}}{(\Gamma\alpha)^n \beta^{\alpha n}},$$

The corresponding log-likelihood function is

$$Log\ L = -\frac{\sum_{i=1}^{n} x_i}{\beta} + (\alpha - 1)\sum_{i=1}^{n} \log x_i - n \log \Gamma(\alpha) - \alpha n\ \log \beta$$

Differentiating this with respect to α and β we have

$$\frac{\partial logL}{\partial \alpha} = \sum_{i=1}^{n} \log x_i - n\frac{\Gamma\alpha^1}{\Gamma\alpha} - n\ log\beta$$

$$\frac{\partial logL}{\partial \beta} = \frac{\sum_{i=1}^{n} x_i}{\beta^2} - \frac{\alpha n}{\beta}$$

$where\ \varphi(\alpha) = \frac{\Gamma\alpha^1}{\Gamma\alpha}$ equation ( ) can be written as

$$\frac{\partial logL}{\partial \alpha} = \sum_{i=1}^{n} \log x_i - n\ \varphi(\alpha) - n \log \beta$$

6.4 Pareto Distribution

The **Pareto distribution**, named after the Italian civil engineer and economist Vilfredo Pareto, is a power law probability distribution that is used in description of social, scientific, geophysical, actuarial, and many other types of observable phenomena.

A random variable X is distributed Pareto with parameters β and K if its pdf is given as

$$f(x) = \frac{K\beta^K}{x^{K+1}},\quad K > \beta,\ x > 0$$

It is interesting to show that

$\int_{\beta}^{\infty} f(x)\ dx = 1$, this is as follows

$$\int_{\beta}^{\infty} \frac{K\beta^K}{x^{K+1}}\ dx$$

$$K\beta^K \int_{\beta}^{\infty} x^{-(K+1)}\ dx$$

$$K\beta^K \frac{x^{-(K+1)+1}}{-(K+1)+1}\Big|_{\beta}^{\infty}$$

$$-K\beta^K \frac{x^{-K}}{K}\Big|_{\beta}^{\infty}$$

$$-\frac{\beta^K}{x^K}\Big|_{\beta}^{\infty} = \frac{-\beta^K}{\infty^K} - -\frac{\beta^K}{\beta^K}$$

$$= 0 + \frac{\beta^K}{\beta^K}$$

$$= 1$$

6.4.1 Moments of the Pareto Distribution

The rth moment of the random variable $X \sim PAR(\beta, K)$

$$\mu_r^1 = E(X^r) = \int_{\beta}^{\infty} x^r f(x)\ dx$$

$$= \int_{\beta}^{\infty} x^r \frac{K\beta^K}{x^{K+1}}\ dx$$

$$= K\beta^K \int_{\beta}^{\infty} x^{r-K-1}\ dx$$

$$= K\beta^K \frac{x^{r-K}}{r-K}\Big|_{\beta}^{\infty}$$

$$\mu_r^1 = \frac{K\beta^K}{r-K} x^{r-K}\Big|_{\beta}^{\infty}$$

$$= \frac{K\beta^K}{K-r}\frac{1}{x^{K-r}}\Big|_{\beta}^{\infty}$$

$$= \frac{K\beta^K}{K-r}\left(\frac{1}{\infty^{K-r}} - \frac{1}{\beta^{K-r}}\right)$$

$$= \frac{-K\beta^K}{K-r}\left(0 - \frac{1}{\beta^{K-r}}\right)$$

$$= \frac{K\beta^K}{(K-r)\beta^{K-r}}$$

$$= \frac{K\beta^K}{(K-r)\beta^K\beta^{-r}}$$

$$E(X^r) = \frac{K\beta^r}{K-r}$$

The above is the *rth* moment of a Pareto distribution.

When $r = 1$, we have

$$E(X) = \frac{K\beta}{K-1}$$

Similarly, when $r = 2$, we have

$$E(X^2) = \frac{K\beta^2}{K-2}$$

From which variance can be obtained as

$$Var(X) = E(X^2) - [E(X)]^2$$

$$= \frac{K\beta^2}{K-2} - \left(\frac{K\beta}{K-1}\right)^2$$

$$= \frac{K\beta^2}{K-2} - \frac{K^2\beta^2}{(K-1)^2}$$

$$= \frac{K(K-1)^2\beta^2 - K^2\beta^2(K-2)}{(K-1)^2(K-2)}$$

$$Var(X) = \frac{K\hat{a}^2[(K-1)^2 - K(K-2)]}{(K-1)^2(K-2)}$$

$$= \frac{K\beta^2[K^2 - 2K + 1 - K^2 + 2K]}{(K-1)^2(K-2)}$$

$$= \frac{K\beta^2(1)}{(K-1)^2(K-2)}$$

$$Var(X) = \frac{K\beta^2}{(K-1)^2(K-2)}$$

6.5 Maxwell Distribution

In physics, particularly statistical mechanics, the Maxwell–Boltzmann distribution or Maxwell speed distribution describes particle speeds in idealized gases where the particles move freely inside a stationary container without interacting with one another, except for very brief collisions in which they exchange energy and momentum with each other or with their thermal environment. Particle in this context refers to gaseous atoms or molecules, and the system of particles is assumed to have reached thermodynamic equilibrium.

A random variable X is said to follow Maxwell distribution if its pdf is defined as

$$f(x) = \sqrt{\frac{2}{\pi}}\frac{x^2e^{-\frac{x^2}{2a^2}}}{a^3}, x > 0, a > 0$$

It is required to show that the above is a time pdf and we proceed as follows. Our expectation is that

$$\int_{\infty}^{\infty} f(x)dx = 1; \ then$$

$$\int_{-\infty}^{\infty} \sqrt{\frac{2}{\pi}}\frac{x^2e^{-\frac{x^2}{2a^2}}}{a^3}\ dx$$

[illegible] dx

$$Let\ y = \frac{x^2}{2a^2} \Rightarrow x^2 = 2a^2 y$$

$$\frac{dy}{dx} = \frac{x}{a^2} \qquad dx = \frac{a^2}{x}\,dy$$

$Since\ x^2 = 2a^2y$

$$x = (\sqrt{2y})a \ \Rightarrow dx = \frac{a^2 dy}{(\sqrt{2y})a} = \frac{a}{\sqrt{2y}}dy$$

$$\frac{\sqrt{\frac{2}{\pi}}}{a^3}\int_0^\infty 2a^2 y\, e^{-y}\frac{a}{\sqrt{2y}}\ dy$$

$$\frac{2\sqrt{2}a^3}{\sqrt{\pi}a^{3\sqrt{2}}}\int_0^\infty y^{\frac{1}{2}}e^{-y}\ dy$$

$$\frac{2}{\sqrt{\pi}}\Gamma\frac{3}{2} = \frac{2}{\sqrt{\pi}}\frac{1}{2}\Gamma\frac{1}{2}$$

$Since\ \Gamma\frac{1}{2} = \sqrt{\pi}$, we have

$$\frac{2}{\Gamma\frac{1}{2}}\frac{1}{2}\Gamma\frac{1}{2} = 1 \qquad Q.E.D.$$

This affirms that Maxwell distribution is a true pdf.

6.5.1 Moments of the Maxwell Distribution

$$E(X^r) = \int_{-\infty}^{\infty} x^r f(x)\ dx$$

$$= \int_0^\infty x^r \sqrt{\frac{2}{\pi}}\frac{x^2 e^{-\frac{x^2}{2a^2}}}{a^3}\ dx$$

$$= \frac{\sqrt{\frac{2}{\pi}}}{a^3}\int_0^\infty x^{2+r}e^{-\frac{x^2}{2a^2}}\ dx$$

Using the notates earlier, we have

$$= \frac{\sqrt{\frac{2}{\pi}}}{a^3}\int_0^\infty [(\sqrt{2y})a]^{2+r} e^{-y}\frac{a}{\sqrt{2y}}dy$$

$$= \frac{\sqrt{\frac{2}{\pi}}}{a^3}\int_0^\infty \left[(\sqrt{2y})^{\frac{1}{2}}a\right]^{2+r} e^{-y}\frac{a}{(2y)^{\frac{1}{2}}}dy$$

$$= \frac{\sqrt{\frac{2}{\pi}}}{a^3}\int_0^\infty \frac{(2y)^{\frac{2+r}{2}}a^{2+r}e^{-y}a}{(2y)^{\frac{1}{2}}}\ dy$$

When $r = 1$

$$E(X) = \frac{2^{1+\frac{1}{2}}a^1}{\sqrt{\pi}}\Gamma\frac{3}{2}+\frac{1}{2}$$

$$= \frac{2^{\frac{3}{2}}a\Gamma 2}{\sqrt{\pi}}$$

$$= \frac{2^{\frac{3}{2}}a}{\sqrt{\pi}} = \frac{2^{\frac{1}{2}}2a}{\sqrt{\pi}}$$

When $r = 2$

$$E(X^2) = \frac{2^{1+\frac{1}{2}}a^2}{\sqrt{\pi}}\Gamma\frac{3}{2}+\frac{2}{2}$$

$$= \frac{4a^2}{\sqrt{\pi}}\Gamma\frac{5}{2}$$

$$E(X^r) = \frac{\sqrt{\frac{2}{\pi}}}{a^3}\int_0^\infty \frac{2^{\frac{2+r}{2}}y^{\frac{2+r}{2}}a^{3+r}e^{-y}}{2^{\frac{1}{2}}y^{\frac{1}{2}}}\ dy$$

$$= \sqrt{\frac{2}{\pi}}\int_0^\infty \frac{2^{1+\frac{r}{2}}y^{1+\frac{r}{2}}a^r e^{-y}}{2^{\frac{1}{2}}y^{\frac{1}{2}}}\ dy$$

$$= \frac{\sqrt{\frac{2}{\pi}}\, 2^{1+\frac{r}{2}} a^{r}}{2^{\frac{1}{2}}} \int_{0}^{\infty} y^{\frac{1}{2}+\frac{r}{2}} e^{-y}\, dy$$

$$= \sqrt{\frac{2}{\pi}}\, 2^{\frac{1}{2}+\frac{r}{2}} a^{r} \Gamma \frac{3}{2} + \frac{r}{2}$$

$$= \sqrt{\frac{2}{\pi}} \sqrt{2}\, 2^{\frac{r}{2}} a^{r} \Gamma \frac{3}{2} + \frac{r}{2}$$

From which the first four moments can be derived

$$Since\ \Gamma \frac{5}{2} = \frac{3}{2} \Gamma \frac{3}{2} = \frac{3}{2}\frac{1}{2} \Gamma \frac{1}{2} = \frac{3}{4} \Gamma \frac{1}{2}$$

$$= \frac{3}{4} \sqrt{ð}$$

$$\therefore E(X^{2}) = \frac{4a^{2}}{\sqrt{\pi}} \frac{3}{4} \sqrt{\pi}$$

$$= 3a^{2}$$

$$Var\ (X) = E(X^{2}) - [E(X)]^{2}$$

$$= 3a^{2} - \left(\frac{2^{\frac{3}{2}} a}{\sqrt{\pi}}\right)^{2}$$

$$= 3a^{2} - \frac{8a^{2}}{\pi}$$

When $r = 3$

$$E(X^{3}) = \frac{2^{1+\frac{3}{2}} a^{3}}{\sqrt{\pi}} \Gamma \frac{3}{2} + \frac{3}{2}$$

$$= \frac{2^{\frac{5}{2}} a^{3}}{\sqrt{\pi}} \Gamma 3$$

$$E(X^{3}) = \frac{2^{\frac{3}{2}} a^{3} 2_{1}^{1}}{\sqrt{\pi}} \qquad Since\ \Gamma a = (a-1)_{1}^{1}$$

$$= \frac{2\, 2^{\frac{5}{2}} a^{3}}{\sqrt{\pi}} = \frac{2^{\frac{n}{2}} a^{3}}{\pi}$$

$$E(X^{4}) = \frac{2^{1+\frac{4}{2}} a^{4}}{\sqrt{\pi}} \Gamma \frac{3}{2} + \frac{4}{2}$$

$$= \frac{2^{3} a^{4} \Gamma \frac{7}{2}}{\sqrt{\pi}}$$

$$= \frac{8a^{4} \frac{5}{2} \Gamma \frac{5}{2}}{\sqrt{\pi}} = \frac{8a^{4} \frac{5}{2}\frac{3}{2} \Gamma \frac{3}{2}}{\sqrt{\pi}}$$

$$E(X^{4}) = \frac{8a^{4} \frac{5}{2}\frac{3}{2}\frac{1}{2} \Gamma \frac{1}{2}}{\sqrt{\pi}}$$

$$= 15a^{4}$$

The third and fourth moments about the mean, i.e. μ_3 *and* μ_4 can then be obtained as

$$\mu_{3} = 2a^{2}\left(\frac{16}{\pi} - 5\right)\sqrt{\frac{2}{\pi}}$$

$$\mu_{4} = a^{4}\left(15 - \frac{8}{\pi}\right)$$

Finally, the coefficient of Skewness and Kurtosis is thus:

$$\delta_{1} = \frac{2\left(\frac{16}{\pi} - 5\right)\sqrt{\frac{2}{\pi}}}{\left(3 - \frac{8}{\pi}\right)^{\frac{3}{2}}} \approx 0.48569$$

$$\delta_{2} = \frac{15 - \frac{8}{\pi}}{\left(3 - \frac{8}{\pi}\right)^{2}} - 3 \approx 0.10818$$

CHAPTER 7
PROBABILITY GENERATING FUNCTIONS (PGF)

7.1 Introduction

The probability generating function (PGF) for a discrete random variable is a power series representation (the generating function) of the probability mass function of a random variable X.

PGFs are often employed for their succinct description of the sequence of probability $P[X = i]$ and to make available the well-developed theory of power series with non-negative coefficients.

Definition 1: The probability generating function (PGF) of a random variable X is defined as:

$$G_x(t) = E[t^x]$$
$$= \sum_x t^x \; P[X = x]$$

where:

$G_x(t)$ is defined only when X take values in the non-negative integers

$P(X=x)$ is the probability mass function of X.

The notation G_x is usually used to emphasize the dependence on X.

7.2 Properties of PGF

1. The probability mass function of X is recovered by taking derivatives of G.

$$P(k) = P(X = k) = \frac{G^{(K)}(0)}{K!}$$

2. If X and Y have identical PGFs, then they are identically distributed. i.e. if there are two random variables X and Y and $G_X = G_Y$, then $f_X = f_Y$.

3. The expectation of X is given by

$$E(X) = G^1(1)$$

Proof:

$$G(t) = E(t^x) = \sum_x t^x P(x)$$

$$G'(t) = \frac{dG(t)}{dt}$$
$$= \sum_x x t^{x-1} P(x)$$
$$G'_{(1)} = \sum_x x P(x) \Rightarrow G'_{(1)} = E(x)$$

4. The variance of X is given by:

$$\text{Var}[X] = G''_{(1)} + G'_{(1)} - [G'_{(1)}]^2$$

Proof: $G'_{(t)} = \sum_x x\, t^{x-1} P_{(x)}$

$$G''_{(t)} = \sum_x x(x-1) t^{x-2} P_{(x)}$$
$$G''_{(t)} = \sum_x (x^2 - x) P_{(x)} t^{x-2}$$
$$G''_{(t)} = \left[\sum_x x^2 P_{(x)} - \sum_x x P_{(x)}\right] t^{x-2}$$
$$G''_{(1)} = \sum_x x^2 P_{(x)} - \sum_x x P_{(x)}$$
$$G''_{(1)} = E(x^2) - E(x)$$
$$G''_{(1)} = E(x^2) - G'_{(1)}$$
$$V(x) = E(X^2) - [E(X)]^2$$
$$V(x) = E(X^2) - \left[G_{(1)}^{1}\right]^2$$

But $E(x^2) = G''_{(1)} + G'_{(1)}$

Therefore $\text{Var}[X] = G''_{(1)} + G'_{(1)} - [G'_{(1)}]^2$

7.3 Probability Generating Functions

1. Bernoulli Distribution

The probability density function (pdf) of a Bernoulli distribution is given by

$$P[X = x] = P^x q^{1-x}$$

(i) Mean:

$G_x(t) = E[t^x]$

$$= \sum_{x=o}^{1} t^x P[X = x]$$

$$= t^0p^0q^{1-0} + t^1p^1q^{1-1}$$

$G_x(t) = q + pt$

$G'_{(t)} = p$

$G'_{(1)} = p = E(x)$

(ii) Variance:

$$G^{11}(t) = \frac{d^2G(t)}{dt^2}$$

Therefore, $G^1(t) = p$

$G^{11}(t) = 0$

But $Var(X) = E(X^2) - [E(X)]^2$

And $Var(X) = G^{11}(1) + G^1_{(1)} - [G^1_{(1)}]^2$

This implies that $0 + p - p^2 = p(1-p)$

$Var(x) = pq$

2. Binomial Distribution

The p.d.f of a binomial distribution is given by

$$\binom{n}{x} p^x q^{n-x}$$

where:

n is the number of observations

p is the probability of success

q is the probability of failure

x is the random variable

(i) Mean:

$G(t) = E[t^x]$

$$= \sum_{x=o}^{n} t^x P[X = x]$$

$$= \sum_{x=o}^{n} t^x \binom{n}{x} p^x q^{n-x}$$

$$= \sum_{x=o}^{n} (pt)^x \binom{n}{x} q^{n-x}$$

$$= \sum_{x=o}^{n} (pt)^x \binom{n}{x} q^{n-x}$$

$$= \sum_{x=o}^{n} (pt)^x \binom{n}{x} (1-p)^{n-x}$$

$$= [pt + (1-p)]^n$$

$G(t) = [pt + q]^n$

$$G^1(t) = \frac{dG(t)}{dt}$$

$$= n[pt + q]^{n-1} p$$

$G^1(t) = n(pt + q)^{n-1} p$

$G^1(1) = n(p + q)^{n-1} p$

$G^1(1) = np = E(x) = Mean$

(ii) Variance :

$G^{11}(t) = n(n-1)(pt + q)^{n-2} p^2$

$G^{11}(1) = n(n-1)P^2$

$\therefore Var(x) = G^{11}_{(1)} + G^1_{(1)} - [G^1_{(1)}]^2$

$= n(n-1)p^2 + np - n^2p^2$

$= n^2p^2 + np - np^2 - n^2p^2$

$= np - np^2$

$= np(1-p)$

$Var(x) = npq$

3. **Poisson Distribution**

(i) Mean :

$G(t) = E(t^x)$

$$= \sum_{x=0}^{\infty} t^x \frac{e^{-\lambda}\lambda^x}{x!}$$

$$= e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda t)^x}{x!}$$

$$= e^{-\lambda}e^{\lambda t}$$

$$= e^{-\lambda+\lambda t}$$

$$G(t) = e^{-\lambda(1-t)}$$

$$G^{I}(t) = \lambda e^{-\lambda+\lambda t}$$

$$G^{I}(t) = \lambda e^{-\lambda+\lambda}$$

$$= \lambda e^{0} = \lambda$$

$$G^{II}(t) = \lambda.\lambda e^{-\lambda+\lambda t}$$

$$= \lambda^2 e^{-\lambda+\lambda t}$$

$$G^{II}(1) = \lambda^2 e^{-\lambda+\lambda} = \lambda^2$$

(ii) Variance :

$Var(X) = G^{II}(1) + [G^{I}(1) - [G^{I}(1)]^2]$

$= G^{II}(1) + G^{I}(1) - [G^{I}(1)]^2$

$= \lambda^2 + \lambda - \lambda^2$

$= \lambda$

CHAPTER 8
MOMENT GENERATING FUNCTIONS

8.1 Moment Generating Function

The moment generating (m.g.f) is one which generates integral moments when these moments exists.

(i) For the univariate random variable X, the mgf is given by

$$M_x(t) = E(e^{tx}) = \sum_{x=0}^{x} e^{tx} \equiv \int_x^x e^{tx} f_{(x)}dx$$

Where t is a dummy variable

(ii) For the bivariate case we have corresponding

$$M_{x_1,x_2}(t_1,t_2) = E(e^{t_1x_1+t_2x_2})$$

Where t_1 and t_2 are dummies and the random variables X_1, X_2 are jointly distributed.

(iii) In general for multivariate case, we have

$$M_{x_1,x_2,\dots,x_n}(t_1,t_2,\dots t_n) = E(e^{t_1x_1+t_2x_2+\dots+t_nx_n})$$

The moment generating function $M_x(t)$ of a random variable X is defined for all real values of t by

$$M_{x(t)} = E(e^{tx})$$

$$= \begin{cases} \sum_x e^{tx}P_{(x)}; & if\ X\ is\ discrete \\ \int_{-\infty}^{\infty} e^{tx}f_{(x)}dx; & if\ X\ is\ continous \end{cases}$$

$M_x(t)$ is m.g.f. because all the moments of X can be obtained by successively differentlaity $M_x(t)$ and then evaluating the result at $t = 0$.

Example: If $f_{(x)} = \frac{1}{4}$, $X = 1,2,3,4$

$$M_x(t) = \sum_{x=1}^{x} e^{tx}f_{(x)}$$

$$= \frac{1}{4}e^t + \frac{1}{4}e^{2t} + \frac{1}{4}e^{3t} + \frac{1}{4}e^{4t}$$

If X_1 and X_x have the same pdf and $Y = X_1 + X_2$

$$M_y(t) = E[e^{t(X_1+X_2)}]$$

$$= E(e^{tx_1}.e^{tx_2})$$

$$M_y(t) = [M_x(t)]^2$$

$$= \frac{1}{6}e^{2t} + \frac{2}{16}e^{3t} + \frac{3}{16}e^{4t} + \frac{4}{16}e^{4t}\frac{5}{16}e^{5t} + \frac{3}{16}e^{6t}\frac{2}{16}e^{7t} + \frac{1}{16}e^{8t}$$

Example:

Let Y be a discrete random variable with pdf $\frac{e^{-\lambda}\lambda^y}{y!}$, $x = 0, 1, 2, ...$

$$M_y(t) = \sum_{y=0}^{\infty} e^{ty}\frac{e^{-\lambda}\lambda^y}{y!}$$

$$= \sum_{y=0}^{\infty}\frac{e^{ty}(\lambda e^t)^y}{y!}$$

$$= e^{-\lambda}\sum_{y=0}^{\infty}\frac{(\lambda e^t)^y}{y!} \quad = e^{-\lambda}e^{\lambda e^t}$$

$$= e^{\lambda(e^t-1)} \qquad = e^{\lambda(e^t-1)}$$

$$Since = \sum_{y=0}^{\infty}\frac{(\lambda e^t)^y}{y!}\lambda e^t = \frac{(\lambda e^t)}{2!} = 1 + \frac{\lambda e^t}{1!}\frac{(\lambda e^t)^2}{2!}$$

$$M_y'(t) = \lambda e^t exp\{(\lambda(e^t - 1)\}$$

$$M_y'(0) = \lambda$$

$$M_y^{"}(t) = (\lambda e^t)^2 exp\{(\lambda e^t - 1)\} + \lambda e^t \exp\{\lambda(e^t - 1)\}$$

$$= \lambda^2 + \lambda$$

$$Var(X) = M_y^{'}(0) - [M_y'(0)]^2 \Longrightarrow \lambda^2 + \lambda - \lambda^2$$

$$= \lambda$$

Obtain the $Var(x)$ given that

$$Var(x) = M_x^{"}(0) - [M_x'(0)]^2$$

For the discrete distribution, If X has a pdf $f_{(x)}$ with support $\{a_1, a_2, ...\}$ then

$$M_x(t) = \sum_R e^{tx}d_{(x)}$$

$$= f_{(a_1)}e^{ta_1} + f_{(a_2)}e^{ta_2} + \cdots$$

Hence, the c.d.f. at e^{ta_1} is $f_{(a_1)} = P(X = a_i)$. Thus, the probability of any value X say a_i is the coefficient of e^{ta_1}.

Example: Let the moments of Y be defined by $E(X^r = 0.8,\ r = 1, 2, 3, ...)$

Then

$$M_x(t) = M(o) + \sum_{r=1}^{\infty} 0.8\left(\frac{t^r}{r!}\right) = 1 + 0.8\sum_{r=1}^{\infty} 0.8\ \frac{t^r}{r!}$$

$$= 0.2 + 0.8\sum_{r=0}^{\infty} 0.8\ \frac{t^r}{r!}\ ??$$

$$= 0.2e^{0t} + 0.8e^{it}$$

Thus, $P * X = 0) = 0.2,\quad P(X = 1+= 0.8$

$$M_x'(t) = \frac{d}{dt}E(e^{tx})$$

$$= E\left[\frac{d}{dt}E(e^{tx})\right]$$

$$= E[Xe^{tx}]$$

Since the interchange at the differentiation and Expectation operator is allowed, we can assume that;

$$\frac{d}{dt}\left[\sum_x e^{tx}P_{(x)}\right] = \frac{d}{dt}\left[e^{tx}P_{(x)}\right]$$

for discrete case

$$\frac{d}{dt}\left[\int e^{tx} f_{(x)} dx\right] = \int \frac{d}{dt} e^{tx} f_{(x)} dx$$

for continuous case

Example 3:

From an Exponential Distribution

$$M_x(t) = E(e^{tx})$$

$$= \int_0^\infty e^{tx} \lambda e^{-\lambda x}\, dx$$

$$= \lambda \int_0^\infty e^{-(\lambda - t)x}\, dx$$

$$= \frac{\lambda}{\lambda - t} \qquad for\ t < \lambda$$

The above function is only defined for $t < \lambda$

$$M_x'(t) = \frac{\lambda}{(\lambda-t)^2} \Longrightarrow M_x''(o) = \frac{1}{\lambda}$$

$$M_x''(t) = \frac{2\lambda}{(\lambda-t)^3} \Longrightarrow M_x''(o) = \frac{2}{\lambda^2}$$

$$Var(X) = M_x''(o) - [M_x'(o)]^2$$

$$= \frac{2}{\lambda^3} - \left(\frac{1}{\lambda}\right)^2$$

$$= \frac{1}{\lambda^2}$$

Example 4: Given prove for (i)Normal Distribution

(ii) Standard Normal Distribution

An important property of m.g.f. is that the m.g.f. of the sum of independent random variables equals the product of the individual m.g.fs.

Let $Z = X + X_2$ where X_1 and X_2 are independent with m.g.fs $M_x(t)$ and $M_y(t)$. The m.g.f of Z is

$$M_x(t) = E\left[e^{t(x+y)}\right]$$

$$= E\left[e^{tx} . e^{ty}\right]$$

$$= E(e^{tx})\ E(e^{ty})$$

$$= \phi_x(t).\ \phi_y(t)$$

Also, the M.g.f. of a random variable uniquely determines the distribution.

Example:

If X and Y are independent random variable with parameters (n, p) and (m, p) respectively. What is the distribution of X + Y.

$$M_x(t) = (Pe^t + q)^n; \quad M_y(t) = (Pe^t + q)^m$$

$$\therefore M_{x+y}(t) = M_x(t).\ M_y(t)$$

$$(Pe^t + q)^n \ (Pe^t + q)^m \Longrightarrow (Pe^t + q)^{n+m}$$

Example

Calculate the distribution of X + Y when X and Y are independent. Poisson random variable with means λ_1 *and* λ_2 respectively.

Solution

$$M_x(t) = e^{\lambda(e^t - 1)}$$

$$M_{x+y}(t) = M_x(t) M_y(t)$$

$$= e^{\lambda_1(e^t-1)} e^{\lambda_2(e^t-1)}$$

$$= e^{(\lambda_1 - \lambda_2)(e^t - 1)}$$

Hence, X + Y is Poisson distributed with mean $(\lambda_1 + \lambda_2)$

Example: If X and Y are independent Normal random variables. The distribution of X + Y is

$$M_{x+y}(t) = M_x(t).\ M_y(t)$$

$$= \left\{e^{\frac{\sigma_1^2 t}{2} + \mu_1 t}\right\}\left\{e^{\frac{\sigma_1^2 t}{2} + \mu_2 t}\right\}$$

$$= \exp\left\{\frac{(\sigma_1^2 + \sigma_2^2)^t}{2} + (\mu_1 + \mu_2)t\right\}$$

If X and Y are independent discrete random variable with non-negative integers $\{0, 1, 2, \ldots\}$ as range with geometric. Distribution function

$$P_{(xi)} = q^j P. \quad with$$

m.g.f $M_x(t) = \frac{P}{1 - qe^t}$

What is the distribution of $Z = X + Y$

Solution

$$M_x(t) = M_x(t)M_y(t)$$
$$= \frac{p^2}{1-2qe^t+q^2e^{2t}}$$

Replace e^t $by\ z$ we have

$$h_z(t) = \frac{p^2}{(1-qz)^2}$$

$$= P^2\sum_{k=0}^{\infty} k+1\ q^kZ^k$$

The distribution of P_Z is a negative binominal distribution.

Examples: Let $f_{(x)} = \dfrac{e^{-\lambda}\lambda^x}{x!}$; $x = 0, 1, 2, ...$

$$M_x(t) = \sum_{x=0}^{\infty} e^{tx}\frac{e^{-\lambda}\lambda^x}{x!}$$
$$= e^{-\lambda}\sum_{x=0}^{\infty}\frac{(\lambda e^t)}{x!}$$
$$= e^{-\lambda}\left[1+\lambda e^t+\frac{(\lambda e^t)^2}{2!}+\ldots\right]$$
$$= e^{-\lambda}e^{\lambda e^t}$$

8.1.1 Moment Generating Function for Normal Distribution

$$f_{(x)} = \frac{1}{\sqrt{2\pi\delta^2}}e^{\frac{1}{2}\left(\frac{x-\mu}{\delta}\right)^2}$$

$$M_x(t) = E(e^{tx}) = \int_{-\infty}^{\infty} e^{tx}f_{(x)}dx$$

$$= \int_{-\infty}^{\infty} e^{tx}\frac{1}{\sqrt{2\pi\delta^2}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\delta}\right)^2}dx$$

$$= \frac{1}{\sqrt{2\pi\delta^2}}\int_{-\infty}^{\infty} e^{tx-\frac{1}{2}\left(\frac{x-\mu}{\delta}\right)^2}dx$$

$$= \frac{1}{\sqrt{2\pi\delta^2}}\int_{-\infty}^{\infty} e^{\left[\frac{1}{2}\left(\frac{x-\mu}{\delta}\right)^2-tx\right]}dx$$

$$= \frac{1}{\sqrt{2\pi\delta^2}}\int_{-\infty}^{\infty} e^{-\left[\frac{\left(\frac{x-\mu}{\delta}\right)^2-2\delta^2tx}{2\delta^2}\right]}dx$$

Note that

$$[(x-\mu)-\sigma^2t]^2 = (x-\mu)^2-2(x-\mu)\sigma^2t+\sigma^2t^2$$
$$= (x-\mu)^2-2x\delta^2+2\mu\,\sigma^2+\sigma^3t^2$$
$$(x-\mu)^2-2\delta^2tx = [(\_-\mu)-\sigma^2t]^2-2\mu\delta^2t-\delta^2t^2$$

$$M_x(t) = \frac{1}{\sqrt{2\pi\delta^2}}\int_{-\infty}^{\infty} e^{\frac{[(x-\mu)-\sigma]^2-2\mu\delta^2t-\sigma^2t^2]}{2\delta^2}}dx$$

$$\frac{e^{\mu t+\frac{\sigma^2}{2}t^2}}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}[(x-\mu)-\sigma^2t]}dx$$

If we left $y = \frac{(x-\mu)-\delta^2t}{\delta}$

$$\frac{dy}{dx} = \frac{1}{\delta} \Rightarrow dx = \delta dy$$

$$M_x(t) = e^{\mu t+\frac{\delta^2t^2}{2}}\int_{-\infty}^{\infty}\frac{1}{\phi\sqrt{2\pi}}e^{-\frac{y^2}{2}}\phi dy$$

$$= e^{\mu t+\frac{\delta^2 t^2}{2}} \int_{-\infty}^{\infty} \frac{1}{\phi\sqrt{2\pi}} e^{-\frac{y^2}{2}}\, dy$$

The function in the integral is a standardized normal distribution.

Therefore,

$$M_x(t) = e^{\mu t+\frac{\delta^2 t^2}{2}}$$

since

$$\int_{-\infty}^{\infty} \frac{1}{\phi\sqrt{2\pi}} e^{-y^2/2}\, dy = 1$$

Let $X \sim N_n(\mu_n, A)$, then the moment generating function of X is given as

$$M_x(t) = e^{t^1\mu + \frac{1}{2}t^1A^{-1}t}$$

Proof

We know from Alternate Integral that

$$\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} e^{-\frac{1}{2}X^1AX}\, dx_1 \dots dx_n = (2\pi)^{n/2}|A|^{-\frac{1}{2}}$$

$$M_x(t) = E(e^{t^1x})$$

Where $A = \Sigma = variance - covariance\ matrix$

$$= (2\pi)^{-\frac{n}{2}}|A|^{-\frac{1}{2}} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \exp\left[t_x^1 - \frac{1}{2}(x-\mu)^2A^{-1}(x-\mu)\right]\ dx_1 \dots dx_n$$

If we let

$L = t_x^1 - \frac{1}{2}(x-\mu)^1A^{-1}(x-\mu)$, then simplifying this we have

$$L = -\frac{1}{2}(x-\mu-At)^1(x-\mu-At) + t^1\mu + \frac{1}{2}t^1At$$

$$M_x(t) = \frac{t^1\mu + \frac{1}{2}t^1At}{(2\pi)^{n/2}|A|^{1/2}} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2}(x-\mu-At)^1V^{-1}(x-\mu - At)\right] dx_1 \dots dx_n$$

If we let $y = x - \mu - At$

$$\frac{dy}{dx} = 1 \Rightarrow |J| = \left|\frac{dx}{dy}\right| = 1$$

$$\therefore M_x(t) = \frac{e^{t^1}\mu + \frac{1}{2}t^1At}{(2\pi)^{n/2}|A|^{1/2}} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} e^{-\frac{1}{2}y^1A^{-1}y}\, dy_1 \dots dy_n$$

By examining

$$\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} e^{-\frac{1}{2}y^1A^{-1}y}\, dy_1 \dots dy_n$$

$$= (2\pi)^{n/2}|A|^{1/2} \quad i.e.\ Ankens\ inTegral$$

$$M_x(t) = \frac{e^{t^1}\mu + \frac{1}{2}t^1At}{(2\pi)^{n/2}|A|^{\frac{1}{2}}} . (2\pi)^{n/2}|A|^{\frac{1}{2}}$$

$$M_{(x)} = e^{t^1}\mu + \frac{1}{2}t^1At$$

8.2 Bivariate Distribution

Let X and Y be jointly distributed as

$f_{(x,y)} = \exp\{-(x+y)\}$

Obtain the joint m.g.f.

Solution: $M_{x,y}(t_1,t_2) = E(e^{t_1x+t_2y})$

$$= \int\int e^{t_1 x + t_2 y} e^{-(x+y)} dy\, dx.$$

$$= \int\int e^{-x(1-t_1)-y(1-t_2)} dx\, dy$$

$$= \ni$$

$$= \frac{1}{(1-t_1)(1-t_2)}$$

Exercise

(1) Obtain the joint m.g.f if

$$f_{(x,y)} = \exp\left\{-\frac{1}{2}\left(x^2+y^2\right)\right\} \qquad -\infty < x, y < \infty$$

(2) Obtain the joint m.g.f. if

$$f_{(x,y)} = \lambda^2 e^{-\lambda y} \qquad 0 < x < y$$

8.3 Obtaining Moments from m.g.f

Since m.g.f. continuous and differentiable in t, it is easy to obtain r^{th} moments $E(X^r)$ from m.g.f.

(i) The Univariate case

$$E(X^r) = \frac{d^{(r)}}{dt^r} M_x(0); \text{ where r is a posterior integral}$$

Example

The m.g.f. for the binomial distribution is

$$M_x(t) = \left(q + Pe^t\right)^n$$

$$E(x) = \frac{d}{dt} M_x(t) = \frac{d}{dt}\left(q + Pe^t\right)^n$$

$$M_x^1(t) = n\left(q + Pe^t\right)^{n-1}\left(Pe^t\right)$$

$$M_x^1(o) = nP$$

$$E(x^2) = M_x^{11}(t) = n(n-1)\left(q + Pe^t\right)^{n-2}\left(Pe^t\right)^2 + n\left(q + Pe^t\right)^{n-1}\left(Pe^t\right)$$

$$M_x^{11}(o) = n(n-1)P^2 + nP$$

$$\tau^2 = M_x^{11}(0) - \left[M_x^1(o)\right]^2$$

$$= n(n-1)P^2 + nP - n^2P^2$$

$$= npq$$

Practice Questions

1. Obtain EX^3, EX^4, hence Kurtosis and Skewness of X

2. The m.g.f. of a random variable is $(i)\ \alpha(\alpha-1)^{-1}$ $\quad (ii)\ \exp\left\{\mu t + \frac{1}{2}\tau^2 t^2\right\}$

Obtain the mean and variance of X for (i) and (ii) above.

3. For the bivariate case

$$E\left(X_1^r X_2^r\right) = \frac{d^{(r+s)}}{dt_1^r\, dt_2^s} M_{x_1,x_2}(0,0)$$

$$E\left(X_1^r\right) = \frac{d^{(r)}}{dt_1^{(r)}} M_{x_1,x_2}(0,0)$$

$$E\left(X_2^3\right) = \frac{d^{(3)}}{dt_2^{(3)}} M_{x_1,x_2}(0,0)$$

For r and s non-negative integers.

4. For $M_{x_1,x_2}(t_1,t_2) = \alpha_1, \alpha_2(\alpha_1 - t_1)^{-1}(\alpha_1 - t_1)^{-1}$

Obtain $E(X_1), E(X_2), E(X_1^2)\ E(X_2^2)\ Var(X_1), Var(X_2)\ and\ Cov(X_1, X_2)$

5. Obtain the m.g.f. if the joint distribution is given by

X_1/X_2	1	2
0	0.2	0.3
1	0.4	0.1

Obtain the estimate of the means as in the above table.

(3) For $M_{x_1,x_2}(t_1,t_2) = \exp\left\{t_1 m_1 + t_2 m_2 + \frac{1}{2}\left(t_1^2 \tau_1 2\right) + 2\rho\tau_1\tau_2 t_1 t_2 + t_2^2 \tau_2^2\right\}$

Obtain the same mean as in (4) above.

CHAPTER 9
CHARACTERISTIC FUNCTIONS

9.1 The Characteristic Function (c.f.)

The characteristic function $\varphi_x(t)$ of a random variable X is the expectation of a complex function of *X*. It is defined as $\varphi_x(t) = E(e^{itx})$

(i) for the bivariate case

$$\varphi_{x_1, x_2}(t_1, t_2) = E(e^{it_1 x_1 + it_2 x_2})$$

(ii) for the multivariate case, through characteristic function is given as

$$\varphi_{x_1, x_2, \dots x_n}(t_1, t_2, \dots t_n) = E(e^{it_1 x_1 + it_2 x_2 + \dots + it_n x_n})$$

Unlike the ordinary expectation, the characteristic function always exists.

This is because

$$|\varphi_x(t)| = |Ee^{itx}| = |E(Cos\, t x + i\, Sin\, tx)| = 1$$

Examples:

The characteristic function for the binomial distribution is given by

$$M_x(t) = E(e^{itx}) = \sum_{x=0}^{n} e^{itx}\binom{n}{x}P^x q^{n-x}$$

$$= \sum \binom{n}{x}(e^{it}P)^x q^{n-x}$$

$$= (Pe^{it} + q)^n$$

Exercise

Obtain the joint characteristic function for the following:

(*i*) $f(x, y) = \exp\{-(x + y)\}$

(*ii*) $f(x, y) = \frac{1}{2\pi}\exp\left\{-\frac{1}{2}(x^2 + y^2\right\} \quad \infty < x,\ y < \infty$

(*iii*) $P(x, y) =$

x/y	0	1	3
1	0.1	0.2	0.3
2	0.1	0.05	0.25

9.1.1 Moments from Characteristic Functions

(i) For **univatiate case:** The r[th] moment can be obtained as: $E(X^r) = \frac{1}{i^r}\varphi_x^{(r)}(0)$

This is obtained by differentiating $\varphi_{(x)}^{(t)} r$ times w.r. to t and evaluate the result at $t = 0$, then divide by $i^{(r)}$.

(ii) For the bivariate case

$$E(X^r Y^s) = \frac{1}{i^{(r+s)}}\frac{d^{(r+s)}}{d_1^{(r)} dt_2^{(s)}}\varphi_{x,y}^{(0,0)}$$

$$E(X^r) = \frac{1}{i^{(r)}}\frac{d^{(r)}}{dt_1^{(r)}}\varphi_{x,y}^{(0,0)}$$

$$E(X^s) = \frac{1}{i^{(s)}}\frac{d^{(s)}}{dt_2^{(s)}}\varphi_{x,y}^{(0,0)}$$

9.2 Exponential Distribution

The p.d.f is given as

$$f_{(x)} = \frac{1}{\theta}e^{-x/\theta} \qquad x < x < \infty$$

The C.F. is

$$\emptyset_{(t)} = \frac{1}{\theta}\int_0^\infty e^{itx}.e^{-x/\theta}dx$$

$$= \frac{1}{\theta}\int_0^\infty e^{-(\frac{1}{\theta}-it)}\,dx$$

$$= \theta^{-1}.\frac{1}{(\theta^{-1}-it)}e^{-x} = \frac{\theta^{-1}}{\theta^{-1}-it}$$

$$= \frac{1}{\left(1-\frac{it}{\theta^{-1}}\right)}$$

$$= \frac{1}{(1-i\theta t)}$$

$$\emptyset^1(t) = +i\theta(1 - i\theta t)^{-2}$$

$$m_1 = \emptyset^1(0) = \frac{i\theta}{i} = \theta$$

$$\emptyset^1(t) = 2i^2\theta^2(1 - i\theta t)^{-3}$$

$$\emptyset^{11}(0) = 2\theta^2 i^2$$

$$m_2 = \emptyset^{11}(0) = 2\theta^2$$

$$Var(r) = m_2 - m_1^2$$

$$= 2\theta^2 - \theta^2$$

$$= \theta^2$$

9.3 Gamma Distribution

A random variable is said to have a gamma distribution with parameters (t, λ), $\lambda > 0$ *and* $t > 0$ its density function is given by

$$f_{(x)} = \begin{cases} \frac{\lambda e^{-\lambda x}(\lambda x)^{K-1}}{\Gamma(k)} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

where

$$\Gamma(t) = \int_0^\infty e^{-y}\, y^{t-1} dy$$

integration by parts yields

$$= -e^{-y}y^{t-1}\Big|_0^\infty + \int_0^\infty e^{-y}(t-1)y^{t-2}dy$$

$$= (t-1)\int_0^\infty e^{-y}y^{t-2}dy$$

$$= (t-1)\Gamma(t-1)$$

If follows that

$$\Gamma(n) = (n-1)\Gamma(n-1)$$

$$= (n-1)(n-2)\Gamma(n-2)$$

$$= (n-1)(n-2)\ldots 3.3\,\Gamma(1) \quad \text{and } \Gamma(1) = \int_0^\infty e^{-x}dx = 1$$

The Characteristic Function of the Gamma distribution is obtained as:

$$\emptyset(t) = E(e^{itx}) = \int e^{itx}f(x)dx$$

$$= \frac{\lambda e^{itx-\lambda x}(\lambda x)^{K-1}}{\Gamma(k)} \Leftarrow \int \frac{e^{itx}.\lambda^{-\lambda x}(\lambda x)^{k-1}}{\Gamma(k)}$$

$$= \frac{\lambda^K}{\Gamma(k)}\int_0^\infty x^{K-1}e^{-(\lambda-it)}dx$$

$$= \int_0^\infty \frac{e^{-(\lambda-it)x}\,\lambda^k x^{k-1}}{\Gamma(K)}dx$$

$$\phi(t) = \frac{\lambda^k}{\Gamma(k)}\int_0^\infty x^{k-1}e^{-(\lambda-it)x}\,dx$$

Using Laplace transformation, we have

$$= \frac{\lambda^k}{(\lambda-it)^k} = \frac{\lambda^k}{\lambda^k(\lambda-\frac{it}{\lambda})^k} = \frac{1}{\left(1-\frac{it}{\lambda}\right)}$$

$$\phi^1(t) = ik\lambda^k(\lambda - it)^{-k-1}$$

$$m_1 = \phi^1(0) = \frac{i\,k\,\lambda^k\lambda^{-k-1}}{i} = \frac{k}{\lambda}$$

$$m_2 = \phi^{11}(0) = k(k+1)\lambda^k\lambda^{-k-2}$$

$$= \left(\frac{k}{\lambda}\right)^2 + \frac{k}{\lambda^2}$$

$$Var(x) = m_2 - m_1^2$$

$$= \left(\frac{k}{\lambda}\right)^2 + \frac{k}{\lambda^2} - \left(\frac{k}{\lambda}\right)^2$$

$$= \frac{k}{\lambda^2}$$

If X is a random variable of the discrete type $[i.e. x = 0, 1, 2, \ldots]$ with probability function. $P(X = x_i) = P_{(x)}$, then the characteristics function of X is define by

$$\phi(t) = E(e^{itx}) = \textstyle\sum_k P_k\, e^{itx_k} \qquad \ldots\ldots\ldots\ldots\ldots\ldots(1)$$

If X is a random variable of the continuous type with pdf$f_{(x)}$

then $\phi(t) = E(e^{itX}) = \int_{-\infty}^{+\infty} f_{(x)} e^{itx} dx$ .... (2)

since $[e^{itx}] = 1$ *and* $\sum_k P_k = 1$ *or* $\int_{-\infty}^{\infty} f_{(x)} dx = 1$

then $\int_{-\infty}^{+\infty} f_{(x)} [e^{itx}]\, dx = 1$

The summation in (1) and the integral in (2) are absolutely and uniforming converged. Thu, the characteristic function $\phi(t)$ is a continuous function for every value of t.

Properties

(i) $\phi(0) = E(e^0) = E(1) = 1$

(ii) $[\phi(t)] = |E(e^{itx})| \leq E|e^{itx}|) = 1$

Hence, $|\emptyset(t)| \leq 1$

(iii) $\phi(-t) = E(e^{-itx}) = E(Cost\, X - i\, Sin\, t\, x) = E(Cost\, x) - iE(Sint\, x)$

$\phi(-t) = E(e^{itx}) = E(Cost\, X + i\, Sin\, t\, x) = E(Cost\, x) - iE(Sint\, x)$

thus, $\phi(-t) = \overline{\phi(t)}$; a conjugate to $\phi(t)$

All c.f. must satisfy the above condition.

Example:

Let X be a random variable from the Bermoulli distribution. Obtain the Characteristic function

Solution

$$\phi(t) = \sum_{k=0}^{1} e^{itx_k} P_{(x_k)}$$

$$= \sum_{k=0}^{1} e^{itx_k} [p^x q^{1-x}], \quad x = 0, 1$$

$$= e^{it0} p^0 q^1 + e^{it(1)} p^1 q^0$$

$$= q + P\, e^{it}$$

$$= 1 - P + Pe^{it}$$

$$= 1 + P(e^{it} - 1)$$

Example: The moments of a characteristics function can be obtained by continuous differentiation of the function (discrete or continuous) r time and dividing the result by i^r

i.e. $\mu_r = \frac{\phi(o)}{i^r}$; r^{th} *moment*

Thus, $\mu_1 = \frac{\phi'(o)}{i}$; 1^{st} *moment*

Second moment $\mu_2 = \frac{\phi^{11}(o)}{i^2}$; $\mu_3 = \frac{\phi^{111}(o)}{i^3}$;

Since $\phi^r(t) = \int_{-\infty}^{\infty} i^r x^r f_{(x)} e^{itx} dx$

and $\phi^r(t) = \sum_k i^r x^r P_{(x_k)} e^{itx_k}$

Example 1:

$$\phi'(t) = \frac{d}{dt}\{(1 + P(e^{it} - 1)\}$$

$$= iPe^{it}p = i^2Pe^{it}$$

$$\phi''(0) = i^2 P$$

$$E(x^2) = \frac{i^2 P}{i^2} = P$$

$$E(X^2) - (E(X))^2$$

$$\mu_2 = Var(x) = P - P^2$$

$$= P(1 - P)$$

$$= pq$$

Example 2: Suppose X is from a Poisson distribution the characteristics function is given by

$$\phi(t) = \sum_{x=0}^{\infty} e^{itx} \frac{\lambda^x e^{-\lambda}}{x!}$$

$$= e^{-\infty} \sum_{x=0}^{\infty} \frac{(\lambda e^{it})^x}{x!}$$

$$= e^{-\lambda} e^{\lambda e^{it}} = e^{\lambda(e^{it} - 1)}$$

$$\phi'(t) = \lambda_i e^{it} . e^{\lambda(e^{it} - 1)}$$

$$\mu_1 = \frac{\phi'(o)}{i} = \frac{\lambda_i}{i} = \lambda$$

$$\phi''(t) = \lambda_i^2 e^{it}.e^{\lambda(e^{it}-1)} + \lambda_i e^{it}.\lambda_i e^{it}.e^{\lambda(e^{it}-1)}$$

$$= \lambda_i^2 e^{it}.e^{\lambda(e^{it}-1)}[\lambda e^{it}+1]$$

$$\phi''(o) = \lambda_i^2[\lambda+1]$$

$$\sigma^2 = \mu_2 = E(X^2) - E(X)^2$$

$$= \lambda(\lambda+1) - \lambda^2$$

$$= \lambda$$

Example 3: The characteristics function. and moments of the standard normal distribution is given as:

$$\phi(t) = \int_{-\infty}^{\infty} e^{itx} f_{(x)} dx$$

Where $f_{(x)} = \frac{1}{\Gamma 2\pi} e^{-x^2/2}$

$$= \phi(t) = \frac{1}{\Gamma 2\pi} \int_{-\infty}^{\infty} e^{itx} e^{-x^2/2}\, dx$$

$$= \frac{1}{\Gamma 2\pi} \int_{-\infty}^{\infty} e^{-\left(\frac{x^2 - itx}{2}\right)}\, dx$$

By completing the square in the experiment

$$= \frac{1}{\Gamma 2\pi} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(\frac{x-it}{2}\right)^2} - \frac{(it)^2}{2}\, dx$$

$$= \frac{1}{\Gamma 2\pi} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(\frac{x-it}{2}\right)^2}.e^{-t^2/2}\, dx$$

$$= \frac{1}{\Gamma 2\pi} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(\frac{x-it}{2}\right)^2} dx\, e^{-t^2/2}$$

$$= e^{-t^2/2}$$

$$\phi'(t) = -t\, e^{-t^2/2};\ \phi'(t) = 0$$

$$E(x) = \frac{\phi^1(o)}{i} = \frac{0}{i} = 0$$

$$m_2 = E(x^2) = \frac{\phi''(t)}{i^2} = t^2 e^{-t^2/2} - e^{-t^2/2}$$

$$= \frac{e^{-t^2/2}(t^2-1)}{i^2}$$

$$\frac{\phi''}{i^2} = \frac{1}{i^2} = 1$$

$$Var(X) = m_2 - m_1^2$$

$$= 1 - 0 = 1$$

Exercise:

Obtain m_3 *and* m_4 whatare your observation(m_2, m_3, m_5)ableto equal to zero

For Binonial distribution

$$\phi(t) = \sum_{x=0}^{\infty} e^{itx} P^x (1-P)^{n-x} \binom{n}{x}$$

$$= \sum_x \binom{n}{x} (P\,e^{it})^x (1-P)^{n-x}$$

$$= (Pe^{it} + q)^n$$

$$\phi'(t) = n(Pe^{it}+q)^{n-1}\, iPe^{it}$$

$$\phi'(0) = inp$$

$$m_1 = \frac{\phi^1(0)}{i} = np$$

$$Var\,(X) = n(n-1)P^2 + np - n^2P^2$$

$$= nP(1-P)$$

$$\phi''(t) = n(n-a)(Pe^{it})^2(Pe^{it}+q)^{n-2} + in\,Pe^{it}(Pe^{it}+q)^{n-1}$$
$$= in\,(n-1)P^2 + in\,P$$
$$m_2 = n(n-a)P^2 + np$$

9.4 Characteristics Function of the Sum of Independent Random Variables

Let X and Y be two independent random variables with characteristics function's e^{itx} *and* e^{ity} respectively. Let Z = X+Y and let $\phi_Z(t), \phi_x(t)$ *and* $\phi_y(t)$ denote their respective C.F.S. then

$$\phi_2(t) = E(e^{itz}) = E[e^{it(x+y)}]$$
$$= E[e^{itx}e^{ity}]$$
$$= E(e^{itx})\,E(e^{ity})$$
$$\phi_x(t)\phi_y(t)$$

This can be extended to any arbitrary number of independent random variable's

i.e. if $Z = X_1 + X_2 + \cdots X_n$ with C.F.S. as $\phi_{z(t)}, \phi_1(t), \phi_2(t) + \cdots, \phi_n(t)$

then

$$\phi_{z(t)} = \phi_1(t)\phi_2(t)\phi_2(t) \ldots \phi_n(t)$$

Example: Suppose two independent random variable X_1 and X_2 have $POI\ (\lambda_1)$ and $POI\ (\lambda_2)$. Determine the characteristics function of Z = $X_1 - X_2$

Solution

$$P(X_1 = r) = \frac{\lambda_1^r e^{-\lambda l}}{r!};\, P(X_2 = r) = \frac{\lambda_2^r e^{-\lambda_2}}{r!}$$
$$\phi_{x_1}(t) = e^{\lambda_1(e^{it}-1)};\ \phi_{x_2}(t) = e^{\lambda_2(e^{it}-1)}$$

But the C.F. of $(-X_2)$ is

$$\phi_{x_2}(t) = e^{\lambda_2(e^{it}-1)}$$
$$\phi_{z(t)} = e^{\lambda_1(e^{it}-1)};\ e^{\lambda_2(e^{it}-1)}$$
$$e^{[\lambda_1 e^{it} + \lambda_2 e^{-it} - \lambda_1 - \lambda_2]}$$
$$\phi_z^1(t) = i\lambda_1 - \lambda_2 e^{[\lambda_1 e^{it} + \lambda_2 e^{-it} - \lambda_1 - \lambda_2]}$$

$$m_1 = \frac{\phi_z^1(0)}{i} = \lambda_1 - \lambda_2$$
$$\phi_z''(t) = i^2(\lambda_1 - \lambda_2)^2 e[\,]$$
$$\mu_2 = \lambda_1 + \lambda_2$$

Example:

Let X, Y and G be two independent random variable with binominal distributions and let the characteristics function of X and Y be respectively.

$$\phi_1(t) = [1 + P(e^{it} - 1)]^{n_1}$$
$$\phi_2(t) = [1 + P(e^{it} - 1)]^{n_2}$$
$$\phi_1(t) = [1 + P(e^{it} - 1)]^{n_3}$$

Consider the r.y z = X + Y + G

Because of independent of X, Y and Z

$$\phi_z(t) = \phi_1(t)\phi_2(t)\phi_3(t)$$
$$= [1 + P(e^{it} - 1)]^{n_1+n_2+n_3}$$

The above is a binominal distribution where the addition theorem for the binominal distribution holds.

9.5 Some Special Probability Distribution

These are probability distribution of special importance in either theory or practices.

9.5.1 The One-Point Distribution

A random variable X has a one-point distribution if there exist a point x_0 such that

$$P(X = x_0) = 1 \qquad \text{(degenerate distribution)}$$

We say the probability mass is concentrated at a point.

The distribution function is given as

$$f_{(x)} = \begin{cases} 0, x \le x_0 \\ 1, x > x_0 \end{cases}$$

The characteristics function is defined as

$$\phi_t(t) = e^{it\,x_0}$$
$$m_1 = \phi'(o) = x_0$$
$$m_k = \phi^{(k)}(o) = x_0^{(k)}$$

$Var\ (x) = m_2 - m_1^2$

$x_0^2 - x_0^2 = 0$

It can be shown that if the variance of a random variable X equals zero, then X has a one-point distribution.

Proof

Since expression $[X - E(X)]^2 \geq 0$ i.e. non-negative

$Var\ (x) = E[(X - E(x)]^2 = 0$

Iff. $P[X - E(X) = 0] = 1\ or$

$$P[X = E(X)] = 1$$

Thus, we find that the random variable X has a one-point distribution.

9.5.2 Two-Point Distribution

A random variable X has a two-point distribution of there exist two values $x_1\ and\ x_2$ set.

$$P(X = x_1) = P,\ \ P(X = x_2) = 1 - P\ \ (0 < P < 1)$$

If we put $x_1 = 1\ and\ x_2 = 0$ we have

$$P(X = 1) = P,\ and\ P(X = 0) = 1 - P$$

Then the above qualities as a zero-one distribution.

A very good example of a zero-one distribution is the Bernoulli Distribution

$$\phi(t) = Pe^{it.1} + (1 - P)e^{it.0}$$

$$= Pe^{it} + (1 - P)$$

$$= 1 + P\left(e^{it} - 1\right)$$

$$\phi'(t) = P$$

$$\phi''(t) = P$$

$$\phi'''(t) = P$$

For every K

$$m_k = P$$

$$\mu_2 = Var(x) = m_2 - m_1^2$$

$$= P - P^2$$

$$= P(1 - P)$$

Show that $\mu_3 = m_3 - 3m_1m_2 + 2m_1^3$

$= P(1 - P)(1 - 2P)$ and

$$\delta = \frac{\mu_3}{\mu_2^{3/2}} = \frac{1-2P}{\sqrt{P(1-P)}}$$

Exercise

Obtain the mean and variance function for each of the following:

(i) $\varphi_x^{(t)} = \exp\left\{i\,\mu t - \frac{1}{2}\tau^2 t^2\right\}$

(ii) $\varphi_x^{(t)} = \alpha(\alpha - it)^{-1}$

(iii) $\varphi_{xy}^{(t_1,t_2)} = \alpha_1,\alpha_2(\alpha_1 - it_1)^{-1}(\alpha_2 - it_2)^{-1}$

9.6 The Inversion Formula

The characteristic function corresponds to a family at distribution which is obtained by adding an arbitrary constant to a d.f. of a random variable. The inversion formula is a tool that can be used to *get back* the original distribution function on the entire real line if the characteristic function is known.

Theorem

Let $F_{(x)}$ and $\phi_x^{(t)}$ be the cumulative distribution and the characteristic function of X respectively, then for given real numbers $a\ and\ b$, the inversion formula is defined as

$$F_{(b)} - F_{(a)} = \lim_{C \to \infty} \frac{1}{2\pi} \int_{-C}^{C} \frac{e^{-ita} - e^{-itb}}{it} \phi_x^{(t)} dt$$

Proof

Let $I_c = \frac{1}{2\pi} \int_{-C}^{C} \frac{e^{-ita} - e^{itb}}{it} \varphi_x^{(t)} dt$

First we need to show that $|\varphi_x^{(t)}| \leq$ that $\left|\frac{e^{-ita} - e^{itb}}{it}\right|$ is bounded.

$$|\varphi_x^{(t)}| = |E(e^{itx})| \leq E|e^{itx}|$$

$$\leq E|Cos tx + i\,Sin\,t x|$$

$$\leq E(Cos t x + i\,Sin\,t x)(Cos t x + i\,Sin\,t x)$$

$$\leq E(Cos^2 t x + i\,Sin^2 t x)$$

$$i.e.\ |\varphi_x^{(t)}| \leq 1$$

$$\left|\frac{e^{-ita} - e^{itb}}{it}\right| \leq 1 \int_a^b e^{-itx} dx|$$

$$\leq 1 \int_a^b dx |e^{-itx}|$$

$$\leq \int_a^b dx \qquad Sin\ is\ |e^{-itx}| \leq 1$$

$$\leq b - a \qquad hence\ bounded.$$

Now it is possible to apply the Fubini's theorem to I_c as

$$= \frac{1}{2\pi} \int_{-C}^{C} \frac{e^{-ita} - e^{-itb}}{it} \int_{-\infty}^{\infty} e^{itx} dF_{(x)} dt; \qquad (a < b)$$

$$= \frac{1}{2\pi} \int_{-C}^{C} \frac{e^{-ita} - e^{-itb}}{it} \int_{-\infty}^{\infty} e^{itx} dF_{(x)} dt$$

$$= \frac{1}{2\pi} \int_{-C}^{C} \frac{e^{it(x-a)} - e^{it(x-b)}}{it} dt) dF_{(x)}$$

Where $e^{it(x-a)} - e^{it(x-b)}$ can be written as

$$Cos t(x-c) + i\,Sin\,t(x-a) - Cos t(x-b) + i\,Sin\,t(x-b)$$

$$\therefore I_c = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-C}^{C} \frac{Cos t(x-a) + i\,Sin\,t(x-a) - Cos t(x-b) + i\,Sin\,t(x-a) dt)}{it} dF_{(x)}$$

multiply numerator and denominator by i

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{2i}{i} \left(\frac{Cos t(x-a) + i\,Sin\,t(x-a) - Cos t(x-b) + i\,Sin\,t(x-a) dt)}{it} \right) dF_{(x)}$$

$as\ C \to \infty$

$$= \frac{1}{\pi} \int_{-\infty}^{\infty} - \int_0^{\infty} i\,Cos t(x-a) - Sin\,t(x-a) + i\,Cos t(x-b) - Sin\,t(x-b) dt \,] dF_{(x)}$$

Since the integration of a Cosine function gives a Sine function which will later vanish

$$= \frac{1}{\pi} \int_{-\infty}^{\infty} \int_0^{\infty} \left(\frac{Sin\,t(x-a) - Sin\,t(x-b)}{t} dt \right) dF_{(x)}$$

By complex integration $\int_0^{\infty} \frac{Sin\ v}{v} dv = \frac{\pi}{2}$

and $\int_0^{\infty} \frac{Sin\ Pv}{v} dv = \frac{\pi}{2} Sgn\,P$ where

$$Sgn\ P = \begin{cases} 1 & if \quad P > 1 \\ 0 & if \quad P = 0 \\ -1 & if \quad P < 1 \end{cases}$$

Corollary: (Modern Probability Theory, (1985)

(1) Distribution function F of a random variable and its characteristic function determines each other.

(2) If $F_{(x)}$ is the d.f. of a random variable then by definition, it determines the characteristic function uniquely.

Proof:

If F and F^1 are the two d.f.s. corresponding to a given characteristic function $\phi_x^{(t)}$, then from the above theorem.

$$F^1_{(b)} - F_{(b)} = F^1_{(a)} - F_{(a)} \quad (b > a)$$

At all the common points of F and F^1.

Allowing b to vary for fixed a

$$F^1_{(b)} - F_{(b)} = F^1_{(a)} - F_{(a)} = a \text{ constant}$$

But $F^1_{(+\infty)} - F(+\infty) = 0$. Allowing b to increase infinitely through continuously points of F and F^1. This implies that $F^1_{(a)} - F_{(a)} = 0$ and hence continuity points of both.

$$\int_0^{t} \frac{Sin\,t(x-a) - Sin\,t(x-b)}{t}dt = \begin{cases} -\frac{\pi}{2}+\frac{\pi}{2}=0; \; x<a,\; x<b \\ 0+\frac{\pi}{2}=\pi/2; \; x=a,\; x<b \\ \pi/2+\pi/2=\pi; \; a<x<b \\ -\pi/2+0=\pi/2; \; x<a, x=b \\ \pi/2-\pi/2=0; \; x>a,\; x>b \end{cases}$$

$$\therefore I_c = \int_{-\infty}^{a} 0.dF_{(x)} + \int_{-a}^{a} \pi/2\, dF_{(x)} + \int_{-\infty}^{b-} 0.dF_{(x)} + \int_{-a}^{b} -\pi/2\, dF_{(x)}$$

$$= -\frac{\pi}{2}\left[F_{(a)} - F_{(a-1)}\right] + \frac{\pi}{2}\left[F_{(b)} - F_{(a-1)}\right]$$

$$= \frac{\pi}{2}\left[F_{(b)} - F_{(a)}\right]$$

$$= F_{(a+0)} - F_{(a-0)} + F_{(b+0)} - F_{(b-0)} + \frac{1}{2}\left[F_{(a+0)} - F_{(a-0)}\right.$$

$$= F_{(b)} - F_{(a)}$$

If a and b are points of continuity of F.

Example:

If $\phi_x^{(t)} = \left(q + pe^{it}\right)^n$, calculate the p.d.f of a random variable X.

Solution

$$f_{(x)} = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-itx}\phi_{(x)}dt$$

$$\pi \to \infty$$

Given $\phi_x^{(t)} = \left(q + pe^{it}\right)^n$

$$\therefore f_{(x)} = \frac{1}{2\pi}\int_{-\pi} e^{itx}\left(q + pe^{it}\right)^n dt$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{itx}e^{itj}\sum_{j=0}^{n}\binom{n}{j}\rho^j q^{n-j}dt$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{itx}e^{itj}\sum_{j=0}^{n}\binom{n}{j}\rho^j q^{n-j}dt$$

$$= \frac{1}{2\pi}\sum_{j=0}^{n}\binom{n}{j}\rho^j q^{n-j}\int_{-\pi}^{\pi} Cos\,t(x-j) - i\,Sin\,t(x-j)dt$$

Let $\mu = t(x-j)$

$$\frac{d\mu}{dt} = (x-j) \Rightarrow dt = \frac{d\mu}{(x-j)}$$

$$= \frac{1}{2\pi}\sum_{j=0}^{n}\binom{n}{j}\rho^j q^{n-j}2\int_0^{\pi} Cos\,t(x-j) - i\,Sin\,t(x-j)dt$$

But

$$= \int_0^{\infty}\frac{Sin\,t(x-j)}{(x-j)}\Bigg|_0^{\pi} - i\int_0^{\infty}\frac{Cos\,t(x-j)}{(x-j)}\Bigg|_0^{\infty}$$

$$= \pi - 0$$

$$\therefore \frac{\pi}{\pi}\sum_{j=0}^{n}\binom{n}{j}\rho^j q^{n-j}$$

$$\ell im_{n\to\infty}$$

$$= \sum_{j=0}^{\infty}\binom{n}{j}\rho^j q^{n-j}$$

$$\therefore f_{(x)} = \binom{n}{j}\rho^j q^{n-j}$$

OR

$$\therefore f_{(x)} = \binom{n}{j}\rho^j q^{n-j}$$

$$= \ell im_{x\to\infty}\frac{1}{\pi}\sum_{j=0}^{n}\binom{n}{j}\rho^j q^{n-j}$$

$$= \ell im_{t\to\infty}\frac{1}{\pi}\int_0^{n}\frac{Sin\,t(x-j)}{(x-j)}\sum_{j=0}^{n}\binom{n}{j}\rho^j q^{n-j}$$

$$= \ell im_{t\to\infty}\frac{1}{\pi}\int_0^{n}\frac{Sin\,t(x-j)}{(x-j)}\sum_{j=0}^{n}\binom{n}{j}\rho^j q^{n-j}$$

PART TWO

CHAPTER 10

INTRODUCTION TO MEASURE THEORY

10.1 Introduction

Probability theory is a part of mathematics which is useful in discovering the regular features of random events or phenomenon. In probability theory, the sigma algebra (which we shall define later) often represents the set of available information about a phenomenon. A function (or a function of a random variable) is measurable if and only if it represents an outcome that is knowable based on the available information about the experiment, the event to which it belongs and the probability function.

For us to understand how a probability measure can be obtained, let us develop an abstract model for the probability of an event particularly for infinite sample space Ω from a specified experiment.

10.2 Abstract Model for Probability of an Event

LetΩ be the sample space such that $\Omega = \{w_i \; i = 1, 2, 3, \dots \dots\}$

w_i are called indecomposable outcome or simple events.

The E_1 is a decomposable or compound events, that is $E_1 = \{w_i \; i = 1, 2, 3\}$

The elementary definition of probability is

$$P(E) = \frac{No\ of\ favourable\ cases}{Total\ number\ of\ cases} \qquad \dots\dots\dots\dots \qquad (1)$$

Since events are subset of Ω , it follows that the union and intersection of a finite number of events and the compliments are also events.

(1) For the model of mirror reality, the operation above can be represented by $A, B, A \cup B, A \cap B, A^{`}, B^{`}$. That is all statements about events can be written in terms of $\cup, .\cap$.

(2) A random for defining probability in term of weights is to allow for the fact, that some events are more likely to occur than others. The weight of a set is just the sum of the weights associated to each point in the set.

Let Ω be sure event, the impossible event will be ϕ. Let $\mathbb{A}$ be a non-empty class of subset of Ω called events. Let P(be the probability) be a real-valued function defined on $\mathbb{A}$. Such that $P(E)$denote the probability of event E.
The pair $\mathbb{A}, P$ is called the probability field and the triplet $(\Omega, \mathbb{A}, P)$ is called the probability space.

10.4 Axiom for Finite Probability Space

(i) If $E_i \in \mathbb{A}$ for $i = 1, 2, \ldots, n$ then

$$\bigcup_{i=1}^{n} E_i \in \mathbb{A} \text{ and } \bigcap_{i=1}^{n} E_i \in \mathbb{A}$$

(ii) If $E \in \mathbb{A}$, then $E' \in \mathbb{A}$

(iii) If $E \in \mathbb{A}$, then $P(E) \geq 0$, also $P(\Omega) = 1$

(iv) If E and F are any two disjoint events, then $P(E + F) = P(E) + P(F)$

(v) If A, B and C are any events, then:

$$P(A_1) + P(A_2) + A_3) = P(A_1) + P(A_2) + (PA_3)$$
$$= P(A_1) + P(A_1^c A_2) + A_1^c A_1^c A_2 + \cdots$$

The number of possible outcomes of an experiment (E) may be finite or infinite.
Let w denote a sample point (an outcome) from the experiment.
Let Ω denote the totality of outcomes of E i.e. $\Omega = \{w_1, w_2, \ldots\}$
Let event A={w: $w_i \in \Omega$}be a subset of Ω. e.g

(i) B={$w_i - \infty < w < \infty$}; all values on $\mathbb{R}$.

(ii) C={w_i: $a < w < b$}; all values in the range (a,b)

(iii) D={w_i: w_0}; a singleton.

(iv) E={w_i: $w_1, w_2,$}; a doubleton.

(v) F={w: $w = \phi$}; an empty set.

The class of all subsets of Ω is called the power set of Ω such that if Ω contains n points, there are 2^n subset of Ω.
Thus, if Ω is finite, the number of all possible subset is also finite.
The power set of Ω when $\Omega = \{w_1, w_2, w_3, w_4\} \Rightarrow 2^4 = 16$

Any collection of events is a class of events. Classes will be denoted by $\mathbb{A}, \mathbb{B}, etc.$

Example

Let Ω be the real line $\mathbb{R}$ containing all the real points $w.i.e.\ \Omega = \{w: -\infty < w < \infty\}$also let

$$A = \{w: w\epsilon(-\infty, a)\} \text{ and}$$
$$B = \{w: w\epsilon(c, d)\}$$

Define:

(i) $A \cap B$; (ii) $A \cup B$; (iii)A^c and B^c and give your assumptions

(iv) Show that the compliment of an interval need not be an interval.

Solution

$A \cap B = \phi$ if $a < c < d$
$= (c, a)$ if $c < a < d$
$= (c, d)$ if $c < d < a$
$A \cup B = \{w$: either WCA or $c < w < d\}$ will not an interval if $(a < c < d)$
$A^c = \{w: a < w < \infty\}$
$B^c = \{w: < w$or $w > d\}$
$A^c \cap B = B$ if $a < c < d$
$A^c \cup B = A^c$ if $a < c < d$
BCA^c if $a < c < d$

On your own, define the above if $c < a < d$ or if $c < d < a$

Sequences and Limits

A sequence of sets is an ordered arrangement of sets in order of magnitude
Monotone increasing sequence: A sequence of { sets $\{A_n\}$ is said to be monotone increasing if $A_n \subseteq A_{n+1}$ for each A_n.
If the sequence $\{A_n\}$ $n = 1, 2, \ldots$ is monotone increasing (non-decreasing) if for every n, we have $A_{n+1} \supset A_n$
Then the limit of $\{A_n\}$ is the 3mm of the sequence i.e.

$$A = \sum_{n\geq 1} A_n = \lim_{n\to\infty} A_n$$

or

$$\bigcup_{k=1}^{n} A_k = A_n; \bigcup_{k=1}^{\infty} A_k = A \quad i.e.\ A_n \uparrow A$$

$$= \{w: w \text{ belonging to all } A_k \text{ except } A_1, \dots A_{n-1}\}$$

$$C_n = \sup_{k\geq n} A_k = \bigcup_{k=n}^{\infty} A_k$$

For any arbitrary monotone increasing sequence $\{A_n\}$, the limit is

$$C = \overline{lin} A_n = li\sup\ A_k = \bigcap_{k=1}^{\infty} A_k \bigcup_{k=n}^{\infty} A_k$$

Monotone decreasing sequence: A sequence of sets $\{A_n\}$ is said to be monotone decreasing if $A_{n+1} \subseteq A_n$ for each A_n.

If $\{A_n\}$ n: 1, 2, ...) of events, is monotone decreasing (non - increasing) and for every n we have $A_n \supset A_{n+1}$, then the limit is the product of event $\{A_n\}$ i.e.

$A = \prod_{n\geq 1} A_n = \lim_{n\to\infty} A_n$ or

$$= \{w: w \text{ belonging to at least one of } A_n,\ A_{n-1} \dots\}$$

$$\bigcap_{k=1}^{n} A_k = A_n; \bigcap_{k=1}^{\infty} A_k = A \quad i.e.\ A_n \downarrow A$$

For any arbitrary monotone decreasing sequence $\{A_n\}$, the limit is

$$B_n = \inf_{k\geq n} A_k = \bigcap_{k=n}^{\infty} A_k$$

$$Limits: B = \underline{lim} A_n = lim\inf\ A_k = \bigcup_{k=1}^{\infty}\bigcap_{k=1}^{\infty} A_k$$

Note that

(i) $\underline{lin} A_n \subseteq \overline{lin} A_n$

(ii) The limit of $\{A_n\}$ is said to exist if $\underline{lim} A_n = \overline{\lim} A_n = A$,

(iii) If $\{A_n\}$ is not monotone and A exists then $A_n \longrightarrow A$ i.e. A_n converges to A.

(iv) Even if $\lim A_n$ does not exist, $\underline{lim} A_n$ and $\overline{\lim} A_n$ will always exist.

Example:

Consider the sequence $\{A_n\}$ where

$$A_n = \left\{w: 0 < w < b + {(-1)^n}/{n}\right\};\ (b > 1)$$

Does the series $\{A_n\}$ converge?

Solution

$$Let\ \ C_n = \begin{cases} \left[w: 0 < w < b + \frac{1}{n}\right]; if\ n\ is\ even, \\ \left(\left[w: 0 < w < b + \left({1}/{(n+1)}\right)\right]; if\ n\ is\ odd\right. \end{cases}$$

$$\overline{lim} A_n = \{w: 0 < w \leq b\}$$

Similarly,

$$B_n = \begin{cases} [w: 0 < w < b - ({1}/{n})]; if\ n\ is\ odd \\ \left(\left[w: 0 < w < b - \left({1}/{(n+1)}\right)\right]; if\ n\ is\ even\right. \end{cases}$$

$$\underline{lim} A_n = \{0 < w < b\}$$

Therefore, $\overline{lim}\ A_n \neq \underline{lim} A_n$

Hence, $\{A_n\}$ does not converge

Exercise:

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*.

Exercise:

Examine the following for convergence, if convergent, derive the limit;

(a) $A_{2^n} = (0, {}^1/_{2n}), A_{2^{n+1}} = \left[-1, {}^1/_{(2_{n+1})}\right]$

(b) $A_n = \left[\text{the set of rational in } \left(1 - {}^1/_{(n+1)}, 1 + {}^1/_n\right)\right]$

(c) $A_n = 2^{-1/n}, 2 + {}^2/_n), \; n \text{ is odd.}$

10.2 Obtaining Countable Class of Disjoint

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 $\bigcup_{i=1}^n \boldsymbol{A_i} = \sum_{i=1}^n \boldsymbol{B_i}$

Proof: By induction

$A_1 \cup A_2 = A_1 + A_1^C A_2$

$= B_1 + B_2 = \sum_{i=1}^2 B_i$ (say)

This is true for n = 2

Suppose it is true for all $n \leq m \geq 2$

Then $\bigcup_{i=1}^{m+1} A_i = (\bigcup_{i=1}^m A_i) \cup A_{m+1}$

$$= \left(\sum_{i=1}^m \boldsymbol{B}_i\right) \cup \boldsymbol{A}_{m+1}$$

$$= \sum_{i=1}^m \boldsymbol{B}_i + (\sum \boldsymbol{B}_i)\boldsymbol{A}_{m+1}$$

$$= \sum_{i=1}^m \boldsymbol{B}_i + \boldsymbol{B}_{m+1}$$

Where B_{m+1} and $\sum_{i=1}^m B_i$ are disjoint. The lemma holds for $n = m + 1$. So $B_i \subset A_i$ $\forall_i$

Corollary:

$$\bigcap_{i=1}^{\infty} A_i = A_1 + A_1^C A_2 + A_1^C A_2^C A_3 + \cdots$$

If

$w \in \bigcup_{i=1}^{\infty} A_i$, then w belongs to some A_i

Thus w may belong to A_1 or A_1^C or A_1^C or A_2 or $A_1^C A_2^C$ i.e. $w \in A_k$ for some k.

$\Rightarrow w \in \bigcup_{k=1}^{\infty} A_k$ establishes equivalent of both sides of (*)

10.4.1 Definition: Additive Set Function

A set function μ is said to be additive if $\forall A, B, s.t. = \varphi(A) + \varphi(B)$ and by finite induction.

$$\varphi\left(\bigcup_{n=1}^{m} A_n\right) = \sum_{n=1}^{m} \varphi(A_n), \; \forall i \neq j \; A_i \cap A_j = \phi$$

Note

- Once the value $+\infty, -\infty$ is not allowed i.e. $\phi \neq -\infty$
- If all the values of φ are finite, then φ is said to be finite i.e. $|\varphi| < \infty$
- If every set in a given class ℓ is countable union of such in ℓ and which is finite, then φ is said to be τ – finite.

10.4.2 Continuity of Additive Set of Function

An additive set function is said to be

(i) Continuous from below if

$$\varphi\left(\ell_{n \to n} E_n\right) = \ell_{n \to n} \varphi(E_n)$$

for every increasing sequence $\{E_n\} \uparrow$

(ii) Continuous from above if

$$\varphi\left(\ell_{n \to \infty} E_n\right) = \ell_{n \to \infty} \varphi(E_n) \forall$$

For every decreasing sequence $\{E_n\}\downarrow$

s.t. $\varphi(E_n)<\infty$ for some values $n=n_0$ and hence for all $n\geq n_0$.

A set function is said to be continuous if it is continuous from above and below.

Theorem

Let φ be finitely additive and continuous from below, then μ *is* $\tau-$*additive.*

Proof

Given a sequence of disjoint sets $\{E_n\}$, then

$$\varphi\left(\bigcup_{n=1}^{\infty}E_n\right)=\sum_{n=1}^{\infty}\varphi(E_n)$$

Let N be a finite number, since φ is finite additive, then

$$\varphi\left(\bigcup_{n=1}^{N}E_n\right)=\sum_{n=1}^{N}\varphi(E_n)$$

$$\sum_{n=1}^{N}E_n\subseteq\sum_{n=1}^{N+1}E_n$$

Let $S_N=\sum_{n=1}^{N}E_n$ be an increasing sequence

$$\varphi\left(\lim_{N\to\infty}S_N\right)=\lim_{N\to\infty}\varphi(S_n)$$

$$\Rightarrow\varphi\left(\lim_{N\to\infty}\sum_{\infty}^{N}E_n\right)=\lim_{n\to\infty}\varphi\left(\sum_{\infty}^{N}E_n\right)$$

$$=\varphi\left(\sum_{n=1}^{\infty}E_n\right)$$

$$=\varphi\left(\sum_{n=1}^{\infty}E_n\right)$$

By finite additively

$$\lim_{N\to\infty}\sum_{n=1}^{N}\varphi(E_n)=\sum_{n=1}^{\infty}\varphi(E_n)$$

$\therefore\varphi$ *is* $\tau-$*additive.*

Theorem

The probability function $P_{(\cdot)}$ is a set function that has $\tau-$additive property and hence is a measurement space.

Example: Let (Ω,F) be a measurable space on which a sequence of probability measure $P_1,P_2,\ldots P_n\ldots$ defines a set function.

Show that $P(E)\underline{\underline{d.f}}\sum_{n=1}^{\infty}\frac{1}{2^n}P_n(E)$ is an additive set function.

Solution

It is required to show that

(i) $0\leq P_{(\cdot)}\leq 1$

(ii) $P_{(\cdot)}$ is countably additive and is $\therefore$ a measure

(iii) Prove that $P(\Omega)=1$

(*i*) $P_{(E)}=\sum_{n=1}^{\infty}\frac{1}{2^n}P_n(E)=\frac{1}{2}P_{(E)}+\frac{1}{2^2}P_{2(E)}+\frac{1}{2^3}P_{3(E)}+\ldots$

but $\frac{1}{2}P_{1(E)}\geq 0,\ \frac{1}{2^2}P_{2(E)}\geq 0\ldots$

and $P_{(E)}\frac{1}{2}P_{(E)}+\frac{1}{2^2}P_{2(E)}+\ldots$

$$\therefore\sum_{n=1}^{\infty}\frac{1}{2^n}P_{n(E)}=S_\infty=\frac{a}{1-r}=\frac{\frac{1}{2}}{1-\frac{1}{2}}=1$$

$$\therefore 0\leq P_{(\cdot)}\leq 1$$

(ii) Let $\{A_i\}$ be a sequence of disjoint set, it is required to prove

$P\left(\bigcup A_i\right)=\sum P(A_i)$

from the L.H.S.

$$P\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{n=1}^{\infty} \frac{1}{2^n} P_n\left(\bigcup_{k=1}^{\infty} A_k\right)$$

$$= \sum_{n=1}^{\infty} \frac{1}{2^n} \sum_{k=1}^{\infty} P(A_k)$$

Since each of P_n is a measure and $0 \le P_{(\bullet)} \le 1$

$$\therefore = \sum_{n=1}^{\infty}\left[\frac{1}{2^n}\sum_{k=1}^{\infty} P(A_k)\right] = \sum_{k=1}^{\infty} P(A_k)\left[\sum_{n=1}^{\infty}\frac{1}{2^n}\right]$$

$$= \sum_{k=1}^{\infty} P(A_k) \bullet 1$$

$$= \sum_{k=1}^{\infty} P(A_k)$$

$\therefore P_{(\bullet)}$ is countably additive.

(iii) $P_{(\Omega)} = \sum_{n=1}^{\infty} \frac{1}{2^n} P_n(\Omega)$

$$= \sum_{k=1}^{\infty} \frac{1}{2^n}(1)$$

$$= \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \ldots$$

$$= 1 \qquad Sin\theta\ S_{\infty} = \frac{a}{1-r} = \frac{1/2}{1-1/2} = 1$$

10.5 The Halley-De-Moivre Theorem

Theorem: Let $\{E_i\}$ be a class of events each of which belongs to a τ − field $\Re$, and each of which may or may not occur. Then

$P\{at\ least\ one\ of\ the\ event\ E_i\ occur\}$

$$= P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{n} P(E_i) - \sum_{1\le i\le j\le n} P(E_i \cap E_j) + \sum_{1\le i<j<k\le n} PE_i \cap E_j \cap E_k + (-1)^{n-1} PP(\cap E_i)$$

Proof: (Using mathematical induction on n)

$for\ n = 1: \qquad P(E_1) = P(E_1)$

$for\ n = 2: \qquad P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$

The result is true for $n = 2$

$n = 3: \ P(E_1 \cup E_2 \cup E_3) = P(E_1) + P(E_2) + P(E_3) - P(E_1 \cap E_2) - P(E_1 \cap E_3) - P(E_2 \cap E_3) + P(E_1 \cap E_2 \cap E_3)$

Assuming it is true for n and also true for $n = m - 1$, we have

$$P(E_1 \cup E_2 \cup \ldots \cup E_{m+1}) = P\left(\bigcup_{i=1}^{m-1} E_i\right) = \sum_{i=1}^{m-1} P(E_i) - \sum_{1<i<j<m+1}^{m-1} P(E_i \cap E_j) + \sum_{1\le i<j<k\le m+1} (E_i \cap E_j \cap E_j)$$
$$+ (-1)^{m-2} P(E_1 \cap E_2 \cap \ldots E_{m-1})$$

$$n = m: P\left(\bigcup_{i=1}^{m} E_i\right) = P\left(\bigcup_{i=1}^{m-1} E_i \cup E_m\right) - \Sigma P(E_i \cap E_j \cap E_m) + \sum_{1\le i\le j\le k\le m-1} (E_i \cap E_j \cap E_k \cap E_m)$$
$$- \ldots + (-1)^{m-2} P(E_1 \cap E_2 \cap \ldots \cap E_{m-1} E_m)$$

Assuming it is true for $n = m$, we need to prove that the theorem is true for $n = m + 1$

$$P\left(\bigcup_{i=1}^{m+1} E_i\right) = P\left(\bigcup_{i=1}^{m+1} E_i \cup E_{m+1}\right)$$

$Let\ E = \bigcup_{i=1}^{m} E_i,\ then$

$$P\left(\bigcup_{i=1}^{m} E_i\right) = P(E \cup E_{m+1})$$

$$= \sum_{i=1}^{m+1} P(E_i) - \sum_{1\le i<j<m} P(E_i \cap E_j) + (-1)^m P\left(\bigcap_{i=1}^{m} E_i\right) + \sum_{1\le i<j<k\le m+1} P(E_i \cap E_j \cap E_k) + \ldots$$
$$+ (-1)^m P\left(\bigcap_{i=1}^{m+1} E_i\right)$$

This implies that the result is true for all positive integers n.

Example:

Let $E_1, E_2, \ldots, E_n$ be events which belongs to a τ − field $\Re$, show that the probability that exactly K events occurred out of n is given by

$$\frac{\sum_{r=0}^{n-k}(-1)^r\binom{n}{k}\binom{n-k}{r}}{\binom{n}{k+r}}S_{k+r}$$

$$=\sum(-1)^r\binom{n+k}{r}S_k$$

Where $S_k=\sum P(E_{i1}\cap E_{i2}\cap \ldots \cap E_{ik})$

From Halley-De Moivie theorem

$$P\left(\bigcup_{i=1}^{n}E_i\right)=S_1-S_2+S_3-\ldots+(-1)^{n-1}S_n$$

If $k=0$, no event occurred.

$$P(E_1\cup E_2\cup \ldots \cup E_n)^c=1-P(E_1\cup E_2\cup \ldots \cup E_n)$$
$$=1-S_1+S_2-S_3+\ldots-(-1)^r S_n$$
$$=P\{\text{non of the events occurred}\}$$

Example 2:

Suppose μ letter and μ corresponding envelopes are typed by a typist. Suppose further that the messenger, who is in a hurry to leave for the post office, randomly insert letters into envelopes, thinking erroneously that all the letters were identical. Each envelope contains one letter, which i.e. equally likely to be any one of the μ letters:

(i) Calculate the probability that at least one of the letters is inserted into its correspondence envelope.

(ii) Find the limit of this probability as $N\to\infty$

Solution

(i) Let E_i denote the event that the i^{th} letter and envelope match

$$\Rightarrow P\left(\bigcup_{i=1}^{N}E_i\right)=\sum_{i=1}^{N}P(E_i)-\sum_{1\le i<j\le N}P(E_i\cap E_j)+(-1)^N P\left(\bigcup_{i=1}^{N}E_i\right)=\sum_{i=1}^{N}\frac{1}{N}-\sum_{1\le i<j\le N}\frac{1}{N(N-1)}+$$

$$\sum_{1\le i<j\le N}\frac{1}{N(N-1)N-2}+\ldots+(-1)^{N-1}\frac{1}{N!}$$

$$\Rightarrow N\cdot\frac{1}{N}-{}^{N}C_2\left(\frac{1}{N(N-1)}\right)+{}^{N}C_3\left(\frac{1}{N(N-1)N-2)}\right)+\ldots+(-1)^{N-1}\frac{1}{N!}$$

$$=1-e^{-1}\left[\text{Since } e^{-x}=1-x+\frac{x^2}{2!}-\frac{x^3}{3!}+\frac{x^4}{4!}-\ldots\right]$$

(i.e. Prob (of 1 or 2 or 3 or … or N match) envelope

$$\text{i.e. } P\left(\bigcup_{i=1}^{N}E_i\right)=1-e^{-1}=0.63212$$
$$\cong 0.6$$

(ii) Taking limit as $N\to\infty$

$$P\left(\bigcup_{i=1}^{\infty}E_i\right)=1-\frac{1}{2!}-\frac{1}{3!}+\frac{1}{4!}+\ldots$$
$$=1-1+1-\frac{1}{2!}-\frac{1}{3!}+\frac{1}{4!}+\ldots$$

10.6.2 Countable Probability Space

Sometimes it is impossible for all the sample points in a Ω to be equally likely. Hence, each P_i is viewed as unit probability mass among the sample points following a certain rule or law. This law is sometimes referred to as probability distribution.

Example: For a geometric distribution

Suppose $\Omega=\{0, 1, 2, \ldots\}$ and

$$P_{(x)}=(1-\theta)\theta^x, \qquad x=0, 1, 2, \ldots, (0<\theta<1)$$

Then $P_x = P_{(x)} \geq 0, \sum_{x \in \Omega} P_{(x)} = 1$

If $A \subset \Omega$; then $P_{(A)} = \sum_{x \in A} P_{(x)}$

Poisson Distribution

Let $\Omega = \{0, 1, 2, ...\}$ *and* A is a class of all subsets of Ω. If P specifies that

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

Then $P_{(x)}$ is a Poisson distribution and X is a Poisson random variable

Definition 3:

A class of sets $\mathbb{A}$ is called a field or σ – field if and only if the following conditions hold true.

1. If $E_i \in \mathbb{A}$, then $\cup_{i=1}^{n} E_i \in \mathbb{A}$
2. If $E \in \mathbb{A}$, then $E' \in \mathbb{A}$

   From the above, it follows that
3. If $E_i \in \mathbb{A}$ implies $\cup_{i=1}^{n} E_i \in \mathbb{A}$

Example 1:

$\mathbb{A} = \{\Omega, \phi\}$is a field.

$\mathbb{B} = \{A, A', \Omega, \phi\}$is a field

$\mathbb{C} = \{A, \Omega, \phi\}$ *is not a field since* $A' \notin \mathbb{C}$

$\in = \{A, B, A'B', A' \cup B, A \cup B', A' \cup B', A \cup B, A'B, AB', A'B', AB, \Omega, \phi\}$ is a field.

The class of all subset of a given set Ω is a field.

Example 2:

(1) Let $\Omega = \{a, b, c, d\}$ and $\xi = \{\{a\}, \{b, c, d\}, \Omega, \phi\}$

i.e. ξ a field? Yes $P(a) = \frac{99}{100}$, $P(b, c, d) = \frac{1}{100}$

$I_{DS}(\Omega, \xi, P)$ is a probability space.

Yes, since ξ forms a field.

(2) From (1) let $\xi = \{\{a\}, \{b, c\}, \{d\}, \Omega, \phi\}$ and $P\{a\} = P\{d\} = 1/4$ $P\{b, c\} = \frac{1}{2}$

$P\{\Omega\} = 1, P\{\phi\} = 0.$

The triplet (Ω, ξ, P) is not a probability space since ξ do not form a field.

Exercise

(1) If $\Omega = \{w_1, w_2\}$and $\mathbb{F} = \{\Omega, \phi\}$

Show that $\mathbb{F}$ is a σ – field.

(2a) Is $\ell = \{E_1, E_2, \dots, \Omega\}$ a field.

(b) Hence or otherwise obtain all the elements of the σ – field of ℓ.

(3) Consider the sample space

$$\mathbb{F} = \{\phi\{w_1, w_2\}, \{w_3, w_4\}, \Omega\}$$

If $A = \{w_1, w_2\}, B = \{w_3, w_4\}$

Show that $\mathbb{F}$ is a field.

Exercise:

Let $E_1, E_2, \dots E_n$ denote an infinite sequence of events in σ – field $\mathbb{A}$.

Define

$$A_n = \bigcup_{m=n}^{\infty} E_m$$

$$B_n = \bigcap_{m=n}^{\infty} E_m$$

(a) Prove that $B_n \subseteq E_n A_n \forall_n$.

(b) Show that $\{A_n\}$ is monotone decreasing.

(c) show that $\{B_n\}$ is monotone increasing.

10.7 Sigma Field (σ – Field)

A non-empty class of sets which is closed under complementation and countable unions (or countable intersection) is called a field.

Note:

- A field containing an infinite number of sets may not be a σ – field.

Intersection of an arbitrary number of σ – fields is a σ – field.

10.8 Borel Field

This is a subset of the real line. Let ℓ be a class of all intervals of the term $(-\infty, x), x \in \mathbb{R}$ as subset of the real line $\mathbb{R}$. Also let $(\ell) = \mathfrak{P}$ be the minimal field generated by ℓ.

Then $\mathfrak{P}$ contains the intervals of the form $[x, \infty)$ (i.e. compliments of $(-\infty, x)$, it also contains the intervals.

$$(*)\begin{cases}(-\infty, a] = \cap\left(-\infty, a + \frac{1}{n}\right), \text{ by coutable intersection} \\ (a, \infty) = (-\infty, a]^{a} \text{ by complimentation} \\ (a, b) = (-\infty, b) \cap (a, \infty), a < b \\ (a, b], [a, b), etc \text{ for } a, b \in \mathbb{R}.\end{cases}$$

Lemma

Let ℓ_1 be the class of all intervals of the term $(\ell, b), (a > b) a, b \in \mathbb{R}$ but arbitrary.

Then $\sigma(\ell_1 = \mathfrak{P})$.

Proof: By (*) (overleaf) $a, b, \in \mathfrak{P}$ for all a, b. Hence, $\ell_1 \mathfrak{P}$.

By definition of minimal field, $\sigma(\ell_1) \subset \mathfrak{P}$.

To prove inclusion

Let $x \in (a, b)$ then.

$\cup_{n=1}^{\infty}(-n, x) \in \sigma(\ell_1), \forall_x$

$\Rightarrow (-\infty, x) \;\sigma(\ell_1) \forall_x$

$\Rightarrow \ell \subset \sigma(\ell_1)$ as defined in the last example.

If is also possible to prove that the Borel field is the minimal field containing any one of the following:

$\ell_2 = \{(-\infty, x], x \in \mathbb{R}\}$

$\ell_3 = \{(a, b], a < b, a, b \in \mathbb{R}\}$

$\ell_4 = \{[a, b], a < b, a, b \in \mathbb{R}\}$

$\ell_5 = \{[a, b), a < b, a, b \in \mathbb{R}\}$

$\ell_6 = \{[x, \infty), x \in \mathbb{R}\}$ etc.

10.8.1 Borel Set

Borel field and Borel sets play a very important role in the study of probability.

Monotone field: A field $\mathbb{A}$ is said to be a monotone field if it is closed under monotone operations, i.e. if $\lim A_n \in \mathbb{F}$ whenever $\{\mathbb{F}\}$ is a monotone sequence of set $\mathbb{F}$.

i.e. $A_n \in \mathbb{F}, A_n \uparrow A \Rightarrow A \in \mathbb{F}$

$A_n \in \mathbb{F}, A_n \downarrow A \Rightarrow A \in \mathbb{F}$

Theorem: A σ – field is a monotone field and conversely.

Proof: Let $\mathbb{A}$ be a σ – field and $A_n \in \mathbb{A}$. If $A_n \uparrow A$, then $A = \cup_n A_n$ is a countable union sets of $\mathbb{A}$. Hence $A \in \mathbb{A}$. Similarly, if $A_n \downarrow A, A = \cap_n A_n$ is a countable intersection of sets of $\mathbb{A}$.

$\therefore A \in \mathbb{A}$, hence, $\mathbb{A}$ is a monotone field.

Conversely, let $\mathbb{A}$ be a monotone field and let $A_1, A_2 \ldots$ be sets belonging to $\mathbb{A}$.

Then $\bigcup_{k=1}^{n} A_k$ and $\bigcap_{k=1}^{n} A_k$ belong to $\mathbb{A}$ since $\mathbb{A}$ 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 $\mathbb{A}$.

Thus $\mathbb{A}$ is a σ – field.

10.9 Random Variable in Measure Space

Let Ω be the sample space with sample points w. Interest is usually in the value $Y_{(w)}$ associated with w.

(a) **Point function:** function on the space Ω to a space Ω' assigns to each point $w \in \Omega$ a unique point in Ω' denoted by $X_{(w)}$. Thus $X_{(w)}$ is the image of the argument w under X. i.e. value of X at w $\underset{domain}{\Omega} \xrightarrow{X} \underset{range}{\Omega'}$

The set $\Omega'' = [X_{(w)}: w \in \Omega]$ which is a subset of Ω' is called the strict range of X.

If $\Omega' - \Omega'' \Rightarrow X$ is a mapping from Ω to Ω.

The symbol $f_{(x)}, X_{(w)}$, etc will be used to denote functions even though they denote values of functions.

Example 1:

Let $\Omega = [0, \pm1, \pm2, ...]$; $\Omega^{'} = [0, 1, 2, ...]$

$$\Omega^{''} = [0, 1, 4, 9 ...,]; \; if \; X_{(w)} = w^2$$

Thus X is a mapping of Ω into Ω'' and onto Ω'' ...

$$\underset{Domain}{\begin{bmatrix} 0 \\ \pm 1 \\ \pm 2 \\ \pm 3 \end{bmatrix}} \rightarrow \underset{Image}{\begin{bmatrix} 0 \\ 1 \\ 2 \\ 3 \end{bmatrix}}$$

$$w_1 = w_2 \Rightarrow X_{(w1)} = X_{(x2)} \text{one} - \text{to} - \text{one}$$

In this case $X_{(w1)} \neq X_{(w2)} \nRightarrow w_1 = w_2$

$X_{(w)} = w^2$ is not $1-1$ function

Since $w_1 = +2, w_2 = -2$ have the same image

$X_{(w1)} = 4 = X_{(w2)}$

If Ω is the real line $(-\infty < w < \infty)$ and $\Omega^{'} = (0 < w^{'} < \infty)$ then $X_{(w)} = \exp(w)$ is a 1-1 onto function from $\Omega \; to \; \Omega^{'}$ and 1-1 from $\Omega \; to \; \Omega$. If the range space is $\mathbb{R}$ or its subset, the function is said to be a .numerical' or 'real-valued' function.

(b) **Set Function**

If the arguments of a function are sets of a certain class, then we have a set function. Suppose $A \in \mathbb{A}$, we associate a value $\mu(A)$, (say) then μ is a set function. μ may represent entity such as weight, length, measure, etc.

The interval (a, b) may be associated with $b - a; f(a, b) \cup f(c, d) = (b - a) + (d - c)$, etc.

Two real valued function X and Y on Ω are said to be equal iff $X_{(w)} = Y(w) \; \forall \; w \in \Omega$.

$i.e. X = Y$

$Or \; X \underset{(<)}{>} Y$ iff. $X_{(w)} \underset{(<)}{>} Y_{(w)}$

If $X_{(w)} = C \; \forall \; w \in \Omega$, then X is degenerate.

(c) **Inverse Function**

The set of all points $w \in \Omega$ whose image under X is w^1 is called the inverse of $\{w^1\}$ denoted by $X^{-1}(\{w\})$. The

$$X^{-1}(\{w\}) = \{\{w \in \Omega\}: X_{(w)} = w^1\}$$

Note that for a point $w^1 \in \Omega^1 \nexists$ one or more than one points in Ω whose image under X is w^1. Let $B^1 \subset \Omega^1$. The set of all point for which $X_{(w)} \in B^1$ is called the inverse of B^1 under X denoted by $X^{-1}(B^1)$.

With every point function X, we associate a set function X^{-1} whose domain is a class $\mathfrak{P}$ of subsets of $\Omega^{'}$ and whose range is a class $\mathfrak{P}^{'}$ (say) of subset Ω. Then, X^{-1} is called the 'inverse function' (or mapping) of X.

$X(B) = [X_{(w)}: w \in B], BC\Omega$

$$X^{-1}(\mathfrak{P}^{'}) = [B^{'}(B^{'}): B^{'} \epsilon \mathfrak{P}^{'}]$$

$$X^{-1}(\Omega^{'}) = [w: X_{(w)} \epsilon \Omega^{'}] = \Omega$$

Lemma:

Inverse mapping preserves all set relations.

Proof: Let $B \subset C \subset \Omega^{'}$, then

$$X^{-1}(B) = [w: X_{(w)} \epsilon B] \subset [w: X_{(w)} \epsilon C] = X^{-1}(C)$$

(d) Indicator Function

A real valued function I_A defined on Ω as

$$I_A = \begin{cases} = 1 \text{ if } w \in A \\ = 0 \text{ if } w \in A^C \end{cases}$$

is called an indicator function (characteristic function by some authors). The strict range I_A is $I_A(\Omega) = \{I_A(w): w\epsilon\Omega\} = \{0, 1\}$. If B is a set function and $B \subset R$, the range space then $\quad I_A^{-1}(B) = \phi$, if B does not contain '0' or '1'

$= A$, if B contain '1' but not '0'

$= A^C$, if B contain '0' but not '1'

$= \Omega$, if B contain both '0' but not '1'

Thus $I_A^{-1}(B) = \{\phi, A, A^C, \Omega\} = \sigma(A)$

Clearly CI_A takes value C on A and 0 on A^C

Hence, $(CI_{(A)}^{-1})B = I_A^{-1}(A)$

Properties

(i) If $A \subset B \Leftrightarrow I_A \leq I_B$

$A \subset= B \Leftrightarrow I_A = I_{(B)}$

$I_A = I^2(A) = I^A(A), I_\Omega = 1$

(ii) $I(A^C) = 1 - I_{(A)}; I_{(B-A)} = I_{(B)} - I_A$

(iii) $I_{A.B} = I_A . I_B, I\bigcap_{i=1}^{n} A_1 = \prod_{i=1}^{n} A_i$

$= \min\{I_{A1}, \ldots, I_{An}\}$

(iv) $I_{(A\cup B)} = I_A + I_B - I_A . I_B = \max\{I_A . I_B\}$

$$I_{(A+B)} = I_A + I\bigcup_{i=1}^{n} A_i = \sum_{i=1} I_{Ai}\Sigma\Sigma I_{Ai}A_{Ai} + \sum_{i=1} I_{Ai}\Sigma\Sigma I_{Ai}A_{Aj} + A_k \ldots$$

Let $B_k \subset \Omega'$, then $w \in X^{-1}(\cap B_k) \Leftrightarrow X_{(w)} \in \cap B_k$

$\Leftrightarrow X_{(w)} \in B_k \forall_k$

$\Leftrightarrow w \in X^{-1}(B_k) \forall_k$

$\Leftrightarrow w \in \cap X^{-1}(B_k)$

Hence, $X^{-1}\left(\bigcap_k B_k\right) = \bigcap_k X^{-1}(B_k)$

Similarly,

$$X^{-1}\left(\bigcap_k B_k\right) = \bigcup_k X^{-1}(B_k)$$

(as above)

(iii) $w \in X^{-1}(B^C) \Leftrightarrow X_{(w)} \in B^C \Leftrightarrow X_{(w)} \notin B$

$\Leftrightarrow w \in X^{-1}(B)$

$\therefore X^{-1}(B^C) = (X^{-1}(B))^C$

Clearly $X^{-1}(\Omega') = [w: X_{(w)} \in \Omega'] = \Omega$

$X^{-1}(\phi) = \phi$. (prove)?

Corollary: (A)

If $\mathbb{A}$ is a field, a class of subset of Ω, then the class of $\mathbb{B}$ of all sets whose inverse images belong to $\mathbb{A}$ is also a σ – field.

Proof:

$B_1, B_2, \ldots, \in \mathbb{B} \Rightarrow X^{-1}_{(B_1)}, X^{-1}_{(B_2)}, \ldots, \in \mathbb{A}$

$\Rightarrow \bigcap_{i=1}^{\infty} X^{-1}_{(B_1)} \in \mathbb{A}$

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

$\Rightarrow \bigcap_i B_i \in \mathbb{B}$

Thus, $\mathbb{B}$ is closed under countable intersection. Hence, $\mathbb{B}$ is a σ – field.

10.9.1 $I_{(A)}$ as a Measurable Function

Since $I_A^{-1}(B) = \{\phi, A, A^C, \Omega\} = \sigma_{(A)}$

Let $\mathbb{A}$ be σ – field in Ω. If $A \in \mathbb{A}$, $\mathbb{A}^C$ also belong to $\mathbb{A}$ and $I_A^{-1}(B) \in \mathbb{A} \Rightarrow A \in \mathbb{A}$

Thus, I_A is $\mathbb{A}$ – measurable iff $A \in \mathbb{A}$.

10.9.2 Induced σ – field

Let X be a real valued function on Ω and $\mathbb{B}$ is a Borel field. By corollary A^* (overleaf) the class of sets $X^{-1}(B) = \{X^{-1}_{(B)}: B \in \mathbb{B}\}$ is called the σ – field induced by X.

10.9.3 Function of Function

If X is a function from Ω to Ω' and X' is a function from Ω' to Ω'', then the function $X'(X_{(w)})$ from Ω' to Ω'' denoted by $X'X$ or $X'(X)$ is said to be a function of function or composition of two functions X and X'.

Its inverse $(X'X)^{-1}$ is a function on the subset of Ω'' to the subset Ω such that for any $B \subset \Omega''$.

$(X^1X)^{-1}(B) = [w: X^1(X_{(w)}) \in B] = X^{-1}(X^{1-1}_{(B)})$

$= [w: X_{(w)} \in X^{1-1}(B) \therefore (X^1X)^{-1} = X^{-1}X^{1-1}$

10.9.4 Measurable Function

Let X be a real valued function on Ω to $\mathbb{R}$.

Definition 1: Let

$\mathbb{B}$ is a σ – field of subset of Ω. If $X^{-1}(B) \in \mathbb{A}$ for ll Borel sets $B \in \mathbb{B}$, then X is said to be function measurable with respect to A, i

Defination2: If Ω is also the real line $\mathbb{R}$ or its subset, and If X is measurable w.r.t. the Borel field $\mathbb{B}$ on the domain; then X is called a Borel function.

10.9.5 Random Variable (Economic Definition)

Suppose Ω be a sample space. Let $\mathbb{A}$aσ – fieldof events associated with a certain fixed experiment. Any real value $\mathbb{A}$ – measurable function defined on Ω is called a random variable. Thus, X is a random variable *iff* B^{-1}, the σ – field induced by X is contained in $\mathbb{A}$.

Suppose we define two non-negative functions

$$X^{+}_{(w)} = X_{(w)}, \text{ if } X_{(w)} \geq 0$$
$$= 0, \text{ if } X_{(w)} < 0$$

and

$$X^{-}_{(w)} = X_{(w)}, \text{ if } X_{(w)} < 0$$
$$= 0 \text{ if } X_{(w)} \geq 0$$

The above are respectively called the positive and negative parts of X. Then X^{+} *and* X^{-} are Borel function of X and will be random variable if X is a random variable

Note:

(1) These functions play an important role in the theory of integration of probability function.

(2) To show whether a function is a random variable, it is not necessary to determine whether $X^{-1}(B) \in \mathbb{A}$ for every B in $\mathfrak{B}$. It is sufficient to verify $X^{-1}(\ell) \subset \mathbb{A}$ where ℓ is any class of subsets of $\mathbb{R}$ given in sub interval on page 8.

Lemma: X is a random variableiff $X^{-1}(\ell) \subset \mathbb{A}$, where ℓ is any class of subsets of $\mathbb{R}$ which generates $\mathfrak{B}$.

Proof

Show that $X^{-1}(\ell) \subset \mathbb{A} \Leftrightarrow X^{-1}(\mathfrak{B}) \subset \mathbb{A}$. Since $\ell \subset \mathbb{A}$ and $X^{-1}(\mathfrak{B}) \subset \mathbb{A}$, $X^{-1}(\ell) \subset \mathbb{A}$ conversely.

Since $\mathbb{A}$ is a σ – field and

$$X^{-1}(\ell) \subset \mathbb{A} \Rightarrow \sigma\left(X^{-1}(\ell)\right) \subset \mathbb{A}$$
$$\Rightarrow X^{-1}\left(\sigma(\ell)\right) \subset \mathbb{A}$$
$$\Rightarrow X^{-1}(\mathfrak{B}) \subset \mathbb{A}$$

10.9.6 Vector Random Variable

Suppose $w \in \Omega$, the associate $Z_{(w)} = \left(X_{(w)}; Y_{(w)}\right)$ a point in the 2-dimensional Euclidian $\mathbb{R}^2$. The Z define a function from $\Omega$to $\mathbb{R}^2$. Consider the class of ℓ of all rectangles bounded by the lines $x_1 = a, x = b, y = c, y = d, a < b, c < d$ arbitrary. The minimal σ – field containing ℓ in Borel field $(\mathfrak{B}_2)$in $\mathbb{R}^2$.

Z is called a 2-dimentional random variable if $Z^{-1}(\mathfrak{B}_2) \subset \mathbb{A}$. $Z^{-1}(\mathfrak{B}_2)$ is a σ – field induced by Z.

Illustration

$$S_n = \sum_{i=1}^{\infty} X_i, \quad E(S_n) = n\lambda, \sigma(S_n) = \sqrt{n\lambda}$$

$$S_n = \frac{n\lambda}{\sqrt{n\lambda}}$$

The moment generating function of $Z_n$is given as

$$M = n^{(t)} = \frac{M_{Sn - n\lambda^{(t)}}}{\sqrt{n\lambda}} = e^{-t\sqrt{n\lambda}} MS_n\left({}^{t}/_{\sqrt{n\lambda}}\right)$$

$$e^{-t\sqrt{n\lambda}}\left(e^{-n\lambda}1 - {}^{t}/_{\sqrt{n\lambda}}\right)$$

$$\log M_Z(t) = -t\sqrt{n\lambda} - n\lambda\left(1 - e^{-t/\sqrt{n\lambda}}\right)$$

$$= -t\sqrt{n\lambda} - n\lambda\left(1 - \{1 + \frac{t}{\sqrt{n\lambda}} + \frac{t^2}{2n\lambda} + \frac{t^3}{3(n\lambda)^{3}/_{2}} + \cdots\}\right)$$

$$\frac{t^2}{2} + \frac{t^3}{n\lambda} + \cdots$$

$$\lim_{n\to\infty} \log M_{Z_n}(t) = \frac{t^2}{2} \Rightarrow M_Z(t) = e^{\frac{1}{2}t^2}$$
$$= mgf \ of \ N(0,1)$$

Problem

Suppose that S_n has the binomial distribution $b(n,p)$ show that $Z_n \xrightarrow{distribution} N(0,1)$

Theorem: Let $Y_n, n \geq 1$ be a sequence of real converging to Y_0. Then the sequence

$$Y_1, \frac{Y_1+Y_2}{2}, \frac{Y_1+Y_2+Y_3}{3}, \ldots \frac{Y_1+Y_2+\ldots,Y_n}{n}$$

Also converges to Y_0 However, the inverse is not true.

Proof:

Let > 0, we find n, s.t.$n \geq n \Rightarrow \left|\frac{1}{n}(Y_1 + \cdots Y_n) - Y_0\right| < \Sigma$

Since $Y_n \to Y_0 \exists \ no \ \ s.t. |Y_n - Y_0| < \mathcal{E}/_2 \forall_n > n_0$

Find $n_1 > n_0 \ s.t. \frac{1}{n_1}\sum_{i=1}^{n_0}|Y_i - Y_0| \leq \epsilon/_2$ for convinence

We claim that $n > n_1 \Rightarrow \left|\frac{Y_i+Y_n}{n} - '''_0\right| < \Sigma$

Then $\left|\frac{Y_i+Y_n}{n} - Y_0\right| = \left|\frac{(Y_1+Y_0)+\cdots+(Y_{n0}+Y_0)}{n}\right|$

$$\left|\frac{(Y_1+Y_0)+\cdots+(Y_{n0}+Y_0)}{n_1} + \frac{(Y_{n0+1}+Y_0)+\cdots+(Y_n+Y_0)}{n}\right|$$

$$\leq \frac{1}{n}\sum_{i=1}^{n0}|Y_i - Y_0| + \frac{1}{n_0+1}\sum_{i=n_0+1}^{n}|Y_i - Y_0|$$

$$\leq \epsilon/_2 + \frac{n-n_0}{n}\epsilon/_2$$

$$\leq \epsilon/_2 + \epsilon/_2$$

$$\leq \epsilon$$

Thus, $(Y_n - Y_0)/_n \xrightarrow{P} 0$ i.e. WLLAN holds.

Cov (XY)= E(XY) – E(X) E(Y)

CHAPTER 11
LIMIT THEOREMS AND LAW OF LARGE NUMBERS

11.1 Introduction

The law of large numbers is concerned with the conditions under which the average of a sequence or random variable converges (in some sense) to the expected average as the sample size increases.

11.2 Concept of Limit

Let x_0 be a point in some intervals of the real line $\Re$. Let f be a function which is defined at every point of f except possibly at x_0. The limit of the function as x approaches x_0 is L written as

$$\lim_{x\to x_0} f_{(x)} = L \ or \ f_{(x)} \to L \ as \ x \to x$$

If for any positive number Σ (no matter how small) there is some δ greater than zero such that

$$\left|f_{(x)} - L\right| < \varepsilon_0 \quad for \ all \ 0 < |x - x_0| < \delta$$

From the above definition, the number $\varepsilon > 0$ is first given, then we try to find a number $\delta > 0$ which satisfy the definition.

Example 1: Prove that $\lim_{x\to 6}(3x-4) = 14$

Solution: Given $\delta > 0$, find $\delta > 0$ [depending on Σ] s.t. $0 < |x-6| < \delta$, we have

$|f - 14| < \varepsilon$

$\Rightarrow |3x - 4 - 14| = |3(x-6)| = 3|x-6| < 3\delta$

Note that $|x - 6| < \delta$

$$\Rightarrow \Sigma = 3\delta \therefore \delta = \varepsilon/_3$$

Example 2:

Prove that $\underset{x\to 2}{\ell}\ \frac{x+1}{3x+4}=\frac{3}{10}$

Solution: Given $\varepsilon_0>0$, we went to final

$0<|x-2|<\delta$, *we have* $\left|f(x)-\frac{3}{10}\right|<\varepsilon_0$

$$\left|f(x)-\frac{3}{10}\right|=\frac{x+1}{3x+4}-\frac{3}{10}=\left|\frac{x+2}{10(3x+4)}\right|=\frac{x-2}{10(3x+4)}<\frac{\delta}{10(3x+4)}$$

If x is sufficiently near to 2 so that

$$3x+4>10,\ thus\ \frac{1}{10(3x+4)}\le\frac{1}{10}$$

Thus $\left|f(x)-\frac{3}{10}\right|<\left|\frac{\delta}{10(3x-4)}\right|<\frac{\delta}{100}$

$$\Rightarrow\delta=100\varepsilon$$

Theorem: Let f be the constant function defined by $f_{(x)}=C$ where C is a constant

$\lim_{x\to a} f(x)=C$

Proof: Given $\varepsilon>0$, find $\delta>0$ such that $0<|x-a|<\delta\Rightarrow|f_{(x)}-C|<\varepsilon$

The distribution of certain statistics of interest are too complicated to derive for differing sample sizes. In many cases, limiting distributions can be obtained as an approximation to the exact distribution, when the number of observation N is large. Thus, most important theoretical results in probability theory are limit theorems. Let consider some useful limit theorem.

11.3 Markov's Inequality

Markov's inequality can be used to obtain approximate probability of an event given that the mean of the probability distribution is known.

Theorem: If X is a random variable that takes only non-negative values then for any value $a>0$

$$P\{X\ge 0\}\le\frac{E(x)}{a}$$

Proof

Suppose X is continuous with density function

$$E(x)=\int_0^\infty xf(x)\,dx$$
$$=\int_0^a xf(x)\,dx+=\int_0^\infty xf(x)\,dx$$
$$\ge\int_0^\infty xf(x)\,dx$$
$$\ge\int_0^\infty af(x)\,dx$$
$$=a\int_0^\infty f(x)\,dx$$
$$\ge aP(X\ge a)$$
$$\therefore aP(X\ge a)\le E(X)\Rightarrow P(X\ge a)\le\frac{E(X)}{a}$$

The above is for a single variable X. Suppose we have a sequence of variable $\{X_n\}$, n = 1, 2, ...n, then we have the Markov's inequality for a sequence of $\{X_n\}$as

PX_n

11.4 Bienayme-Chebyshev's Inequality

Theorem: If X is a random variable with mean μ and variance σ^2, then for any value $\in>0$:

$$P\{|X-\mu|\ge K\}\le\frac{\sigma^2}{\sigma^2}$$

Proof:

Since $(X-\mu)^2$ is a non-negative random variable applying the Markov's inequality with $a=k^2$; we have $P\{(X-\mu)\ge K\}\le\frac{E(X-\mu)^2}{\sigma^2}$

but since $(X-\mu)^2=K^2$ *iff* $|X-|\ge k$, then the above $(*)$ is equivalent to

$$P\{|X-\mu|\geq k\}\leq\frac{E(X-\mu)^2}{k^2}=\frac{\sigma^2}{k^2}$$

The above inequalities are important in that they enable us:

(i) derive bounds on probability when only the mean, or mean and variance of the probability distribution are known.

(ii) Determine the convergence of a sequence of random variables or sum of independent probability distribution.

(iii) Prove important results in statistical theory

Example:

(i) Suppose it is known that the number of eggs sold in a poultry farm in a month is a random variable with mean 75 crates.

(ii) What is the probability that the sales for next month is greater than 100 crates.

(iii) IIf the variance of the sales for the month is >5, determine the bounds on the probability that sales in the coming month will be between 50 and 100 crates.

Solution: Let X be the number of eggs sold in a month

(i) by Markov's inequality

$$P(X>100)\leq\frac{75}{100}=\frac{3}{4}$$

(ii) by Chebyshev's inequality

$$P\{|X-75|\geq 25\}\leq\frac{25}{25^2}=\frac{1}{25}$$

$$\therefore P\{|X-75|<25\}\geq 1-\frac{1}{25}=\frac{24}{25}$$

So the probability of sales of eggs for this month is at least $\frac{24}{25}$

Definition: The sequence $\{X_n\}$ of a random variable is said to be *stochastically convergent* to zero if for every $\in>0$ the relation

$\lim_{n\to\infty} P\{|X_n|\geq\in\}=0$ is satisfied.

Note: It is only the probability of the event $|X_n|\geq\in$ tends to zero as $n\to\infty$, it does not follow that for every $\in>0$, we can find a finite n_0 such that for all $n>n_0$ the relations $|X_n|<\in$ will be satisfied.

Example 1:

Let $\{X_n\}$ be a sequence of binomial random variable with

$$P(X_n=x_n)=\binom{n}{r}P^n(1-P)^{n-r}$$

To show that $\lim_{n\to\infty} P\{|X_n|\geq\in\}=0$

Solution:

By Chebyshev's inequality we have

$$E(X_n)=n.p;\quad \text{Var}(X)=\sigma^2=\frac{\sqrt{npq}}{\sqrt{n}}$$

But Chebyshev's inequality states $\sigma=\frac{\sqrt{npq}}{\sqrt{n}}$

$$P\{|X-\mu|\geq\in\}\leq\frac{\sigma^2/n}{2}\quad\text{for}\in>0.$$

or $$P\{|X_n|>K\sigma\}\leq\frac{\sigma^2/n}{k^2\sigma^2}=\frac{1}{nK^2}$$

$$P\{|X_n|\geq k\sigma\}\leq\frac{1}{k^2}$$

Letting $\in=\sqrt{\frac{pq}{n}}$ we have

$$P\{|X_n|\geq\in\}\leq\frac{pq}{n\in}=$$

$$\Pr\{|X_n - np| > \in\} \leq \frac{pq}{\in^2}$$

Chebyshev's Inequality

This theorem is often used as a statistical tool in proving important results in statistics.

For example:

If Var (x) = 0 prove that

$PX = E(x)=1$

Proof by Chebyshev's inequality, for any ▯ ≥ **1.**

$$Pr\left\{|X - \mu| > \frac{1}{n}\right\} = 0$$

as $n \longrightarrow \infty$ and using the continuity property of probability and Chebyshev inequality.

$$\lim_{n\to\infty} P\left\{|1 - \mu| > \frac{1}{n}\right\} \leq \frac{\theta}{n(Y_n)^2}$$

$$= P\left\{\lim_{n\to\infty}\left\{|1 - \mu| > \frac{1}{n}\right\}\right\} = 0$$

$\Longrightarrow P[X \neq \mu] = 0$

This implies strong convergence (Strong convergence)

11.5 Convergence of Random Variables

Convergence in law denoted by

$X_n \xrightarrow{L} X$ if at every continually of X through distribution function F of

$$\text{Lim}_{n\to\infty} F_n(x) = F_{(x)}$$

Where $F_n(x)$ denotes the distribution function of X_n

11.6 Laws of Large Number

This refers to the weak or strong convergence of sample mean $\overline{X}_{▯} = \frac{1}{n}\sum_{i=1}^{n} X_i$ to a corresponding population mean (μ).

11.6.1 Weak Law of Large Number (WLLN)

Let $X_1, X_2, \ldots$ be a sequence of iid random variable's each having finite mean $E(X_i) = \mu$. Then for any $\in > 0$

$$P\left\{\left|\frac{X_1+X_2\ldots X_n}{n} - \mu\right| > \in\right\} \to 0$$

$$as\ n \to \infty$$

This implies that $\overline{X}_n \xrightarrow{P} \mu$

Proof: suppose the random variable has a finite variance σ^2

$$E\left[\frac{X_1+X_2\ldots X_n}{n}\right] = \mu, Var\left(\frac{X_1+X_2\ldots X_n}{n}\right) = \frac{\sigma^2}{n}$$

It follows from the Chebyshev's inequality that

$$P\left\{\left|\frac{X_1+X_2\ldots X_n}{n} - \mu\right| > \in\right\} \leq \frac{\sigma^2}{n\epsilon^2}$$

Thus, as $as\ n \to \infty$

$$\lim_{n\to\infty} P\left\{\left|\frac{X_1 + X_2 \ldots X_n}{n} - \mu\right| > \in\right\} = 0$$

$$\therefore \overline{X}_n \xrightarrow{P} \mu$$

Convergence Almost Surely

X_n is said to converge to X *almost surely, almost certainly* or *almost strongly* denoted by $X_n \xrightarrow{a.s} X$ if $X_n(w) \to X_{(w)}$ for al w except for those belonging to a null set N.

Thus $X_n \xrightarrow{a.s} X$ iff $X_n(w) \to X_{(w)} < \infty$

Thus, the set of convergence of $\{X_n\}$ has probability unity.

Lemma:

$X_n \xrightarrow{a.s} X$ iff as $n \to \infty$

$$P\left(\bigcup_{K=n}\left[w : |X_n - X| \geq \frac{1}{r}\right]\right) \to 0\ \forall_r \text{ an integer}$$

Proof:

Now $X_n(w) \to X(w)$, if for arbitrary $r > 1$, there exist some $n_0(w,r)$ s.t. $\forall K \geq n_0(w,r)$, $|X_n(w) - X(w)| < 1/r$

Moreover $X_n \xrightarrow{a.s} X$ implies that $P[X_n - X] = 0$

Using de-Morgan rules,

$$P\left(\bigcup_r \bigcap_n \bigcup_{k=n}^{\infty} \left[w : |X_n(w) - X(w)| \geq 1/r\right]\right) = 0$$

i.e. for each r

$$P\left(\bigcap_n \bigcup_{k=n}^{\infty} \left[w : |X_K(w) - X(w)| \geq 1/r\right]\right) = 0$$

Equivalently for each r

$$\lim_{n\to\infty} P\left(\bigcup_{k=n}^{\infty} \left[|X_K - X| \geq 1/r\right]\right) = 0$$

Replacing the above by the complimenting condition we have

$$P\left(\bigcap_{K=1}^{\infty} \left[|X_K - X| < 1/r\right]\right) \to 1$$

Note:

Lemma 1: A sequence of random variable's converges a.s. to a random variable *iff* the sequence converges mutually almost surely.

Lemma 2: If $X_n \xrightarrow{P} X$, then there exist a subsequence $\{X_{nk}\}$ of $\{X_n\}$ which converges a.s. to X.

Convergence in Distribution

If $F_n(x)$ is the d.f. of a random variable X_n and $F_{(x)}$ the d.f of random variable X, then $\{X_n\}$ is said to converge in *distribution* or *in law* or *weakly*. It is denoted as $X_n \xrightarrow{L} X$, $F_n \to F$ weakly or $F_n(x) \to F(x)$.

Theorem 1: If $X_n \xrightarrow{P} X$, then $F_n \to F(x)$, $x \in C(f)$

Theorem 2: If $X_n \xrightarrow{P} C$ implies that $F_n(x) \to 0$ for $x < C$, $F_n(x) \to 1$ for $x > C$ and conversely.

Proof: If $X_n \xrightarrow{P} C$, $F_n(x) \to F(x)$ where $F(x)$ is the d.f. of the degenerate random variable which takes a constant value C since

$$F_{(x)} = \begin{cases} 0, & x < C \\ 1, & x \geq X \end{cases}$$

Conversely, let $F_n(x) \to F(x)$ as defined above.

Then $\Pr[|X_n - C| \geq \in] = P[X_n \geq C + \in] + \Pr[X_n \leq C]$

$$= 1 - P[X_n < C + \in] + \Pr[X_n \leq C]$$
$$= 1 - F_n(C + \in - 0) + F_n(C - \in)$$

which tends to zero as $n \to \infty$.

Hence $X_n \xrightarrow{P} C$

Suppose X_n's are discrete random variables taking values $0, 1, 2, \ldots$ s.t. $P(X_n = i) = P_{in}$. If $P_{in} \to P_i$ as $n \to 0$ and X takes value i with probability P_i $(i = 0, 1, 2, \ldots)$ and hence $\sum P_i = 1$. then

$$F_n(x) = \sum_{i<x} P_{in} \to \sum P_i = F_{(x)}$$

So that X_n converges to X in distribution.

Example: Let X_n be a binomial random variable with index n and parameter P_n s.t. as $n \to \infty$, "$P_n \to \lambda > 0$ and finite. Then we can verify

$$P[X_n = K] \to \frac{e^{-\lambda}\lambda^k}{K!}; \quad (K = 0, 1, 2, \ldots)$$

The binomial random variable leads to Poisson random variable with parameter λ in distribution as $n \to \infty$.

Convergence in r^th Mean

A sequence of random variables is said to converge to X in the r^th mean, denoted by $X_n \xrightarrow{r} X$ if $E|X_n - X|^r \to 0$ as $n \to \infty$.

For $r = 2$, it is called the convergence quadratic mean or mean square.

For $r = 1$, it is called convergence in the first mean.

Lemma: If $X_n \xrightarrow{r} X \Rightarrow E|X_n|^r \to E|X|^r$

Proof: For $(r \le 1)$ put $(X_n - X)$ and X for X in tyheinequlaity

$$E|X_n|^r \le E|X_n - X|^r + E|X|^r$$

Interchanging X_n and X in inequality and combining, we have

$$E|X_n|^r - E|X|^r \le E|X_n - X|^r$$

Thus $X_n \xrightarrow{r} X \Rightarrow E|X_n|^r \Rightarrow E|X|^r$

If $X_n \to X$

Then $P\{|X_n - X| > \epsilon\} \le \dfrac{E|X_n - X|^2}{\epsilon^2}$

$\therefore X_n \xrightarrow{m.s.} X$ as $n \to \infty$.

Lemma

The binominal distribution $b(k; n, p)$ approaches the normal distribution as $n \leftarrow \infty$

i.e. $b(k; n, p) \sim \dfrac{n!}{\sqrt{2\pi\, npq}} e^{-x_k^2/2}$

Proof

Let $b(k; n, p) = \dfrac{n!}{K!(n-K)!} p^k q^{n-K}$ for large values of n.

The above represent $P(S_K = K)$ where S_k is a random variable which denote the number of successes in n. Bernoulli trials with probability success for success in each trial.

If we let $n \to \infty$ and keep P fixed then $P\{|S_n - np| > nE\} \to 0 \ \forall \epsilon > 0$ by the law of large number. Accordingly $|K - np|/n \to 0$.

Now let $x_k = \dfrac{k - np}{\sqrt{npq}}$ ....(**)

Since n and k are large, we can expressed the factorials in (*) above by means f the stirlings formula/ approximation as

$$m! = (2\pi)^{1/2} m^{m+1/2} e^{-m} e^{\theta(m)}$$

$$b(k, m, p) = \frac{(2\pi)^{1/2} n^{n+1/2} e^{-n} e^{\theta(n)} p^k q^{n-k}}{(2\pi)^{1/2} k^{k+1/2} e^{-n} e^{\theta(k)} (2\pi)^{1/2} (n-k)^{n-k+1/2} e^{-(n-k)} e^{\theta(n-k)}}$$

$$= \frac{1}{(2\pi)^{1/2}} \cdot \frac{n^{n+1/2}}{(2\pi)^{1/2} k^{k+1/2} (n-k)^{n-k+1/2}} p^k q^{n-k} e^{\theta}$$

$$= \frac{1}{(2\pi)^{1/2}} \sqrt{\frac{n}{k(n-k)}} \left(\frac{np}{k}\right)^k \left(\frac{nq}{n-k}\right)^{n-k} e^{\theta}$$

Where $\theta = \theta_{(n)} - \theta_{(k)} - \theta(n-k)$

Using $|\theta| \le \dfrac{1}{12}\left(\dfrac{1}{n} + \dfrac{1}{k} + \dfrac{1}{n-k}\right)$

Substituting for k from (**) the above can be rewritten in the form

$$|\theta| \le \frac{1}{12n}\left(1 + \frac{1}{P} \cdot \frac{1}{1 + x_k\sqrt{q/np}} + \frac{1}{q} \cdot \frac{1}{1 - x_k\sqrt{q/np}}\right)$$

If we assume that $x_k \sqrt{n} \to 0$ as $n \to \infty$, then $\theta \to 0$ and $e^{\theta} \to 1$.

The factor $\sqrt{\dfrac{1}{k(n-k)}}$ can be replaced by

$$\frac{1}{\sqrt{npq}} \cdot \frac{1}{\left(1 + x_n\sqrt{q/np}\right)\left(1 - x_k\sqrt{q/np}\right)}$$

which can be approximated by

$\dfrac{1}{\sqrt{npq}}$ for large n

To estimate the quantity $\left(\frac{np}{k}\right)^k\left(\frac{nq}{n-k}\right)^{n-k}$ ....(*)

Taking logarithm of the above gives

$$K\log\left({}^{np}\!/_{k}\right)+(n-K)\log\left({}^{nq}\!/_{n-k}\right)$$

Which can be rewritten in the form

$$-np\left(1+x_k\sqrt{\frac{q}{np}}\right)\log\left(1+x_k\sqrt{\frac{q}{np}}\right)$$

$$nq\left(1+x_k\sqrt{\frac{q}{np}}\right)\log\left(1+x_k\sqrt{\frac{p}{np}}\right)$$

Upon substitution for K.

Since $x_k n^{-1/2}$ is small, we cam expand the Logarithmic function in power series.

Using the Taylor expansion then a reminder

$$\log(1+x)=x-\frac{x^2}{2}+\frac{\in^3}{3!};\qquad \left(0<\left|\in^3\right|<x\right)$$

(* *) above becomes

$$-\frac{1}{2}x_k^2+C\Sigma^3 n^{-\frac{1}{2}}$$

Where C is a constant

If we assume $\Sigma^3/_{n^{1/2}}\to 0$ then can be approximated by

$$-\frac{1}{2}x_k^2$$

Hence $(*\text{ii})$ is asymptotic to $e^{-1/2 x_k^2}$

Gathering the estimates (i), (ii) and (iii) above, we have

$$b(k;n,p)\sim\frac{1}{\sqrt{2\pi}\,npq}e^{x_k^2/2}$$

The normal approximation to the binominal distribution.

11.6.2 Criterion for Convergence in Probability

The following lemma gives the necessary and sufficient condition for convergence in probability.

Lemma

$X_n \xrightarrow{P} 0$ iff $E\left(\frac{|X_n|}{|t||X_n|}\right)\to 0$ as n ∀

$$X_n \xrightarrow{P} 0 \text{ iff } E\left(\frac{|X_n|}{|t|X_n|}\right)\to 0 \text{ as } n\to\infty.$$

Proof

For any X, the r.v. $\frac{|X_n|}{|t|X_n|}$ is bounded by unity. Taking $g(x)=\frac{|X_n|}{|t|X_n|}\to 1$ for $\in>0$

$$E\left(\frac{|X_n|}{1+|N_n|}\right)-\frac{\in}{1+\in}\le P\left[|X_n|\ge\in\right]\le E\left(\frac{|X_n|}{1+|X_n|}\right)\Big/\frac{\in}{1+\in}$$

From RHS $E\left(\frac{|X_n|}{1+|X_n|}\right)\to 0\Rightarrow P\left[|X_n|\ge\in\right]\to 0$

From LHS $P\left[|X_n|\ge\in\right]\to 0\Rightarrow E\left(\frac{|X_n|}{1+|X_n|}\right)\to 0$

But $\frac{|X_n|}{1+|X_n|}$ is a non-negative r.v. so

$$\therefore E\left(\frac{|X_n|}{1+|X_n|}\right)\ge 0$$

Theorem: If $f_{(x)}$ is a continuous real valued function and $X_n \xrightarrow{P} X$, then $f_{(x)} \xrightarrow{P} f_{(x)}$

11.6.3 De Moivre-Laplace Limit Theorem

If S_n is the number of occurrence of an event in n independent Bernoulli trial, with probability P for success in each trial, then

$$P\left\{a \le \frac{S_n - np}{\sqrt{npq}} \le C\right\} \sim \frac{1}{\sqrt{2\pi}} \int_a^c e^{-x^2/2} dx$$

If a an b vary so that $a^3 n^{-1/2} \to 0$ and $C^2 n^{-1/2} \to 0$ as $n \to \infty$.

11.6.4 The Weak Law of Large Number

Let $X_1, X_2, \ldots X_n$ be a sequence of independent and identically distributed random variables each having mean $E(X_i) = \mu$ and finite variance σ^2. Then for any $\in > 0$

$$P\{|\overline{X} - \mu| > \in\} \to 0, \qquad as\ n \to \infty$$

Proof

It follows that

$$E[\overline{X}] = \mu \text{ and } \operatorname{Var}(\overline{X}) = \frac{\sigma^2}{n}$$

From Chebyshev's inequality we have

$$P\{|\overline{X} - \mu| \ge \in\} \le \frac{\sigma^2}{n \in^2}$$

$$\therefore \lim_{n\to\infty} P\{|\overline{X} - \mu| \ge \in\} = 0$$

This theorem was first proved by Jacob Bernowlli

11.6.5 Bernoulli's Law of Large Number

Let $\{Y_n\}$ be a sequence at random variable with pdf

$$P\left(Y_n = \frac{r}{n}\right) = \binom{n}{r} P^r (1-P)^{n-r} \text{ for } 0 < P < 1 \text{ and } r = 0, 1, 2, \ldots$$

Further let $X_n = Y_n - P$ sequence of random variables $\{X_n\}$ is stochastically convergent to 0 for any $\in > 0$. i.e. $\lim_{n\to\infty} P(|X_n| > \in) = 0$

Proof:

We have $E(X_n) = 0$

$$\sigma^2 = \frac{P(1-P)}{n}; \quad \sigma = \sqrt{\frac{P(1-P)}{n}}$$

Now using the Chebyshev's inequality

$$\therefore P\{|X_n| > \in\} \le \frac{\sigma^2}{\in^2}$$

$$= P\{|X_n| > \in\} \le \frac{P(1-P)}{n \in^2}.$$ It follows that $\lim_{n\to\infty} P(|X_n| > \in) = 0$ as $n \to \infty$.

11.6.6 Strong Law of Large Number (SLLN)

This refers to the strong convergence of the sample mean to the population mean.

i.e. $\overline{X}_n \xrightarrow{a.s.} \mu \Rightarrow E(X_i) = \mu$

i.e. $\lim_{n\to\infty} P\{sup\ |\overline{X}_n - \mu| > \in\} = 0$

Or

$P \lim_{n\to\infty} \{\overline{X}_n = \mu\} = 1$

Note that SLLN holds iff the population mean exist.

Theorem: Let $X_1, X_2, \ldots, X_n$ be a sequence of independent and identically distributed random variables each having a finite mean $\mu = E(X_i)$. Then with probability 1

$$\frac{X_1 + X_2 + \ldots + X_n}{n} \to \mu \quad \text{as} \quad n \to \infty$$

Or $\quad \Pr\{\lim_{n\to\infty} (X_1 + X_2 + \ldots + X_n)/n = \mu\} = 1$

Theorem: Let $\{X_k\}, k = 1, 2, \ldots$ be an arbitrary sequence of random variables with various σ_k^2 and first moment M_k. If the Markov's condition (i.e. $\lim_{n\to\infty} \sigma_k^2 = 0$) is satisfied then the sequence $\{X_k - M_k\}$ is stochastically convergent to zero.

Proof

Suppose X_k are pairwise untouched (i.e. independent). Consider the r[th] variable

$$Y_m = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

We have

$$E(Y_n) = \frac{1}{n}\sum_{k=1}^{n} M_k$$

Such X_k are pairwise uncorrelated, we have

$$\sigma^2(Y_n) = \frac{1}{f^2}\sum_{k=1}^{n} \sigma_k^2$$

If

$$\text{If} \lim_{n\to\infty} \frac{1}{n^2}\sum_{k=1}^{n} \sigma_k^2 = 0$$

Then by Chebyehev's inequality (theorem) it follows that

$$\lim_{n\to\infty} P[|Y_n - E(Y_n)| \geq \in] = 0$$

Thus, the sequence $\{X_k - M_k\}$ is stochastically convergent to zero.

CHAPTER 12

PRINCIPLES OF CONVERGENCE AND CENTRAL LIMIT THEOREM

12.1 Introduction

The Central limit theorem is concerned with determining conditions under which the sum of a large number of random variables has a probability distribution that is approximately normal.

12.2 Convergence of Random Variable

A sequence of random variables $\{X_n\}$ is said to be converge to a random variable X if $\{X_n(w)\}$ converges to $X(w) < \infty \; as\, n \to \infty$ for all $w \in \Omega$. Thus $\{X_n\}$ is said to converge to X everywhere.

If $X_n(w)$ converges to $X(w)$ only for $w \sum \Omega \not\vdash w \in A$, then C is called the set of convergence of X_n. If $C \in A$, then $\ell im\ X_n$ is a random variable clearly, C is the set of all $w \in \Omega$, at which whatever be $\varepsilon > 0$, $|X_n(w) - X(w)| < \varepsilon$ for all n greater than $n = N_0(w)$ sufficiently large symbolically for $= n + m,\ m \geq 1$

$$C = [w : X_n(w) \to X(w)]$$

$$= \bigcap_{\varepsilon>0}\bigcup_{n}\bigcap_{m}\left[w : |X_{n+m}(w) - X_{(w)}| < \varepsilon\right]$$

Equivalently, replacing "for every $\varepsilon > 0$ by for every $\frac{1}{k}$ $k = 1, 2, \ldots$ e

$$C = \bigcap_{k}\bigcup_{n}\bigcap_{m}\left[w : |X_{n+m}(w) - X_{(w)}| < \frac{1}{k}\right]$$

Since C is obtained from countable operation on measurable set, C is measure from $C \in A$.

Now $|f(x) - C| = |C - C| = 10] < \varepsilon$

Hence $|f(x) - C| < \varepsilon \; \forall \; x$

This theorem tells us that the limit of a constant δ that constant.

Remark:

If $f_{(x)}$ has the limit L *as* $x \to x$ then $f_{(x)}$ is said to converge to l.

OR

If C is the limit of $f_{(x)}$ as $x \to a$ then $f_{(x)}$ is said to converge to a constant C written as $f_{(x)} \to C$ *as* $x \to a$

Note that the constant L *or* C can also be a random variable.

Convergence in Probability

A sequence of random variables $\{X_n\}$ is said to converge to X in probability, denoted by $X_n \xrightarrow{P} X$, If for every $\varepsilon > 0$, *as* $n \to \infty$

$$P[|X_n - X \geq \varepsilon|] \to 0$$

equivalently, if for $\forall \varepsilon > 0$, *as* $n \leftarrow \infty$

$$P[|X_n - X < \varepsilon] \to 1.$$

Note:

This concept plays an important role in statistics, i.e. consistency of estimations, weak laws of large numbers.

Equivalentrandom variables: Two random variables X *and* X^1 are said to be equivalent if $X \to X^1$ a.s [almost surely]

Lemma: $X_n \xrightarrow{P} X$ *and* $X_n \xrightarrow{P} X^1 \Rightarrow X$ *and* X^1 are equivalent.

This lemma shows that a sequence of random variables cannot converge in probability to two essentially different random variables.

Lemma: $X_n \xrightarrow{P} 0$, *if* $\varepsilon|X_n|^r \to 0$

Replacing X_n *by* $(X_n - X)$ we have:

$$X_n \to X < \infty \quad iff \quad X_n - X \xrightarrow{P} 0$$

$$\therefore \Sigma|X_n - X|^r \to 0 \ \ implies \ \ X_n \xrightarrow{P} 1$$

This lemma provides us with sufficient evidence/condition for the convergence in probability.

The proof of the above follows from Markov's inequality.

$$i.e.\ P[|X_n| \geq \varepsilon] \leq \varepsilon^{|\varepsilon_{m}|^r} \big/ \varepsilon^r \to 0$$

$$as\ n \to \infty$$

Theorem: Let X be a k-dimensional r. vector and $g \geq 0$ be a valued (measurable) function defined on $\Re^k$, so that $g(x)$ is a vector random and let $C > 0$, then

$$P[g(x) \geq C] \leq \frac{E[g(x)]}{C}$$

Poof: Assume X is continuous with pdf Then

$$E[g(x)] = \int_{-\infty}^{\infty} \cdots \int_{-w}^{\infty} g(x_1, x_2, \ldots, x_k) f(x_1, x_2, \ldots, x_k) dk_1$$

$$= \int_A g(x_1, x_2, \ldots, x_k)\ f(x_1, \ldots, x_k) dx_1, \ldots, dx_k + \int_{A^c} (gx_1, \ldots, x_k)\ f(x_1, \ldots, x_k) dx_1, \ldots, dx_k$$

$$\text{Where } A = \left[\int_A g(x_1, x_2, \ldots x_k)' \in \Re^k_; g(x_1, \ldots, x_k) \geq 0\right.$$

$$E[g(x)] \geq \int_A g(x_1, \ldots, x_k) f(x_1, \ldots, x_k) dx_1, \ldots, dx_k$$

Using the result from Markov inequality

$$\geq C \int_A f(x_1, \ldots, x_k) dx_1, \ldots, dx_k$$

$$= C\hat{P}[g(x) \in A] = CP[g(x) \geq C]$$

$$\therefore P[g(x) \geq C] \leq \frac{E[g(x)]}{C}$$

Note:

If X is of discrete type, the proof is initial analogous.

Special Case I: (•)

Let X be a random variable and take $g(x) = |X - \mu|^r$, $\mu = E(X)$, $r > 0$. Then

$$P[|X-\mu| \geq C] = P\left[|X-\mu|^r \geq C^r \leq \frac{E|X-\mu|^r}{C^r}\right.$$

The above is known as Markov's Inequality

Special Case II:

If r *in* (*) above is replaced by 2 (*i.e.* $r=2$) we have

$$P[|X-\mu| \geq C] = P\left[|X-\mu|^2 \geq C^2\right] \leq \frac{E|X-\mu|^2}{C^2} \leq \frac{\sigma^2_{(x)}}{C^2} = \frac{\sigma^2}{C^2}$$

This is known as Tchebichev's Inequality.

In particular, if $C = K\sigma$ *then*

or $P[|X-\mu| \geq K\sigma] \leq \dfrac{1}{K^2}$

Remark

Let X be a random variable with mean μ and variance $\sigma^2 = 0$. Then the above gives

$$P[|X-\mu| \geq C] = 0 \text{ for every } C > 0$$

This implies that $P(X=\mu) = 1$

12.3 Cauchy-Schwarz Inequality

Let X and Y be two random variables with means μ_1, μ_2 and positive variance σ^1 *and* σ^2 respectively. Then

$$E^2[(X-\mu_1)(Y-\mu_2)] \leq \sigma_1^2\sigma_2^2$$

Or equivalently,

$$-\sigma_1^2\sigma_2^2 \leq E[(X-\mu_1)(Y-\mu_2)] \leq \sigma_1^2\sigma_2^2$$

and $E[(X-\mu_1)(Y-\mu_2)] = \sigma_1^2\sigma_2^2$

$$iff\ P\left[Y = \mu_2 + \frac{\sigma_2^2}{\sigma_1^2}(X=\mu_1)\right] = 1$$

Proof:

Set $X_1 = \dfrac{X-\mu_1}{\sigma_i^\theta}$. $Y_1 = \dfrac{Y-\mu_2}{\sigma_2}$

Then X_1 *amd* Y_1 are standardized variables hence

$$E^2(X_1Y_1) \leq 1$$

$$iff\ -1 \leq E(X_1Y_1) \leq 1$$

Which becomes (replacing X_1 *and* Y_1 by their standardized variable)

$$\frac{E^2[(X-\mu_1)(Y-\mu_2)]}{\sigma_1^2\sigma_2^2} \leq 1$$

$$\Rightarrow E^2[(X-\mu_1)(Y-\mu_2)] \leq \sigma_1^2\sigma_2^2$$

Note:

A more familiar term of Canchy-Schwarz inequality is

$$E^2(X\,Y) \leq E(X^2)E(Y^2)$$

12.4 Borel-Cantelli Lemma

In the study of sequences of events $A_1, A_2 ...$ with $P_k = P(A_K)$; a significant role is played by Borel-Cantelli Lemma.

i.e. (i) If the series $\sum P_{(k)}$ converges, then a finite number at events A_k occurs with probability 1.

(ii) If the events are (completely) independent, the series $\sum P_k$ diverges, then an infinite number of event A_k occur with probability 1.

Theorem:

Let $\{A_n\}$ $n = 1, 2, ...$ be a sequence of events and $P(A_n)$ denote the probability of the event A_n where $0 < P(A_n) < 1$. Then if

(i) If $\sum_{n=1}^{\infty} P(A_n) < \infty$ with probability one only a finite number of event A_n occur.

(ii) If the events $\{A_n\}\, n=1,2,\ldots$ are independent $\sum_{i=1}^{\infty} P(A_n)=\infty$ with probability one, an infinite number of event A_n occur.

Proof: Suppose $\sum_{i=1}^{\infty} P(A_n)<\infty$

Let A denote an infinite number of event A_n occur

$$A=\bigcap_{r=1}^{\infty}\bigcup_{n=r}^{\infty} A_n$$

$$\Rightarrow A \subset \bigcup_{n=r}^{\infty} A_n$$

Meaning that $P(A)\le P\left(\bigcup_{n=r}^{\infty} A_n\right)\le \sum_{n=r}^{\infty} P(A_n)$

as $n\to\infty$; $\sum_{n=r}^{\infty} P(A_n)\to 0$

hence, $P(A)=0 \Rightarrow \sum_{n=1}^{\infty} P(A_n)<\infty$

(ii) If A_n are independent and

$$\sum_{n=r}^{\infty} P(A_n)=\infty$$

then $\bar{A}=1-A$ *iff* at most finite number of events A_n occur.

hence, $\bar{A}=\bigcup_{r=1}^{\infty}\bigcap_{n=1}^{\infty}\bar{A}_n$

In view of the independence of A_n, we have

$$1-P(A)=P(\bar{A})=P\left(\bigcup_{r=1}^{\infty}\bigcap_{n=r}^{\infty}\bar{A}_n\right)$$

$$=\sum_{r=1}^{\infty} P\left(\bigcap_{n=1}^{\infty}\bar{A}_n\right)=\Sigma$$

$$=\sum_{r=1}^{\infty}\prod_{n=1}^{\infty}[1-PA_n]=\infty$$

i.e. The infinite product of the r.h.s of the above is divergent. Hence, $P(A)=1$ which shows that $\sum_{n=1}^{\infty} P(A_n)=\infty$

12.5 The Central Limit Theorem

Theorem: Let $X_1, X_2, \ldots$ be a sequence of independent and *identically* distributed random variable's each having mean μ and variance σ. Then the distribution of

$$\frac{X_1+X_1+\cdots+X_n-n\mu}{\sigma\sqrt{n}}$$

what intends to the standard normal as $n\to\infty$. That is

$$P\left[\frac{X_1+X_1+\cdots+X_n-n\mu}{\sigma\sqrt{n}}\le a\right]\to\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{a} e^{-x^2/2}dx$$

$$as \ \langle \to\infty$$

In simple language, the theorem states that a large number of independent random variables has a distribution that is approximate normal. It provides a simple method for computing approximate probability for sum of independent random variable's and explain the fact that many natural populations are normally distributed.

12.6 The Central Limit Theorem

Let $\{X_n, n\ge 1\}$ is a sequence at random variable Define

$S_n = X_1 + X_2 + \cdots X_n$, $\sigma(S_n)$ as the standard deviation of S_n *and* Z_n

$$=\frac{S_n-E(S_n)}{\sigma(S_n)}$$

Then Z_n converges in distribution to $N(0,1)$. This is an example of SLLN.

Example

Suppose X_s above are i.i.d each with the Poisson distribution with parameter λ. Show that the SLLN holds.

12.6.1 Central Limit Theorem for Independent Random Variables

Let $X_1, X_2, \ldots$ be a sequence of *independent* random variable's having means $\mu_i = E(X_i)$ and variance $\sigma_i^2 = Var\,(X_i)$. If (a) the X_i are uniformly bounded, that is for some M, $P\{|X_i| < M\} = 1$ for all i and

$(b) \sum_{i=1}^{\infty} \sigma_i^2 = \infty,\ then$

$$P\left\{\frac{\sum_{i=1}^{n}(X_i - \mu)}{\sqrt{\sum_{i=1}^{\infty}\sigma_i^2}} \le a\right\} \to \emptyset\,(a) \quad as \quad n \to \infty$$

Kroncker's Lemma (Proposition)

If $a_1, a_2, \dots$ are real number such that

$$\sum_{i=1}^{\infty}\frac{a_i}{i} < \infty, \text{converges, then } \lim_{n\to\infty}\sum_{n\to\infty}^{n}\frac{a_i}{n} = 0$$

12.7 Strong Law of Large Numbers for Independent Random Variables

Let $X_1, X_2, \dots$ be independent random variable's with $E(X_i) = 0,\ \mathrm{Var}\,(X_i) = \sigma_i^2 < \infty$.

If $\sum_{i=1}^{\infty}\frac{\sigma_i^2}{i^2} < 0$, then with probability i.

$$\frac{X_1 + X_1 + \cdots + X_n}{n} \to 0 \quad as \ n \to \infty$$

Note: It can be observed that Kolmogorov's inequality is a generalization of Chebyshev's inequality. If X has a mean μ and variance σ^2, then by letting $n = 1$ in Kolmogorov's inequality we obtain

$$P\{|X - \mu| > a\} \le \frac{\sigma^2}{a^2}\{\text{which is Chebyshev's inequlaity}\}$$

Where $X_1, X_2, \dots X_n$ are independent random variable's with $E(X_i) = 0,\ \mathrm{Var}\,(X_i) = \sigma_i^2$; then Chebyshev's inequality yields

$$P\{|X_1 + \cdots + X_n| > a\} \le \sum_{i=1}^{n}\frac{\sigma_i^2}{a_i^2}$$

Kolmogorov's inequality gives the same bound for the probability of larger set of variables. The theorem (Kolmogorov's) is used as a basis for the proof of the strong law of large numbers in the case where the random variable's are assumed to by independent but not necessarily identically distributed.

Proof: (Of strong law of large numbers for independent random variable's)

We will show that with probability I,

$$\sum_{i=1}^{\infty} X_i/_i$$

Let $X_1, X_2, \dots X_n$ be independent random variable's with $E(X_i) = 0,\ Var\,(X_i) = \sigma_i^2$ we have for some $a > 0$ by Kolmogoro's inequality

$$P\left\{\max_{j\le k<n}\left|\sum_{i=j}^{k} X_i/_i\right| \ge a\right\} \le \frac{\sum_{i=j}^{n} Var\left(X_i/_i\right)}{a_i^2} \le \frac{1}{a^a}\sum_{i=j}^{n}\sigma_i^2/_{i^2}$$

Since $\sum_{i=1}^{\infty}\sigma_i^2/_{i^2} < \infty$ implies that

$$\lim_{k\to\infty}\sum_{i=j}^{k}\sigma_i^2/_{i^2} = 0$$

By Knonecker's proposition, we have that

$$P\left\{\lim_{k\to\infty}\sum_{i=1}^{n} X_i/_n = 0\right\} = 1$$

$$Pr\left\{\underset{j\le k\le n}{\mathrm{Max}}\left|\sum_{i=j}^{k} X_i/_i\right| \ge a\right\} = 0$$

This implies that

$$\lim_{n\to\infty}\sum_{i=1}^{n} X_i/_n = 0 \text{ or equivalently that}$$

$$\lim_{n\to\infty}\sum_{i=1}^{n}\frac{(X_i - \text{i})}{n} = 0$$

and that

$$P\left\{\lim_{n\to\infty}\overline{X}_k = 0\right\} = 1$$

Definition 1:

Two sequences $h\{X_n(w)\}, \{Y_n(w)\}$ of random variables are said to be "*Trial Equivalent*" a finite number of terms,

i.e. for almost all $w \cdot \Omega, X_n(w) = Y_n(w)$

for all but a finite number at n

Lemma

If $\sum_{n=1}^{\infty} \Pr\{w : X_n(w) \neq Y_n(w)\} < \infty$ *i.e. trials*

Then $\Pr\{w : X_n(w) \neq Y_n(w) \text{ inf } \textit{initely after}\} = 0$

Proof

Let $E_0 = \{X_n(w) \neq Y_n(w)\} \Rightarrow \Pr\{E_n \textit{ occurs inf initely after}\} = 0$

Since $\sum P\{E_n\} \rightarrow \infty$ *in converges*

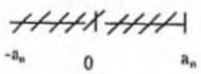
$\Pr\left\{\ell im \sup_{n\to\infty} E_n\right\} = 0$ *or* 1

Definition 2

A sequence $\{Y_n\}$ is said to be a *truncation* at the sequence $\{X_n\}$ at $\{a_n\}$ where $\{a_n\}$ is a sequence of positive real number if

$$Y_n = \begin{cases} X_n, \textit{ it } |X_n| < a_n \\ 0, \textit{ it } |X_n| > a_n \end{cases}$$

We know: $|X_n| < a_n \Rightarrow -a_n < X_n < a_n$



cut off the $\{a_n\}$ at $\{X_n\}$ in order to obtain $\{Y_n\}$

Lemma:

Let the sequence $\{Y_n\}$ be a truncation of sequence $\{X_n\}$ at sequence $\{a_n\}$ be finite i.e.

$Y_n = \sum \Pr\{|X_n| > a_n\} < \infty$

then

$$Y_n(w) \rightarrow X_n(w) \qquad as\ n \rightarrow \infty$$

$$\Rightarrow X_n(w) \rightarrow X(w) \qquad as\ n \rightarrow \infty$$

Proof

Given $Y_n = \sum_n \Pr\{|X_n| > a_n\} < \infty$

$\Rightarrow \Pr\{\text{En occur infinitely often}\} = 0$

$\Pr\{\lim_{n\to\infty} \sup E_n\} = 0$

$\Pr\left(\bigcap_{m=1}^{\infty} \bigcup_{m=r}^{\infty} E_n\right) = 0$

$\Rightarrow \Pr\left[\left(\bigcap_{m=1}^{\infty} \bigcup_{m=p}^{\infty} E_n\right)^c\right] = 1$ by de - Morgasloa

$\therefore \Pr\left[\bigcup_{m=1}^{\infty}\left(\bigcap_{n=m}^{\infty} E_n^C\right)\right] = 1$

$\Rightarrow \Pr\left\{\lim_{n\to 0} \inf E_n^c\right\} = 1$

$\Pr\{X_n(w) = Y_n(w) \forall \text{ except a finite number of n}\}$

$\Rightarrow \{X_n\}$ and $\{Y_n\}$ are Tail Equivalent.

Examples

Let $E_n = \{w : Y_n(w) \rightarrow X_n(w) \text{ as } n \rightarrow \infty\}$

$\Rightarrow P(E) = 1$

i.e. $A = \{X_n = Y_n \ \forall n \text{ except finitly many n}\}$

$\Rightarrow P(A) = 1$

If $w \in E$ amd $w \in A$

$w \in E \cap A = B = \{w : Y_n(w) \rightarrow X_n(w) \text{ as } n \rightarrow \infty \text{ amd } X_n = Y_n \ \forall \text{ except finitrly many n}\}$

$\Rightarrow w \in E \cap A \Rightarrow w \in B$

$(E \cap A) \leq B$

thus $P(E) = 1, P(A) = 1 \Rightarrow P(B) = 1$

where E and A are defined on the same ample space.

Hence, they are tail equivalent.

12.8 Bolzano-Cauchy Criterion for Convergence

Lemma: Let C be a fixed real number. If $|C| < K\epsilon$ for some $K > 0$ and every $\epsilon > 0$, it follows that $C = 0$.

Proof: suppose not. Then $C \neq 0$

Since ϵ is chosen arbitrary, put ϵ into let $\epsilon = \dfrac{|C|}{2k} > 0$, since K is given.

Then $|C| < \dfrac{|c|}{2} = K\epsilon$ which is clearly a contradiction except for $|C| = 0$

12.9 First Borel-Cantelli Lemma

Theorem: Let $\{E_n\}$ be a sequence of events each of which is a subset of Ω such that $E_n \in \mathbb{F}$ where $\mathbb{F}$ is a $\sigma - field$ of sub-events of Ω defined on the probability space (Ω, F, P), then $\sum_{n=1}^{\infty} P(E_{En}) < \infty \Rightarrow P\{\limsup_{n\to\infty} E_n\} = 0.$

[If $\{E_n\}$ is a sequence of events we are often interested in how many of the event occured]

OR

If $\{E_n\}$ is a sequence of events in a $\sigma - field$ $\mathbb{F}$, where $\Omega, \mathbb{F}, P$ is a probability space. Then $\sum_{n=1}^{\infty} P(\{E_{En}\}) < \infty \Rightarrow P\{\limsup_{n\to\infty} E_n\} = 0$

Proof:

Let $E = \limsup_{n\to\infty} E_n$

Then $E = \bigcap_{m=1}^{\infty} \bigcup_{m=n}^{\infty} E_n$; clearly $E \subseteq \bigcup_{n=m}^{\infty} E_n \ \forall m \in \mathbb{N}$

$$\therefore P(E) \leq P\left(\bigcup_{m=n}^{\infty} E_n\right)$$

$$\leq \sum_{n=m}^{\infty} P(E_n)$$

By Bolzano-Cauchy criterion for convergence

$$\sum_{n=1}^{\infty} P(E_n) < \infty$$

given any $\epsilon > 0, \exists$ an $N_0(\epsilon)$

Such that $\forall_n > N_0$.

$\left|\sum_{n=m}^{m+K} P(E_n)\right| < \epsilon$ and letting $K \to \infty$

$|P(E_n) + P(E_{n+1}) + \ldots + P(E_{n+K})| < \infty$

Now $\sum_{n=m}^{\infty} P(E_n) = \lim_{N\to\infty} \sum_{n=1}^{\infty} P(E_n) \forall_m > m$

$$\therefore P(E) \leq \left|\sum_{n=m}^{\infty} P(E_n)\right| \leq \epsilon \to 0$$

Where ϵ can be taken arbitrarily close to zero.

i.e. only finitely many E_n occurred.

1[st] BC-Lemma does not require independence of the event E_n.

12.10 Second Borel-Cantelli Lemma

Let $\{E_n\}$ be a sequence of independent events on the same probability space $(\Omega, \mathbb{F}, P)$ then if $E = \limsup_{n\to\infty} E_n$;

$$\sum_{n=1}^{\infty} P(E_n) = \infty$$

$\Rightarrow P(E_n$ occur infintely often$) = 1$

i.e. $P\{\limsup_{n\to\infty} E_n\} = 1$.

Proof:

Recall that $\lim_{\to\infty} \sup E_n = \bigcap_{m=1}^{\infty}\bigcup_{n=m}^{\infty} E_n$ s

$$P\left\{\left(\limsup_{n\to\infty}(E_n)\right)^C\right\} = P\left\{\liminf_{n\to\infty} E_n^C\right\} = P\left\{\lim_{n\to\infty}\left(\cap E_n^C\right)\right\}$$

For any $N > 0$ and every $K > N$

$$\text{Or } P\left\{\bigcap_{n=m}^{\infty}(\Omega - E_n)\right\} = P\left(\bigcap_{n=m}^{\infty} E_n^C\right) \le P\left(\bigcap_{n=m}^{K} E_n^C\right)$$

Since E_i's are independent $\Rightarrow E_i^C$'s are independent too.

$$\Rightarrow P\left(\bigcap_{n=N}^{K} E_n^C\right) = \prod_{n=N}^{K} P(E_n^C) = \prod_{n=N}^{K}[1 - P(E_n)]$$

Since $\prod_{n=N}^{K}(1 - P(E_n)) \le \prod_{n=N}^{K} e^{-P(E_n)} = e^{-\sum_{n=N}^{K} P(E_n)}$ by exponential property

$$as\ K \to \infty; \sum_{n=N}^{\infty} P(E_n) \to \infty \text{ i.e. } \sum_{n=N}^{\infty} P(E_n) = \infty$$

$$\Rightarrow \lim_{K\to\infty} e^{-\sum_{n=N}^{k} P(E_n)} \to 0$$

$$\Rightarrow 1 - P(E) = 0 \Rightarrow P(E) = 1.$$

Theorem:

$$\Pr\left\{\limsup_{n\to\infty} E_n\right\} = \begin{cases}0\\1\end{cases}; \text{ according as } \sum_{n=1}^{\infty} P(E_n) = \begin{cases}< \infty\\ \uparrow\infty\end{cases}$$

Whenever $E_1, E_2, ..., E_n, ...$ are independent

i.e. (i) If $\sum_{n=1}^{\infty} P(A_n) < \infty$, $P(\overline{\lim}\, A_n) = 0$

(ii) If $\sum_{n=1}^{\infty} P(A_n) = \infty$ and A_i's are independent

$$P(\overline{\lim}\,(A_n)) = 1.$$

Corollary to 2nd B.C. Lemma

If X_n's are independent and $X_n \to 0$ (a.s)

Then $\sum_{n=1}^{\infty} P[|X_n| \ge C] < \infty$

Whatever be $C > 0$, finite

Proof:

If X_n's are independent random variables $A_n = [|X_n|] \ge C$ are independent. Since $X_n \to 0$ a.s. *iff.*

$$P\left(\bigcup_{n=1}^{\infty}\{X_n| \ge C\right] < \infty \quad \text{as n} \to 0 \text{ and for any } C > 0$$

We have $P(\limsup_{n\to\infty} A_n) = 0$

Since $P\sum(A_n) < \infty$.

Note:

The converse of Borel-Cantelli lemma is not true if A_n's are not independent.

12.11 The Zero-One-Law

Let $A_1, A_2, ...,$ be events and let A be the smallest $\sigma - field$ containing each of these events. Suppose E is an event in A with the property that, for any integer $j_1, j_2, ..., j_k$ that events.

$$E \text{ and } Aj_1 \cap Aj_2, \cap ... \cap Aj_k \text{ are independent.}$$

Then $P(E)$ is either 0 or 1.

Exercise 1: If $P(A_n) = \frac{1}{2}\ \forall\, n \in \Omega;\ \sum_{n=1}^{\infty} PA_n = \infty$

Does $\{A_n\}$ converge??

$$\Rightarrow P(\limsup A_n) = 1$$

But $\sum_{n=1}^{\infty} P(A_n) = \infty \Rightarrow P(\liminf_{n\to n} A_n) = 0$

Thus $\lim_{n\to\infty}\sup A_n \neq \lim_{n\to\infty}\inf A_n$

Hence $\{A_n\}$ dos not converge.

Exercise 2: Let X have the uniform distribution $(X \sim \mu(0,1))$ consider the sequence of events $\{A_n\}$.

Where $A_n = \left\{w : X(w) < \frac{1}{n}\right\}$. Are $\{A_n\}$ independent.

Proof: $f_{(x)} = 1, \qquad 0 < x < 1$

$$P\left(x < \frac{1}{n}\right) = \int_0^{\frac{1}{n}} 1\,dx = [x]_0^{\frac{1}{n}} = \frac{1}{n}$$

Then $\sum_{n=1}^{\infty} P(A_n) = \sum_{n=1}^{\infty} \frac{1}{n} = +\infty$ Harmonic Series Diverges

But $E_n \supset A_{n+1} \supset A_{n+1} \supset \ldots$

$$\Rightarrow P\left\{\lim_{n\to\infty}\sup A_n\right\} = P\left(\bigcap_{m=1}^{\infty}\bigcup_{n=m}^{\infty} A_n\right) = P(\phi) = 0$$

$$= P\left(\bigcap_{m=1}^{\infty} A_n\right) ==$$

Clearly the above violates the 2nd B-C lemma as the sequence $\{A_n\}$ of events is overlapping and the therefore not independent.

Proof:

$$\text{Let } IA_j = \begin{cases} 0 & \text{on } \Omega - A_j \\ 1 & \text{on } A_j \end{cases}$$

Then $\int_E IA_{j1}, IA_{j2}, \ldots, IA_{jk}\, dP = P(Aj_1 \cap Aj_2 \cap \ldots \cap Aj_k \cap E)$

$$= P(Aj_1 \cap Aj_2 \cap \ldots \cap Aj_k) P(E)$$

By independence of E and $Aj_1 \cap Aj_2 \cap \ldots \cap Aj_k$

$$= \int_\Omega Aj_1 \cdot Aj_2 \ldots Aj_k dP \cdot P(E)$$

But $(\Omega, \mathbb{A}, P)$ is a complete probability space.

$$\therefore P(A \cap E) = P(A) \cdot P(E)$$

For all $A \in \mathrm{A}$, in particular $A = E$.

$$\therefore P(E) = \{P(E)\}^2$$

Then $P(E) = 0$ or 1

Completeness

A measure space $(\Omega, \mathbb{A}, P)$ is said to be *complete* if $\mathbb{A}$ contains *all* subsets of sets of measure zero.

Note

(i) A non-empty event with zero probability is negligible

(ii) Every subset of a negligible event have zero probability.

Lemma

(1) Given a probability space $(\Omega, \mathbb{A}, P)$ and a sequence $\{E_n, n = 1, 2, \ldots\}$ of event where $E_n \subseteq \Omega$ and $E_n \in \mathbb{T}\ \forall_n$

Prove

(i) $\lim_{n\to\infty}\inf E_n \subseteq \lim_{n\to\infty}\sup E_n$

(ii) $\lim_{n\to\infty}\inf P(E_n) \leq \lim_{n\to\infty}\inf P(E_n)$

(2) Let (Ω, F) be a measure space on which a sequence of probability measure is defined. The set function $P(E) \underline{\partial.f} \sum_{n=m}^{\infty} \frac{1}{2^n} P_n(E)$

(i) Show that $0 \leq P_{(\cdot)} \leq 1$.

(ii) $P(\cdot)$ is countably additive and is therefore a measure

(iii) Prove that $P_1(\Omega) = 1$.

Solution

(i) $P(E)=\frac{1}{2n}\,P_n(E)\geq 1$. Since $P_n(E)\geq 1$ and $\frac{1}{2^n}>0$ $\lim_{n\to\infty}(E_n)=0$

(ii) Show $\lim_{n\to\infty}\sum_{n=1}^{N}\frac{1}{2^n}P_n(E)=\sum_{n=1}^{N}\frac{1}{2^n}P_n(E)$

(iii) $P(\Omega)=\sum_{n=1}^{N}\frac{1}{2^n}P_n(\Omega)=\sum_{n=1}^{N}\frac{1}{2^n}(1)$

$$=\frac{1}{2}+\frac{1}{4}+\ldots$$

$$S_\infty=\frac{a}{1-r};\ a=\frac{1}{2},\ r=\frac{1}{2}$$

$$\frac{\frac{1}{2}}{1-\frac{1}{2}}=1$$

12.12 Limit Theorems for Sums of Independent R.V's

Lindeberg-Levy Theorem

Let $X_1, X_2, \ldots$ be a sequence of *i.i.d.r.v* each with mean θ and variance $\tau^2\ (0<\tau^2<\infty)$ Then $\frac{X_1+X_2+\ldots+X_n}{\tau\Gamma n}\to N(0,1)$

Proof:

Consider an array X_1

X_1, X_2

X_1, X_2, X_3

$\vdots \quad \vdots \quad \vdots$

Condition 1, 2 and 3 of Lindeberg's Theorem are satisfied.

We only need to verify condition 4.

Let $\in>0$, $B_n^2=n\tau^2$, then

$$\frac{1}{n\tau^2}\sum\int_{\{w:|X_i(w)|\geq\in\tau\Gamma n\}}X_n dP=\frac{n}{n\tau^2}\int_{\{w:|X_i(w)|\geq\in\tau\Gamma n\}}X_n dP$$

Since $X_{\cdot s}$ are i.i.d.

Now let $A_n\{w:|X_i(w)|\geq\in\tau\Gamma n\}$, *then* $A_n\downarrow\varphi$

and $\lim_{n\to\infty}P(A_n)=0$

and $\lim_{n\to\infty}\int_{A_n}X_n^2 dP=0$

This verifies the 4[th] L.T. so by Lindeberg's theorem

$$\frac{X_1+X_2+\ldots+X_n}{\tau\Gamma n}\to N(0,1)\text{ in distribution}$$

Lindeberg's Theorem (The Conditions of Lindeberg Theorem)

Let $X_1, X_2, \ldots X_{1k_1}$

$X_{21}, X_{22}, \ldots X_{2k_2}$

$\vdots$

$X_{n1}, X_{n2}, \ldots X_{nk_n}$

Be a rectangular array of random variable satisfying the following condition:

1. $\forall n\geq 1\ X_{n_1}, X_{n_2}, \ldots X_{nk_n}$ are independent
2. $E(X_{nk})=a_k^0$
3. $B_n^2=\tau_m^2+\tau_{n_2}^2+\ldots+\tau_{nk_2}^2$ *with* $B_n^2>0$
4. $\frac{1}{B_n^2}\sum_{k=0}^{k_n}\int|(x_k-a_k)|^2 dF_k(x)\to 0$ *as* $n\to\infty$

$\forall\in>0.$

Let $S_n=X_{n_1}+X_{n_2}+\ldots+X_{nk_n}$ *and* N a random variable with standard normal distribution

Then $\frac{S_n-a_k}{B_k}\to\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-t^2/2}dt$

The above statement is basic to the central limit theorem

However, if $E(X_{nk})=0$

The $\frac{S_n}{B_k}\to\frac{1}{\sqrt{2\pi}}\int e^{-t^2/2}dt$

and

Lyapunov's Theorem

Let X_i be a sequence of independent random variables. If a positive number δ can be found such that as $n \to \infty$; $i \geq 1$.

$$\frac{1}{B_n^{2+\delta}} \sum_{k=1}^{n} E|X_k - a_k|^{2+\delta} \to 0$$

$$P\left[\frac{1}{B_n}\sum_{k=1}^{n}(X_k - a_k) \leq x\right] \to \frac{1}{\Gamma 2\pi}\int_{-\infty}^{\infty} e^{-t^2/2}\,dt$$

Proof:

The random variables. define above satisfies 1, 2, and 3 of Lindeberg's theorem. It also satisfies the following:

(i) for some fixed $\delta > 0,\ E|X_{nk}|^{2+\delta} < \infty$

(ii) $\lim\limits_{n\to\infty} \frac{1}{B_n^{2+\delta}} \sum_{k=1}^{kn} E|K_{nk}|^{2+\delta} = 0$ then

$$\frac{S_n}{B_n} \to N(0, 1).$$

We now need to show that condition 4 of Lindeberg's theorem is satisfied.

Let $Var(X_i) = b$ for $i = 1, 2, \ldots, n,$ then $B_n = b\Gamma n$

Setting $E(X_i) = a_i$; condition 4 becomes

$$\left(1/nb^2\right)_n \int_{|x-a|>\in B_n} (x-a)^2\, dF_{(x)} = \int_{a+\in b\Gamma p}^{\infty} + \int_{-\infty}^{a+b\Gamma n} (x-a)^2\, dF_{(x)}.$$

which approaches zero since the $Var(X_i) < \infty$ and $b = 0$

Now we need to show that condition (2) implies condition (4)

This follows from the inequality

$$\frac{1}{B_n^2}\sum_{k=1}^{\infty} \int_{|x-ak|>\in B_n} (x-a)^2\, dF_k^{(x)} \leq \frac{1}{B_n^2(\in B_n)^2}\sum_{k=1}^{\infty} \int_{|x-a|>\in B_n} |X_{nk} - a_k|^{2+\delta}\, dF_k^{(x)}$$

$$\leq \frac{1}{\in\delta}\left(\sum_{k=1}^{\infty}\int |X_{nk} - a_k|^{2+\delta}\, dF_k^{(x)}\right) B_n^{-2-\delta}$$

$$\leq \frac{1}{\in \delta B_n^{2+\delta}}\left(\sum_{k=1}^{\infty}\int |X_{nk} - a_k|^{2+\delta}\, dF_k^{(x)}\right)$$

By hypothesis (ii) above

$$\lim_{n\to\infty} \frac{1}{\in\delta\, B_n^{2+\delta}}\left(\sum_{k=1}^{n} E|X_{nk}|^{2+\delta}\right) = 0$$

$$\therefore \frac{S_n}{B_n} \Leftrightarrow \frac{S_{n_k} - a_k}{B_n} \to N(0, 1).$$

CHAPTER 13
INTRODUCTION TO BROWNIAN MOTION

13.1 Brownian Motion (Weiner Process)

Brownian motion describes the macroscopic picture of a particle emerging in random system defined be a host of microscopic random effects in d-dimensional space, Peter & Yuval (2008). At any step on the microscopic level, the particle receives a displacement caused by other particles hitting if or by an external forces so that it's posterior at time-zero is So, its posterior at time n is given by $S_n = S_0 + \sum_{i=1}^{n} x_i$ where the displacements $X_1, X_2,...$ are assumed to be independent, identically distributed random variables with value in TR^d. The process $\{S_n, : n \geq 0\}$is a random walk, the displacements represent the microscopic inputs. Thus Brownian motion is a kind of stochastic process.

Any continuous time stochastic process $\{B(t) : t \geq 0\}$ describing the macroscopic feature of a random walk should have the following properties:

(i) For all time $0 \leq t_1 \leq t_2 \leq ... \leq t_n$.

$\beta(t_n) - \beta(t_{n-1}), \beta(t_{n-1}) - \beta_{m-2}),...,(\beta_{t2}) - \beta(t_1)$ are independent, we say the process has independent decrements.

(ii) the distribution of the increment $\beta_{(t+h)} - \beta_{(t)}$does not depend on t, we say the process has stationary increments.

(iii) the process $\{\beta_{(t)} : t \geq 0\}$ has almost surely continuous paths.

(iv) It follows from the CLT that these feature implies the existence of $\mu \in \Re^d$ and a matrix $\sum \in \Re^{dxd}$such that for every $t \geq 0$ *and* $h \geq 0$, the increment $\beta_{(t+h)} - \beta_{(t)}$ is multivariate normally distributed with mean $h\mu$ and covariance matrix $h \sum \sum^T$

Any process $\{X_t\}$ with the above feature seem be represented by

$$\bar{\beta}_{(t)} = \bar{\beta}_{(0)} + \mu_t + \sum \beta_{(t)}, \quad for\ t \geq 0$$

Where $\bar{\beta}_{(0)}$is the initial distribution

μ_t is the drift vector μ

$\sum \beta_{(t)}$ is the diffusion matrix

13.2 Brownian Process

If the drift vector is zero and the diffusion matrix is the identity, then $\{\beta_{(t)} : t \geq 0\}$ is termed/referred to as the Standard Brownian Motion. Hence, the macroscopic picture emerging from a random walk can be fully described by a *Standard Brownian Motion.*

13.3 Multinomial Distribution and Gaussian Process

The most important joint distribution is the multivariate normal (or the multinormal) distribution. It arises in many applications and has some properties that makes its manipulation very simple.

If A is any $(n\,x\,n)$symmetric matrix, consider the quadratic form

$$Q_{(x)} = \underline{X}^T A \underline{X}$$
$$= \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}\ x_i x_j$$

Where $\underline{x} \in \Re$ is the point which has coordination x_j and a column vector with transpose $\underline{x}^T$. If A is positive-definite, then $(2\pi)^{-\frac{1}{2}n}(def\ A)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\underline{x}^T A \underline{x}\right\}$ is a probability density on $\Re^n$.

Let $V = A^{-1}$, then V is also positive-definite and symmetric.

Definition 1:

A collection $(X_1, X_2,..., V_n)$which has the joint density.

$$(2\pi)^{-\frac{1}{2}n}(def.V)^{\frac{-1}{2}} \exp\left\{\frac{-1}{2}\underline{x}^T V^{-1} \underline{x}\right\}$$

is said to have the multinomial distribution $N(0, V)$

Definition 2: If $\mu_1, \mu_2, \ldots, \mu_n$ are finite real numbers then $\underline{X} = (X_1 + \mu_1\ X_2 + \mu_2, \ldots, X_n + \mu_n)$joint p.d.f

$(2\pi)^{-1/2}(\det.V)^{-1/2} \exp\left\{-\frac{1}{2}(X - \underline{\mu})^T V^{-1}(\underline{x} - \underline{\mu})\right\}$ and $\underline{x}$ is said to have the multinomial distribution $N(\underline{\mu}, V)$

Definition 3: Let τ be any set (usually a subset of the real axis). For every $t\ t \in \tau$, let $X_t^{(w)}$be a random variable defined on a probability space (Ω, A, P). Then the family $\{X(t,w): t \in \tau\}$ at random variables is called a *Stochastic process.*

Definition 4: Let $V(s,t) = E\{[X_{(t)} - \mu_t][X_{(s)} - \mu_s]\}$ be the autocovariance function at $X(t,w)$ for all relevant values of t *and* s and $\mu_t = E[X_{(t)}]$, $\mu_s = E[X_{(s)}]$.

Definition 5: A stochastic process $X(t,w)$with the property that all its finite-dimensional distribution are multinomial and $E(X_t) = 0$,

$$E(X_s, X_t) = V(s,t)$$

Where $V(\cdot,\cdot)$ is a positive-definite function on τ, is called a *Gaussian Process* with autocovariance function $V(\cdot,\cdot)$.

Remark:

- Two Gaussian processes with the same autocovariance function have the same finite-dimensional distribution
- The most important example of a Gaussian process is the Weiner (or Brownian motion) process.

Definition 6: A Gaussian process is said to be a wiener (Brownian) process

if (i) $\tau = (0, \infty)$

(ii) $X_{(0)} = 0$ *and*

(iii) $V_{(s,t)} = \min(s,t)$

13.4 Properties of a Brownian motion (B. M)

The following are the properties of a Brownian motion.

1. The Brownian motion is a Gaussian process with autocovariance function.

$$V(s,t) = E(X_s, X_t)$$
$$= \min(s,t)$$

2. The autocovariance function

$$V(s,t) = \min(s,t)$$

i.e. symmetric for $\tau = (0, \infty)$

3. Let $X_{(t)}$ be a B.M. process and define $X(s,t) = X_{(t)} - X_{(s)}$, the increment process on the interval (s,t). Then $X(s,t) \sim N(0, t-s)$

4. Given the Brownian motion process

$$B(t) = E\left(\left|B(t+h) - B_{(t)}\right|^4\right)$$
$$= 3h^2$$

5. The Brownian motion process is continuous everywhere but is nowhere differentiable.

Definition 7: Let τ be any set (usually infinite) and possibly uncountable) and let $V(\cdot,\cdot): TXT \to \Re$ be a function with the two properties.

(i) $V(t_1, t_2) = V(t_2, t_1)\ \forall t_1, t_2. \in T$

(ii) for any finite subset $\{t_1, t_2, \ldots, t_n\} \in \tau$ and any real numbers $Z_1, Z_2, \ldots, Z_n$ not all zero

$$\sum_{i=1}^{n}\sum_{j=1}^{n} V(t_i, t_j) Z_i, Z_j > 0$$

then $V(\cdot,\cdot)$ is called a *positive-definite* function on τ

Lemma:

$V(t_1, t_2) = \min(t_1, t_2)$ is a positive definite function on τ

Proof:

(i) *Clearly* $V(t_1,t_2)=V(t_2,t_1)$

(ii) *If* $0<t_1<t_2<\ldots<t_n$ *then*

$$\sum_{i=1}^{n}\sum_{j=1}^{n}V(t_i,t_j)Z_iZ_j=\sum_{i=1}^{n}\sum_{j=1}^{n}\min(t_i,t_j)Z_iZ_j$$

$$=\sum_{i=1}^{n}t_iZ_i^2+2\sum_{i=1}^{n}\sum_{j=1}^{n}t_iZ_iZ_j$$

$$(1<j)$$

Since $\min(t_i,t_j)=t_i$ *for* $i=j$ and

Since by symmetry, we may interchange i and j to cover cases in which $t_j<t_i$

$$=\sum_{i=1}^{n}t_i\left[Z_i^2+2Z_i\sum_{j=i+1}^{n}Z_j\right]$$

Expanding the square bracket gives

$$=Z_i^2+2Z_i\sum_{j=i+0}^{n}Z_j$$

$$=Z_i^2+2Z_i(Z_{i+1})+2Z_i(Z_{i+2})+\ldots+2Z_i(Z_{n-1})+2(Z_i,Z_n)$$

$$=\left(\sum_{j=i}^{n}Z_j\right)^2-\left(\sum_{j=i+1}^{n}Z_j\right)^2$$

$$\therefore\sum_i\sum_j V(t_i,t_j)Z_iZ_j=\sum_{i=1}^{n}t_i\left[\left(\sum_{j=i}^{n}Z_j\right)^2-\left(\sum_{j=i+1}^{n}Z_j\right)^2\right]$$

Writing the expression in full, we have

$$=t_1\{(Z_1+Z_2+\ldots+Z_n)^2-(Z_2+Z_3+\ldots+Z_n)^2\}$$
$$+t_2\{(Z_2+Z_3+\ldots+Z_n)^2-(Z_3+Z_4+\ldots+Z_n)^2\}$$
$$+\ldots+t_{n-1}\{(Z_{n-1}+Z_n)^2-Z_n^2\}+t_nZ_n^2$$

But $t_i>t_{i-1}$, by hypothesis, rewriting

$\therefore$ The above expression we have

$$t_1\left[\sum_{j=1}^{n}Z_j\right]^2+(t_2-t_1)\left[\sum_{j=2}^{n}Z_j\right]^2+(t_3-t_2)\left[\sum_{j=3}^{n}Z_j\right]^2+\ldots+(t_n-t_{n-1})\left[\sum_{j=1}^{n}Z_j\right]^2+(t_n-t_{n-1})Z_n^2$$

$$=\sum_{i=1}^{n}(t_i-t_{i-1})\left[\sum_{j=1}^{n}Z_j\right]^2$$

Where $t_0=0, t_i>t_{i-1}$

Clearly the last expression in the "curly bracket] is a positive number.

$\therefore V(s,t)=\min(s,t)$ is positive definite.

Theorem:

Let $X(t)$ be a Wiener process and let $X(s,t)=X_{(t)}-X_{(s)}$ denote the increment of the process on an interval (s,t), then

(i) $X(s,t)\sim N(0,t-s)$

(ii) *If* (s_1,t_1) *and* (s_2,t_2) *are disjoint intervals, then* $X(s_1,t_1)$ *and* $X(s_2,t_2)$ *are stochastically independent.*

Proof

(i) $X_{(t)}$ is Gaussian, therefore the joint distribution of $X_{(s)}$ *and* $X_{(t)}$ is multinormal and so $X_{(t)}-X_{(s)}=X_{(s,t)}(s,t)\sim N(0,\tau^2)$

where $\tau^2=Var(X(s,t))$

$$=E[X_{(t)}-X_{(s)}]^2$$
$$=E[X_{(t)}^2-2X_tX_s+X_{(s)}^2]$$
$$=V(t,t)-2V(s,t)+V(s,s)$$
$$=t-2s+s$$
$$=t-s$$

(iii) We assume without loss of generality that $s_1<t_1\le s_2<t_2$. Then $X(s_1,t_1)$ and $X(s_2,t_2)$ have multinormal joint distribution with covariance given by

$$E[X(s_1,t_1)(X(s_2,t_2)] = E[X(t_1) - X(s_1)][X(t_2) - Xs_2]$$
$$= E[X_{(t_1)}X_{(t_2)} - X_{(t_1)}X_{(s_2)} - X_{(s_1)}X_{(t_2)} + X_{(s_1)}X_{(s_2)}]$$
$$= V(t_1,t_2) - V(t_1,s_2) - V(s_1,s_2) + V(s_1,s_2)$$
$$= t_1 - t_1 - s_1 + s_1$$
$$= 0$$

Since $Cov(s,t) = 0$, there exist stochastic independence.

Exercise 1: For any real set of number $C_1, C_2, \ldots, C_n$ and real values random variable $\{X_i\}^n$, show that $\sum_{i=1}^{n}\sum_{j=1}^{n} C_i C_j E(X_i - \mu)(X_j - \mu)$ is positive semi-definite for $E(X_i) = \mu$.

Hint for Solution: Let $Y_i = X_i - \mu$, *then* $E(Y_i) = 0$ and $Var\left(\sum_{i=1}^{n} C_i Y_i\right) = E\left(\sum_{i=1}^{n} C_i Y_i\right)^2$

$$\therefore 0 \le E\left(\sum_{i=1}^{n} C_i Y_i\right)^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} C_i Y_i E(Y_i, Y_j)$$
$$= \sum_{i=1}^{n}\sum_{j=1}^{n} C_i Y_i V(i,j)$$

Example 2: Calculate the autocovariance function of the Gauss-Markov Process

$$Y_{(t)} = e^{-t}x(e^{2t}) \qquad \infty < t < \infty$$

Hint for Solution: Assuming $E(Y_i) = 0$;

$$V_y(s,t) = E[(e^{-s} X e^{2s})(e^{-t} X e^{2t})]$$
$$= e^{-(s+t)} E[(e^{2s})(e^{2t})]$$

$$V_y(s,t) \begin{cases} e^{-(s-t)} e^{2s} & for\ t > s \\ e^{-(s+t)} e^{2t} & for\ t < s \\ = 1 & for\ t = s \\ = e^{-|t-s|} & \forall s,t \end{cases}$$

Example 3:

Let $X_{(t)}$ be a Wiener process (B.M., then consider the B.M. process

$$Y_{(t)} = \begin{cases} 0, & t = 0 \\ tx\frac{1}{t}, & 0 < t < 1 \end{cases}$$

Show that $Y_{(t)}$ is a Brownian process.

Hint for Solution

Calculate $V(.,.)$ *of* $Y_{(t)}$

Note that for a B.M. process $\mu = 0$

$$E(Y_{(s)} - \mu)(Y_t - \mu) = E[(Y_{(s)})(Y_{(t)})]$$
$$= E\left[\left(t\left(\frac{1}{t}\right)\right)\right]\left[s \cdot \frac{1}{s}\right]$$
$$= st \min\left(\frac{1}{t}, \frac{1}{s}\right)$$
$$= \min(s,t)$$

Exercise 4:

Let $X_{(t)}$ be a Brownian motion process and let $Z_{(t)} = X_{(t)} - tX_{(1)}$; $\quad 0 < t < 1$

Find the $Cov(Z_{(s)}, Z_{(t)})$, *or* $V_Z(s,t)$

Hint for Solution

Note that $E(Z_{(t)}) = 0$ *and*

$$V_Z(s,t) = Cov(Z_{(s)} - Z_{(t)})$$

$$
\begin{aligned}
Cov\left(Z_{(s)}, Z_{(t)}\right) &= E\left[Z_{(s)}, Z_t\right] \\
&= E\left[Z_{(s)} - s X_{(1)}\left(X_{(t)} - t\left(X_{(1)}\right)\right)\right] \\
&= E\left[X_{(t)}X_{(s)} - sX_{(1)}X_{(t)} - X_{(s)}t X_{(1)} + tsX_{(s)}X_{(1)}\right] \\
&= V(s,t) - sV(1,t) - tV(1,s) + ts V(1,s) \\
&= Min(s,t) - S \min(1,t) - t \min(1,s) + ts \min(1,s) \\
&= s - st - ts + st; \qquad for\ s \le t \\
&= t - ts - st + st; \qquad for\ t \le s \\
&= Min(s,t) - st; \qquad \forall s,t
\end{aligned}
$$

PART THREE

CHAPTER 14
INTRODUCTION TO STOCHASTIC PROCESSES

14.1 Basic Concepts

Researchers in science, engineering, computing, business studies and economics quite often need to model real-world situations using stochastic models in order to understand, analyze, and make inferences about real-world random phenomena. Finding a model usually begins with fitting some existing simple stochastic process to the observed data to see if this process is an adequate approximation to the real-world situation.

Stochastic models are used in several fields of research. Some models used in the engineering sciences are models of traffic flow, queuing models, and reliability models, spatial and spatial-temporal models. In the computer sciences, the queuing theory issued in performance models to compare the performance of different computer systems.

Learning stochastic processes requires a good knowledge of the probability theory, advanced calculus, matrix algebra and a general level of mathematical maturity. Nowadays, however, less probability theory, calculus, matrix algebra and differential equations are taught in the undergraduate courses. This makes it a little bit difficult to teach stochastic processes to undergraduate students.

The mathematical techniques and the numerical computation used in stochastic models are not very simple. In an introductory course, the hope is to teach students a small number of stochastic models effectively to enable them to start thinking about the applications of stochastic processes in their area of research. These small numbers of stochastic models are the core topics to be taught in an introductory course on stochastic processes directed to researchers in the physical sciences, engineering, operational research and computing science. These researchers have a stronger background in mathematics and probability than researchers in the biological sciences.

Definition

A stochastic process is any process that evolves with time. A few examples are data on weather, stock market indices, air-pollution data, demographic data, and political tracking polls. These also have in common that successive observations are typically not independent, such collection of observations is called a stochastic process. Therefore, a stochastic process is a collection of random variables that take values in a set S, the state space. The collection is indexed by another set T, the index set.

The two most common index sets are the natural numbers $T = \{0, 1, 2, ...\}$, and the nonnegative real numbers which usually represent discrete time and continuous time, respectively. The first index set thus gives a sequence of random variables $(X_0, X_1, X_2, ...)$and the second, a collection of random variables $\{X_{(t)}, t \geq 0\}$, one random variable for each time t. In general, the index set does not have to describe time but is also commonly used to describe spatial location.

The state space can be finite countable infinite, or uncountable, depending on the application.

14.1.1 Applications of Stochastic Processes

The followings are some areas of Stochastic Processes:

(i) **Marketing**: To study customers or consumer buying behaviour and forecast.

(ii) **Finance**: To study the customer's account recordable behaviour and forecast.

(iii) **Personnel**: To study and determine the manpower requirement of an organization.

(iv) **Production:** To study and evaluate alternative maintenance policies, inventory, and so on, in industries.

(v) **Transport**: To effectively control flow and congestion in the transport industry.

14.2 Discrete-Time Markov Chains

You are playing a lotto, in each round betting N13 on odd. You start with N30 and after each round record your new fortune. Suppose that the first five rounds give the sequence loss, loss, win, win, win, which gives the sequence of fortunes, 9, 8, 9, 10, 11, and that you wish to find the distribution of your fortune after the next round, given this information. Your fortune will be 12 if you win which has probability $\frac{18}{38}$ and 10 if you lose, with probability $\frac{20}{38}$. One thing we realize is that this depends only on the fact that the current fortune is N11 and not the values prior to that. In general, if your fortunes in the first of rounds are the random variables $X_1, \ldots, X_n$, the conditional distribution of X_{n+1} given $X_1, \ldots, X_n$ depends only on X_n. This is a fundamental property and we state the following general condition.

Definition

Let $X_0, X_1, X_2, \ldots$ be a sequence of discrete random variables, taking values in some set S and that are such that

$$P(X_{n+1} = J|X_{1=i1}, \ldots, X_{n-1=in-1}, X_n = 1) = P(X_{n+1} = j|X_n = i)$$

For all $i, j, i_1, \ldots, i_{n-1}$ and all n, the sequence $\{X_n\}$ is then called a Markov chain. In general, the probability $P(X_{n+1} \; j|X_{n=i})$ depends on i, j and n. It is however, often the case that there is no dependence on n. We call such chains time-homogeneous and restrict or attention to these claims. Since the conditional probability in the definition thus depends only i and j, we use the notation $P_{ij} = P(X_{n+1} = j|X_n = i)$ i, j ∈ S and call these the transition probabilities of the Markov chain. This, if the chain is in state i, the probabilities p_{ij} describe how the chain chooses which state to jump to next. Obviously, the transition probabilities have to satisfy the following two criteria:

(i) $P_{ij} \geq 0$ (ii) $\sum P_{ij} = 1$, for $i \in S$

for all $i \in S$

14.2.1 The Transition Matrix

In changing from one state to another in any Markov system, a measure of probability is always attached. It is the collection of all such probabilistic measures which are arrange din rows and columns that ids called the transition matrix.

For a transformation matrix, a 2-level change of state will produce 2 by a matrix, a 3-level change produces 3 by 3 matrix and so on

14.3 Classification of General Stochastic Processes

The main elements of distinguishing stochastic process are in the nature of the *state space*, the index parameter T, and the dependence relations among the random variables X_t.

14.3.1 State Space S

This is the space in which the possible values of each X_t lie. In the case that S = (0, 1, 2, ...), we refer to the processes as integer valued, or alternatively as a discrete state process.

If S the real line $(-\infty, \infty)$, then we call X_t a real-valued stochastic process. If S is the each decision K space then X_t is said to be a k vector process.

Remarks:

The choice of state space is not uniquely specified by the physical situation being described, although usually one particular choice may sand out as most appreciate.

14.3.2 Index (Parameter) Set T

If $T = (0, 1, \ldots)$ then we state that X_t is a discrete time stochastic process. Often when T is discrete we should write X_n instead of X_t. If $T = [0, \infty]$, then X_t is called a continuous time process.

14.4 Classical Type of Stochastic Processes

We now describe (first briefly) then in details some of the classical types of stochastic processes characterized by different dependence relationships among X_t. Unless random stated, we take $T = [0, \infty]$ and assume the random variables X_t are real valued.

14.4.1 Process with Stationary Independent Increment

If the random variables $X_{t2}, -X_{t1}, X_{t3}-X_{t2},\ldots, X_{tn}-X_{tn-1}$ are independent for all choices of $t_1, t_2, \ldots, t_n$ satisfying $t_1 < t_2 < \ldots < t_n$ then we say that X_t is a process with independent increments.

If the index set contains a smallest index t_0, it is also assumed $X_{t0}, X_{t1}-X_{t0}\ X_{t2}-X_{t1},\ldots, X_{tn}-X_{tn-1}$ are independent. If the index set is divided, what is $T = (0, 1, \ldots)$, then a process with independent were reduces to a sequence of independent random variables $Z_0 = X_0, Z_i = X_i - X_{i-1}, i = 1, 2, 3, \ldots$in the sense that knowing the individual probabilities/distributions of $Z_0, Z_1, \ldots$ enable us to determine the joint distributions of any finite set of X_t, in fact that of $X_i = Z_0 + Z_1 + \ldots + Z_i$, $1 = 0, 1, 2, \ldots$

Remarks/Definition

1. If the distribution of the increments $X(t_1+h)-X(t_1)$ depends only on the length h of the interval and not on the time t, the process is said to have *stationary increment.*
2. For a process with stationary increments, the distribution of $X(t_2+h)-X(t_2)$, no matter what the values of t_1, t_2 and h.
3. We now state a theorem;

   If a process $\{X_t, t\in T\}$, where $T=[0,\infty]$ or $T=(0,1,2,\ldots)$ has stationary independent increments and has a finite mean, then it true that:

   $E(X_t) = M_0 + M_1$ where $M_0 = E(X_0)$ and $M_1 = E(X_1) - M_0$

   $\sigma_t^2 = \sigma_0^2 + \sigma_t^2$ where

   $\sigma_0^2 = E[(X_0 - M_0)]$ and $\sigma_1^2 = E[(X_1 - M_1)^2] - \sigma_0^2$

(4) Both the Brownian motion process and the Poisson process have stationary independent increments.

(5) We now prove remark 3(a)

$E(X_t) = E(X_0) + [E(X_t) - E(X_0)]$

Let $f(t) = E(X_t) - E(X_0)$

Then for any t and s we have

$$\begin{aligned} f(t+s) &= E\,[X_{t+s} - X_0] \\ &= E\,[X_{t+s} - X_S + X_S - X_0] \\ &= E[X_{t+S} - X_s] + E[X_S - X_0] \\ &= E[X_t - X_0] + E[X_S - X_0] \end{aligned}$$

Using the property of stationary increments

$$= f(t) - f(s)$$

The only solution to the functional equation $f(t+s = f(t) + f(s) = f(i)t$. differentiating with respect to t and independently with respect to s we have $f'(t+s) = f'(t) = f'(s)$.

Therefore for $S = 1$, we find $f'(t) = \text{constant} = f'(i) = c$. Integrating this elementary differential equation yields $f(t) = ct + d$.

But $f(0) = 2, f(0)$ *implies* $f(0) = 0$ and therefore $d = 0$.

Therefore expression $f(t) = f(1)t$ is

$$E(X_t) - M_0 = (E[X_1] - M_0)t$$

$$\Rightarrow E[X_t] = M_0 + M_i\,t \text{ as requires}.$$

14.5 Markov Processes

A Markov process is a process with the property that, given the value of X_t, the values of X_s, $S > t$, do not depend on the value if $X_u < t$; that is, the probability of any particular future behaviour of the process, when the present state is known exactly, is not altered by additional knowledge concerning the past behaviour, (provided our knowledge of the present state is precise).

Definition 1

In formal terms, a process is said to be Markov if

$\Pr\{a < X_t \le b | X_{t1} = X_1, X_{t2} = X_2, \ldots X_{tn} = X_n\}$

Whenever $t_1 < t_2 < \ldots t_n < t$ $P\{a < X_t \le b | X_m = X_n\}$

Definition 2

Let A be an interval of the real line. The function

$p\{x; s; t, A\} = \Pr\{X_t \in A | X_s = x\}$ is called the transition probability function t > s and is basic to the study of the structure of Markov process. We may express the condition (1) as follows:

$\Pr\{a < X_t \le b | X_{t1} = x_1, X_{t2} = x_2, \ldots X_m = n_n\} = P(x_n, t_n, t\ A)$ where $\{\sum | A < \sum \le b\}$

14.5.1 Martingales

Let $\langle X_t \rangle$ be a real-valued stochastic process with discrete or count parameter set. We say that $\langle X_t \rangle$ is a Martingale if, $E[|X_t|] < \infty$ for all t, and if for any $t_1 < t_2 < \ldots < t_{n-1}$ $E(X_{m-1} | X_{t1} = a_1, \ldots X_m = a_n) = c$ for all values of $a_1, a_2, \ldots a_n$.

14.5.2 Renewal Process

A renewal process is a sequence T_k of independent and identically distributed $(i.i.d)$ positive random variables, repressing the lifetimes of some "units". The first unit is placed at time zero; it falls at time T_1 and is immediately replaced a new unit which then fails at time $T_1 + T_2$ and so on, the motivating the name "renewal process". The time of the nth renewal is $S_n = T_1 + T_2 + \ldots + T_n$.

A renewal counting process N_t counts the number of renewals in the interval $[0, t]$. Formally $N_t = n$ for $S_n \le t < S_{n+1}$, $n = 0, 1, 2, \ldots$

Remarks: The Poisson process with parameter λ is a renewal counting process for which the unit lifetimes have exponential distribution with common parameter λ.
Other examples such as Poisson process, birth and death processes and Branching Process will be consider5ed in small details.

Practice Questions

1. Define and explain the concept of Stochastic Processes, and give three areas of application.
2. Explain the concept of a simple Markov Chain.
3. Define the following:
   (a) State Space (S)
   (b) Index Set (T)
   (c) Renewal Process

CHAPTER 15

GENERATING FUNCTIONS AND MARKOV CHAINS

15.1 Introduction

Generating function is of central importance in the handling of stochastic processes involving integral-valued random variables not only in theoretical analysis that also in practical appreciations. Stochastic process involves all process dealing with individuals' populations, which may be biological organisms, radioactive atoms, or telephone calls.

15.2 Basic Definitions and Tail Probabilities

Suppose we have a sequence of real numbers $a_0, a_1, \ldots$. Involving the doming variable x, we may define a formula

$A(x) = a_0x^0 + a_1x^1 + a_2x^2 + \cdots = \sum_{i=0}^{n} a_i x^j$

If the series converges in some real inference $-x_0 < x < x_0$, then the function $A(x)$ is known as the generating functions of the sequence $\{a_i\}$. We may also see this as a transformation that carries the sequence unit the function $A(x)$. If the sequence $\{a_i\}$ is bounded, then a comparison with the geometric series shows that $A(x)$ converge at least for $/x/x/$.

If the following restriction is introduced

$$a_i \geq 0, \sum_{i=0}^{n} a_i = 1$$

Then the corresponding function $A(x)$ is viewed as a probability-generating function. Specifically, consider the probability distribution given by

$P\{x = i\} = p_i$

Where x is an integral valued random variable assuming the values 0,1,2 ….

Consequently, we define the tail probabilities as

$P\{x > i\} = q_i$

But the usual distribution function is

$P\{x \leq i\} = 1 - q_i$

So that the probability generating function follows

$P(x) = \sum_{i=0} p_i\, x^i = E(x^i)$

Also for the joint probability, we have the generating function as

$Q(x) = \sum_{i=0} q_i\, x^i$

We can see that $Q(x)$ is not the same as $P(x)$

$Q(x)$ do not in general constitute probability distribution despite the fact the coefficients are probabilities.

Note that

$$\sum p_i = 1, \qquad sothat P(i) = 1,$$

$$and\ /P(x)/ \leq \sum /p_i x^i/$$

$$\leq \sum p_{j,}\ _{if}\ /x/ \leq 1$$

$$\leq 1$$

This means that $P(x)$ is absolutely convergent at least for $/x/ < 1$. But for $Q(x)$, all coefficients are less than unity, this making $Q(x)$ to converge absolutely at least in the open interval $/x/ < 1$.

Converting $P(x)\ and Q(x)$, we have

$$(1 - x)Q(x) = 1 - P(x)$$

which is easily seen when the coefficient of both sides are compared.

For the mean and variance of p_i, we have

$$U \equiv E(x) = \sum_{i=0} ip_i = p^i(1)$$

$$= \sum_{i=0} q_i = Q(1)$$

then $E[x\,(x - 1)] = \sum i(i - 1)p_i = p^{''}(1) = 2Q^{'}(1)$

So that $\sigma^2 \equiv var(x) = p''(1) + p'(1) - (P^1(1))^2$

$$= 2Q'(1) + Q(1) - (Q(1))^2$$

In the same vein rth factorial moment $\mu'_{(r)}$ about the origin to be

$$E[X(x-1)(x-2)\dots(x-r+1)] = \sum_i (i-1)(i-2)\dots\dots(i-r+1)p_i$$

$$= p^{(r)}(1) \equiv rQ^{(r-1)}(1)$$

From these result, several other generating function could be obtain such as the moment generating function, characteristics function, cumulative generating function.

15.3 Moment-Generating Function

This is define as

$M_x(t) = E(e^{tx})$

for X discrete with probability p_i, we have

$$M_x(t) = \sum_i e^{ti} p_i \equiv P(e^t)$$

for X continues with frequency function $f(u)$, we have

$$M_x(t) = \int_{-\infty}^{\infty} e^{tu} f(u)du$$

obtaining the Taylor series expansion of $M_x(t)$

we have

$$M(t) = 1 + \frac{\sum_{r=1}^{\infty} U_r^1 t^r}{r!}$$

where μ'_r is the rth moment assume the original.

Because of the limitation of the moment generation function (in that it does not always exist) the characteristics function become appropriate which is define by

$$\phi(t) = E(e^{itx})$$

The Taylor expansion is similar

$$\phi(t) = 1 + \sum_{r=1}^{N} \frac{u_r^1(it)^{r!}}{r!}$$

the characteristics function exist always both for discrete function and continues function.

$$\phi_x(t) = \sum_{t=1}^{\infty} e^{itx} f(x)$$

and

$$\phi_x = \int_{-\infty}^{\infty} e^{itx} f(x)dx$$

where the Fourier transform of $f(x)$ is

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \phi(t)d(t)$$

A range simpler generating function is that of the cumulants. When the natural logarithm of either the *mgf* or the *cf* is generated, it results into the cumulant-generating function, which is simpler to handle than the former two.

This is given by

$$K_x(t) = logM_x(t)$$

$$\equiv \frac{\sum_{r=1}^{\infty} k_r t^r}{r!}$$

where K_r is the rth cumulant.

In handling discrete variables, the functional moment generating-function is also useful, which is defined as

$$Q(g) = P(1+y) = E[(1+y)^i]$$

$$\equiv 1 + \frac{\sum_{r=1} u'_{(r)} y^r}{r!}$$

where $u'_{(r)}$ is the rth factorial moment about the origin.

15.4 Convolutions

Let there be two non-negative independent integral-valued random variables X, Y with *pdf*

$P(x = i) = a_i$

and

$P(y = j) = b_j$ the probability of the joint event $(x = y, = j)$ is given as $a_i b_j$.

Let there be a new random variable $S = x + y$ the event $(s = k)$ is made up of the mutually exclusive events $(X = 0, Y = k), (X = 1, Y = k - 1), .., (X = k, Y = 0)$

Given the distribution of S as

$$P(s = k) = c_k$$

then it can be shown that

$$C_k = a_0 b_k + a_1 b_{k-1} + \cdots + a_r b_0$$

When two sequence of numbers which may not be probabilities are compounded, then it is called a ***convolution*** which can be represented generally as

$$\{C_k\} = \{a_k\} * \{b_k\}$$

Given the following general functions

$$\left.\begin{aligned} A(x) &= \sum_{i=0}^{\infty} a_i x^i \\ B(x) &= \sum_{i=0}^{\infty} b_i x^i \\ C(x) &= \sum_{i=0}^{\infty} c_i x^i \end{aligned}\right\}$$

we can then write

$$C(x) = A(x)B(x)$$

this is because, multiplying the two series $A(x)$ and $B(x)$, and given the coefficients of x^k as c_k.

When considering probability distribution functions, the probability-function of the sum, S, of two independent non-negative integrated-valued random variable X and Y is simply the product of the letters probability-generating functions.

Just as the case of two sequences, several sequences can also be combining together. The generating function of the convolution is simply the product of the individual generating functions. That is, if we have the sequence $\{a_i\} * \{b_i\} * \{c_i\} * \{d_i\} * \ldots$, the generating function becomes $A(x)\, B(x)\, C(x)\, D(x) \ldots$..

Given the sum of several independent random variables,

$$S_n = x_1 + x_2 + x_3 + \cdots + X_n$$

Where X_k have a common probability distribution given by p_i, with *pgf* $P(x)$, then the *pgf* of S_n is $\{(P(x)\}^n$. Further, the distribution of S_n is given by a sequence of probabilities which is the n-fold convolution of $\{p_i\}$ with if its written as

$$\{p_i\} * \{p_i\} * \ldots\ldots * \{p_i\} = \{p_i\}]^n$$

15.5 Compound Distributions

Suppose the number of random variables contributing to the sum is itself a random variable. That is

$$S_N = x_1 + x_2 + \cdots + x_n$$

where

$$\left.\begin{aligned} P\{x_k = i\} &= f_i \\ P\{N = n\} &= g_n \\ P\{S_n = i\} &= h_i \end{aligned}\right\}$$

and the corresponding *pdf* be given as

$$\left.\begin{aligned} F(x) &= \sum f_i x^i \\ Q(x) &= \sum g_n x^n \\ H(x) &= \sum h_i x^i \end{aligned}\right\}$$

Simple probability consideration show that we can write the probability distribution of S_n as

$$h_i = p\{s_n = l\}$$

$$= \sum_{n=0} p\{N = n\}\, P\{S_n = l/N = n\}$$

$$= \sum_{n=0} g_n p\{s_n = l/N = n\}$$

For fixed n, the distribution of S_n is the n-fold convolution of $\{F_i\}$ with itself, that is $\{F_i\}.^n$. Thus

$\sum_{l=0} P\{S_n = l/N = n\}x^l = \{F(x)\}^n$

Thus the probability generating function $H(x)$ can be expressed as

$$H(x) = \sum_{l=0} h_i x^i$$

$$= \sum_{i=0} x^i \sum_{n=0} g_n p\{S_n = l/N = n\}$$

$$\sum_{n=0} g_n \sum_{i=0} p\{S_n = l/N = n\}x^l$$

$$= \sum_{n=0} g_n \{F(x)\}^n$$

$$= G(f(x))$$

Thus gives a functionally simple form for the *pgf* of the compound distribution $\{h_i\}$ of the sum S_N.

15.6 Markov Chain

It would be of interest to define the joint probability of the entire experiment. This will be a very complicated or intricate problem.Early in the 20th century, a Russian Mathematician A.A. Markov, provide a simplification of the problem by making the assumption that the outcome of a trial X_t depends on the outcome of the immediate proceeding trial X_{t-1} (and on if only) and effects X_{t+1} (next trial) only. The resulting process is known as Markov Chain.

15.6.1 Transition Problem

If a_i denote the state of the process X_t and a_i, i not equal to j denotes the state of the process X_{t+1}, then there is a problem of going from a_i to a_j denoted by p_{ij} define as,

$p_{ij} = P(X_{t+1} = a_j / X_t = a_i)$

Set $t = 0$, $a_i = i$, and $a_j = j$

$p_{ij} = P(X_1 = j / Xo = i)$

The above is known as transition problem. The entire process is defined by $\{p_{ij}\}$.

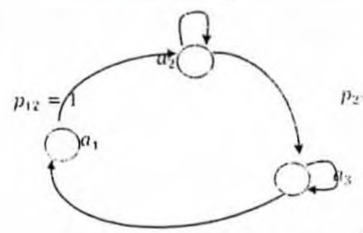
15.6.2 Transition Diagram

A transition diagram is a graphical representation of the process with arrows from each state to indicate the possible direction of movement together with the corresponding transition probabilities against the arrows.

Example 15.1

Consider a process with three possible states $a_1, a_2, and a_3$. Let $p_{ij}: i = 1,2,3,\ j = 1,2,3$, denote the transition from one state to the other.

The corresponding transition diagram is as follows:



The diagram above represents a square matrix

$P = (p_{ij})\ i = 1,2,---n, \quad j = 1,2,----,n$

15.6.3 Transition Matrix

To every transition diagram, there exist a transition matrix and vice versa. For the example 16.1, the transition matrix is as given below:

$$p = \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{bmatrix}$$

This is a one-step transition matrix for every given $i, \{p_{ij}\}$ indicate the branch problem in a tree diagram. In general.

$$p = \begin{bmatrix} p_{11} & p_{12} \cdots\cdots & p_{1n} \\ p_{21} & p_{22} \cdots\cdots & p_{2n} \\ p_{n1} & p_{n2} \cdots\cdots & p_{nn} \end{bmatrix}$$

and.

$$\sum_{j=1}^{n} Pij = 1$$

For any given i, p_{ij} is the probability of transition to a, given that the process was in state a_i.

In this section. we consider a stochastic process $\{X_n, n = 0, 1, 2, \ldots\}$ that takes in a finite or countable number of possible values unless otherwise mentioned, his set of possible values of the process will be tested by the set of non-negative integers (0, 1, 2, ...). If $X_n = 1$. the process is said to be in state i at time n. We suppose that whenever the process is in state i, there is a fixed probability p_{ij}that it will set be in state j. That is we suppose that $\{X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \ldots, X_1 = i_1, X_0 = i_0\} = P_{ij}$ or all states$i_0, i_1, \ldots, i_{n-1}$, $i, j$and $\forall n \geq 0$. Such a stochastic process is known as a Markov chain. The value p_{ij} represents the probability that process will, when in state i, next make a transition into state j. Since probabilities are non-negative and since the process must make a transition into some state, we have that $P_{ij} \geq 0$, i, j ≥ 0; $\sum_{j=0}^{\infty} P_{ij} = 1, i = 0, 1, \ldots$

P denote the matrix of one-step transition probabilities p_{ij}, so that

$$P = \begin{vmatrix} P_{00} & P_{01} & P_{02} & \cdots \\ P_{10} & P_{11} & P_{12} & \cdots \\ \vdots & & & \\ P_{i0} & P_{i1} & P_{i2} & \cdots \\ \vdots & \vdots & \vdots & \end{vmatrix}$$

Example 15.2 (Forecasting the Weather)

Suppose that the chance of rain tomorrow depends on the previous weather conditions only through whether or not it is raining today and not on past weather conditions. Suppose also that if it rains today, then it will rain tomorrow with probability α; and if it does not rain today, then it will rain tomorrow with probability β.

If we say that the process is in state 0 when it rains and state 1 when it does not rain, then the preceding is a two state Markov chain whose transition probabilities are given by

$$P = \begin{bmatrix} \alpha & 1-\alpha \\ \beta & 1-\beta \end{bmatrix} \quad P = \begin{pmatrix} \alpha & 1-\alpha \\ P & 1-\beta \end{pmatrix}$$

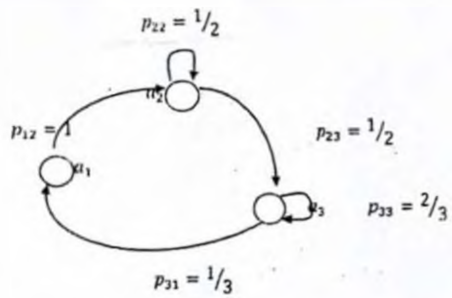
Example 15.3

Suppose that company XYZ has three departments a_1, a_2 and a_3. The employees lean to be transferred to another department at the end of the year as follows:

i) A man who is in a_1, must be transferred only to a_2

ii) A man who is in a_2 cannot be transferred to a_1, but can be transferred to a_2 or a_3 with equal probability.

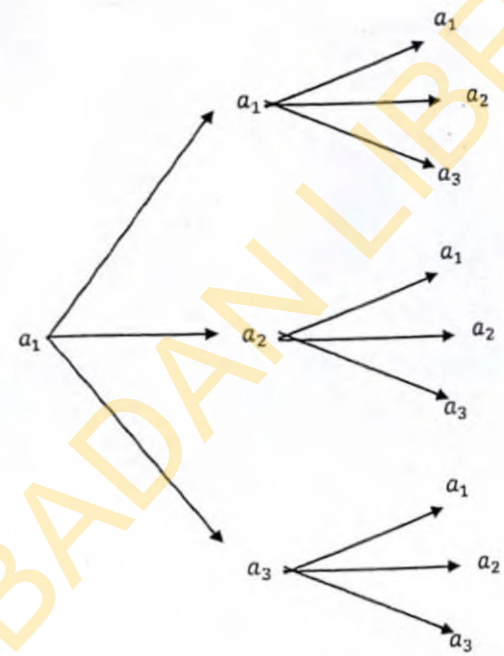
iii) A man who is in a_3 cannot be transferred to a_2 but can be transferred either to a_3 with probability $^2/_3$ or a1 with probability $^1/_3$. Draw a one state transition diagram, and matrix.



$$P = \begin{array}{c} \\ a_1 \\ a_2 \\ a_3 \end{array} \begin{array}{c} \begin{array}{ccc} a_1 & a_2 & a_3 \end{array} \\ \begin{bmatrix} 0 & 1 & 0 \\ 0 & 1/2 & 1/2 \\ 1/3 & 0 & 2/3 \end{bmatrix} \end{array}$$

First problem: Suppose the process state in other I, what is the probability that after n-steps it will be in state j? Consider a process with only three states a1, a2 and a3. What is the probability that after two steps the process will be in state $j, for\ j = 1,2,3$, given that the initial state of the process is $i, for\ i = 1,2,3$.

By assuring that $i = 1$, we obtain a probability tree for the process as follows:



$P\{X_2 = a_1/X_0 = a_1\} = P_{11}.P_{11} = PP_{11}{}^{(2)}$

$P\{X_2 = a_2|X_0 = a_1\} = P_{12}.P_{21} = P_{21}{}^{(2)}P_{12}{}^{(2)}$

$P\{X_2 = a_3|X_0 = a_1\} = P_{13}.P_{31} = P_{31}{}^{(2)}P_{13}{}^{(2)}$

$p_{11}p_{11} + p_{12}p_{21} + p_{13}p_{31}p_{11}p_{12} + p_{12}p_{22} + p_{13}p_{32}p_{11}p_{13} + p_{12}p_{23} + p_{13}p_{33}$

$p_{21}p_{11} + p_{22}p_{21} + p_{23}p_{31}p_{21}p_{12} + p_{22}p_{22} + p_{23}p_{32}p_{21}p_{13} + p_{22}p_{23} + p_{23}p_{33}$

$p_{31}p_{11} + p_{32}p_{21} + p_{33}p_{31}p_{31}p_{12} + p_{32}p_{22} + p_{33}p_{32}p_{31}p_{13} + p_{32}p_{23} + p_{33}p_{33}$

Assume that $i = 2,\ then$

$P\{x_2 = a_1 / x_o = a_2\} = p_{21}.p_{12} = p_{21}^{(2)}$

$P\{x_2 = a_2 / x_o = a_2\} = p_{22}.p_{22} = p_{22}^{(2)}$

$P\{x_2 = a_3 / x_o = a_2\} = p_{23}.p_{32} = p_{32}^{(2)}\ p_{23}^{(2)}$

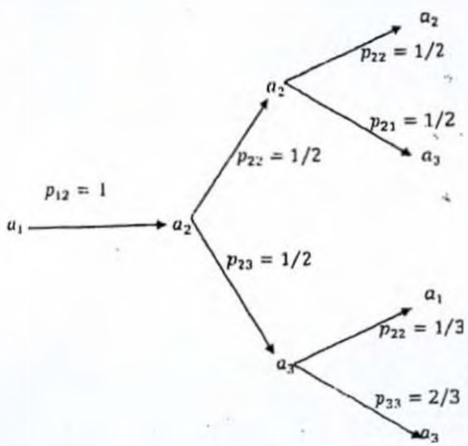
Assume that $i = 3$, *then*

$$P\{x_2 = a_1 / x_o = a_3\} = p_{31}.p_{13} = p_{31}^{(2)}$$
$$P\{x_2 = a_2 / x_o = a_3\} = p_{32}.p_{23} = p_{32}^{(2)}$$
$$P\{x_2 = a_3 / x_o = a_3\} = p_{33}.p_{33} = p_{33}^{(2)}$$

It could be seen that $p^{(n)} = p^n$

Example 15.4

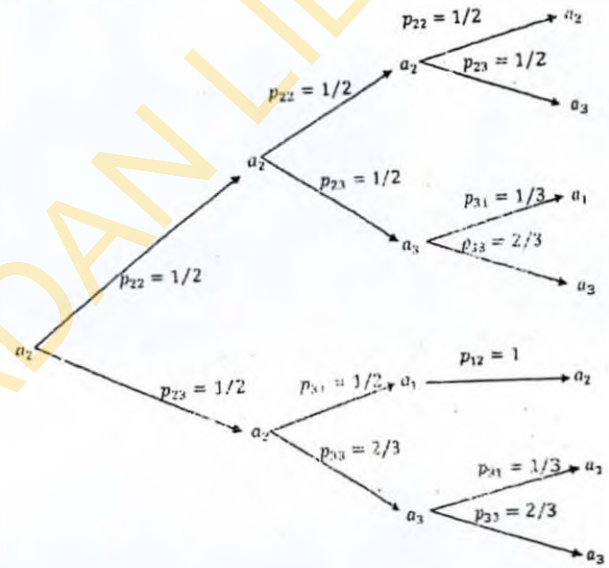
Use a probability tree to find$p^{(3)}$ in example 16.3



$$\{x_3 = a_2 / x_o = a_1\} = 1.{}^1/_2.{}^1/_2 = {}^1/_4$$
$$P\{x_3 = a_3 / x_o = a_1\} = 1.{}^1/_2.{}^1/_2 = {}^1/_4$$
$$P\{x_3 = a_1 / x_o = a_1\} = 1.{}^1/_2.{}^1/_3 = {}^1/_6$$
$$P\{x_3 = a_3 / x_o = a_1\} = 1.{}^1/_2.{}^2/_3 = {}^1/_3$$
$$P_{13}^{(3)} = {}^1/_4 + {}^1/_3 = {}^7/_{12}$$

$$P_{12}^{(3)} = {}^1/_4$$
$$P_{11}^{(3)} = {}^1/_6$$

In the same vein



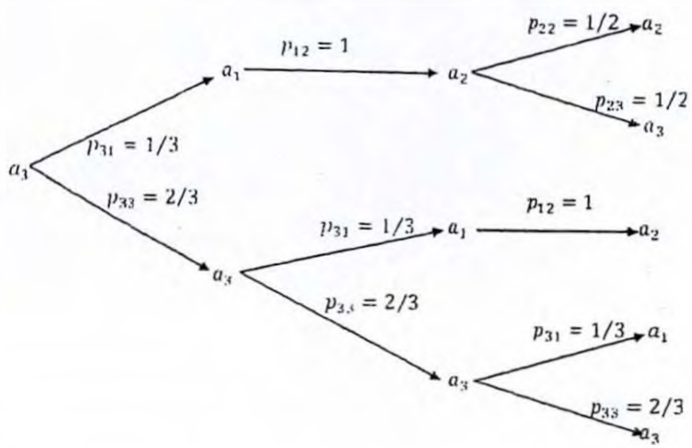
$$P\{x_3 = a_2 / x_o = a_2\} = {}^1/_2.{}^1/_2.{}^1/_2 = {}^1/_8$$
$$P\{x_3 = a_3 / x_o = a_2\} = {}^1/_2.{}^1/_2.{}^1/_2 = {}^1/_8$$
$$P\{x_3 = a_1 / x_o = a_2\} = {}^1/_2.{}^1/_2.{}^1/_3 = {}^1/_{12}$$
$$P\{x_3 = a_3 / x_o = a_2\} = {}^1/_2.{}^1/_2.{}^2/_3 = {}^1/_6$$
$$P\{x_3 = a_2 / x_o = a_2\} = {}^1/_2.{}^1/_3.1 = {}^1/_6$$
$$P\{x_3 = a_2 / x_o = a_1\} = {}^1/_2.{}^1/_2.{}^1/_2 = {}^1/_6$$
$$P\{x_3 = a_1 / x_o = a_2\} = {}^1/_2.{}^2/_3.{}^1/_3 = {}^1/_9$$
$$P\{x_3 = a_3 / x_o = a_2\} = {}^1/_2.{}^2/_3.{}^2/_3 = {}^2/_9$$

$P_{21}^{(3)} = {}^1/_{12} + {}^1/_9 = {}^7/_{36}$

$P_{22}^{(3)} = {}^1/_8 + {}^1/_6 = {}^7/_{24}$

$P_{23}^{(3)} = {}^1/_{18} + {}^1/_6 + {}^2/_9 = {}^{37}/_{72}$

Also



$$P\{x_3 = a_2 / x_o = a_3\} = {}^1/_3 . 1 . {}^1/_2 = {}^1/_6$$

$$P\{x_3 = a_3 / x_o = a_3\} = {}^1/_3 . 1 . {}^1/_2 = {}^1/_6$$

$$P\{x_3 = a_2 / x_o = a_3\} = {}^2/_3 \, {}^1/_3 . 1 = {}^2/_9$$

$$P\{x_3 = a_1 / x_o = a_3\} = {}^2/_3 . {}^2/_3 . {}^1/_3 = {}^4/_{27}$$

$$P\{x_3 = a_3 / x_o = a_3\} = {}^2/_3 . {}^2/_3 . {}^2/_3 = {}^8/_{27}$$

$p_{31}^{(3)} = {}^4/_{27}$

$p_{32}^{(3)} = {}^1/_6 + {}^2/_9 = {}^7/_{18}$

$p_{33}^{(3)} = {}^1/_6 + {}^8/_{27} = {}^{25}/_{54}$

Therefore,

$$p^{(3)} = \begin{bmatrix} p_{11}^{(3)} & p_{12}^{(3)} & p_{13}^{(3)} \\ p_{21}^{(3)} & p_{22}^{(3)} & p_{23}^{(3)} \\ p_{31}^{(3)} & p_{32}^{(3)} & p_{33}^{(3)} \end{bmatrix} = \begin{bmatrix} {}^1/_6 & {}^1/_4 & {}^7/_{12} \\ {}^7/_{36} & {}^7/_{24} & {}^{37}/_{72} \\ {}^4/_{27} & {}^7/_{18} & {}^{25}/_{54} \end{bmatrix}$$

Note that $p^{(n)} = p^{(n-1)}p = p^{(0)}p^n$

at $n = 1,\ p^{(1)} = p^{(0)}p$

at $n = 2,\ p^{(2)} = p^{(1)}p = p^{(0)}p^2$

at $n = 3,\ p^{(3)} = p^{(2)}p = p^{(0)}p^3$

This implies that

$$p_{ij}^{(n)} = \sum_k p_{ik}^{(n-1)} p_{kj}$$

Definition

Let $\{X_n, n = 0, 1, 2 \ldots\}$ denote a square of real valued variable index by n. The value of x for given n is the state of the process at the n th step.

$P\{x_n = j / x_{n-1} = i\}$ is a one-step transition probability matrix. The index n denote something close to time and therefore x_n *depend on* $x_{n-1}, x_{n-2}, \ldots, x_0$ and not on $x_{n+1}, x_{n+2}, \ldots$

The Markov assumption is that

$$P\{x_n = j_n / x_{n-1} = j_{n-1}, x_{n-2} = j_{n-2}, \ldots, x_0 = j_0\} = P\{x_n = j_n / x_{n-1} = j_{n+1},\}$$

The conditional distribution of x_n given the whole past history of the process must equal to the conditional distribution of x_n *given* x_{n-1}.

Example 15.5: (A Communication System)

Consider a communications system which transmits in digits 0 and 1. Each digit transmitted must pass through several stages, at each of which more is a probability p that the digit entered will be unchanged when it leaves. Letting δ_n denote the digit entering the nth stage, then $\{X_n, n = 0,1...\}$ is a two-state Markov chain having a transition probability matrix.

$$P = \begin{bmatrix} P & 1-P \\ 1-P & P \end{bmatrix} \quad P = \begin{pmatrix} P & 1-P \\ 1-P & P \end{pmatrix}$$

Example 15.6

On any given day Gary is either cheerful (C), so-so (S), or glum (G). If he is cheerful today, then he will be C, S, or G tomorrow with respective probabilities 0.5, 0.4, 0.1. If he is feeling so-so today, then he will be C, S, or G tomorrow with probabilities 0.3, 0.4, 0.3. If he is glum today, then he will be C, S, G tomorrow with probabilities 0.2, 0.3. 05.

Letting X_n denote Gary's mood on the nth day, then $\{X_n, n \geq 0\}$ is a three states Markov chain (State $0 = C$, state $1 = S$., State $2 = G$) with transition probability matrix.

$$P = \begin{bmatrix} 0.5 & 0.4 & 0.1 \\ 0.3 & 0.4 & 0.3 \\ 0.2 & 0.3 & 0.5 \end{bmatrix}$$

Example 15.7: (Transforming a process into a Markov chain)

Suppose that whether or not it rains today depends on previous weather conditions through the last two days. Specifically, suppose that if it has rained for the past two days. Even it will rain tomorrow with probability 0.7. if it rained today but not yesterday, then it will rain tomorrow with probability 0.5; if it rained yesterday but not today, then it will rain tomorrow with probability 0.4; if it has not rained in the past two days, then it will rain tomorrow with probability 0.2.

If we let this state at time n depend only on whether or not it is raining at time n, then the preceding model is not a Markov chain (why not?). However, we can transform this model into a Markov chain by saying that the state at any time is determined by the weather conditions during both that day and the previous day.

In order words, we can say that the process is in

State 0 if it rained both today and yesterday;

State 1 if it rained today but not yesterday;

State 2 if it rained yesterday but not today;

State 3 if it did not rain either yesterday or today.

The preceding would then represent a fair-state Markov chain having a transition on probability matrix.

$$P = \begin{bmatrix} 0.7 & 0 & 0.3 & 0 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0.4 & 0 & 0.6 \\ 0 & 0.2 & 0 & 0.8 \end{bmatrix}$$

You should carefully check the matrix P, and make sure you understand how it was obtained.

15.7 Stationarity Assumption

A Markov chain is stationary if for $m \neq n$

$$P\{x_n = j_n / x_{n-1} = j_{n-1},\} = P\{x_m = j_m / x_{m-1} = j_{m-1},\}$$

or simply,

$$P\{x_n = j / x_{n-1} = i,\} = P\{x_m = j / x_{m-1} = i\}$$

In this case the one-step transition probability does not depend on the step number. It is therefore sufficient

For us to state only the one-step transition probabilities.

We therefore set $n = 1$ and obtain

$p_{ij}^{(1)} = P\{x_1 = j / x_0 = i\}$

$p_{ij}^{(n)} = P\{x_n = j / x_0 = i\}$

For, $n = 0$, this leads to

$p_{ij}^{(1)} = 1$ for $j = i$

$p_{ij}^{(1)} = 0 \qquad$ for $j \neq i$

$p_{ij}^{(n)} = P\{x_n = j/x_0 = i\}$ by definition

$= \sum_k p\{x_n = j, x_{n-1} = k/x_0 = i\}$ marginal from joint

$$= \sum_k p\{x_n = j/x_{n-1} = k, x_0 = i\} P\{x_{n-1} = k/x_0 = i\}$$

$$= \sum_k p\{x_n = j/x_{n-1} = k\} P\{x_{n-1} = k/x_0 = i\}$$

$$= \sum_k p_{kj} p_{ik}^{(n-1)}$$

$$= \sum_k p_{ik}^{(n-1)} p_{kj}$$

15.8 Absorbing Markov Chain

A state in a Markov chain is absorbing if it is impossible to move out of that state. That is, the process stays there. A Markov chain is absorbing if it can't least one absorbing state. That is,

$$p_{jj} = 1.0$$

A state in a Markov chain is transient or non-absorbing if it is possible to get out of that state. That is

$$p_{jj} \neq 1.0 \quad for\ state\ j.$$

15.8.1 Probability of a Markov Process ending in a Given Absorbing State

This depend on the given in that state. Let a_{ij} denote the probability that an absorbing chain will be absorbed in state a_j if it states in the non-absorbing state a_i.

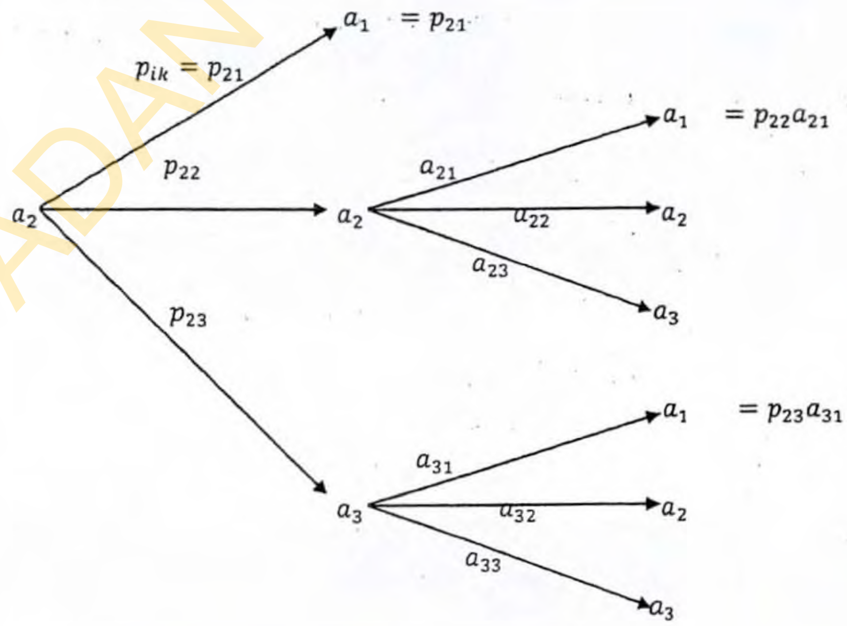
Method 1

There are two possibilities, either the first transition is to state a_j (in which case the chain is immediately absorbed) or the first transition is to some transient or non-absorbing state a_k, $k \neq j$, and then the process immediately enters states a_j $from a_k$. These are two mutual exclusive events.

p_{ij} is the probability of the first event, and that of the second is

$\sum_k p_{ik} a_{kj}$

Consider a process with the following three states; $a_1, a_2, a_3, where\ a_1$ is an absorbing state, and others are transient.



Then

$a_{ij} = P\{x_i = a_j/x_0 = \}$

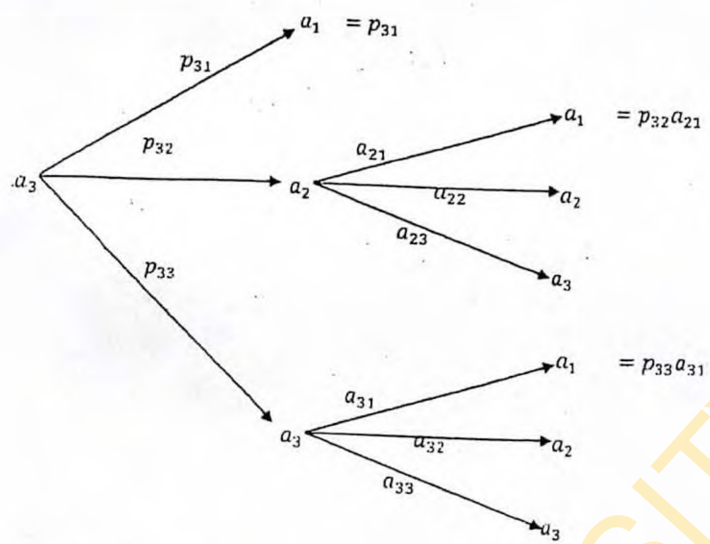
By substitution

$$a_{21} = p_{21} + p_{22}a_{21} + p_{23}a_{31}$$

a_{ij} is a one-linear equation in several unknowns. Construct a corresponding linear equation by using each of the other transit state as initial state.

In the given example, a_{21} is a linear equation in two unknowns. Note that P_{ij} is obtained from the given one step transition matrix. The only unknown are $a_{kj}, all\ k \neq j$

The corresponding a_{21} is given by



$$a_{31} = p_{31} + p_{32}a_{21} + p_{33}a_{31}$$

Then solve all values of equation ii for all $i \neq j$ (such as a_{21} and a_{31}) simultaneously

As an example consider the following transition matrix for absorbing Markov chain with four states. Note that an absorbing state is indicated by probability 1

$$P = \begin{bmatrix} 1/4 & 1/4 & 1/2 & 0 \\ 1/3 & 1/3 & 0 & 1/3 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Note that the absorbing state are a_3 and a_4.

Suppose that we want a_{13}, that is the probability starting from a, will get absorbed in state a_3 . In other word, we want the probability that the chain will enter a_3 from a_1 .

Then a_{ij} will give us

$$a_{13} = p_{13} + p_{11}a_{13} + p_{12}a_{23}$$

$$a_{23} = p_{23} + p_{21}a_{13} + p_{22}a_{23}$$

Substitute for p_{ij}, noting that a_{kj} is unknown.

$$a_{13} = 1/2 + 1/4\, a_{13} + 1/4\, a_{23}$$

$$a_{23} = 0 + 1/3\, a_{13} + 1/3\, a_{23}$$

Solving the simultaneous equation, we obtain

$a_{13} = 4/5$ and $a_{23} = 2/5$

The matrix becomes

$$\begin{bmatrix} a_{13} & a_{14} \\ a_{23} & a_{24} \end{bmatrix} = \begin{bmatrix} 4/5 & 1/5 \\ 2/5 & 3/5 \end{bmatrix}$$

Alternatively.

$$for\ \ a_{ij} = p_{ij} + \sum_k p_{ik}a_{kj}$$

We can write this matrix form, Let A denotes the matrix of a_{ij} R denotes the matrix of p_{ij}, Q denotes

The matrix of p_{ik}. That is

$A = \{a_{ij}\} = \{a_{kj}\} \quad --- s \times r$

$R = \{p_{ij}\} \quad --- s \times r$

$Q = \{p_{ik}\} \quad --- s \times s$

Then a_{ij}can be written as

$A = R + QA$

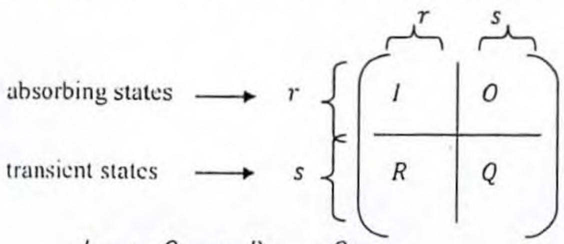
Where

$r = number\ of\ absorbing\ states$

$s = number\ of\ transit\ states$

Step 1: Arrange the rows and columns of the one-step transition matrix in which a way that the absorbing states appear first in the rows and first in the columns.

Step 2: partition the new one-step transition matrix as follows



$I_{(r\times r)},\ O_{(r\times s)},\ R_{(s\times r)},\ Q_{(s\times s)}$

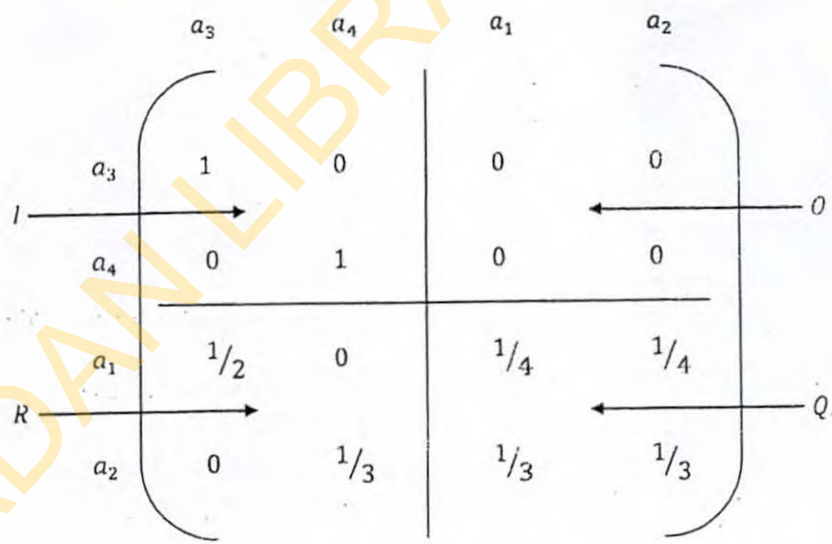
Step 3: Solve for A, the matrix of the unknown, as follow

$(I - Q)A = R$

$A = (I - Q)^{-1}R$

Since (I-Q) is non-singular and so has an inverse. $(I - Q)^{-1}$ is known as the fundamental matrix.

For example, the above 4 × 4 matrix incan be rearranged as follows:



Step 4: Find I-Q and hence $(I - Q)^{-1}$

$$(I - Q) = \begin{vmatrix} 3/4 & -1/4 \\ -1/3 & 2/3 \end{vmatrix}$$

$$|I - Q| = (3/4)(2/3) - (1/4)(1/3) = (5/12)$$

$$\text{cof}(I - Q) = \begin{bmatrix} 2/3 & 1/3 \\ 1/4 & 3/4 \end{bmatrix}$$

$$\text{cof}^{T}(I - Q) = Adj(I - Q) = \begin{bmatrix} 2/3 & 1/3 \\ 1/3 & 3/4 \end{bmatrix}$$

$$since\ (I - Q)^{-1} = \frac{Adj(I - Q)}{set(I - Q)} = \begin{bmatrix} 2/3 & 1/4 \\ 1/3 & 3/4 \end{bmatrix} 12/5$$

$$= \begin{bmatrix} 8/5 & 3/5 \\ 4/5 & 9/5 \end{bmatrix}$$

$$Therefore\ A = (I - Q)^{-1}R = \begin{bmatrix} 8/5 & 3/5 \\ 4/5 & 9/5 \end{bmatrix}\begin{bmatrix} 1/2 & 0 \\ 0 & 1/3 \end{bmatrix}$$

$$= \begin{bmatrix} 4/5 & 1/5 \\ 2/5 & 3/5 \end{bmatrix}$$

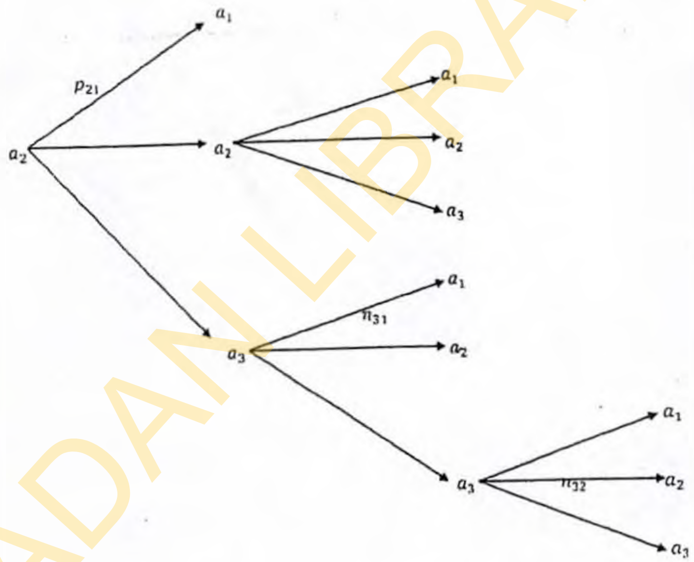
15.8.2 The Expected Number of Times a Markov Process will be in each Possible Starting Transient (Absorbing) State

Let $N = (n_{ij})$ *where* n_{ij} is the number of times the chain is in transient state a_j given the initial state is a_i.

Let n_{ij} denote the mean number of time that the chain is in transient state a_i.

Let N denote the matrix of n_{ij}, which is a square matrix since i *and* j range over the transient stateds.

Consider a chain with the three states in (a), a_1, a_2, a_3 where a_1 is the absorbing state. Assume that the initial state process is a_2.



Consider the state at time 1. That is, the first time interval is spent in state a_i (a_i is transient state). If $i \neq j$ and the transition probability p_{ik} given the probability that the process will be in a_K from a_i. Then

$n_{ij} = \sum_k p_{ik} n_{kj}$

$n_{ii} = p_{ik} n_{ki} = 1$

$= d_{ii} + \sum_k p_{ik} n_{ki}$

Which is combined into

$$n_{ij} = d_{ij} + \sum_k p_{ik} n_{kj}, \qquad d_{ij} = 1, \quad \text{for } i - j$$

$$= 0, \quad \text{for } i \neq j$$

In matrix form this can be written as

$$\begin{array}{c|cccc|} & d_1 & d_2 & \dots & d_n \\ \hline d_1 & 1 & 0 & \dots & 0 \\ d_2 & 0 & 1 & & 0 \\ \vdots & \vdots & & & \\ d_n & 0 & 0 & & 1 \end{array}$$

15.8.3 The Length of Time (Expected Number of Steps) Required before Absorbtion Occurs

For any given initial transient state a_i the expected number of step required before absorption is given by the elements of the rector

$$t = \sum_j n_{ij}$$

Let c be a column rector with the same number of elements as the columns of N and every element of C is unity.

Then

$$t = NC$$

$$\text{where } c = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

By using the above fundamental matrix we find

$$t = \begin{bmatrix} 1.6 & 0.6 \\ 0.8 & 1.8 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2.2 \\ 2.6 \end{bmatrix}$$

15.8.4 The Number of Transitions that will occur before a particular Absorbing State is reached

Letm_{ij}. denote the number of transition that will occur before a particular absorbing state j is entered given the initial state i.

Recall

$a_{ij} = p_{ij} + \sum_k p_{ik} a_{kj}$

$$\begin{array}{c|cccc|} & d_1 & d_2 & \dots & d_n \\ \hline d_1 & 1 & 0 & \dots & 0 \\ d_2 & 0 & 1 & & 0 \\ \vdots & \vdots & & & \\ d_n & 0 & 0 & & 1 \end{array}$$

$\{p_{ik}\} = Q$

$\{n_{ij}\} = N$

$N = I + QN$

$N = (I - Q)^{-1} I = (I - Q)^{-1}$

Thus the element of the fundamental matrix give the expected number of times the process will spend in given transient states for any given initial transient state.

$$N = (I - Q)^{-1} = \begin{matrix} a_1 \\ a_2 \end{matrix} \begin{bmatrix} 8/5 & 3/5 \\ 4/5 & 9/5 \end{bmatrix} = \begin{bmatrix} 1.6 & 0.6 \\ 0.8 & 1.8 \end{bmatrix}$$

Interpretation:

Starting in state a_1, the expected number of times in state a1 before absorption occurs is 1.6. Similarly, starting from a_1 the expected number of times in state a_2 before absorption is 1.8.

The expected number of transition before absorption is

$a_{ij} m_{ij} = p_{ij} m_{ij} + \sum_k p_{ik} a_{kj}\, m_{ij}$

summing over j

$$M = \sum_k a_{ij} m_{ij}$$

Multiply both side of bym_{ij}

$$a_{ij} m_{ij} = p_{ij} m_{ij} + \sum_k p_{ik} a_{kj}\, m_{ij}$$

$= a_{ij} + \sum_k p_{ik} a_{kj} m_{kj}$

Note that $m_{ij} = 1$ *and* $m_{ij} \neq 1$

In the four state process or chain given earlier, suppose we want to compute m_{13}. Thus, $a_{ij}m_{ij}$ becomes.

$a_{13}m_{13} = a_{13} + p_{11}a_{13}m_{13} + p_{12}a_{23}m_{23}$

Constructing another equation by using the other initial transient state,

$a_{23}m_{23} = a_{23} + p_{21}a_{13}m_{13} + p_{22}a_{23}m_{23}$

this is because $k = 1, 2$ *and* $j = 3$

Substitute for the known values, that is the a's and p's.

Note: The p's are from the given one-step transition matrix and a's are earlier solutions.

Therefore

$${}^4/_5\, m_{13} = {}^4/_5 + {}^1/_4 . {}^4/_5\, m_{13} + {}^1/_4 . {}^2/_5\, m_{23}$$

$${}^2/_5\, m_{23} = {}^2/_5 + {}^1/_3 . {}^4/_5\, m_{13} + {}^1/_3 . {}^2/_5\, m_{23}$$

This result in

$6m_{13} - m_{23} = 8$

$4m_{13} - 4m_{23} = -6$

Thus, $m_{13} = 1.9$ and $m_{23} = 3.4$

So that we have

$$\begin{matrix} & a_3 & a_4 \\ a_1 & \\ a_2 & \end{matrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1.9 \\ 3.4 \end{pmatrix}$$

The transition matrix P of a Markov chain is

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{matrix} 1 & 2 & 3 & 4 & 5 \end{matrix} \\ \begin{pmatrix} 0 & 1/2 & 1/4 & 1/4 & 0 \\ 1/2 & 0 & 1/4 & 1/4 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 2/5 & 1/10 & 1/2 \\ 0 & 0 & 1/2 & 0 & 1/2 \end{pmatrix} \end{array}$$

(*The University of Sydney, 2009*)

Practice Questions

1. Define a generating function, $A(x)$.
2. Define the following:
   (a) moment generating function
   (b) characteristic function
   (c) cumulant generating function
3. Given two random variables X and Y, with probabilities $P(x = i) = a_i$ and $P(y = j) = b_j$. Write an expression for their convolution c_k where $(i = 0, 1, , ..., r)$ and $(j = 0, 1, ..., k)$
4. Given the sum of random variables S_N, show that the probability generating function $H(x)$ is given as $G\{F(x)\}$.
5. On any given day Bruce is either cheerful (C), or so-so (S), or glum (G). If he is cheerful today, then he will be C or S tomorrow with respective probabilities 0.5, 0.4. If he is feeling so-so today, he will be C or S tomorrow with probabilities 0.3, 0.4. If he is glum today, he will be S or G tomorrow with probabilities 0.3, 0.5.

(a) Write down the transition matrix P which describes Bruce's mood oscillations over time.

(b) Bruce is currently in a cheerful mood. What is the probability that he is not in a glum mood on any of the following two days?

(c) Obtain $P^n, n = 3,4,5.$ (*The University of Sydney, 2011*)

6. Suppose that whether or not it rains today depends on weather conditions of the last three days. If it has rained in the past three days then it will rain today with probability 0.8; if it did not rain for any of the past three days, then it will rain today with probability 0.2; and in any other case the weather today will, with probability 0.6, be the same as the weather yesterday. Denote the states by triples of the kind RRR, RRF, etc., and write down the transition matrix P of this Markov chain. Obtain $P^n, n = 3,4,5.$ (*The University of Sydney, 2011*)

7. Define a Markov Chain and state its assumptions.

8. (i) State the stationarity assumption. (ii) Show that $p_{ij}^{(n)} = \sum_k p_{ik}^{(n-1)} p_{kj}$

9. (a) What is the absorbing state?

(b) Assuming the process starts from states $\{1,2\}$, what is

(i) the probability that a Markov process will end up in the given absorbing state.

(ii) the expected number of times a Markov process will be in each transient state for each possible transient state.

(iii) the length of time that it will take for a Markov process to be absorbed. That is the number of steps required to reach an absorbing state for the first time.

(iv) the number of transition that will occur before the absorbing state is reached.

CHAPTER 16
STEADY STATE AND PASSAGE TIME PROBABILITIES

16.1 Introduction

This chapter introduces the student to the process of determining the equilibrium state of a Markov chain. That is, after a long process, the probability of a process being in a steady situation.

Consider the formula for the rector, $P^{(n)}$,of state probabilities for the time n given by

$$P^{(n)} = P^{(0)}P^{(n)}$$

Where $P^{(0)}$ is a vector of initial state probabilities and $P^{(n)}$ is the n-step transition matrix. The interest is to find out what happen to a Markov chain with Pas n becomes large.

We will approach this problem by considering the following example.

Example 16.1

An Engineering company has three departments. Engineering (a_1), production (a_2), and sales (a_3).A man in the Engineering dept. cannot be assigned to sales but may be transferred to production or Engineering with equal probability. A man in the production dept. cannot be transferred to engineering but can be transferred to production or sales with equal probability.A man in sale can be transferred to Engineering or production and his probability of going to Engineering is 3-times that of going to production.

The associated P is as follows:

$$P = \begin{bmatrix} 0.5 & 0.5 & 0 \\ 0 & 0.5 & 0.5 \\ 0.75 & 0.25 & 0 \end{bmatrix}$$

The following transition probabilities can be deduced from P

$$P^{(2)} = P^2 = \begin{bmatrix} 0.250 & 0.500 & 0.250 \\ 0.375 & 0.375 & 0.250 \\ 0.375 & 0.500 & 0.125 \end{bmatrix}$$

$$P^{(3)} = P^3 = \begin{bmatrix} 0.3125 & 0.4375 & 0.2500 \\ 0.3750 & 0.4375 & 0.1875 \\ 0.28125 & 0.46875 & 0.2500 \end{bmatrix}$$

$$P^{(4)} = P^4 = \begin{bmatrix} 0.34375 & 0.4375 & 0.21875 \\ 0.328125 & 0.453125 & 0.21875 \\ 0.328125 & 0.4375 & 0.234375 \end{bmatrix}$$

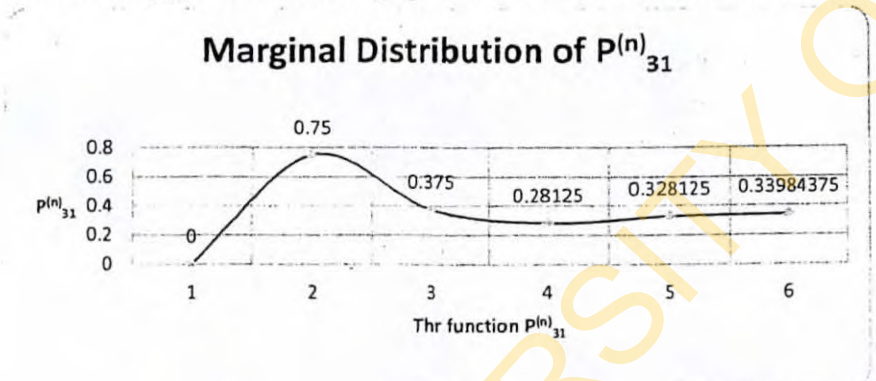
$$P^{(5)} = P^5 = \begin{bmatrix} o.3359375 & 0.4453125 & 0.21875 \\ 0.328125 & 0.4453125 & 0.2265625 \\ 0.33984375 & 0.44140625 & 0.21875 \end{bmatrix}$$

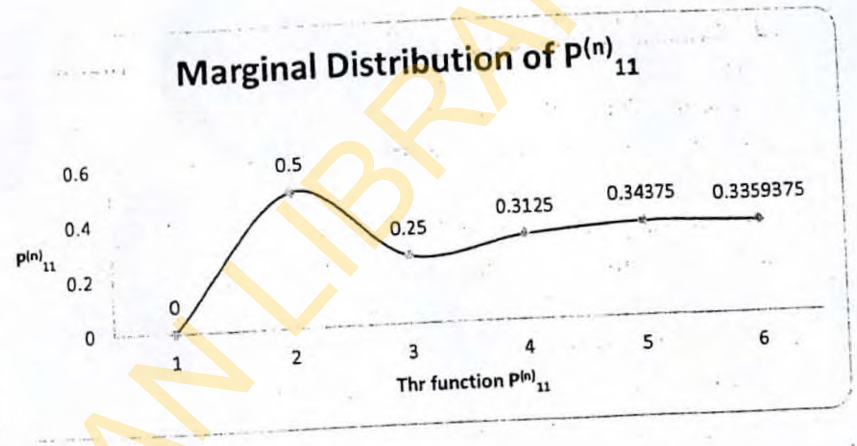
We can go on and on, until the n-step is reached. It should be noted that as the steps increase, the

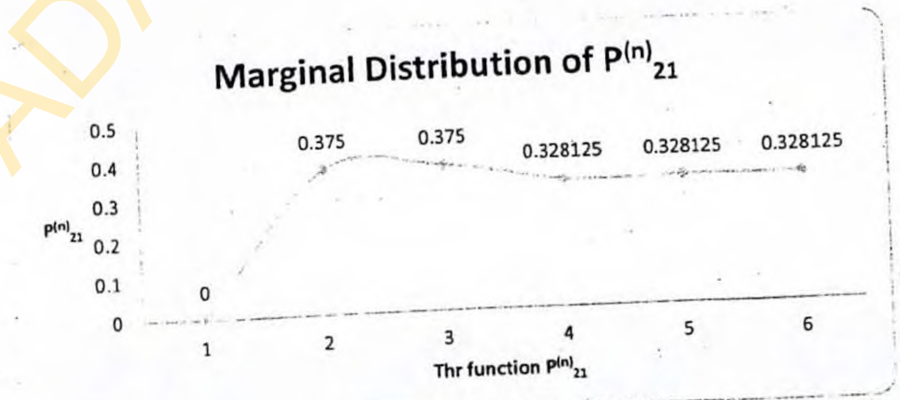
Probabilities tend to be steady. This can readily be seen in the following graph.

16.2 Graph of Marginal Distribution of $P^{(n)}$

For fixed fair (i, j) we can draw the graph of $P^{(n)}$ for various values of n as follows.







Summary of Solution to Problem

1. The function $P_{i1}^{(n)}$ gives the probability of getting to dept. a_1, from dept. a_i in step n, $n = 0, 1, 2, \ldots$
2. When $n = 0, P_{i1}^{(n)} = 0$, since the person will certainly not be transferred.
3. As $n \rightarrow \infty$, $P_{i1}^{(n)}$ tends to converge at $^1/_3$ interpretation of i. similarly, $P_{i1}^{(n)} \rightarrow$ 0.44 or $^4/_9$ irrespective of i, and $P_{i1}^{(n)} \rightarrow 0.22$ *or* $2/9$ irrespective of i.

4. All the rows of the matrix $P(n)$ are identical but the columns are not as $n \to \infty$.

5. The state probabilities that satisfy the above criterion are called steady-state probabilities or equilibrium probabilities, or limit value, or stationary distribution.

6. A Markov chain is said to approach equilibrium as n tens to infinity if its transition probabilities approach limit values.

16.3 Stationary Distribution

The steady-state distribution is define as the set (u_i) where

$\mu_j = \lim_{n\to\infty} p_{ij}^{(n)} = \lim_{n\to\infty} P\{X_n = j\}$

and is independent of i. Furthermore, $\mu_j > 0$.

$\sum_j \mu_j$

$\mu_k = \sum_i \mu_i p_{ik}$

A probability distribution which satisfies μ_k is called invariant or stationary distribution (for a given Markov Chain). In this case row of$P^{(n)}$ is the probability vector$\mu = (\mu_1, \mu_2, \ldots)$. Hence, given

$$P^{(n)} = P^{(n-1)}P$$

$$\lim_{n\to\infty} P^{(n)} = \lim_{n\to\infty} P^{(n-1)}P$$

$$\begin{bmatrix} \mu_1\mu_2 & \cdots \\ \mu_1\mu_2 & \cdots \\ \vdots & \end{bmatrix} = \begin{bmatrix} \mu_1\mu_2 & \cdots \\ \mu_1\mu_2 & \cdots \\ \vdots & \end{bmatrix} [P]$$

This can be written as

$\mu = \mu P$

or

$\mu^T = P^T\mu^T$

This represents a dependent set of equation (since each row elements must sum up to unity). One of the infinite numbers of solutions can be found to represent a probability solution by imposing the condition

$$\sum_i \mu_i = 1$$

This is known as a normalizing equation.

Example 16.1:

From example 15.1 , find $\lim_{n\to\infty} P^{(n)}$

<u>Solution</u>

$$|\mu_1\mu_2\mu_3| = |\mu_1\mu_2\mu_3| \begin{bmatrix} 0.5 & 0.5 & 0.0 \\ 0.0 & 0.5 & 0.5 \\ 0.75 & 0.25 & 0.0 \end{bmatrix}$$

or

$$\begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} = \begin{bmatrix} 0.5 & 0.0 & 0.75 \\ 0.5 & 0.5 & 0.25 \\ 0.0 & 0.5 & 0.0 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix}$$

$\mu_1 = 0.5\mu_1 \qquad + 0.75\mu_3$

$\mu_2 = 0.5\mu_1 + 0.5\mu_2 + 0.25\mu_3$

$\mu_3 = \qquad 0.5\mu_2$

Thus

$\mu_3 = \frac{2}{3}\mu_1$

and

$\mu_2 = 2\mu_3$

Substituting we have

$\mu_2 = \frac{4}{3}\mu_1$

By imposing the normalizing condition on the sum u_i we obtain

$\mu_1 + \mu_2 + \mu_3 = 1$

$\mu_1 + \frac{4}{3}\mu_1 + \frac{2}{3}\mu_1$

Therefore

$\mu_1 = \frac{1}{3}$

This means therefore that

$\mu_2 = \frac{4}{9}$, and $\mu_3 = \frac{2}{9}$

Thus $\mu = (\mu_1 \mu_2 \mu_3) = ({}^1/_3 \, {}^4/_9 \, {}^2/_9)$

This gives a sample method of obtaining$P^{(n)}$ than raising $P$to power n.

Interpretation:

$\mu_j$can be interpreted as follows:

1. Probability of a distant state: if a point in time is fixed in the distant future , μ_j is the probability that the process will be as state j.
2. As a time average: if the process is operated for a long time, μ_j is the fraction of time that the process we be at state j.
3. As a fraction of process: if many identical processes are operated simultaneously, μ_j is a fraction of the process that can be found in state j after a long time.
4. Reciprocal of mean number of transition: μ_j is the reciprocal of the mean number of transition between recurrence of the state, that is, average number of transition before a man ina_j will come back to a_i.

Example 16.2

An individual of unknown genetic character is crossed with a hybrid. The offspring is again crossed with a hybrid, and so on. The states are dominant(D), hybrid (H) and recessive (R). The transition probabilities are

$$P = \begin{array}{c|ccc} & D & H & R \\ \hline D & 1/2 & 1/2 & 0 \\ H & 1/4 & 1/2 & 1/4 \\ R & 0 & 1/2 & 1/2 \end{array}$$

Find$\lim_{n\to\infty} P^{(n)}$ and give all possible interpretation of the result.

16.4 First-Passage and First-Return Probabilities

We shall approach this topic by way of asking certain questions.

Q1: What is the probability that in a process stating from a_i, the first entry to a_j occurs at the $n$thstep?

Q2: What is the number of steps$n$, required to reach state a_j for the first time?

For Q1, consider the function$P_{ij}^{(n)}$ which is the probability that the process will enter state j at the nth step given that it is in state i of the initial step. That is,

$$P_{ij}^{(n)} = P\{X_n = j | X_0 = i\}$$

(a) In this case, the process would enter state a_j, after only$k, 1 \le k \le n-1$, steps.

(b) After that is called either stay three is a_j or change to another state and then return to a_j. For Q2, the probability$f_{ij}^{(n)}$ that the process will reach state a_j for the first time at the nth step given that it stated from a_i is called first-passage probability and is define as

$$f_{ij}^{(n)} = P\{X_n = j, X_{n-1} \ne j, X_{n-2} \ne j, \ldots, X_1 \ne j | X_0 = i\}$$

Definition: First-Passage Probability

This is the probability that the process is in state a_j at time $n$and not before, given that it was in state a_i at time 0.

This implies that the probability that n steps are required to reach state a_j for the first time given that the process startsfrom state a_i.

Clearly

$f_{ij}^{(0)} = 0$, the process is still at a_i

$f_{ij}^{(1)} = p_{ij}$, the one-step transition probability, $i \neq j$

Also,

$p_{jj}^{(0)} = 1, \quad i = j$

Then,

$$p_{ij}^{(n)} = \sum_{k=1}^{n} f_{ij}^{(k)} p_{jj}^{(n-k)}$$

$$= f_{ij}^{(n)} p_{jj}^{(0)} + \sum_{k=1}^{n-1} f_{ij}^{(k)} p_{jj}^{(n-k)}$$

$$= f_{ij}^{(n)} + \sum_{k=1}^{n-1} f_{ij}^{(k)} p_{jj}^{(n-k)}$$

or,

$f_{ij}^{(n)} = p_{ij}^{(n)} - \sum_{k=1}^{n-1} f_{ij}^{(k)} p_{jj}^{(n-k)}$

N/B $F^{(k)} p^{(n-k)}$ =joint probability of reaching state a_j in $1 \leq k \leq n-1$steps only. given that it started from a_i.

$f_{ij}^{(n)}$, can be obtained iteratively if $\{p_{ij}^{(n)}\}$ are known.

Example 16.2

Consider the problem of departmental transfer in chapter 17.

$$F^{(1)} = P = \begin{bmatrix} 0.5 & 0.25 & 0 \\ 0 & 0.5 & 0.5 \\ 0.75 & 0.25 & 0 \end{bmatrix}$$

$$p_{jj}^{(1)} = \begin{bmatrix} 0.5 & 0 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$f_{ij}^{(1)} p_{jj}^{(2-1)} = \begin{bmatrix} 0.5 & 0.25 & 0 \\ 0 & 0.5 & 0.5 \\ 0.75 & 0.25 & 0 \end{bmatrix} \begin{bmatrix} 0.5 & 0 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0.25 & 0.25 & 0 \\ 0 & 0.25 & 0 \\ 0.375 & 0.125 & 0 \end{bmatrix}$$

$$p_{ij}^{(2)} p^{(2)} = \begin{bmatrix} 0.250 & 0.500 & 0.250 \\ 0.375 & 0.375 & 0.250 \\ 0.375 & 0.500 & 0.125 \end{bmatrix}$$

$$F^{(2)} = p_{ij}^{(2)} - f_{ij}^{(1)} p_{jj}^{(1)} = \begin{bmatrix} 0.250 & 0.500 & 0.250 \\ 0.375 & 0.375 & 0.250 \\ 0.375 & 0.500 & 0.125 \end{bmatrix} - \begin{bmatrix} 0.25 & 0.25 & 0 \\ 0 & 0.25 & 0 \\ 0.375 & 0.125 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 0.25 & 0.25 \\ 0.375 & 0.125 & 0.25 \\ 0 & 0.375 & 0.125 \end{bmatrix}$$

$$F^{(3)} = p_{ij}^{3} - \sum_{k=1}^{2} f_{ij}^{(k)} p_{jj}^{(2-k)} = p_{ij}^{(3)} - f_{ij}^{(1)} p_{jj}^{(2)} - f_{ij}^{(2)} p_{jj}^{(1)}$$

$$= p_{ij}^{(3)} - f_{ij}^{(1)} p_{jj}^{(2)} - f_{ij}^{(2)} p_{jj}^{(1)}$$

16.5 Distribution of Number of Steps for First Passage

(i) For any fixed fair (i, j), the set $\{f_{ij}^{(n)}; n = 1, 2, \ldots\}$gives the distribution of the number of steps to get from i to j(the first passage). That is, the number of steps required to reach a_j for the first time.

(ii) The number of steps required to get from i to j is therefore a random variableN_{ij}, with

$$P\{N_{ij} = n\} = f_{ij}^{(n)}$$

16.6 First Return (Recurrence)

(i) If $j = i$, $f_{ii}^{(n)}$ gives the probability of the first return to state a_i. For example, the probability that the person transferred from department a_i will return to a_i for the first time at time n.

Corresponding to $f_{ij}^{(n)}$

$$f_{ii}^{(n)} = P\{X_n = i, X_{n-1} \neq i, X_{n-2} \neq i, \ldots, X_1 \neq i | X_0 = i\}$$

(iii) The equation relating $f_{ii}^{(n)}$ to $p_{ii}^{(n)}$ would also be the same.

But

$$f_{ii}^{(n)} = P\{X_i = i | X_0 = i\} = 1$$

(ii) Then N_{ii} is a random variable whose value is the recurrence of state a_i.

(iii) Since $\{f_{ii}^{(n)}\}$ for fixed i, j gives the distribution of N_{ij}, the mean first passage time from a_i to a_j denoted by m_{ij} is given by

$$m_{ij} = E(N_{ij}) = \sum_{n=1}^{\infty} n f_{ij}^{(n)}$$

(iv) where $i = j$, m_{ii} is the mean first recurrence time.

16.6.1 Calculation of m_{ij}

(1) The formula in (6.6) for m_{ij} would required the complete first passage time distribution for solution to be obtained.

(2) A simplification of the problem is obtained by conditioning the formula for m_{ij} on the state at step 1. That is, on one value of i at a time.

(3) Given that the process is in state a_i at time 0, either the next state is a_j in which case $N_{ij} = 1$, or it is in some other state a_k after which it enter state a_j, in which case the passage time will be $m_{kj} = 1 + N_{kj}$, the passage time from a_k to a_j.

(i) Thus

$$m_{ij} = 1p_{ij} + \sum_{k \neq j} (1 + m_{kj}) p_{ik}$$

$$= p_{ij} + \sum_{k \neq j} p_{ik} + \sum_{k \neq j} p_{ik} m_{kj}$$

$$= \sum_{\text{all } k} p_{ik} + \sum_{k \neq j} p_{ik} m_{kj}$$

$$= 1 + \sum_{k \neq j} p_{ik} m_{kj}$$

since $\sum_k p_{ik} = 1$, This expresses m_{ij} as a linear function of m_{kj} as the unknowns.

(ii) By using the same relation for other m_{ij}'s a complete set of linear equation (equation to the number of unknowns) can be expressed.

(iii) A solution of the linear equation gives the mean first passage time from any state into state j.

(iv) Mean first recurrence times are obtained in the same way.

Example 16.3

Consider the three-department job assignment. How many assignments will occur, in the average, before a man who is first assignment to a_1 (engineering) will be assigned to a_3 (sales)? That is, what is m_{13}?

Solution to Example 16.3

Using the formula for m_{ij}

$m_{13} = 1 + p_{11} m_{13} + p_{12} m_{23}$

There are two unknowns. Hence we form a similar equation for m_{23} as follows.

$m_{23} = 1 + p_{21} m_{13} + p_{22} m_{23}$

Now, recall that

$$P = \begin{bmatrix} 0.5 & 0.5 & 0 \\ 0 & 0.5 & 0.5 \\ 0.75 & 0.25 & 0 \end{bmatrix}$$

By substitution we obtain

$m_{13} = 1 + 0.5 m_{13} + 0.5 m_{23}$

$m_{23} = 1 \qquad\qquad + 0.5m_{23}$

Solving the simultaneous equation, we found that

$$m_{13} = 4 \text{ and } m_{23} = 2$$

Practice Questions

1. Define the term, ***steady state probability***
2. Write an expression for a limiting distribution.
3. Solve completely the problem in example 5.1, and draw all the graphs.
4. Use matrix multiplication and limiting probabilities to solve Problem 5.2.
5. In the post test in lecture four, obtain the stationary probabilities.
6. Define and write an expression for
   (a) First-passage probability.
   (b) First-return probability.
7. Using the post test of lecture four, find the mean first passage time from state 5 to state 4 by making state 4 absorbing. (This has nothing to do with states {1, 2}.) (*The University of Sydney, 2009*)
8. The transition matrix P of a Markov chain $X = (X_n : n \geq 0)$ is

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

(a) Specify the classes of this chain and determine whether they are transient, null recurrent or positive recurrent.

(b) Find all stationary distributions for this chain.

(c) Find the mean recurrence time m_{jj} for all positive recurrent states.

(*The University of Sydney, 2010*)

CHAPTER 19
CHAPMAN-KOLMOGOROV EQUATIONS AND CLASSIFICATION OF STATES

17.1 Introduction

The n[th]-step transition probabilities P_{ij}^n is the probability that a process in state i will be in state j after n additional transitions that is,

$$P_{ij}^n = P\{X_{n+m} = J | X_m = i\}, n \geq 0, i, j \geq 0.$$

The Chapman-Kolmogorov equations provide a method for computing these n-step transition probabilities. These equations are:

$$P_{ij}^{x} P_{ik}^{n} P_{kj}^{m} \quad \textit{for } \text{all n, m} \geq 0, \text{all i, j}$$

and are established by observing that

$$P_{ij}^{n+m} = P\{X_{n+m} = J | X_0 = i\}$$

$$= \sum_{K=0}^{\infty} P\{X_{n+m} = J, X_n = K | X_0 = i\}$$

$$= \sum_{K=0}^{\infty} P\{X_{n+m} = J, | X_n = K, X_0 = i\} P\{X_n = K | X_0 = i\}$$

$$= \sum_{K=0}^{\infty} P_{kj}^{m} P_{ik}^{n}$$

If we let $P^{(n)}$ denote the matrix of n-step transition probabilities, P_{ij}^n, then it can be asserted that

$$P^{(n+m)} = P^{(n)}.P^{(m)}$$

where the dot represents matrix multiplication.

Hence,

$$P^{(n)} = P.\ P^{(n-1)} = P.P.P^{(n-2)} = \ldots = P^n,$$

and thus $P^{(n)}$ may be calculated by multiplying the matrix P by itself n times.

p_{oj} is said to be Accessible from state i if for some $P_{ij} > 0$. Two states i and j accessible to each other is said to communicate and we write $i \leftrightarrow j$.

Let P_{ij} denotes the one-step transition probabilities, and $P_{ij}^{i} = P_{j}$

Observe that P_{ik}^{n} P_{kj}^{m} represents the probability that starting in i the process will go to state j in $n + m$transitions through a path which takes it into K at the nth transition.

17.1.1 Proof of $C - K$ Equations

Using remark (3) above, summing over all intermediate states $k$yields the probability that the process will be in state j after n + m transitions. We have $P_{rj}^{n+m} = P\{X_{n+m} = J|X_0 = i\}$

$$P_{ij}^{n+m} = \sum_{k=0}^{\infty}\{X_{n+m} = J, X_n = K|X_0 = i\}$$

$$= \sum_{k=0}^{\infty} P\{X_{n+m} = J,|X_{n=k}, X_0 = i\} P\{X_n = K|X_0 = 1\}$$

$$= \sum_{k=0}^{\infty} P_{kj}^{m} \ P_{ik}^{n}$$

Matrix of n-step transition probabilities: $P^{(n)}$

Let $P^{(n)}$ denote the matrix of n -step transition probabilities P_{ij}^{n} then the C-K Equation asserts that

$$P^{(n+m)} = P^{(m)} \ P^{(m)}$$

By induction

$$P^{(n)} = P^{(n-1+k)} = P^{n-1} \ P^{*} = P^{n}$$

That is the n -step transition matrix may be obtained by multiplying the matrix P by itself n times.

Example 17.1

Consider example in which the weather is considered as a two-state Markov chain. If $\alpha = 0.7$ and $p = 0.4$, the calculate the probability that it will rain four days from today given that it is raining today.

Solution:

The one-step transition on probability matrix is given by

$$P = \begin{pmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{pmatrix}$$

$$Hence, P^{(2)} = P^2 = \begin{pmatrix} 07 & 0.3 \\ 0.4 & 0.6 \end{pmatrix} \begin{pmatrix} 0.7 & 0.3 \\ 0.4 & 0.5 \end{pmatrix}$$

$$= \begin{pmatrix} 0.61 & 0.39 \\ 0.52 & 0.48 \end{pmatrix}$$

$$P^{(4)} = (P^{(2)})^2 = \begin{pmatrix} 0.61 & 0.39 \\ 0.52 & 0.48 \end{pmatrix} \begin{pmatrix} 0.61 & 0.39 \\ 0.52 & 0.48 \end{pmatrix}$$

$$\begin{pmatrix} 0.579 & 0.4251 \\ 0.5668 & 0.4332 \end{pmatrix}$$

Hence. the required probability P_{00}^{4} equal 0.5749.

Example 17.2

Consider Example 2.4. Given that it rained on Monday and Tuesday, what is the probability that it will rain on Thursday?

Solution

The two-step transition matrix is given by

$$P^{(2)} = P^2 = \begin{pmatrix} 0.7 & 0 & 0.3 & 0 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0.4 & 0 & 0.6 \\ 0 & 0.2 & 0 & 0.8 \end{pmatrix} \begin{pmatrix} 0.7 & 0 & 0.3 & 0 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0.4 & 0 & 0.6 \\ 0 & 0.2 & 0 & 0.8 \end{pmatrix}$$

$$== \begin{pmatrix} 0.49 & 0.12 & 0.21 & 0.18 \\ 0.35 & 0.20 & 0.15 & 0.30 \\ 0.20 & 0.12 & 0.20 & 0.48 \\ 0.10 & 0.16 & 0.10 & 0.64 \end{pmatrix}$$

Since rain on Thursday is equivalent to the process being in either state 0 or state 1 on Thursday, the required probability is given by $P_{00}^2 + P_{01}^2 = 0.49 + 0.12 = 0.61$

17.2 Classification of States

In order to analyze precisely the asymptotic behaviour of the Markov chain process, we need to introduce some principles of classifying state of a Markov chain.

Properties to be classified include: Accessible, Communicate, A periodic, Recurrent, Transient, and Irreducible. Definitions of these properties now follow:

17.2.1Irreducible Property

We say that the Markov chain is irreducible if there is only one class– i.e. if all states communicate with each other.

Proposition

Communication is a exultance relation. That is

(*i*) $i \leftrightarrow i$;

(*ii*) *If* $i \leftrightarrow j$, then $j \leftrightarrow i$;

(iii) *If* $i \leftrightarrow j$, then $j \leftrightarrow i$; then $i \leftrightarrow k$.

Proof: the 1[st] two parts follow trivially from the definition of communication. To prove (iii) suppose that $i \leftrightarrow j$., and $j \leftrightarrow k$ then there exists m, n such that $P_{ij}^m > 0$, $P_{ij}^n > 0$. Hence,

$$P_{ik}^{m+n} = \sum_{r=0}^{\infty} P_{ir}^m \; P_{rk}^n \geq P_{ij}^m \; P_{jk}^n > 0$$

Similarly, we may show there exists an S for which $P_{ki}^s > 0$. Two states that communicate are said to be in the same class and by the proposition any two classes are either disjoint or identical. We say that the Markov chain is *Irreducible*. If there is only one class-- that is, if all states communicate with each other.

State is said to have period d if $P_{ii}^n = 0$, whenever n is not divisible by d and d is the greatest integer with this property. (If $P_{ii}^n = 0$, for all n > 0; then define the period of i to be infinite). A state with period 1 is said to be *A periodic*. Let d (i) denote the period of i, we can show that periodicity is a class property.

17.2.2Recurrent (or Persistent) State

A state $i \in S$ is said to be Recurrent if $\Pr(T_i < \infty) = 1$ where T_i is the number of steps it takes for the chain to finally visit i.

17.2.3Transient State

A state $i \in S$ is said to be transient if $\Pr(t_i < \infty) < 1$ where T_i is the number of stops it takes for the chain to finally visit i.

Example 17.3

Suppose that the weather on any day depends on the weather condition for the previous two days. To be exact, suppose that if it was sunny today and yesterday, then it will be sunny tomorrow with probability 0.8; if it was sunny today but cloudy yesterday, then it will be sunny tomorrow with probability 0.6; if it was cloudy today but sunny yesterday, then it will be sunny tomorrow with probability 4; if it was cloudy for the last two days, then it will be sunny tomorrow with probability 1.

Definitely. the model above is not a Markov chain. However, such a model can be transformed into a Markov chain.

(a) Transform this into a Markov chain

(b) Obtain the transition probability matrix

(c) Find the stationary distribution of this Markov chain.

Solution

(a) Suppose we say that the state at any time is determined by the weather conditions during both that day and the previous day. We say the process is in:

State (S, S) if it was sunny both today and yesterday;

State (S. C) if it was sunny both yesterday but cloudy today;

State (C, S) if it was cloudy yesterday but sunny today;

State (C, C) if it was cloudy both today and yesterday

(b) The transition probability matrix is

Today's state

Yesterday's state	(S, S)	(S, C)	(C, S)	(C, C)
(S, S)	.8	.2	0	0
(S, C)	0	0	.4	.6
(C, S)	.6	.4	0	0
(C, C)	0	0	.1	.9

2. An airline reservation system has two computers only one of which is in operation at any given time. A computer may brake down on any given day which probability p. there is a single repair facility which takes at least 2 days to restore a computer to normal. The facilities are such that only one computer and a time can be dealt with.

(a) Form a Markov chain by taking as states the pairs (x, y) where x is the number of machines in operating condition at the end of a day and y is 1 if a day's labour has been expended on a machine not yet repaired and 0 otherwise.

(b) Obtain the transition matrix

(c) Find the stationary distribution in terms of p and q where p + q = 1

Solution

The state are (2, 0), (1, 0), (1, 1), (0, 1).

The transition matrix is

From state / *To* state→	(2, 0)	(1, 0)	(1, 1)	(0, 1)
P = (2, 0)	q	p	0	0
(1, 0)	0	0	q	p
(1, 1)	q	p	0	0
(0, 1)	0	1	0	0

17.3 Discrete Time Process

1. Consider a series of events E resulting from the repetition of the same experiment and occurring consecutively. The common examples are telephone calls, average customers at a service point, chromosomes breakages and radiation, and so on
2. The occurrence are assumed to be of the same kind γ. The number n of events in a given interval t is a random variable.
3. Let $z(t)$ denote the total number of occurrences within an arbitrary time interval t.
4. Let $P_n(t) = P(z(t) = n)$.

Assumptions

i. $P_n(t)$ depends only on the time interval of duration t, and does not depend on the initial instant.

ii. The probability that E will occur more than once is the time interval that is infinitesimally small (that is, negligible).

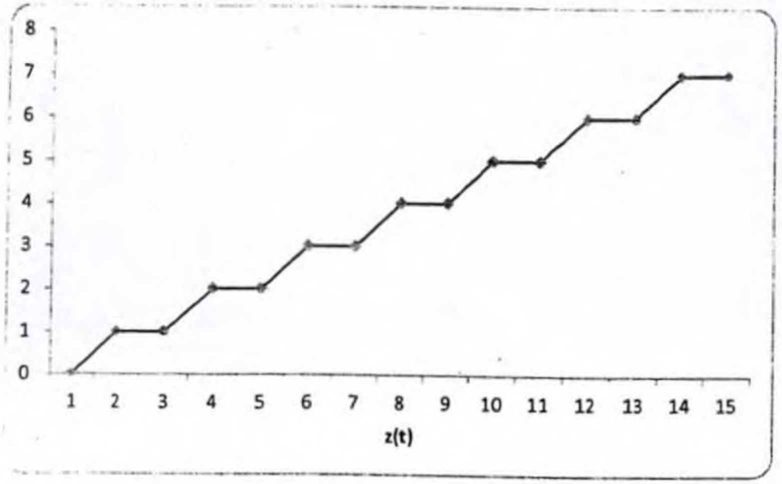
iii. The probability that E occur once in the interval dt is proportionaly to that interval and is written As λdt.

Assumptions on $z(t)$

i. the initial $z(t)$ is 0.

ii. $z(t))$ increases by 1 when E occurs.

iii. $z(t)$ remains constant when E does not occur.



iv. $z(t) = 0, 1, 2, \ldots, n, \ldots$ .At random instants $t_1, t_2, \ldots, t_n$, ...it jump abruptly from 0 to 1 , 1 to 2, and 2 to 3, … . The increment of $z(t)$ at time interval t is equal to the number n of events that have occurred.

v. If we know the value of $z(t_0)$att_0(the initial instant) we can find the value at $t = t_0 + \Delta t$.

$z(t) = z(t_0 + \Delta t)$

$$= z(t_0) + z(\Delta t)$$

$$= z(t_0) + n$$

The increment $z(\Delta t) = n$ is characterized by the properties of the probability of occurrence in the time interval if

i. its probability is $P_n(\Delta t)$

ii. it is independent of the values of $z(t)$prior to t_0.

Thus if $z(t_0)$ is known, the value $z(t)$ (which is determined by the probability of occurrence in the interval z) depends solely on the law of probability that governs the increment $n$after t_0.

The random variable$z(t)$ follows a defined poisson process and constitute an example of a Markov chain. It is defined completely by the probability

$$p_n(t) = \frac{(\Box t)^n e^{-\Box t}}{n!} \qquad n = 0, 1, \ldots$$

17.4 The Poisson Process

Under the assumption of the continuity of time we can expand $p_1(\Delta t)$ in Maclaurin series

$p_1 \Delta t = p_1(0) + p_1'(0)\Delta t + \frac{1}{2!}p_1''(0)(\Delta t)^2 + \cdots$

$= p_1(0) + p_1'(0)\Delta t + o(\Delta t)^2$

But $p_1(0) = 0$

The probability that Ecan occur 0 time or once is

$p\{z(\Delta t) = 0 \text{ or } z(\Delta t) = 1\} = p_0\Delta t + p_1 \Delta t$

Now consider $p_n \Delta t$

$p_0\Delta t + p_1\Delta t + \cdots \ = 1$

which is necessary.

$p_0 \Delta t = 1 - p_1\Delta t - p_2 \Delta t - \cdots$

$= 1 - p_1 \Delta t$

This is so because of the assumption $1 - p_1 \Delta t - o(\Delta t)^2$. But our only interest is in $p_n(t)$.

Event E can occur precisely n times during the interval $t + \Delta t$ if the following mutually exclusive events are time.

i. E occurs n times in the interval t, 0 times in the interval Δt.

ii. E occurs $n - 1$ times in interval t, once in the interval Δt.

iii. E occurs $n - 2$ times in the interval t, twice in the interval Δt.

And so on.

These lead to the following:

$$p_n(t + \Delta t) = p_n(t)p_0\Delta t + p_{n-1}(t)p_1\Delta t + p_{n-2}(t)p_2\Delta t + \cdots$$

$$p_1\Delta t = 1 - p_1'(0)\Delta t - o(\Delta t)^2$$

since $p_1(0) = 0$

Let $p_1'(0) = \lambda$, then

$$p_0' \Delta t = 1 - \lambda\Delta t - o(\Delta t)^2$$

$p_1\Delta t$ can be written as

$$p_1\Delta t = \lambda\Delta t + 0(\Delta t)^2$$

So that

$$p_n(t + \Delta t) = p_n(t)[1 - \lambda\Delta t] + p_{n-1}(t)\lambda\Delta t + o(\Delta t)^2$$

$$= p_n(t) - p_n(t)\lambda\Delta t + p_{n-1}(t)\lambda\Delta t + o(\Delta t)^2$$

$$p_n(t + \Delta t) - p_n(t) = [p_{n-1}(t) - p_n(t)]\lambda\Delta t + o(\Delta t)^2$$

Divide both side by Δt

$$\frac{p_n(t + \Delta t) - p_n(t)}{\Delta t} = \lambda[p_{n-1}(t) - p_n(t)] + o(\Delta t)^2$$

As $\Delta t \to 0$

$$\frac{d}{dt}p_n(t) = P_n'(t) = \lambda p_{n-1}(t) - \lambda p_n(t), \qquad n = 0, 1, 2, \ldots$$

This equation does not hold for $n = 0$. We can use the forward chapman-kolmogorov equation.

$$p_0(t + \Delta t) = p_0(t)p_0(\Delta t)$$

$$= p_0(t)[1 - \lambda\Delta t]$$

$$\frac{p_0(t + \Delta t) - p_0(t)}{\Delta t} = -\lambda p_0(t)$$

In the limit as $\Delta t \to 0$

$$p_0'(t) = -\lambda p_0(t)$$

At the beginning of the interval t, we have

$$p_0(0) = 1 \text{ and } p_n(0) = 0, n \neq 0$$

Divide the limit result by $p_0(t)$, we have

$$\frac{p_0'(t)}{p_0(t)} = -\lambda = \frac{1}{p_0(t)}p_0'(t)$$

$$\text{or} \frac{d}{dt} log_e p_0(t) = -\lambda$$

$$\int \frac{d}{dt} log_e p_0(t) = -\lambda \int dt$$

$$log_e p_0(t) = -\lambda t$$

$$p_o(t) = e^{-\lambda t}$$

$\frac{d}{dt}p_n(t)$ can still be written as

$$Dp_n(t) = \lambda p_{n-1}(t) - \lambda p_n(t) \quad , \quad n > 0$$

At $n = 1$

$$Dp_1(t) = \lambda p_0(t) - \lambda p_1(t)$$

At $n = 2$

$$Dp_2(t) = \lambda p_1(t) - \lambda p_2(t)$$

At $n = 3$

$$Dp_3(t) = \lambda p_2(t) - \lambda p_3(t)$$

$Dp_1(t)$ can be rewritten as

$$(D+\lambda)p_1(t) = \lambda p_0(t)$$

So that we have

$$(D+\lambda)p_1(t) = \lambda e^{-\lambda t}$$

Divide through by $(D+\lambda)$

$$p_1(t) = \frac{\lambda e^{-\lambda t}}{D+\lambda} = \frac{\lambda e^{-\lambda t}t^r}{(r+1)!\,1!}$$

This is a general solution.

Notice that $\frac{\lambda e^{-\lambda t}t^r}{(r+1)!\,1!}$ may also be written as $\frac{\lambda e^{-\lambda t}t^r}{r!\,1!}$

$$Now\ p_n(t) = \frac{\lambda^n e^{-\lambda t}}{D+\lambda} = \frac{\lambda^n t^j e^{-\lambda t}t^r}{(r+j)!\,1!}, \qquad j = 0, 1, 2, \ldots$$

so that

$$p_1(t) = \frac{\lambda e^{-\lambda t}}{D+\lambda} = \frac{\lambda t^j e^{-\lambda t}t^r}{(r+j)!\,1!} \quad , j = 0, 1, 2, \ldots$$

put $r = 1$, $j = 0$

$$p_1(t) = \frac{\lambda t^0 e^{-\lambda t}t^1}{(1+0)!\,1!} = \lambda t e^{-\lambda t} \quad = (\lambda t)e^{-\lambda t}$$

Then

$$(D+\lambda)p_2(t) = \lambda^2 t e^{-\lambda t}$$

$$p_2(t) = \frac{\lambda^2 t e^{-\lambda t}}{D+\lambda} = \frac{\lambda^2 t^j e^{-\lambda t}t^r}{(r+j)!\,1!}$$

put $r = 1$, $j = 1$

$$p_2(t) = \frac{\lambda^2 t^2 e^{-\lambda t}}{2!} = \frac{(\lambda t)^2 e^{-\lambda t}}{2!}$$

Consequently

$$p_3(t) = \frac{\lambda^3 t e^{-\lambda t}}{D+\lambda} = \frac{\lambda^3 t^j e^{-\lambda t}t^r}{(r+j)!\,1!}$$

put $r = 1$, $j = 2$

$$p_3(t) = \frac{\lambda^3 t^3 e^{-\lambda t}}{3!} = \frac{(\lambda t)^3 e^{-\lambda t}}{3!}$$

In general,

$$p_n(t) = \frac{(\lambda t)^n e^{-\lambda t}}{n!} \quad , n = 0, 1, 2,$$

If we fix t, λt is a fixed parameter for the distribution and the set $p_1(t), p_2(t)\ldots$ then gives a probability distribution of the process at the fixed time interval which is a Poisson distribution. In terms of a counts of events the above results shows that the member of events occurring in a fixed time interval t is distributed as a Poisson with parameter λt.

Also since the mean of the Poisson distribution is equal to the parameter λt, λt can be interpreted as the expected number of events that can occur in time t. the quantityλ is the average or mean rate of occurrence of E.

17.5 Continuous Time Process

A continuous-time Markov chain is a stochastic process having the Markovian property that the conditional distribution of the future state at time$t + s$, given the present state at t and all past states depends only on the present state and is independent of the past. Thus, this lecture establishes the fact that a continuous time process is also distributed as an exponential probability

17.5.1 Definition and Properties

Consider a continuous-time stochastic process $\{X_{(t)}, t \geq 0\}$ taking on values in the set of non-negative integers. In analogy with the definition of a discrete-time Markov chain, given earlier, we say that the process $\{X_{(t)}, t \geq 0\}$ is a continuous-time Markov chain if for all $s, t \geq 0$ and non-negative integers $i, j, X_{(u)}$ $0 \leq u \leq s$,

$$P\{X_{(t+s)} = j | X_{(s)} = i, X_{(u)} = x_{(u)}, 0 \leq u < s\}$$
$$= P\{X_{(t+s)} = j | X_{(s)} = 1\}$$

If, in addition $P\{X_{(t+s)} = j | X_{(s)} = 1\}$ is independent of s, then the continuous-time Markov chain is said to have stationary or homogeneous transition probabilities. All Markov chains we consider will be assumed to have stationary transition probabilities.

Suppose that a continuous-time Markov chain enters state i at some time, say time 0, and suppose that the process does not leave state i (that is, a transition does not occur) during the next s time units. What is the probability that the process will not have state i during the following t time units?

To answer this, note that as the process is in state i at time s, it follows, by the Markovian property, that the probability it remains in that state during the interval $[s, s + t]$is just the (unconditional) probability that it stays in state i for at least t time units. That is, if we test t_i denote the amount of time that the process stays in state i before making a transition into a different state, then

$$P = \{T_i > S + t | T_i > S\} = P\{T_i > t\}$$

For all $s, t \geq 0$. Hence, the random variable T_i is memoryless and must thus be exponentially distributed.

The above gives us a way of construction a continuous-time Markov chain, namely, it is a stochastic process having the properties that each time it enters state i:

(i) the amount of time it spends in that state before making a transition into a different state is exponentially distributed with rate say v_i; and

(ii) when the process leaves state i, it will next enter state j with same probability, call it p_{ij}, where $\sum j \neq i \, p_{ij} = 1$.

A state i for which $v_i = \infty$ is called an instantaneous state since when entered it is instantaneously left. Whereas such states are theoretically possible, we shall; assume throughout that $0 \leq v_i < \infty$ for all i. (If $v_i = 0$, then state i is called absorbing since once entered it is never life).

Hence, for our purposes or continuous-time Markov chain is a stochastic process that moves from state to state in accordance with a (discrete-time) Markov chain, but is such that the amount of time it spends in each state, before proceeding to the next state is exponentially distributed. In addition, the amount of time one process spends in state i and the next state visited, must be independent random variables. For if the next state visited were dependent on T_i, then information as to how long the process has already been in state i would be relevant to the prediction of the next state-and this would contradict the Markovian assumption.

A continuous-time Markov chain is said to be regular if, with probability 1, the number of transitions in any finite length of time is finite. An example of a non-regular Markov chain is the one having.

$$p_i, i+1 = 1, v_i = i^2$$

It can be shown that this Markov chain-which always goes from state i to $i+1$, spending an exponentially distributed amount of time with mean $^1/_i{}^2$ in state i – will, with positive probability, make an infinite number of transitions in any time interval of length $t, t > 0$. We shall assume from now on that all Markov chains considered are regular.

Let q_{ij} be defined by

$$q_{ij} = v_i p_{ij}, \ \nabla i \neq j$$

Since v_i is the rate at which the process leaves state i and p_{ij} is the probability that it then goes to j, it follows that q_{ij} is the rate when in state i that the process makes a transition into state j; and in fact we call q_{ij} the transition rate from i to j.

Let us denote by $p_{ij(t)}$ the probability that a Markov chain, presently in state i, will be in state j after an additional time t

$$p_{ij(t)} = P\{X_{(t+s)} = j | X_{(s)} = i\}$$

17.6 The Exponential Process

Let us consider a finite state but continuous time process. Let $X(t)$ denote a random variable. The value of $X(t)$ at fixed t is the state of the process at time t.

A time dependent process is the set ($X(t)$for given $t \geq 0$. $X(t_1)$depends on $t_1 > t_0$, and not on $t_2, t_1, < t_2$. The process is continuos if t can take value on the t-axis.

Definition

A continuous time stochastic process is said to have Markov property and is called a continuous time Markov process if for all $t_n > t_{n-1} > \cdots . t_1 > t_0$ satisfying the condition $t_n > t_{n-1} > \cdots . t_1 > t_0$.

1. $$P(X(t_n) = j_n / X(t_{n-1}) = j_{n-1}, X(t_{n-2}) = j_{n-2}, \ldots, X(t_0) = j_0)$$

$$= P(X(t_n) = j_n / X(t_{n-1}) = j_{n-1})$$

This is the independent probability and it state that all that is needed to predict the state of the process at time n is the state of the process at the immediately preceding time.

2. A Markov process is said to be time-homogeneous or stationary if

$$P\{X(t_2) = j / X(t_1) = i\} = P\{X(t_2 - t_1) = j / X(0) = i\} \nabla i \text{ and } j, t_1 < t_2$$

In words, the process is stationary or time homogeneous if the conditional probability in (2) depends only on the time interval between the events considered, rather on the absolute time. Note that 'time-homogeneous' and 'stationary' denote sameness in time. We can also know that a stationary Markov process is defined completely by the transitional probability function which we defined as

$$p_{ij}(t) = p\{x(t) = j / x(0) = i\}$$

The fundamental equation for stationary Markov process is Chapman-Kolmogorov equation for $p_{ij}(t+\tau)$. By definition,

$$p_{ij}(t+\tau) = P\{X(t+\tau) = j / X(0) = i\}$$

$$= \sum_k P\{X(t+\tau) = j, X(t) = k / X(0) = i\} \text{ Marginal from joint}$$

Using Markov assumption

$$= \sum_k P\{X(t+\tau) = j / X(t) = k, X(0) = i\} P\{X(t) = k, X(0) = i\}$$

But $P\{X(t) = k, X(0) = i\} = P\{X(t) = k/X(0) = i\}P\{X(0) = i\}$

Therefore,

$$p_{ij}(t+\tau) = \sum_k P\{X(t+\tau) = j/X(t) = k, X(0) = i\}\, P\{X(t) = k/X(0) = i\}$$

This is because $P\{X(0) = i\} = 1$

Thus

$$\sum_k P\{X(t+\tau) = j/X(t) = k\}P\{X(t) = k/X(0) = i\}$$

By the stationary assumption in (2)

$$p_{ij}(t+\tau) = \sum_k P\{X(t) = j/X(0) = k\}\, P\{X(t) = k/X(0) = i\}$$

$$= \sum_k p_{kj}(\tau)p_{ik}(t) \text{ (By definition)}$$

This is the general form of Chapman-Kolmogorov equation.

A specified form of this is:

$$p_{ij}(t+\Delta t) = \sum_k p_{ik}(t)p_{kj}(\Delta t)$$

The above is forward Chapman-Kolmogrov equation.

The forward Chapman-Kolmogrov equation is given as

$$p_{ij}(\Delta t + t) = \sum_k p_{ik}(\Delta t)p_{kj}(t)$$

We expect the following to hold

i) $0 \le p_{ij}(t) \le 1 \quad for\ all\ t$

ii) $p_{ij}(0) = P\{X(0) = j/X(0) = i\} = 1 \qquad , i = j$

$= 0 \qquad , i \ne j$

And for any given i

iii) $\sum_j p_{ik}(t) = \sum_j P\{X(t) = j/X(0) = i\} = 1$

Under the assumption that $p_{ij}(t)$ is a continuous function of t, we can express $p_{ij}(\Delta t)$ by the use of Maclaurin series.

$$p_{ij}(\Delta t) = p_{ij}(0) + p'_{ij}(0)(\Delta t) + \frac{1}{2!}p''_{ij}(0)(\Delta t)^2 + \frac{1}{3!}p'''_{ij}(0)(\Delta t)^3 + \cdots$$

$$= p_{ij}(0) + p'_{ij}(0)\Delta t + 0(\Delta t)^2$$

Let $p'_{ij}(0) = \lambda_{ij}$

$p_{ij}(\Delta t) = p_{ij}(0) + \lambda_{ij}\Delta t + o(\Delta t)^2 \qquad$ for $i \ne j$

$p_{ij}(\Delta t) = \lambda_{ij}\Delta t + o(\Delta t)^2 \qquad$ for $i = j$

Also, let $p'_{jj}(0) = \lambda_{jj}$

$$p_{ij}(\Delta t) = 1 + p'_{jj}\Delta t + o(\Delta t)^2$$

$$= 1 + \lambda_{jj}\Delta t + o(\Delta t)^2$$

Since $p_{ij}(0) = 0 \quad for\ i \ne j$ is a minimum, λ_{ij} is positive. Also, since $p_{ij}(0) = 1$ For $i = j$ is a maximum, λ_{ij} is non-positive.

We can unite the forward Chapman-Kolmogorov.

$$p_{ij}(t+\Delta t) = \sum_k p_{ik}(t)p_{kj}(\Delta t)$$

$$= p_{ij}(t)p_{jj}(\Delta t) + \sum_{k \ne j} p_{ik}(t)p_{kj}(\Delta t)$$

Which we can write as

$$p_{ij}(t+\Delta t) = p_{ij}(t)\left[1+\lambda_{ij}\Delta t + 0(\Delta t)^2\right] + \sum_{k\neq j} p_{ik}(t)\left[\lambda_{kj}\Delta t + 0(\Delta t)^2\right]$$

$$= p_{ij}(t) + p_{ij}\lambda_{jj}\Delta t + p_{ij}(t)0(\Delta t)^2 + \sum_{k\neq j}\left[p_{ik}(t)\lambda_{kj}\Delta t + p_{ik}(t)0(\Delta t)^2\right]$$

$$\frac{p_{ij}(t+\Delta t) - p_{ij}(t)}{\Delta t}$$

$$= \left[p_{ij}(t)\lambda_{jj} + \sum_{k\neq j} p_{ik}(t)\lambda_{kj}\right] + \left[\frac{\sum_{k\neq j} p_{ij}(t)0(\Delta t)}{\Delta t} + \frac{\sum_{k\neq j} p_{ik}(t)0(\Delta t)^2}{\Delta t}\right]$$

$$\sum_k p_{ik}(t)\lambda_{kj} + \frac{\sum_k p_{ik}(t)0(\Delta t)^2}{\Delta t}$$

The limit as $\Delta t \to 0$

$$\frac{dp_{ij}(t)}{dt} = \sum_k p_{ik}(t)\lambda_{kj}$$

In matrix form,

$$\frac{dP(t)}{dt} = P(t)\wedge \qquad (\wedge = diagonal\ element)$$

$$\frac{dP(t)}{dt} = \left\{\frac{dp_{ij}(t)}{dt}\right\}, \qquad \wedge = \{\lambda_{ij}\}$$

$$P(t) = \{p_{ij}(t)\}$$

But, $\sum_j p_{ij}(t) = 1$

$$\frac{d}{dt}\left[\sum_j p_{ij}(t)\right]_{t=0} = 0$$

$$\sum_j \left.\frac{dp_{ij}(t)}{dt}\right|_{t=0} = 0$$

$$\sum_j p'_{ij}(0) = 0$$

$$\sum_j \lambda_{ij} = 0$$

$$\sum_j \lambda_{ij} = \lambda_{jj} + \sum_{j\neq i}\lambda_{ij} = 0$$

$$\lambda_{jj} = -\sum_{j\neq i}\lambda_{ij}$$

Thus since every of $\wedge$ (diagonal element) is non-negative, the diagonal element λ_{jj} must be equal in magnitude and opposite in signal to the sum of the other element in the same row. λ_{ij} is called the transition rate from i to j for $i \neq j$. λ_{ij} can be interpreted as the parameter of negative exponential distribution. For each λ_{ij}, the exponential distribution gives the distribution of time spent in a state i, given that j is the next step. Thus if T_{ij} is the random variable with λ_{ij}

$$E(T_{ij}) = \frac{1}{\lambda_{ij}}$$

So that λ_{ij} can be estimated as the inverse of a sample mean.

$$f(x) = \lambda e^{-\lambda x}$$

or $f(t) = \lambda_{ij}e^{-\lambda_{ij}t}, \qquad x > 0$

with mean

$$E(T_{ij}) = \int_0^\infty tf(t)dt = \int_0^\infty t\lambda_{ij}e^{-\lambda_{ij}t}dt$$

$$= \frac{1}{\lambda_{ij}} \int_0^\infty (\lambda_{ij} t) e^{-\lambda_{ij} t} \, d(\lambda_{ij} t)$$

$$\frac{1}{\lambda_{ij}} \Gamma(2) = \frac{1}{\lambda_{ij}} \text{ since } \Gamma(2) = 1$$

Suppose that we have the likelihood

$$L = \prod_{i=1}^{n} f(t) = \lambda_{ij}^{n} e^{-\sum \lambda_{ij} t}$$

$$= n \log \lambda_{ij} - \sum \lambda_{ij} t$$

$$\frac{d \log L}{d\lambda_{ij}} = \frac{n}{\lambda_{ij}} - \sum t = 0$$

$$\frac{1}{\lambda_{ij}} = \frac{\sum t}{n} = \bar{t}$$

This implies that,

$$\lambda_{ij} = {}^{1}/_{\bar{t}}$$

Practice Questions

1. Considered a two-state process such as the operation of a loom for weaving cloth. The two-state for the looms are 0, the loom is shut off and the operator is repairing it. And 1, the loom is operating and the operator is idle. Consider the operating and repair time as continuous. Assume that the constant proportionality is 3 for repair transition and 2 for breakdown transition. Find the probability distribution of the repair and the operation time.
2. Obtain the general form of the Chapman-Kolmogorov $(C - K)$ equation.
3. Show that λ_{ij} transition from i to j, $\forall\, i \neq j$, is ${}^{1}/_{\bar{t}}$.

CHAPTER 20

INTRODUCTION TO THE THEORY OF GAMES AND QUEUING MODELS

18.1 Games Theory

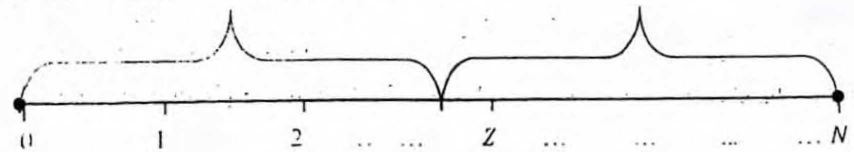
Games theory is a branch of Stochastic Processes that can be applied to a situation such as business, stock trading, politics, and so on, where the person involved can be referred to as a ***player*** or simply a ***gambler***

18.2 Gambler's Ruin

Consider a gambler who plays a game of chance against an adversary. Suppose that at the start of the game, the gambler deposits an amount in naira Z. The adversary deposit $N - Z$ in naira where N is the cumulated initial capital.

1. The role of the game is that if the gambler wins a game he takes ₦1 from the adversary and loses same to the adversary otherwise.
2. The game terminates. If either player loses all his deposits. When the gambler loses all his deposit, he is said to be ruined.
3. No game is jumped.

We can put the money on a number scale.



The gain or loss is represented by movement along the scale. Gambler's gain is represented by movement to the right observed and its loss represented by movement to the lefty observed.

No point on the scale is jumped. Movement in either direction on the scale is by pure chance. The movement along the scale can be seen as that of a particle that moves at

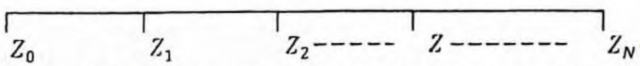
random forward and backward. Because of that the process is known as ***random walk***. The points on the scale represent the state of the process.
Movement to a point on the scale depends on the point the gambler (or adversary) is at currently. It is therefore a Markov Chain.

We shall approach this problem by attempting the following questions:
Q1 - What is the probability that a gambler with the initial capital Z will be ruined?
Q2 - What is the expected gain of the gambler?
Q3 - What is the expected duration of the game?

18.2.1 Probability of Gambler's Ruin

Let p denote the probability that the gambler will move to the right of Z. That is, the probability of winning a game.
Let $q = 1 - p$, the probability of moving to the left of Z, that is losing a game (by the gambler). Let the points on the scale be denoted by $Z_0, Z_1, \ldots, Z_N$ and q_{Z_j}, the probability of ruin given the initial capital Z_j.

For simplicity, let $N = 5$ (five naira). Assume that the initial capital by the gambler is Z_2



The probability that the gambler will be ruined if his initial capital is Z_2 is

$$P\{R|Z_2\} = P\{Z_3, R\} + P\{Z_1, R\}$$

$$= P\{R|Z_3\}P\{Z_3\} + P\{R|Z_1\}P\{Z_1\}$$

We can write this as

$$q_{Z_2} = pq_{Z_3} + qq_{Z_1}, \qquad 1 \le Z_j \le N - 1$$

Since $P\{R|Z_j\} = q_{Z_j}$, $P\{Z_k\} = p$ and $P\{Z_i\} = q$

Generally (10.1) can be written as

$$q_{Z_j} = pq_{Z_k} + qq_{Z_i}, \qquad 1 \le Z \le N - 1$$

Systems like these are known as difference equation. We can write the unit factor on the left as $p + q = 1$. That is.

$$(p+q)q_{Z_2} = pq_{Z_3} + qq_{Z_1}$$
$$pq_{Z_2} + qq_{Z_2} = pq_{Z_3} + qq_{Z_1}$$
$$q(q_{Z_1} - q_{Z_2}) = q(q_{Z_2} - q_{Z_3})$$

This implies that

$$q_{Z_1} - q_{Z_2} = r(q_{Z_2} - q_{Z_3}), \quad \text{where } r = {}^{p}/_{q}$$

Thus we can have the following system of equation.

$$q_{Z_0} - q_{Z_1} = r(q_{Z_1} - q_{Z_2})$$

$$q_{Z_1} - q_{Z_2} = r(q_{Z_2} - q_{Z_3})$$

$$q_{Z_2} - q_{Z_3} = r(q_{Z_3} - q_{Z_4})$$

$$q_{Z_3} - q_{Z_4} = r(q_{Z_4} - q_{Z_5})$$

To unify these equations we define

$q_{Z_0} = 1$, $q_{Z_N} = q_{Z_5} = 0$

These are boundary conditions on q_Z. This becomes

$$q_{Z_3} - q_{Z_4} = rq_{Z_4}$$

This extends to other equation in the system

$$q_{Z_2} - q_{Z_3} = r^2 q_{Z_4}$$

$$q_{Z_1} - q_{Z_2} = r^3 q_{Z_4}$$

$$q_{Z_0} - q_{Z_1} = r^4 q_{Z_4}$$

$$q_{Z_4} = r^0 q_{Z_4}$$

Adding the equations, the result is gotten

$$q_{Z_0} = q_{Z_4}(r^0 + r + r^2 + r^3 + r^4)$$

This implies that

$$q_{Z_0} = r q_{Z_4}\left(\frac{1}{r} + 1 + r + r^2 + r^3\right)$$

$$\equiv q_{Z_4}(r^4 + r^3 + r^2 + r + 1)$$

If we sum up the identity we have

$$1 - r^5 = (1 + r + r^2 + r^3 + r^4)(1 - r)$$

Thus

$$1 + r + r^2 + r^3 + r^4 = \frac{1 - r^5}{1 - r}$$

Meaning that

$$q_{Z_0} = \left(\frac{1 - r^5}{1 - r}\right) q_{Z_4}$$

and

$$q_{Z_4} = \left(\frac{1 - r}{1 - r^5}\right) q_{Z_0}$$

By addition

$$q_{Z_3} = (1 + r) q_{Z_4}$$

$$= \left(\frac{1 - r^2}{1 - r}\right) q_{Z_4}$$

$$= (1 + r)\left(\frac{1 - r}{1 - r^5}\right) q_{Z_0}$$

$$= \left(\frac{1 - r^2}{1 - r^5}\right) q_{Z_0}$$

$$= \frac{1 - r^2}{1 - r^5}$$

Thus, we have

$$q_{Z_2} = (1 + r + r^2) q_{Z_4}$$

Now,

$$(1 + r + r^2)(1 - r) = 1 - r^3$$

So that,

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

Substituting for q_{Z_4}, the result follows,

$$q_{Z_2} = \frac{1 - r^3}{1 - r^5}$$

And also,

$$q_{Z_1} = (1 + r + r^2 + r^3) q_{Z_4}$$

Solving in the same manner we did for q_{Z_2}, we see that

$$(1 + r + r^2 + r^3) = \frac{1 - r^4}{1 - r}$$

Such that,

$$q_{Z_1} = \frac{1 - r^4}{1 - r^5}$$

In general,

$$q_{Z_A} = \frac{1 - r^{N-A}}{1 - r^N}$$

or

$$q_Z = \frac{1 - r^{N-Z}}{1 - r^N}$$

This is the probability that the gambler will be rained, given his initial capital$Z$.

Method of Difference Equation

$$q_Z = pq_{Z+1} + pq_{Z-1}$$

This is the same as

$$q_{Z+1} = pq_{Z+2} + pq_Z$$

The particular solution of 10.11 can be written as

$$q_Z = X^Z$$

This becomes

$$X^{Z+1} = pX^{Z+2} + q\,X^Z$$

Divide by X^Z

$$X = pX^2 + q$$

$$-pX^2 + X - q = 0$$

Thus,

$$X = \frac{1}{2} \pm \sqrt{3i}$$

So that 10.15 can be solved by

$$\frac{-1 \pm (1-4pq)^{1/2}}{-2p}$$

To simplify, we multiply the solutions

$$\frac{-1 + (1-4pq)^{1/2}}{-2p} \cdot \frac{-1 + (1-4pq)^{1/2}}{-2p} = \frac{q}{p}$$

$$X = \frac{q}{P} \qquad \text{if } p \neq q$$

$$= 1 \qquad \text{if } p \neq q$$

Then by substitution, equation 10.12 becomes

$$q_Z = \left(\frac{q}{P}\right)^Z \qquad \text{if } p \neq q$$

The general solution can be written as

$$q_Z = A + B\left(p/q\right)^Z$$

The boundary conditions

$$q_0 = 1 \quad \text{and} \quad q_N = 0$$

That is, when Z= 0, $q_Z = 1$

and when Z= N, $q_Z = 0$

This implies that

$$A + B = 1, \qquad Z = 0$$

$$A + B\left(p/q\right)^N = 0, \qquad Z = N$$

Solving the system of equations, we obtain

$$B = \frac{1}{1-\left(\frac{q}{p}\right)^N} \qquad A = \frac{\left(\frac{q}{p}\right)^N}{\left(\frac{q}{p}\right)^N - 1}$$

Substitute for A and B, we have

$$q_Z \frac{1-\left(\frac{q}{p}\right)^{Z-N}}{1-\left(\frac{p}{q}\right)^N} = \frac{1-\left(\frac{q}{p}\right)^{N-Z}}{1-\left(\frac{p}{q}\right)^N}$$

$$= \frac{1-r^{N-Z}}{1-r^N}$$

18.2.2 Gambler's Expected Gain (G)

Possible Values

Gain	N - Z	with	probability	$1-q_Z$
Loss	Z	with	probability	q_Z

The expected gain is

$$E(G) = (\textit{Combined}\ \text{Capital})\ (\text{Probability of gain}) - (\textit{Initial}\ \text{Capital})$$
$$= N(1-q_Z) - Z$$

That is

$$E(G) = N(1-q_Z) - Z$$

If $p = q = \frac{1}{2}$ $\quad$ *or* $q + p \neq 0$

We can write q_Z as a function of Z

$q_{Z+1} = pq_{Z+2} + qq_Z = f(Z)$ a constant

Then the solution from the result of differential equation with constant coefficient is

$$q_Z = Z$$

In general

$$q_Z = A + BZ$$

Under the boundary conditions

$q_0 = 1,\ q_N = 0$ $\quad$ *at* $Z = 0$ and N respectively.

Thus,

$A = 1$ $\quad$ at $Z = 0$

and

$A + B^N = 0$ $\quad$ at $Z = N$

Thus,

$$B = -\frac{1}{N}$$

Substitute for A and B

$$q_Z = 1 - \frac{1}{N}Z$$

Substituting for q_Z

$$E(G) = N[1-(1-Z/N)] - Z$$
$$= N(Z/N) - Z$$
$$= 0$$

18.2.3 Expected Duration of the Game

Assume that the expected duration of the game has a known value D_Z. If the first trial results in a success, the game continues as if the initial position was $Z + 1$.

Now, the initial position is Z, so that

$$D_Z = pD_{Z+1} + qD_{Z-1}$$

Under the condition that the first trial in a success

$$D_Z = pD_{Z+1} + qD_{Z-1} + 1$$

With boundary conditions

$D_0 = \phi, \quad D_N = 0$

But (10.28) is the same as

$D_{Z+1} = pD_{Z+2} + qD_{Z+1}$

The complete solution is

$D_Z = U_Z + V_Z$

Where U_Z is the general; solution and V_Z is the particular solution.

General Solution

Any difference U_Z between any two solutions can be written as

$U_Z = pU_{Z+1} + qU_{Z-1}$

This is the same as

$U_{Z+1} = pU_{Z+2} + qU_Z$

Let $U_Z = X^Z$

So that we have

$X^{Z+1} = pX^{Z+2} + qX^Z$

Dividing through X^Z

$X = pX^2 + q$

This becomes a quadratic equation, which can be written as

$-pX^2 + X - q = 0$

$$X = \frac{-1 \pm \sqrt{1-4pq}}{-2p}$$

Multiplying the solutions results in

$$X = \frac{q}{p}$$

U_Z is given as

$U_Z = A + BX^Z$

So that,

$$U_Z = A + B\left(\frac{q}{p}\right)^Z$$

Particular Solution

Let the particular solution be

$V_Z = aZ$

This means that we can write

$a(Z+1) = pa(Z+2) + qaZ + 1$

So that

$$a = \frac{1}{Z + 1 - pZ - 2p - qZ}$$

The denominator becomes

$1 - 2p = q - p \quad (\text{Since } p + q = 1)$

Therefore,

$$a = \frac{1}{q-p}$$

Substituting for a

$$V_Z = \frac{Z}{q-p}$$

The complete solution is

$D_Z = U_Z + V_Z$

$$= A + B\left(\frac{q}{p}\right)^Z + \frac{Z}{q-p}$$

The required boundary conditions are

$$A + B = 0, \quad Z = 0, \quad D_Z = 1$$

$$A + B\left(\frac{q}{p}\right)^N = \frac{-N}{q-p}; \quad Z = N, \ D_N = 0$$

For $Z = 0, \ D_Z = 0$

$$A + B = 0$$

For $Z = N, D_N = 0$

$$A + B\left(\frac{q}{p}\right)^N + \frac{N}{q-p} = 0$$

So that

$$A + B\left(\frac{q}{p}\right)^N = \frac{-N}{q-p}$$

This results in

$$B\left[\left(\frac{q}{p}\right)^N - 1\right] = \frac{-N}{q-p}$$

$$B = \frac{N}{q-p}\left[1 - \left(\frac{q}{p}\right)^N\right]$$

So that we have

$$\frac{A + \frac{N}{q-p}}{1 - \left(\frac{q}{p}\right)^N} = 0$$

Solving for A,

$$A\left[1 - \left(\frac{q}{p}\right)^N\right] = \frac{-N}{q-p}$$

So that

$$A = -\frac{\frac{N}{q-p}}{\left[1 - \left(\frac{q}{p}\right)^N\right]}$$

Substituting for A and B in (10.39) we have

$$D_Z = -\frac{\frac{N}{q-p}}{\left[1 - \left(\frac{q}{p}\right)^N\right]} + \frac{\frac{N}{q-p}\left(\frac{q}{p}\right)^Z}{\left[1 - \left(\frac{q}{p}\right)^N\right]} + \frac{Z}{q-p}$$

$$= \frac{Z}{q-p} - \frac{N}{q-p}\left[\frac{1 - \left(\frac{q}{p}\right)^Z}{1 - \left(\frac{q}{p}\right)^N}\right]$$

18.3 Queuing Theory

The principal pioneer of queuing system was A.R. Erlang, who began in 1908 to study problems of telephone congestion for the Copenhagen Telephone Company. He was concerned with problems such as the following: A manually operated telephone exchange has a limited number (one or more) of operations when a subscriber attempts to make a call, the subscriber must wait if all the operations are already busy making connections for other subscribers. It is of interest to study the waiting time of subscribers e.g. the average waiting time and the chance that a subscriber will obtain service immediately without waiting and to examine how much the waiting times will be affected if the number of operations is affected or conditions are changed in any other way. If there are more or if service can be speeded up, subscribers will be pleased because waiting will be reduced, but the improved facility will become expensive to maintain, therefore, a reasonable balance must be struck.

18.3.1 Applications of Queuing Theory

When persons or things needing the services of a facility or persons arrive at a service channel or counter on the account that the facility or persons cannot serve all at a time, a queue or waiting line is formed. Examples of this include:

(i) cars arriving at a fuel station waiting to be served.
(ii) persons waiting at a bus station waiting to be checked in.
(iii) books arriving at a librarians desk.
(iv) patients waiting to see a doctor or community health dispenser.
(v) customers arriving at a departmental store (supermarket).
(vi) clients waiting to see the Customer Service Executive or Officer.

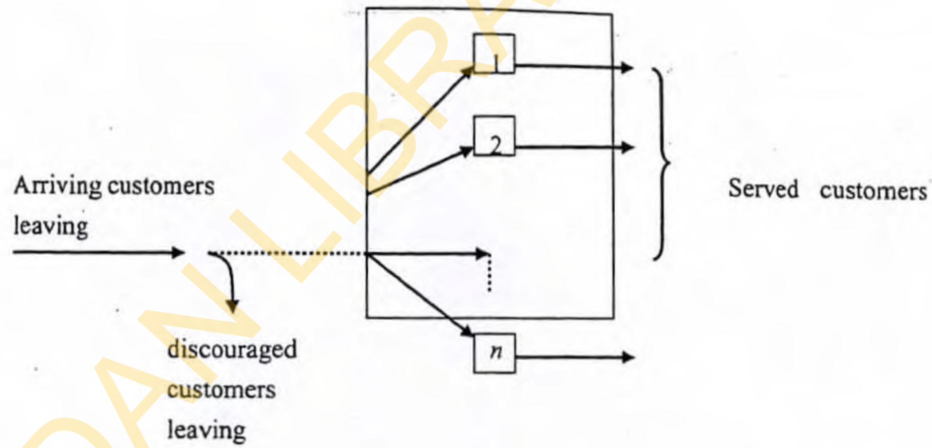
Queuing theory is applied into every field of human endeavour. This is because there is no perfect service or treatment that can be meted out. Below are some of the fields of application:

(i) Business – banks, supermarket, booking offices, and so on.
(ii) Industries – servicing of automatic machines, production lines, storage, and so on.
(iii) Engineering – telephony, communication networks, electronic computers, and so on.
(iv) Transportation – airports, harbours, railways, traffic operations in cities, postal services, and so on.
(v) Others – elevators, restaurants, barber shops, and so on.

18.3.2 Concept and Definition

Queuing theory is concerned with the design and planning of service facilities to meet a randomly fluctuating demand for service in order to minimize congestion and maintain economic balance between service cost and waiting cost. The *cost* here refers to *time*.

A queuing system is composed of customers arriving at a service channel and is attended to by any one or more of the service attendants. If a customer is not served immediately he may decide to wait. In the process, however, a few customers may leave the line if they cannot wait. At the end of the process, served customers leave the system.



A QUEUING SYSTEM
(OR by Swarup et al. 1978, p505)

18.3.3 Components of the Queue System

A queue situation can be divided into five elements. These are:

(i) Arrival mode
(ii) Service mechanism
(iii) Service channels
(iv) System capacity
(v) Queue discipline

(i) **Arrival Mode** – this refers to the rate at which customers arrive at a service centre and the statistical law which governs the pattern of arrival.
Certain definitions pertaining to the arrival of customers:
- ***bulk*** or ***batch*** arrival: more than one arrival allowed to enter into the system simultaneously.
- ***balk***: customers deciding not o enter a queue because it is long or lengthy.
- ***renege***: customer leaving a queue due to impatience.

jockey: customer jostling among parallel queues.

stationary: arrival pattern which does not change with time.

transient: a time-dependent arrival process.

The arrival mode is always denoted by M.

(ii) **Service Mechanism** – this refers to the number of service points that are available and the duration of service. When the service points or servers are infinite, the service will be instantaneous, which will result in no queue. In case of finite points, queue is inevitable. Customers can be served according to a specific order, which may be in batches of fixed size or of variable size. This system is called ***bulk service system***.

(iii) **Service Channels** – where there are more than one channel of service, then arrangement of service may be in parallel or series, or a combination of both, depending on the system design.

(iv) **System Capacity** – most queuing system are limited in such a way that waiting rooms are all accommodating. This gives limit to the number of customers that can be accepted to the waiting line at any given time. Such situation gives rise to ***finite source queues***, and results in ***forced balk***.

(v) **Queue Discipline** – this is a method of customer selection for service when a queue has been formed. The different forms of discipline include:

(ai) First Come, First Served (**FCFS**), or

(aii) First In, First Out (**FIFO**)

(b) First In, Last Out (**FILO**)

(c) Last In, First Out (**LIFO**)

(d) First in, First Out with Priority (**FIFOP**)

(e) Selection for Service In Random Order (**SIRO**)

Symbols and Notations

We shall employ the following symbols and notations this lecture:.

$n =$ number of customers in the system, both waiting and in service,

$\lambda =$ average number of customers arriving per unit of time

$\mu =$ average number of customers being served per unit of time

$\frac{\lambda}{\mu} = \rho =$ traffic intensity

$C =$ number of parallel service channels (servers)

$E(n) =$ average number of customers in the system, both waiting and in service

$E(m) =$ average number of customers within in the queue

$E(v) =$ average waiting time of customers in the system, both waiting and in service.

$E(w) =$ average waiting time of a customer in the queue

$P_n(t) =$ probability that there are n customers in the system at any time t, both waiting and in service.

$P_n =$ time independent probability that there are n customers in the system, both waiting and in service.

18.4 The Basic Queuing Process

The statistical pattern by which customers arrive over a period of time must be specified.

It is usually assumed that they are generated according to a Poisson process that is, the number of customer who arrives until any specific time has a Poisson distribution. The Poisson distribution involves the probability of occurrence of an arrival and is independent of what has occurred in the preceding observation. This Poisson assumption indicates the number of arrivals per unit time(λ) (or mean arrival rate), while $1/_{\lambda}$ on the lengthy of interval between two consecutive arrivals. This time between two consecutive arrivals is referred to as "inter-arrival time."

The mean service rate μ is the number of customers served per unit time whole average service time $(1/_{\mu})$ is the time units per customer service time delivered is given by an experiment distribution where the servicing of a customer takes place between the time$t$ and$t + \Delta t$.

18.5 Poisson Process and Exponential Distribution

In queuing theory, the arrival rate and service rate follow a Poisson distribution. However, it should be noted that the number of occurrences in some time interval is a Poisson random variate, and the time between successive occurrences is an exponential distribution. Both are equivalent

18.5.1 Axioms of the Poisson Process

Given an arrival process $[N(t), t \geq 0]$, where $N(t)$ denotes the total number of arrivals up to time t, $N(0) = 0$, an arrival characterized by the following assumptions (axioms) can be described as a Poisson process;

AXIOM 1 - the number of arrivals in non-overlapping intervals are statistically independent. This means there is independent increment in the process.

AXIOMS 2 - the probability of more than one arrival between time t and time $t + \Delta t$ is $o(\Delta t)$; this means there is negligibility in the probability of two or more arrivals during the small time interval Δt. This implies that

$$p_0(\Delta t) + p_1(\Delta t) + o(\Delta t) = 1$$

AXIOMS 3 - the probability that an arrival occurs between time t and time $t + \Delta t$ is $\lambda\Delta t + o(\Delta t)$. This implies that

$$p_1(\Delta t) = \lambda\Delta t + o(\Delta t)$$

Where λ, a constant, is independent of $N(t)$, Δt is an incremental element, and $o(\Delta t)$ represents the terms such that

$$\lim_{\Delta t \to 0} \frac{o(\Delta t)}{\Delta t} = 0$$

18.6 Classification of Queuing System

Queuing systems, generally, may be completely specified in the following symbolic forms:

$$(a|b|c):(d|e)$$

Description

First symbol (**a**) – type of distribution of inter-arrival times
Second symbol (**b**) – type of distribution of inter-service times
Third symbol (**c**) – number of servers
Fourth symbol (**d**) – system capacity
Fifth symbol (**e**) – queue discipline

For the first and second symbols, the following letters may be used:

$M \equiv$ Poisson arrival or departure distributions
$E_k \equiv$ Erlangian or Gamma inter-arrival or service distribution
$GI \equiv$ General input distribution
$G \equiv$ General service time distribution

An example of a queue system is

$$(M|E_k|C):(N|SIRO)$$

Queuing system is classified into

(i) Poisson Queues
(ii) Non-Poisson Queues

Definitions

Transient State: When a queuing system ahs its operating characteristic (e.g. input, output, mean queue length, etc) dependent upon time, then it is said to be in transient state.

Steady State: This is a queue system that is independent of time.

Assume $P_n(t)$ to be the probability that there are n customers in the system at time i, then the steady state use becomes

$$\lim_{t\to\infty} \mathrm{P}_n(t) = \mathrm{P}_n \text{ (Independent of t)}$$

Meaning that

$$\lim_{t\to\infty} \frac{d}{dt} \mathrm{P}_n(t) = 0$$

18.7 Poisson Queues

18.7.1 The $M|M|1$ System

This deals with the process where arrivals and departures occur randomly over time generally known as birth-death process.

1. **Model 1: $(M|M|1):(\infty|FIFO)$**

In this model, we have Poisson input, exponential service, single channel, infinite system capacity and first in first out basis.

If $P_n(t)$, be the probability that there are $n$customers in the system at time t, then in order to write the difference equation for $P_n(t)$, we first consider how the system can get to state E_n at time $t+\Delta t$. To be in state E_n of time $t+\Delta t$, the system could have been in the state $E_n$at time t and have no arrivals or service competitions in Δtor be in state E_{n-1} of time t and have, during Δ_t, one service completion and no arrivals. If we assume that $n \geq 1$ (having arrivals and service independent of each other), it can be easily seen that

$$\begin{aligned} P_n(t+\Delta t) &= P_n(t).P(no\ arrivals\ in\ \Delta t).P(no\ service\ completions\ in\ \Delta t) \\ &+P_n(t).P(one\ arrival\ in\ \Delta t).P(one\ service\ in\ \Delta t) \\ &+P_{n+1}(t).P(one\ service\ completed\ in\ \Delta t).P(no\ arrivals\ in\ \Delta t) \\ &+P_{n-1}(t).P(one\ arrival\ in\ \Delta t).P(no\ service\ completions\ in\ \Delta t) + o(\Delta t) \\ &\quad n \geq 1 \end{aligned}$$

This can be re-written as

$$\begin{aligned} P_n(t+\Delta t) = P_n(t)[1-\lambda\Delta t + 0(\Delta t)][1 - \mu\Delta t + 0(\Delta t] + P_n(t)[\lambda\Delta t][\mu\Delta t] + P_{n+1}(t)[\mu\Delta t + 0(\Delta t)] \\ [1-\lambda\Delta t + 0(\Delta t)] + P_{n-1}(t)[\lambda\Delta t + 0(\Delta t][1-\mu\Delta t + 0(\Delta t] + o(\Delta t) \end{aligned}$$

This leads to

$$P_n(t) = P_n(t)[1-\lambda\Delta t - \mu\Delta t] + P_{n+1}(t)[\mu\Delta t] + P_{n-1}(t)[\mu\Delta t] + o(\Delta t) \qquad n \geq 1$$

Suppose n = 0, we have

$$\begin{aligned} P_0(t+\Delta t) &= P_0(t)[1-\lambda\Delta t + o(\Delta t)] + P_1(t)[1-\lambda\Delta t + o(\Delta t)] \\ &\quad [\mu\Delta t + o(\Delta t)] + o[\Delta t] \\ &= P_0(t)[1-\lambda\Delta t] + P_1(t)\mu\Delta t + o(\Delta t) \end{aligned}$$

We can record the difference equation

$$P_n(t+\Delta t) - P_n(t) = -(\lambda+\mu)\Delta t P_n(t) + \mu\Delta t\, \mathrm{P}_{n+1}(t) + \lambda\Delta t\, \mathrm{P}_{n-1}(t) + o(\Delta t); \qquad n \geq 1$$

and

$$\mathrm{P}_0(t+\Delta t) - P_0(t) = -\lambda\Delta t P_0(t) + \mu\Delta t\, \mathrm{P}_1(t) + o(\Delta t)$$

Then,

$$\lim_{\Delta t\to\infty} \frac{\mathrm{P}_n(t+\Delta t - P_n(t)}{\Delta t} = -(\lambda+\mu)P_n(t) + \mu P_{n+1}(t) + \lambda P_{n-1}(t) + o(\Delta t)$$

and

$$\lim_{\Delta t\to 0} \frac{\mathrm{P}_0(t+\Delta t - P_0(t)}{\Delta t} = -\lambda P_0(t) + \mu P_1(t) + o(\Delta t)$$

So that we have

$$\frac{d}{dt} P_n(t) = P_n'(t) = -(\lambda+\mu)P_n(t) + \mu P_{n+1}(t) + \lambda P_{n-1}(t) \quad n \geq 1$$

and

$$\frac{d}{dt}P_0(t) = P_0^{'}(t) = -\lambda P_0(t) + \mu P_1(t)$$

The above are known as difference equations in n and t. The steady-state solutions for P_n in the system at an arbitrary point of time is obtained by taking the limit as $t \to \infty$..

If the steady-state exists $(\lambda < \mu, \text{ as } t \to \infty)$, then

$$P_n(t) \to P_n \text{ and } \frac{d}{dt}P_n(t) \to 0 \text{ as } t \to \infty$$

If $\lambda = \mu$ there exist no queue

If $\frac{\lambda}{\mu} > 1$, we have an explosive state

Using the condition of steady state,we have

$$0 = -(\lambda + \mu)P_n + \mu P_{n+1} + \lambda P_{n-1}; \qquad n \geq 1$$

and

$$0 = -\lambda P_0 + \mu P_1$$

Using iterate procedure we have

$$P_1 = \frac{\lambda}{\mu}P_0$$

$$P_2 = \left(\frac{\lambda + \mu}{\mu}\right)P_1 - \frac{\lambda}{\mu}P_0 = \left(\frac{\lambda}{\mu}\right)^2 P_0$$

$$P_3 = \left(\frac{\lambda + \mu}{\mu}\right)P_2 - \frac{\lambda}{\mu}P_1 = \left(\frac{\lambda}{\mu}\right)^3 P_0$$

In general we have

$$P_n = \left(\frac{\lambda}{\mu}\right)^n P_0 \qquad \forall n$$

Proof

By mathematical condition, we have

$$P_{n+1} = \frac{\lambda + \mu}{\mu}P_n - \frac{\lambda}{\mu}P_{n-1}, \qquad n \geq 1$$

$$= \left(\frac{\lambda + \mu}{\mu}\right)\left(\frac{\lambda}{\mu}\right)^n P_0 - \frac{\lambda}{\mu}\left(\frac{\lambda}{\mu}\right)^{n+1} P_0$$

$$= \left[\frac{\lambda^{n+1} + \mu\lambda^n}{\mu^{n+1}} - \frac{\lambda^n}{\mu^n}\right]P_0$$

$$= \left(\frac{\lambda}{\mu}\right)^{n+1} P_0$$

Using the boundary condition; $\sum_{n=0}^{\infty} P_n = 1$, then 6.5 becomes

$$1 = \sum_{n=0}^{\infty}\left[\frac{\lambda}{\mu}\right]^n P_0 = P_0\sum_{n=0}^{\infty}\left[\frac{\lambda}{\mu}\right]^n$$

$$= P_0\left[\frac{1}{1 - \frac{\lambda}{\mu}}\right] \qquad \textit{Sum of geometric series where } \frac{\lambda}{\mu} < 1$$

$$= P_0\left[\frac{1}{1-\rho}\right]$$

This implies that

$$P_0 = 1 - \rho$$

Resulting in the steady-state

$$P_n = \rho^n(1-\rho). \qquad \rho < 1 \quad \textit{and } n \geq 0$$

This is the probability distribution of queue length.

Characteristics of Model 1

(i) Probability of queue size greater or equal to n.

$$P(\geq n) = \sum_{K=n}^{\infty} P_K = \sum_{K=n}^{\infty}(1-\rho)\rho^k$$

$$= (1-\rho)\rho^n \sum_{k=n}^{\infty}\rho^{k-n}$$

$$= (1-\rho)\rho^n \sum_{K-n=0}^{\infty}\rho^{k-n}$$

$$= \frac{(1-\rho)\rho^n}{1-\rho} = \rho^n$$

(ii) Average number of customers in the system

$$E(n) = \sum_{n=0}^{\infty} n\,P_n = \sum_{n=0}^{\infty} n(1-\rho)\rho^n$$

$$= (1-\rho)\sum_{n=0}^{\infty} n\,\rho_n = \rho(1-\rho)\sum_{n=1}^{\infty} n\,\rho^{n-1}$$

$$= \rho(1-\rho)\sum_{n=0}^{\infty}\frac{d}{de}\rho^n$$

$$= \rho(1-\rho)\frac{d}{de}\sum_{n=0}^{\infty}\rho^n, \qquad \text{Since } \rho < 1$$

$$= \rho(1-\rho)\left[\frac{1}{(1-\rho)^2}\right]$$

$$= \frac{\rho}{1-\rho} = \frac{\lambda}{\mu-\lambda}$$

(iii) Average queue length

$$E(m) = \sum_{m=0}^{\infty} m\,P_n;$$

where m = n - 1 (that is number of customers in queue minus customer in service)

$$= \sum_{n=1}^{\infty}(n-1)P_n = \sum_{n=1}^{\infty} n\,P_n - \sum_{n=1}^{\infty} P_n$$

$$= \sum_{n=0}^{\infty} n\,P_n - \left[\sum_{n=0}^{\infty} P_n - P_0\right]$$

$$= \frac{\rho}{1-\rho} - [1-(1-\rho)]$$

$$= \frac{\rho}{1-\rho} - \rho$$

$$= \frac{\rho^2}{1-\rho} = \frac{\lambda^2}{\mu(\mu-\lambda)}$$

(iv) Average length of non-empty queue

$$E(m/m>0) = \frac{E(m)}{P(m>0}$$

$$= \frac{\lambda^2}{\mu(\mu-\lambda)} \quad \frac{1}{\left(\frac{\lambda}{\mu}\right)^2} = \frac{\mu}{\mu-\lambda}$$

This is because $P(m>0) = P(n>1) = \left[\sum_{n=0}^{\infty} P_n - P_0 - P_1\right] = \left(\frac{\lambda}{\mu}\right)^2$

(v) The fluctuation (variance) of queue length

$$V(n) = \sum_{n=0}^{\infty}[n - E(n)]^2 P_n$$

$$= \sum_{n=0}^{\infty} n^2 P_n - [E(n)]^2$$

By algebraic transformations,

$$V(n) = (1-\rho)\frac{\rho+\rho^2}{(1-\rho)^2} - \left[\frac{\rho}{1-\rho}\right]^2$$

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

$$= \frac{\lambda\mu}{(\mu-\lambda)^2}$$

Example 18.1

A TV repairman finds that the time spent on his jobs has an exponential distribution with mean 30 minutes. If he repairs sets in the order in which they come in and if the arrival of sets is approximately Poisson with an average rate of 10 per day.

(i) What is the repairman's expected idle time each day?

(ii) How many jobs are ahead of the average set just brought in?

Solution

$\lambda = \frac{10}{8} = \frac{5}{4}$ sets per hour

$\mu = \frac{1}{30} X 60 = 2$ sets perf hour

(i) The probability of no unit in the queue is

$$\rho_0 = 1 - \frac{\lambda}{\mu} = 1 - \frac{5}{8} = \frac{3}{8}$$

Hence the idle time for repairman in 8 hour days

$$= \frac{3}{8} X 8 = 3 \text{ hours}$$

(ii) $E(n) = \frac{\lambda}{\mu - \lambda} = \frac{5/4}{2 - 5/4} = 1\frac{2}{3}$ jobs

18.7.2 Waiting Time Distribution for Model 1

Waiting time is mostly a continuous random variable and there is a non-zero probability of delay being zero. Denote time spent in queue by w. Let $\psi_w(t)$ be the cumulative probability distribution so that from a complex randomness of the Poisson, we have

$$\psi_w(0) : P(w = 0)$$
$$= P \text{ (No customer on the systemn upon arrival)}$$
$$= P_0 = 1 - \rho$$

To find $\psi_{w(t)}$ for $t > 0$, we suppose there be n customers in the system upon arrival. For a customer to go into service at time between 0 and t, it means all the customers must have been served at time t.

Therefore,

$$\psi_n(t) = P[(n-1) \text{ customers are served at time t}]\ P[\text{one customer being served in time dt}]$$
$$= \frac{(\mu t)^{n-1} e^{-\mu t}}{(n-1)!} \mu dt$$

The waiting time w is therefore

$$\psi_w(t) = P[w \le t]$$
$$= \sum_{n=1}^{\infty} P_n \int_0^t \psi_n(t) + \psi_w C_0)$$
$$= \sum_{n=1}^{\infty} (1-\rho)\rho^n \int_0^t \frac{(\mu t)^{n-1}}{(n-1)} e^{-\mu t} . \mu dt + (1-\rho)$$
$$= (1-\rho)\rho \int_0^t \mu e - \mu^t \sum_{n=1}^{\infty} \frac{(\mu t\rho)^{n-1}}{(n-1)!} dt + (1-\rho)$$
$$= (1-\rho)\rho \int_0^t \mu e - \mu t (1-\rho) dt + (1-\rho)$$
$$= 1 - \rho e^{-\mu t(1-\rho)}; \qquad t \ge 0$$

The distribution of waiting time in queue is

$$\psi_w(t) = \begin{cases} 1 - \rho & t = 0 \\ 1 - \rho e^{-\mu t(1-\rho)} & t > 0 \end{cases}$$

Characteristics of Waiting Time Distribution for Model 1

(i) Average waiting time of a customer (in the queue)

$$E(w)=\int_0^\infty t.d\psi_w(t)$$

$$=\int_0^\infty t\rho\mu(1-\rho)e^{-\mu(1-\rho)t}dt$$

$$=\rho\int_0^\infty \frac{x e^{-x}}{\mu(1-\rho)}dx \quad \text{for } \mu(1-\rho)t=x$$

$$=\frac{\rho}{\mu(1-\rho)}=\frac{\lambda}{\mu(\mu-\lambda)}$$

(ii) Average waiting time of an arrival that has to want

$$E(w/w>0)=\frac{E(w)}{\rho(w>0)}$$

$$=\left\{\frac{\lambda}{\mu(\mu-\lambda)}\right\}\Big/\frac{\lambda}{\mu}$$

$$=\frac{1}{\mu-\lambda}$$

We note that $P(w>0)=1-P(w=0)=1-(1-\rho)=\rho$

(iii) For the busy period distribution, suppose v is the random variable denoting the total time that a customer had to spend in the system including service. This makes the cumulative density function to be

$$\psi(w/w>0)=\frac{\psi(w)}{P(w>0)};\ \text{where } \psi(w)=\frac{d}{dt}[\psi_w(t)]$$

$$=\left[\lambda\left(1-\frac{\lambda}{\mu}\right)e^{-(\mu-\lambda)t}\right]\Big/\left(\frac{\lambda}{\mu}\right)$$

$$(\mu-\lambda)e^{-(\mu-\lambda)t},\quad t>0$$

(iv) Average waiting time that a customer spends in the system including service

$$E(v)=\int_0^\infty t.\psi(w/w>0)dt$$

$$=\int_0^\infty t(\mu-\lambda)e^{-(\mu-\lambda)t}dt$$

$$=\frac{1}{\mu-\lambda}\int_0^\infty x e^{-x}dx, \qquad \text{for } (\mu-\lambda)t=x$$

$$=\frac{1}{\mu-\lambda}$$

Relation between Average Queue Length and Average Waiting Time

(Little's Formula)

$$E(n)=\frac{\lambda}{\mu-\lambda}; \qquad E(m)=\frac{\lambda^2}{\mu(\mu-\lambda)}$$

$$E(w)=\frac{\lambda}{\mu(\mu-\lambda)} \qquad E(v)=\frac{1}{\mu-\lambda}$$

It can be seen that $E(n)=\lambda\,E(v)$, $E(n)=\lambda\,E(w)$ and $E(v)=E(w)+\frac{1}{\mu}$

Example 18.2

Arrivals at a telephone both are considered to be Poisson with an average time of 10 minutes between one arrival and the next. The length of a phone call is assumed to be distributed exponentially with mean 3 minutes.

(i) What is the probability that a person arriving at the booth will have to wait?

(ii) The telephone department will initial a record booth when convinced that an arrival would expect waiting far at least 3 minutes for phone. By how much should the flow of arrivals increase in order to justify a record booth.

Solution

We are given

$\lambda = \frac{1}{10} = 0.10$ person per minute

and

$\mu = \frac{1}{3} = 0.33$ person per minute

(*i*) $P(w > 0) = 1 - P_0 = 1 - \left(1 - \frac{\lambda}{\mu}\right)$

$$= \frac{\lambda}{\mu} = \frac{0.01}{0.33}$$

$$= 0.33$$

(ii) The installation of record booth will be justified if the arrival rate is greater than the waiting time. Then the length of queue will go on increasing.

Now, $E(w) = \frac{\lambda}{\mu(\mu - \lambda)} = 3$

$$= \frac{\lambda^1}{0.33\,(0.33 - \lambda^1)}$$

Where E(w) = 3 and $\lambda = \lambda^1(w)$ for record booth. On simplification this yields $\lambda^1 = 0.16$. hence the arrival rate should become 0.16 person per minute to justifies the record booth.

18.7.3 ModelII$(M|M|1):(\infty|SIRO)$

This model is similar to model I. The only difference is in the service discipline. The first follow the FIFO rule, while this follows the SIRO rule. We recall that the derivation of P_n for model I does not depend on any specific queue discipline, it may then be concluded that for the SIRO rule case, we must have.

$$\rho_n = (1 - \rho)\,\rho^n,\ n \geq 0$$

The average number of customer in the system, $E(n)$ remains the same irrespective of cases, FIFO or SIRO. Provided $P_n$remains unchanged, $E(n)$ remain the same in all queue discipline, thus

$$E(v) = \frac{1}{\lambda}\,\mathrm{E(n)} = \frac{1}{\mu - \lambda}$$

This result applies to the FIFO SIRO and LIFO cases. These three queue discipline sonly differ in the distribution of waiting time when the probabilities of along and short waiting times change depending upon the discipline used. When the waiting time distribution is not required, the symbol GD(general discipline) can be used to represents the three queue disciplines above.

18.7.4 ModelIII $(M|M|1):(N|FIFO)$

There is a deviation from the previous model I (especially I) because the number of customers is now finite (N). As long as $n < N$, the difference equated of model remains valid for this model. If the system is in state E_N, then the probability of an arrival into the system is zero.

Thus, the additional difference equation for $n = N$ becomes

$$P_N(t + \Delta t) = P_N(t)\left[1 - \mu\Delta t\right] + P_{N-1}(t).\left[\lambda\Delta t\right]\left[1 - \mu\Delta t\right] + o(\Delta t)$$

resulting in the differential-difference equation.

$$\frac{d}{dt}\mathrm{P_N}(t) = -\mu\,\mathrm{P_N}(t) + \lambda\,P_{N-1}(t)$$

and gives the resultant steady state difference equation

$$0 = -\mu\,\mathrm{P_N} + \lambda\,\mathrm{P_{N-1}}(t)$$

Given the interval $1 \leq n \leq N - 1$, the complete set of steady–state difference equations for this model is as follows.

$$\mu P_1 = \lambda P_0$$

$$\mu P_{n+1} = (\lambda + \mu)P_n - \lambda\,P_{n-1}$$

$$\vdots$$

$$\mu P_n = \lambda P_{N-1}$$

As in model I, by iterative procedure, the first two difference equations are

$$P_N = \left(\frac{\lambda}{\mu}\right)^n P_0; n \le N-1$$

In the same manner, the value of P_n holds for the last difference equation if $n = N$.

Thus, we have

$$P_n = \left(\frac{\lambda}{\mu}\right)^n P_0$$

$$= \rho^n P_0; \quad n \le N$$

Using the boundary condition, we can obtain the value of P_0.

Boundary condition is $\sum_{n=0}^{N} P_n = P_1$

Thus

$$1 = P_0 \sum_{n=0}^{N} \rho^n$$

$$= \begin{cases} P_0\left(\dfrac{1-\rho^{N+1}}{1-\rho}\right) \\ P_0\ (N+1) \end{cases}$$

Thus,

$$P_0 = \begin{cases} \dfrac{1-\rho}{1-\rho^{N+1}} \\ \dfrac{1}{N+1} \end{cases}$$

Hence

$$P_n = \begin{cases} \dfrac{(1-\rho)\rho^n}{1-\rho^{N+1}}, & \rho \ne 1 \\ \dfrac{1}{N+1} \quad (\rho = 1) \end{cases} \quad ; 0 \le n \le N$$

Note that the steady-state solution exists even for $\rho \ge 1$. Intuitively, there is sense in this since the process is prevented from blowing up by the maximum limit.

Thus, given $N \to \infty$, the steady-state solution results in

$$P_n = (1-\rho)\rho^n \quad n < \infty$$

Which is the same as that in model I.

Characteristics of Model III

(i) Average number of customers in the system is given by

$$E(n) = \sum_{n=0}^{N} nP_n = P_0 \sum_{n=0}^{N} n\rho^n$$

$$= P_0\,\rho \sum_{n=0}^{N} \frac{d}{d\ell}\,\rho^n = P_0\,\rho \frac{d}{dp} \sum_{n=0}^{N} P_n$$

$$= P_0\,\rho \frac{d}{dp}\left[\frac{1-\rho^{N+1}}{1-\rho}\right]$$

$$= \frac{P_0\,\rho\left[1-(N+1)\rho^N + N\rho^{N+1}\right]}{(1-\rho)^2}$$

$$= \frac{\rho\left[1-(N+1)\rho^N + N\rho^{N+1}\right]}{(1-\rho)(1-\rho^{N+1})}$$

(ii) Average queue length

$$E(m) = \sum_{n=1}^{N}(n-1)\,P_n = E(n) - \sum_{n=1}^{N} P_n$$

$$= E(n) - (1 - P_0)$$

$$= E(n) - \frac{\rho(1-\rho^N)}{1`-P^{N+1}}$$

$$= \frac{\rho^2\left[1 - N\rho^{N-1} + (N-1)\rho^N\right]}{(1-\rho)\,(1-\rho^{N+1})}$$

(iii) Average waiting time.

Using Little's formula:

$E(v) = \dfrac{E(n)}{\lambda^1}$ where λ^1 is the mean rate of customers entering the system and is equal to $\lambda(1-P_N)$

Thus, $E(w) = E(v) - \dfrac{1}{\mu} = \dfrac{E(m)}{\lambda^1}$

Example 18.3

At a railway station, only one train is handled at a time. The railway yard is sufficient only for two trains to wait while the other is given signal to leave the station. Trains arrive at the station at an average rate of 6 per hour and the railway station can handle them on an average of 12 per hours. Assuming Poisson arrivals and exponential service distribution,

(a) Find the steady-state probabilities for the various numbers of trains in the system.

(b) Also, find the average waiting time of a new train coming into the yard.

Solution

$$\lambda = 6 \quad \mu = 12, \rho = \frac{6}{12} = 0.5$$

Probability of no train in the system (both waiting and in service is

$$P_0 = \frac{1-\rho}{1-\rho^{N+1}} = \frac{1-0.5}{1-(0.5)^{3+1}} = 0.53$$

We know that $P_n = P_0\,\rho^n$, thus

(a) $P_1 = (0.53) \quad (0.5) = 0.27$

$P_2 = (0.53) \quad (0.5)^2 = 0.13$

$P_3 = (0.53) \quad (0.5)^3 = 0.07$

(b) $E(n) = 1(0.27) + 2(0.12) + 3(0.07) = 0.74$

Hence, the coverage number of trains in the queue is 0.74, and each train takes on an average $^1/_2$ (0.085) hours for getting service. As the arrival of new train expects to find on average of 0.74 trains in the system before it.

$$E(w) = (0.74)\,(0.085) \text{ hours}$$
$$= 0.0629 \text{ hours or 38 minutes}$$

18.7.5 Model IV (Birth- Death Process)

Assume the system to be in date En, the probability of a birth occurring in a small time interval Δt is considered as $\lambda_n \Delta t + o(\Delta t)$; and that of the death is considered as $\mu_n \Delta t + o(\Delta t), n \geq 1$. The system being in E_n at time t means it will remain in E_n at time$t + \Delta t$ provided there is no birth and no death/on birth and one death, or the system might have been in $E_{n-1}$and had a birth, or in $E_{n+1}$and had a death. Thus, this result in

$$P_n(t+\Delta t) = P_n(t)\,(1 - \lambda_n\Delta t - o(\Delta t))(1 - \mu_n\Delta t - o(\Delta t)) + P_{n+1}(t)\,(\mu_{n+1}\Delta t + o(\Delta t))(1 - \lambda_{n+1}\Delta t - o(\Delta t) + P_{n-1}(t)\,(\lambda_{n-1}\Delta t + o(\Delta t))(1 - \mu_{n-1}\Delta t - o(\Delta t) + o(\Delta t), \qquad n \geq 1$$

$$P_0(t+\Delta t) = P_0(t)(1 - \lambda_0\Delta t - o(\Delta t) + P_1(t)\,(\mu_1\Delta t + o(\Delta t)) + o(\Delta t), \quad n = 0$$

Dividing by Δt, and taking limit as $\Delta t \to 0$, the diffential – difference equations results

$$\frac{d}{dt}P_n(t) = -(\lambda_n + \mu_n)P_n(t) + \mu_{n+1}P_{n+1}(t) + \lambda_{n-1}P_{n-1}(t), \qquad n \geq 1$$

and

$$\frac{d}{dt}P_0(t) = -\lambda_n P_0(t) + \mu_1 P_1(t)$$

Since $P_n(t)$ is independent of time, the steady-state solution $\frac{d}{dt}P_n(t) = 0$ and the differential-difference equation reduce to

$$0 = -(\lambda_n + \mu_n)P_n + \mu_{n+1}P_{n+1} + \lambda_{n-1}P_{n-1}, \qquad n \geq 1$$

and

$$0 = -\lambda_0 P_0 + \mu_1 P_1$$

Consequently by interactive procedure as in model 1

$$P_1 = \frac{\lambda_0}{\mu_1}P_0$$

$$P_2 = \frac{\lambda_1 + \mu_1}{\mu_2}P_1 - \frac{\lambda_0}{\mu_2}P_0$$

$$= \frac{\lambda_1\lambda_0}{\mu_2\mu_1}P_0$$

$$P_3 = \frac{\lambda_2 + \mu_2}{\mu_3}P_2 - \frac{\lambda_1}{\mu_3}P_1$$

$$= \frac{\lambda_2\lambda_1\lambda_0}{\mu_3\mu_2\mu_1}P_0$$

So that in general

$$P_n = \frac{\lambda_{n-1}\lambda_{n-2}\cdots\lambda_0}{\mu_n\mu_{n-1}\cdots\mu_1}P_0$$

$$= \prod_{i=0}^{n-1}\frac{\lambda_i}{\mu_{i+1}}P_0, \qquad n \geq 1$$

By mathematical induction, one can prove that this formula is correct

$$P_{n+1} = \frac{\lambda_n + \mu_n}{\mu_{n+1}}P_n - \frac{\lambda_{n-1}}{\mu_{n+1}}P_{n-1}$$

$$= \prod_{i=0}^{n-1}\frac{\lambda_i}{\mu_{i+1}}P_0$$

Making use of the boundary condition, we obtain P_0

$$\sum_{n=0}^{\infty} p_n = 1 \; or p_0 + \sum_{n=0}^{\infty} p_n = 1$$

$$\text{thus } p_0 = \left[1 + \sum_{n=1}^{\infty}\prod_{i=0}^{n-1}\left(\frac{\lambda_i}{\mu_{i+1}}\right)\right]^{-1}$$

If the R.H.S in a divergent series, $p_0 = 0$. If the R.H.S in a divergent series, p_0 will have its value defining on λ_i's and μ_i's.

Special case

I. When $\lambda_n = \lambda for n \geq 0, and \mu_n = \mu for n > 1$

then

$$p_0 = \left[1 + \sum_{n=1}^{\infty}\left(\frac{\lambda}{\mu}\right)^n\right]^{-1}$$

$$= 1 - \rho$$

Thus

$$p_n = \rho^n(1-\rho), \qquad for n \geq 0$$

(same as model 1)

II. When $\lambda_n = \frac{\lambda}{n+1} for n \geq 0, \quad and \mu_n = \mu for n > 1$

Then

$$p_0 = \left[1 + \sum_{n=1}^{\infty}\frac{\lambda^n}{n!\,\mu^n}\right]^{-1}$$

$$= \left[1 + \rho + \frac{1}{2!}\rho^2 + \frac{1}{3!}\rho^3 + \cdots\right]^{-1}$$

$$= e^{-\rho}$$

Thus

$$p_n = \left(\frac{1}{n!}\rho^n\right)e^{-\rho} \qquad for n \geq 0$$

Here we can see that p_n follows the Poisson distribution where $\rho = \frac{\lambda}{\mu}$. But, $\rho > 1$ or $\rho < 1$ most be finite.

III. When $\lambda_n = \lambda for n \geq 0$, $and \mu_n = n\mu$. $for n > 1$

Then

$$p_0 = \left[1 + \sum_{n=1}^{\infty} \frac{\lambda^n}{n!\,\mu^n}\right]^{-1}$$

$$= e^{-\rho}$$

Thus

$$p_n = \left(\frac{1}{n!}\rho^n\right)e^{-\rho} \qquad for n \geq 0$$

Here, service rate increase with increase in queue length. Hence it is known as the queuing problem with infinite number of channels= $(\boldsymbol{M}|\boldsymbol{M}|\infty):(\infty|\boldsymbol{FIFO})$

Example 18.4

Problems arrive at a computing center in a Poisson fashion at an average rate of five per day. The rules of the computing center are that any man waiting to get his problem solved must aid the man whose problem is being solved. If the time to solve a problem with one man has an exponential distribution with mean time of ½ day, and if the average solving time is inversely proportional to the number people working on the problem, approximate the expected time in the center for a person entering the line.

Solution

$\lambda = 5$ problem per day, $\mu = 3$ problems per day

It is given that the service increases with increase in the number of persons.

Thus, $\mu_n = n\mu$. where there are n persons.

$$p_n = \left(\frac{1}{n!}\rho^n\right)e^{-\rho}$$

$$E(n) = \sum_{n=0}^{\infty} np_n = \sum_{n=0}^{\infty} n.\left(\frac{1}{n!}\rho^n\right)e^{-\rho}$$

$$e^{-\rho}.\rho.e^{\rho} = \rho$$

$$= {}^{5}/_{3} \text{ persons}$$

The average solving time is inversely proportional to the number of people solvingon the problem is given by ${}^{2}/_{5}$ day problem.

Expected time for a person entering the line is

$\frac{1}{5}E(n) = \frac{1}{3}$ day or 8 hours.

Practice Questions

1. Derive, using both methods, the probability that a gambler will be ruined given that his initial capital is Z.
2. Show that gambler's expected gain is given as$N(1 - q_Z) - N$.
3. Under what condition can the expected gain be zero?
4. Company A enters into a project deal with another company B. A's initial deposit is $N5m$, while B's initial deposit is $N4m$. For every success, A gains more naira from B, otherwise it loses same to B. If the probability of success is 0.7, what is the probability of losing the entire deal?
5. A gambler's initial fortune is i. On each play of the game the gambler wins 1 with probability p, or loses 1 with probability $1 - p$. He or she continues playing until he/she is n ahead (that is, the fortune is $i + n$), or losing by m. Here $0 < i - m$ and $i + n < N$. What is the probability that the gambler quits as a winner?

6. Given an initial capital, Z, show that expected duration of the game is

$$D_z = \frac{Z}{q-p} - \frac{N}{q-p}\left[1 - \frac{\left(\frac{q}{p}\right)^z}{\left(\frac{q}{p}\right)^N}\right]$$

7. Describe the model 1 of the $M|M|1$queue discipline, and show that

(a) the average number of customers in the system is given as $\frac{\lambda}{\mu-\lambda}$.

(b) the average queue length is given as $\frac{\lambda^2}{\mu(\mu-\lambda)}$.

8. In the $M|M|1$ system of a queuing process, show that the

(a) steady state probability of model 1 is $P_n = \rho^n(1-\rho)$, where $\rho < 1$ and $n \geq 0$.

(b) the waiting distribution is given as

$$\psi(t) = \begin{cases} 1-\rho, & t = 0 \\ 1-\rho e^{-\mu(1-\rho)t}, & t > 0 \end{cases}$$

9. SAO Super market has one cashier at its counter. The service discipline of the cashier is FIFO. It is observed that the supermarket has 18 arrivals on average of every 10 minutes while the cashier can serve 12 customers in 6 minutes. If the distributions of arrivals and service rates are poisson and exponential respectively. Calculate.

(a) The traffic intensity and interpret the figure obtained

(b) The average number of customers in the system

(c) The average queue length

(d) The average time a customer spends in the system

(e) The average time a customer waits before being served

10. Customers arrive at an ATM where there is room for three customers to wait in line. Customers arrive alone with probability $\frac{2}{3}$ and in pairs with probability $\frac{1}{3}$ (but only one can be served at a time). If both cannot join, they both leaver call a completed services or an arrival an "event" and let the state

be the number of customer in the system (serviced and waiting) immediately after an event. Suppose that an event is equally likely to be an arrival or a completed service.

(a) State the transition graph and transition matrix and find the stationary distribution.

(b) If a customer arrivers, what is the probability that he finds the system empty? Full?

(c) If the system is empty, the time until it is empty again is called a "busy period". During a busy period, what is the expected number of times that the system is full?

(d) Show that a limit distribution is a stationary distribution.

REFERENCES

Amahia, G. N.(2007). *STA 211- Probability II. Ibadan Distance Learning Centre Series*. Distance Learning Centre, University of Ibadan.

Alawode, O. A. andShittu O. I. (2011). Probability III. An Unpublished lecture note developed for the Ibadan Distance Learning Programme.

Bhat, B. R. (1985). Modern Probability Theory: An Introductory Textbook. 2[nd] Edition. Wiley Eastern Ltd., Bombay.

Bhat U. N. (1971). Elements of Applied Stochastic Processes. John Wiley & Sons Ltd., London.

Attenborough, M. (2003). *Mathematics for Electrical Engineering and Computing*. Newnes. Elsevier. Linacre House, Jordan Hill, Oxford.

Brualdi, R. A. (1999). *Introductory Combinatorics*. Pearson Education Asia Limited and China Machine Press.

Banneheka BMSG, Ekanayake GEMUPD (2009) "A new point estimator for the median of gamma distribution". *Viyodaya J Science*, 14:95–103

Barry C. Arnold (1983). *Pareto Distributions*. International Co-operative Publishing House. ISBN 0-89974-012-X.

Dan Musa (2010): http://probabilityandstats.wordpress.com/2010/02/18/the-matching-problem/

Grimaldi, R. P. (1999). *Discrete and Combinatorial Mathematics*. Pearson Addison-Wesley.

Gupta, B. D. (2001). *Mathematical Physics*. Vikas Publishing House PVT Ltd.

Hogg, R. V. and Craig, A. T. (1970). *Introduction to Mathematical Statistics*. New York: Macmillan Publishing Co.

Krishnamoorthy, K. (2006). *Handbook of Statistical Distributions with Applications*. Chapman& Hall/CRC.

Merris, R. (2003). Combinatorics. Wiley-Interscience.

Mood, A. M., Graybill, F. A. and Boes, D. C. (1974). *Introduction to theory of Statistics*. McGraw-Hill Inc.

Rosen, K. H., Michaels, J. G., Gross, J. L., Grossman, J. W. and Shier, D. R. (2000). *Handbook of Discrete and Combinatorial Mathematics*. CRC Press.

Odeyinka, J. A. and Oseni, B. A. (2008). *Basic Tools in Statistical Theory*. Highland Publishers.

Olubusoye, O. E. (2000). Unpublished*Lecture notes on STA 311*.

Roussas, G. G. (1973). *A First Course in Mathematical Statistics*. Addison-Wesley Publishing Company.

Shittu, O. I. (2011).Unpublished*Notes on Probability II*.

www.wikipedia.org

Encyclopaedia of Physics (2nd Edition), R.G. Lerner, G.L. Trigg. VHC publishers, 1991, ISBN (Verlagsgesellschaft) 3-527-26954-1, ISBN (VHC Inc.) 0-89573-752-3

Feller W. (1970). An Introduction to Probability theory and its applications. 3[rd] Edition. John Wiley and Sons Inc, New York.

Fisz M. (1963): Probability Theory and Mathematical Statistics. John Wiley and Sons, New York.

Fisher Snedecor (1938). The true characteristic function of the F distribution. *Biometrika*, 69, 261-264.

Ilori S. A. and Ajayi O. O. (2000). Algebra. University Mathematics Series. Y-Books (a

Johnson NL, Kotz S, Balakrishnan N (1994) Continuous univariate distributions Vol 1. Wiley Series in Probability and Statistics.

Krishnamoorthy K. (2006): Handbook of Statistical Distributions with Applications. *Chapman & Hall/CRC*

Lord, Nick. (July 2010). "Binomial averages when the mean is an integer", The Mathematical Gazette 94, 331-332. division of Associated Book Makers Nig. Ltd.)

Mandl F (2008) *Statistical Physics* (2nd Edition). F. Mandl, Manchester Physics, John Wiley & Sons. ISBN 9780471915331

Maxwell, J.C. (1860) Illustrations of the dynamical theory of gases. *Philosophical Magazine* 19, 19-32 and *Philosophical Magazine* 20, 21-37.

Mood, A. M., Graybill, F. A. and Boes, D. C. (1963). *Introduction to theory of Statistics.* McGraw-Hill Books Coy.

Morters P. and Peres Y. (2008). Brownian Motion. An Unpublished lecture note on the web.

Nadarajah, S. and Kotz, S (2006).The beta-exponential distribution. Reliability Engineering and System Safety, 91(1):689-697.

Neumann, P. (1966). "Über den Median der Binomial- and Poissonverteilung". *Wissenschaftliche Zeitschrift der Technischen Universität Dresden* (in German) **19**: 29–33.

Odeyinka J.A., and Oseni B.A. (2008): Basic Tools in Statistical Theory. *Highland Publishers*

Ollofsson P. (2005): Probability, Statistics and Stochastic Processes. John Wiley and Sons, U.S.A.

Rogers L. C. G. and Williams D. (1994): Diffusions, Markov processes and Martingles, Vol. 1, second ed., Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, *John Wiley & Sons Ltd., Chichester, 1994. Foundations.*

Rogers L. C. G. and Williams D. (1994): Diffusions, Markov processes and Martingles. Vol. 2, *Cambridge Mathematical Library, Cambridge University Press, Cambridge, 2000,* Ito calculus, Reprint of the second (1994) EDITION.

Roussas G.G. (1973): A First Course in Mathematical Statistics. *Addison-Wiley Publishing Company*

Ross S. (1988). A First Course in Probability. Macmillan Publishing Coy. New York.

Ross, Sheldon M. (2009). *Introduction to probability and statistics for engineers and scientists* (4th ed.). Associated Press. p. 267. ISBN 978-0-12-370483-2.

Shangodoyin, D. K., Olubusoye, O. E., Shittu, O.I. and Adepoju. A. A. (2002). Statistical Theory and Methods. Joytal Printing Press, ISBN: 978-2906-23-9

Swarup K., Gupta P.K. and Mohan M. (1978): Operation Research. *Sultan Chand and Sons, New Delhi.* Second Ed. Reprinted.

Udofia G. (1997): Lecture Series on Stochastic Processes. *University of Uyo, Uyo, Nigeria.* (Unpublished)

Udomboso, C. G. and Shittu O. I. (2011). Stochastic Processes. A lecture note developed for the Ibadan Distance Learning Programme.

Ugbebor, O. O. and Bassey U. K. (2003). Mathematics for Users. University Mathematics Series. Y-Books (a division of Associated Book Makers Nig. Ltd.)

Ugbebor, O. O. (2010). Unpublished Lecture notes on Probability in Mathematics and Statistics Departments, University of Ibadan.

Wadsworth, G. P. (1960). *Introduction to probability and random variables.* USA: McGraw-Hill New York. p. 52.

Young, G.A. and Smith R. L. (2005). Essentials of Statistical Inference. Cambridge University Press

Young, H.D and Freedman R.A. (2008): University Physics – With Modern Physics (12th Edition), (Original edition), Addison-Wesley (Pearson International), 1st Edition: 1949, 12th Edition:, ISBN (10-) 0-321-50130-6, ISBN (13-) 978-0-321-50130-1