STATISTICAL THEORY AND METHODS

$$\int_{-\infty}^{\infty} x\,dx$$

$$\sum_{i=0}^{n}$$

D. K. Shangodoyin
O. E. Olubusoye
O. I. Shittu
A. A. Adepoju

STATISTICAL THEORY AND METHODS

D.K. Shangodoyin
O.E. Olubusoye
O.I. Shittu
A. A. Adepoju

First Edition 2002

ISBN: 978 – 2906 – 23 – 9

Printed by:
Joytal Printing Press
2, Adegbite Street opposite Sade Stores,
Agbowo U.I. Area, Ibadan.

Preface

This book is intended as a text for both undergraduate and graduate students and as a reference for researchers in statistical theory and methods. Although no prior knowledge of statistics is assumed, but a good grasp of real analysis and calculus is required for the advanced part of this textbook.

The primary purpose of the book is didactic, methods are emphasized and the book is subdivided into:

* Probability: the introductory part on the background of the concepts and problems are treated without advanced mathematical tools in chapters one to three

* Distribution theory: the general concepts and tools of random sequences; joint distributions as well as functions of random variables are emphasized with ease in chapters four and five.

* Inference: the basic ideas of staitstical inference that every probabilist and statistician requires in estimation and test of hypothesis are given in chapters six and seven.

* Advanced theory: the notion of advanced probability calculus and limit theorems; law of Numbers; generating functions and inversion theorem are discussed in chapter 8 through chapter 10.

D.K. Shangodoyin
O.E. Olubusoye
O.I. Shittu
A. A. Adepoju

Acknowledgements

Thanks to all our colleagues and students who have influenced the wri[illegible] of this book and have therefore contributed indirectly to its success. [illegible] are grateful to Mr. 'Kunle Gbenro of the Department of Mathema[illegible] University of Ibadan for producing the typescript of this book.

Dedication

To the memory of **Late Professor Samuel Oseni Adamu**, the Pioneer Acting Head of the Department of Statistics, University of Ibadan. (The first Department of Statistics in Nigeria).

Contents

Chapter One: Introduction

Chapter Two: Counting Technique

Chapter Three: Random Variables and Their Distributrions

Chapter Four: Moment-Generating Functions

Chapter Five: Sampling Theory

Chapter Six: Introduction to Statistical Inference

Chapter Seven: Tests of Hypotheses

Chapter Eight: Probability Calculus and Limit Theorems

Chapter Nine: Law of Large Numbers

Chapter Ten: Generating Functions and Inversion Theorem

Chapter 1

Introduction

1.1 Introduction

Probability can be 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 view point. 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. 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".

Section I of this book will be devoted to the study of the concept of probability, and its distribution together with the generating functions.

Since probability is a means to an end, a tool to enable us make valid statistical inferences.

Later, in Section II, we shall look into the concept of Statistical inference via the estimation procedure and hypothesis testing.

1.2 Basic Definitions

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 an outcome which can not be predicted. A trial usually results into only one of the possible outcomes e.g., A toss of a coin once, will lead to either a Head (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 can not be predetermined. e.g. A toss of two coins at a time; draw of two cards from a deck oneafter the other; a random selection of a ball from a box and examine the colour.

(c) **An outcome:** 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 collection of all possible outcomes of an experiment. It is a set of all finite or countably infinite number of elementary outcomes e_1; $e_2, \ldots$. It is usually represented by

$$S = e_1, e_2, \ldots, e_n.$$

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

Coin	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

$$S = \{11, 1, 2, 1, 3, 1, 4, 1, 5, 1, 6, 1, 2, 1, 22, \ldots, 6.6\}$$

(e) **An Event:** is a subset of a sample space.
It consist of one or more possible outcomes of an experiment. It is usually denoted by capital letters $A, B, C, D, \ldots$. 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
$A = \{\text{set of even number}\} = \{2, 4, 6\}$
$B = \{\text{set of prime number}\} = \{1, 3, 5\}$
$C = \{\text{set of number greater than } 7\} = \{\phi\}$.

(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 Exhurstive Events:** Events $A_1, A_2, A_3, A_4, \ldots, A_n$ 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 exhurstive. 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 experiments 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.

1.3 The Concept of Probability

The probability associated with an event is a measure of believe that an event will occur.

However, there are 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 approach:** 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.,

$$P(E_i) = \frac{1}{\text{Total number of outcomes of experiment}}$$

and

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

$$P(A) = \frac{n(A)}{n(S)}$$

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

Example 1.1: Find the probability of obtaining of obtaining any number in a simple thrown of a die.
Solution: The experiment has six outcomes 1, 2, 3, 4, 5, 6.

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

Example 1.2: Find the probability of obtaining an event number in one roll of a die.
Solution: Let A be the event of an even number.

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

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

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. Afterall we are not equally endowed.

(b) **The Limiting Frequency Approach:** 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 if 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 determine aposteriori. It should be noted that some events in real life can not be repeated before the probability is determined. Even if it can be determined the limit may not converge.

Example 1.3: Twenty of the 500 cars that enters the University of Ibadan on a graduation day are found to be Lexus. Assuming different cars comes into the campus randomly, what is the probability that the next car is a Lexus?

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

$$N = 500, \quad n = 20$$

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

$$P(\text{Lexus}) = \frac{n}{N} = \frac{20}{500} = 0.04$$

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

(d) **The Axiomatic or Bayesian Approach:** To circumvent the difficulties posed by the earlier approaches to the definition of probability, some 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).

1.4 Probability of an event

A in a relation to an experiment with sample space S is defined as a real valued function $P(A)$ which satisfy the following axioms:

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

(2) $P(S) = 1$

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

$= \sum_{i=1}^{\infty} P(A_i)$ for every finite or infinite sequence of disjoint event $A_1, A_2, \ldots$

1.5 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:
$S \cup \phi = S$.
$P(S \cup \phi) = S = 1$ by axiom (2)
$P(S) + P(\phi) = 1$ since $P(S) = 1$
$1 + P(\phi) = 1$
$\Rightarrow P(\phi) = 0$.

(c) **Theorem III - Addition Rule:**
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

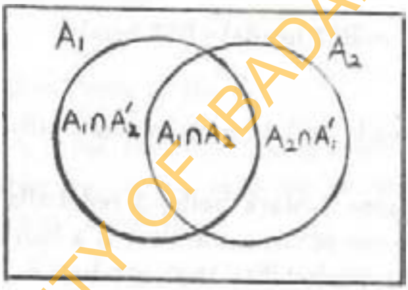


Fig 1.1

$P(A_1 \cup A_2) = P(A_1) \cup P(A_2) = 1$
$P(A_1 \cup A_2) = P(A_1) + P(A_2 \cap A_1^c)$
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

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

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

we have

$$\boxed{P(A_1 \cup A_2) = P(A_1) + P(A_2) \text{ Special Addition rule}}$$

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

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

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(at least 2 heads) = P(2 heads)+P(3 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 balls respectively. Total number of balls = 20.

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

(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) $P(B') = 1 - P(B) = 1 - \dfrac{8}{20} = 0.6$

(iv) $$\begin{aligned} P(B \cup G)^c &= 1 - P(B \cup G) \\ &= 1 - [P(B) + P(B)] \\ &= 1 - [\frac{8}{20} + \frac{4}{20}] \\ &= \frac{8}{20} \\ &= 0.4 \end{aligned}$$

Alternatively,

$$\begin{aligned} \text{P(neither Black nor Green)} &= \text{P(Yellow or Red)} \\ &= P(Y) + P(R) \\ &= \frac{5}{20} + \frac{3}{20} \\ &= \frac{8}{20} = 0.4 \end{aligned}$$

(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

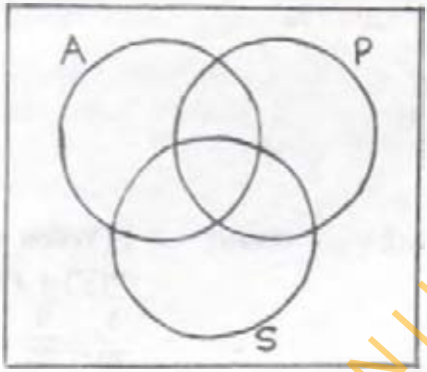
(ii) Statistics but not Physics

(iii) Statistics but not Physics and Algebra

(iv) Statistics, Algebra but not Physics

(v) Algebra or Physics

Solution: Let A, P and S denote the event of a student taking Algebra, Physics and Statistics respectively.
Presenting the information in a Venn diagram we have



$$n(A \cap A \cap B^c) = n(A \cap S) - n(A \cap P \cap S) = 10$$

$$n(P \cap S \cap A^c) = n(P \cap S) - n(A \cap P \cap S) = 164$$

$$n(A \cap P \cap S^c) = n(A \cap P) - n(A \cap P \cap S) = 30$$

Using the addition rule, we can find the number of students that takes all the three subjects.
$n(A \cup P \cup S) = n(A) + n(P) + n(S) - n(A \cap P) - n(A \cap S) + n(A \cap P \cap S)$
$500 = 186 + 329 - 83 - 217 - 63 + n(A \cap P \cap S)$
$\therefore\ n(A \cap P \cap S) = 53$
$\therefore$ P(All three subjects) $= \frac{53}{500} = 0.106$

(ii) P(Statistics but not Physics)

$= P(S \cap P^c)$
$= P(S) - P(S \cap P)$
$= \dfrac{329}{500} - \dfrac{217}{500}$
$= \dfrac{112}{500}$
$= 0.224$

(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)

$$\equiv P(A \cup P)$$
i.e. $P(\cup P) = P(A) + P(P) - P(A \cup P)$
$$= \frac{186}{500} + \frac{295}{500} - \frac{83}{500}$$
$$= \frac{398}{500}$$
$$= 0.796$$

1.6 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. $P(A/B) = \dfrac{P(A \cap B)}{P(B)}$ provided $P(B) \neq 0$

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

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

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

$$\begin{aligned} P(A_1 \cap A_2 \cap A_3) &= P(A_1) \cdot P(A_2/A_1) \cdot P(A_3/A_1 \cap A_2) \\ &= \frac{4}{52} \times \frac{3}{51} \times \frac{2}{50} \\ &= \frac{24}{132600} \\ &= 0.00018 \end{aligned}$$

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) $P(B \cap W) = P(B) \cdot P(W)$

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

(ii) $P(B_1 \cap B_2) = P(B) \times P(B)$

$$= \left(\frac{15}{25}\right)^2 = 0.36$$

(iii) P(both black or both white)

$$= P(B_1 \cap B_2) + P(W_1 \cap W_2)$$
$$= \left(\frac{15}{25}\right)^2 + \left(\frac{10}{25}\right)^2$$
$$= 0.36 + 0.16$$
$$= 0.52$$

(iv) P(both are of different colours)

$$= P(B \cap W) + P(W \cap B)$$
$$= \left[\frac{15}{25} \times \frac{10}{25}\right] + \left[\frac{10}{25} \times \frac{15}{25}\right]$$
$$= 2(0.24)$$
$$= 0.40$$

(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) $$P(B \cap W) = P(B) \cdot P(W/B)$$
$$= \frac{15}{25} \times \frac{10}{25} = 0.25$$

(ii) $$P(B_1 \cap B_2) = P(B_1) \cdot P(B_2/B_1)$$
$$= \frac{15}{25} \times \frac{14}{24} = 0.35$$

(iii) P(both black or both white)

$$= P(B_1)P(B_2/B_1) + P(W_1)P(W_2/W_1)$$
$$= \frac{15}{25} \times \frac{14}{24} + \frac{10}{25} \times \frac{9}{24}$$
$$= 0.35 + 0.15$$
$$= 0.50$$

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

$$= \frac{15}{25} \times \frac{10}{25} + \frac{10}{24} \times \frac{15}{24}$$
$$= 0.25 + 0.25$$
$$= 0.50$$

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

$$= \frac{15}{25} \times \frac{10}{24} / \frac{10}{25}$$
$$= \frac{0.25}{0.4}$$
$$= 0.625$$

1.7 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

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

Three events are said to be mutually independent if

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

$$P(A \cap B) = P(A) \cdot P(B);\ P(A \cap C) = P(A) \cot P(C);$$
$$P(B \cap C) = P(B) \cdot P(C) \text{ and}$$

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

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

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

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

(ii) $$\text{P(2 games and in ties)} = P(D.D.D^c) + P(D^c.D.D) + P(D.D^c.D)$$
$$= \left(\frac{1}{6} \times \frac{1}{6} \times \frac{5}{6}\right) + \left(\frac{5}{6} \times \frac{1}{6} \times \frac{1}{6}\right) + \left(\frac{1}{6} \times \frac{5}{6} \times \frac{1}{6}\right)$$
$$= \frac{5}{72}$$

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

$$= P(ABA) + P(B.A.B)$$
$$= \left(\frac{1}{2} \times \frac{1}{3} \times \frac{1}{2}\right) + \left(\frac{1}{3} \times \frac{1}{2} \times \frac{1}{3}\right)$$
$$= \frac{5}{36}$$

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

$$= 1 - P(B_1'B_2'B_3')$$
$$= 1 - \left(\frac{2}{3} \times \frac{2}{3} \times \frac{2}{3}\right)$$
$$= \frac{19}{27}$$

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:

$$P(\text{a six in a throw}) = \frac{1}{6}$$
$$P(\text{no six in a throw}) = \frac{5}{6}$$
$$(i) P(\text{at least 1 six in } n \text{ trials}) = 1 - P(\text{no six in } n \text{ trials})$$
$$= 1 - (\frac{5}{6})^n$$

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

$$\frac{1}{2} = 1 - (\frac{5}{6})^n$$
$$\Rightarrow (\frac{5}{6})^n = \frac{1}{2}$$
$$n\log(\frac{5}{6}) = \log(\frac{1}{2})$$
$$n = \frac{\log(1/2)}{\log(5/6)}$$
$$n = 4$$

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) $P(\text{king}) = \frac{4}{52}$, $P(\text{an ace}) = \frac{4}{52}$

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

$P(\text{Queen of diamond}) = \frac{1}{52}$

P(a king, 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

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

(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 one head Is events A, B and C mutually independent?

Solution:

$$S = \{HH, HT, TH, TT\}$$
$$A = \{HH, HT\}, \; B = \{HH, TH\}$$
$$C = \{HT, TH\}$$

$$A \cap B = \{HH\},\ A \cap C = \{HT\},\ B \cap C = \{TH\},\ A \cap B \cap C = \phi$$
$$\therefore P(A) = P(B) = P(C) = \frac{2}{4} = 0.5$$
$$P(A \cap B) = P(A) \cdot P(B) = \frac{1}{4}; P(B \cap C) = P(B) \cdot P(C) = \frac{1}{4}$$
$$P(A \cap C) = P(A) \cdot P(C) = \frac{1}{4};\ P(A \cap B \cap C) \neq P(A) \cdot P(B) \cdot P(C)$$

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 $c + d + 1$

$$\begin{aligned} & P(\text{white ball in the 2nd urn}) \\ &= P(W)P(W/B) + P(W)P(W/W) \\ &= \frac{c}{c+d+1}\left(\frac{p}{p+d}\right) + \frac{c}{c+d+1}\left(\frac{q}{p+q}\right) \\ &= \frac{c(p+q)}{(c+d+1)(p+q)} \\ &= \frac{c}{c+d+1} \end{aligned}$$

1.8 Total Probability rule and Baye's Theorem: If there are two or more causes of an outcome, it is often desirable to determine the probability that the outcome was due to a particular one of the possible causes. 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.

1.9 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 when the experiment was performed. Such that $P(A_i) \neq 0$, $i = 1, 2, \ldots, n$ then,

$$P(B) = \sum_{i=1}^{n} P(A_i)P(B/A_i)$$

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

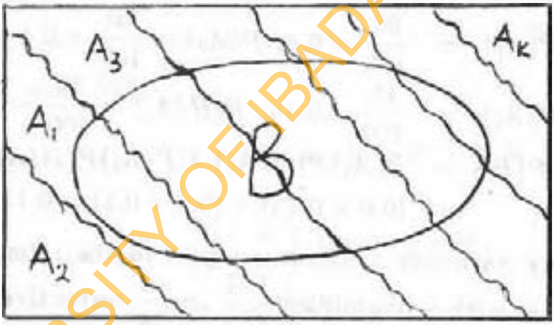


Fig 1.3

By definition of conditional probability, we have

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

$$P(A_i \cap B) = P(A_i)P(B/A_i) \qquad (1)$$

We know that

$$P(A_i/B) = \frac{P(A_i \cap B)}{P(B)} \qquad (2)$$

But total probability is

$$P(B) = P(A_i \cap B) + P(A_2 \cap B) + P(A_2 \cap B) + \ldots + P(A_n \cap B) \qquad (3)$$

Using (1) in (3) we have

$$\begin{aligned} P(B) &= P(A_1)P(B/A_1) + P(A_2)P(B/A_2) + \ldots + P(A_n)P(B/A_n) \\ &= \sum_{i=1}^{n} P(A_i)P(B/A_i) \qquad (3) \end{aligned}$$

Using (3) in (2) we have

$$P(A_i/B) = \frac{P(A_i)P(B/A_i)}{\sum_{i=1}^{n} P(A_i)P(B/A_i)} = \text{Baye's formula}$$

Example: Suppose 15% of apple and 10consignment were toxic. If the consignment consist 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.

$$\begin{aligned} P(A_1) &= \frac{60}{100} = 0.6; P(A_2) = \frac{40}{100} = 0.4 \\ P(B/A_1) &= \frac{15}{100} = 0.15; \ P(B/A_2) = \frac{10}{100} = 0.1 \\ P(B) &= P(A_1)P(B/A_1) + P(A_2)P(B/A_2) \\ &= (0.6 \times 0.15) + (0.4 \times 0.1) = 0.13 \end{aligned}$$

Example: 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.

$$\begin{aligned} P(S) &= \frac{1}{2}; P(F/S) = \frac{4}{5} \\ P(R) &= \frac{1}{4}; P(F/R) = \frac{2}{5} \\ P(L) &= \frac{1}{4}; P(F/L) = \frac{3}{5} \end{aligned}$$

(a) Using the idea of total probability,

$$\begin{aligned} P(F) &= P(S)P(F/S) + P(R)P(F/R) + P(L)P(F/L) \\ &= \frac{1}{2} \times \frac{4}{5} + \frac{1}{4} \times \frac{2}{5} + \frac{1}{4} \times \frac{3}{5} \\ &= \frac{13}{20} = 0.65 \end{aligned}$$

(b) Let the number of Saturdays on which he catches fish be a random variable X with $B(5, \frac{13}{20})$.

$$P(X \geq 3) = P(X = 3) + P(X = 4) + P(X = 5)$$

$$= \binom{5}{3}(0.65)^3(0.35)^2 + \binom{5}{3}(0.65)^4(0.35)^1 + \binom{5}{5}(0.65)^5(0.35)^0$$

$$= 0.3364 + 0.3124 + 0.116$$

$$= 0.765$$

c Here we need to calculate the probability that he goes to each of the locations without catching fish

$$\begin{aligned} P(S/F') &= \frac{P(S \cap F')}{P(F')} \\ &= \frac{P(S)P(F'/S)}{P(F')} = \frac{\frac{1}{2} \times \frac{1}{5}}{\frac{7}{20}} = \frac{2}{7} = 0.286 \end{aligned}$$

Similarly,

$$P(R/F') = \frac{P(R)P(F'/R)}{P(F')} = \frac{\frac{1}{4} \times \frac{3}{5}}{\frac{7}{20}} = \frac{3}{7} = 0.429$$

$$P(L/F') = \frac{P(L)P(P'/R)}{P(F')} = \frac{\frac{1}{4}\times\frac{2}{5}}{\frac{7}{20}} = \frac{2}{7} = 0.286$$

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

$$\begin{aligned} & P(S_1 \cap S_2) + P(R_1 \cap R_2) + P(L_1 \cap L_2) \\ = & \frac{1}{2}\times\frac{1}{3}+\frac{1}{4}\times\frac{1}{3}+\frac{1}{4}\times\frac{1}{3} \\ = & \frac{1}{3} = 0.33 \end{aligned}$$

Probability that they fail to meet on a Saturday is

$$(1-\frac{1}{3}) = \frac{2}{3} = 0.666$$

The probability that they fail to meet on three consecutive Saturdays is

$$(1-\frac{1}{3})^3 = \frac{8}{27} = 0.296$$

The probability that they meet at least once in three weekends is

$$\begin{aligned} &= 1 - P(\text{failed to meet}) \\ &= 1 - 0.296 \\ &= 0.703 \end{aligned}$$

Exercises:

1. If A_1, A_2 and A_3 be any three events, prove that

$$P(A_1 + A_2 + A_3) = \sum_{i=1}^{3} P(A_i) - \sum_{i=j} P(A_i A_j) + P(A_1 A_2 A_3).$$

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 Comment and the Times, 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 news agent 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 Comment, but not the Times?

Chapter 2

Counting Techniques

In simple experiments such as a roll of two dice or a roll of three coins, it is easy to determine the sample space. But when an experiments is such that it can be treated in three or more stage, it becomes tedious to determine the sample space as well as the number of outcomes in favour of a particular event.
Counting techniques are those methods developed to solve this problems.

2.1 Tree Diagram

The problem of counting the sample space and the points corresponding to various events is simplified by the use of a tree diagram, especially if the experiment can be treated as not more than three stages. If there are more than three stages, the tree becomes unmanageable.

Example 2.1: A bag contains 8 balls, identical except for colour, of which 5 are red and 3 are white. A man draws two balls at random, what is the probability that

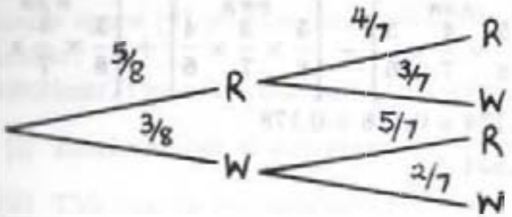
a(i) one of the balls shown is white and the other red.

(ii) both balls are of the same colour

(b) If three balls are drawn at random what is the probability that exactly 2 ball are red.

(c) What would be the probability in (a) if the first ball drawn is replaced before the second one is drawn?

Solution: Since two balls are drawn without replacement



The possible outcomes is easily seen from the diagram with the corresponding probabilities

$$\text{a(i)}\quad P(RW) = \left[\overset{RW}{\frac{5}{8} \times \frac{3}{7}}\right] \overset{\text{or}}{+} \left[\overset{WR}{\frac{3}{8} \times \frac{5}{7}}\right]$$

$$= 0.268 + 0.268$$

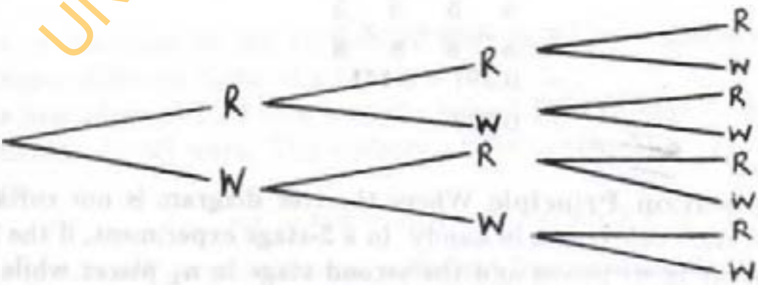
$$= 0.536$$

$$\text{(ii)}\quad \text{P(Same Colour)} = \left[\overset{RW}{\frac{5}{8} \times \frac{4}{7}}\right] \overset{\text{or}}{+} \left[\overset{WW}{\frac{3}{8} \times \frac{2}{7}}\right]$$

$$= 0.357 + 0.107$$

$$= 0.107 = 0.404$$

(b) If three balls are drawn

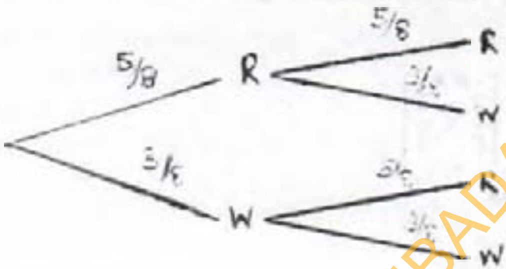


$$P(\text{exactly 2 red balls}) = \left[\frac{5}{8} \overset{RRW}{\times \frac{4}{7} \times} \frac{3}{6}\right] + \left[\frac{5}{8} \overset{RWR}{\times \frac{3}{7} \times} \frac{4}{6}\right] + \left[\frac{3}{8} \overset{WRR}{\times \frac{5}{7} \times} \frac{4}{6}\right]$$

$$= 0.178 + 0.178 + 0.178$$

$$= 0.534$$

(c) Balls are drawn with replacement, hence,



(i) $P(RW) = RW \, or \, WR$

$$= \frac{5}{8} \times \frac{3}{8} + \frac{3}{8} \times \frac{5}{8}$$

$$= 0.234 + 0.234$$

$$= 0.468$$

(ii) $P(\text{Same Colour}) = RR \, or \, WW$

$$= \frac{5}{8} \times \frac{5}{8} + \frac{3}{8} \times \frac{3}{8}$$

$$= 0.391 + 0.141$$

$$= 0.532$$

.2 Multiplication Principle Where the tree diagram is not suitable, e multiplication rule comes in handy. In a 3-stage experiment, if the first ge can occur in n_1 places and the second stage in n_2 places while the ird stage can occur in n_3 stages, then the total number of possible outmes becomes $n_1 \times n_2 \times n_3$. This can be extended to any number of stages.

Example 2.2: Two coins and a die are tossed at the same time (i) the sample space (ii) what is the probability of obtaining a head and an even number?

Solution: Two coins can occur in 4 ways, and one die can occur in 6 ways.

(i) Total number of outcomes is $4 \times 6 = 24$ ways.

(ii) This can be represented in

	1	2	3	4	5	6
HH						
HT		x		x		x
TH		x		x		x
HH						

X denotes the number of outcomes in favour of our event

$$P(H \text{ and even number}) = \frac{6}{24} = 0.25$$

2.3 Permutation The ordered arrangement of n distinct items taking all or r of them at a time is called permutation. The items are usually assume 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 filled 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 $n(n-1)(n-2)\dots(n-1r+1)!$ ways.

$$\therefore \quad n_{P_r} = n(n-1)(n-2)\dots(n-r+1)$$

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

$$n_{P_r} = n(n-1)(n-2)\dots 3.2.1 = n! \quad \text{ways}$$

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

Example 2.3: Evaluate 5_{P_3}
Solution:

$$\begin{aligned} 5_{P_3} &= \frac{5!}{(5-3)!} \\ &= \frac{5 \times 4 \times 3! \times 2!}{2!} \\ &= 5 \times 4 \times 3 \\ &= 60 \text{ ways} \end{aligned}$$

Example 2.4: If $13_{P_r} = 17160$, find r.
Solution:

$$\begin{aligned} 13_{P_r} &= 13(12)(11)\ldots(12-r+1) = 13(12)(11)(10) \\ &= 13 - r \\ \therefore \; r &= 3 \end{aligned}$$

Example 2.5: How many different words of three letters can be formed with letters A, B, C, D, E and F no letter is being 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 $6 \times 5 \times 4 \times 3$.
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 2.6: 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 are

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

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

(i) 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 $n - s_{P_{r-s}} \times r_{P_s}$.

(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 2.7: A committee of 7 representative of a class consist of class captain and his deputy. On a visit to the Head-teaching 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 were enough seats.

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

Solution:

(i) When there is no restriction

$$n = 7, \quad r = 4$$

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

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

$$\begin{aligned} 4_{P_2} \cdot (n-2)P_{r-2} &= 5P_2 \times 4P_2 = \frac{5!}{(5-3)!} \times \frac{4!}{(4-2)!} \\ &= 5 \times 4 \times 3 \times 12 \\ &= 12 \times 60 \\ &= 720 \text{ ways} \end{aligned}$$

(iii) Leave out the criminal then we have

$$\begin{aligned} {}^{n-1}P_4 = {}^6P_4 &= \frac{6!}{(6-4)!} \\ &= 6 \times 5 \times 4 \times 3 \\ &= 360 \text{ ways} \end{aligned}$$

(iv) $P(\text{event (i)}) = \dfrac{720}{840} = 0.857$

$P(\text{event (ii)}) = \dfrac{360}{840} = 0.428$

(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 2.8: 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 $219F_9 = 2 \times 9! = 72560$ ways. Number of arrangement without restriction is ${}^{10}P_{10} = 10! = 3628800$ ways so the permutation when the two books are not to stand together is

$$\begin{aligned} &10! - 2 \times 9! \\ &= 3628800 - 725760 \\ &= 2903040 \text{ ways} \end{aligned}$$

Example 2.9: 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.

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

(ii) If two R's and two A's do occur together we have (A,A)(R,R)N G E i.e., ${}^5P_5 = 5! = 120$ ways.

$$P(\text{R's and A's occur together}) = \frac{120}{1260} = 0.095$$

(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 2.10: 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 ${}^6P_4 = 360$ ways. The required number of ways of parking 5 blue cars and 4 red cars is $5! \times {}^6P_4$

$$= 120 \times 360$$
$$= 43200 \text{ ways}$$

(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 2.11: 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 2.12: 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 $= {}^5P_5 = 5! = 120$ 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 ${}^4P_4 = 4! = 24$ 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	$\times 2 = 24$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

$$^4P_4 = 4! = 24$$

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

(d) Suppose we have to form numbers which begin with 1 and end with 3. Here the first and the last places are fixed.
Then, the remaining 3 digits can be filled in

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

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 or the end. Then the two digits can be arranged among themselves in 2! ways. Hence total number of arrangement will be $^3P_3 \times 2 = 12$ 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

$$3 \times {}^4P_4$$
$$= 3 \times 24 = 72$$

Example 2.13: 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 ${}^4P_1$. The number of two digit number is ${}^5P_2$. 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 ${}^5P_2 - {}^4P_1$. Similarly, the number of three digit number is ${}^5P_3 - {}^4P_2$.
The total number of numbers is

$$\begin{array}{ccccccccc} {}^4P_1 & + & ({}^5P_2 - {}^4P_1) & + & ({}^5P_3 - {}^4P_2) & + & ({}^5P_4 - {}^4P_3) & + & ({}^5P_5 - {}^4P_4) \\ 4 & + & 16 & + & 48 & + & 96 & + & 96 \end{array}$$

260 numbers.

Example 2.14:

(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 ${}^5P_5 = 5! = 120$. 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 as $120(29) \times 1 = 3480 \times 1$.
Similarly the sum of numbers in the 10th place is also

$$\begin{aligned} 120(1+3+4+5+7+9) &= 3480 \times 10 \\ &= 34800 \end{aligned}$$

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

$$^6P_6 = 6! = 720$$

The number of numbers when 3 occupy the tenth place is $^5P_5 = 120$

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

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

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

Example 2.15: Suppose the letters of the word STAPLER is given to form words.

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

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

(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 $^2P_2 = 2$ ways. The other 6 digits can be filled in $^6P_6 = 6! = 70$ ways.
Hence total number of words is $2 \times 120 = 240$ 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.

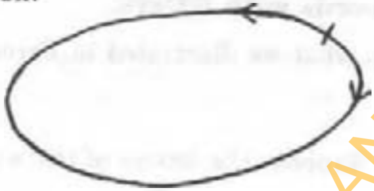
(e) Suppose two vowels A and E are to stand together. Regard A and E as one

$$a, E, STPLR$$

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 $2 \times 120 = 240$ 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 $6 \times 24 = 144$

(H) **Ordered:**
Arrangement of items round a circle:
Things can be arranged round a circle in (i) clockwise and (ii) anti-clockwise direction.



Example 2.16: 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 <u>not</u> 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 2.18: 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

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

2.4 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 $r(r-n)$ is denoted by

$$\binom{n}{r} \text{ or } {}^nC_r = \frac{n!}{(n-r)!r!}$$

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

Example 2.19: 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 2.20: 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 balls.

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 $\binom{5}{2}\binom{15}{2}$

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

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

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

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

$$= 0.983$$

(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 2.21: 15 players were invited for a crucial 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 2.22: A certain examination consist 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 $2 \times 2 \times \cdots \times n$ 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 2.23: 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

$$\begin{aligned} \binom{8}{1} + \binom{8}{2} + \cdots + \binom{8}{8} &= 2^n - 1 \\ &= 256 - 1 \\ &= 255 \text{ ways} \end{aligned}$$

<u>Note:</u>
If it must include a case where none of the questions is solved, then the required number is

$$\begin{aligned} \binom{8}{0} + \binom{8}{1} + \binom{8}{2} + \cdots + \binom{8}{8} &= 2^8 \\ &= 256 \text{ ways} \end{aligned}$$

Example 2.24: 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 include 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 2.25: How many factors has 2160?
Solution: The factors of 2160 are i.e.

$$\begin{aligned} 2160 &= 16 \times 27 \times 5 \\ &= 2^4 \times 3^3 \times 5^1 \end{aligned}$$

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 $5 \times 4 \times 2 = 40$.

(D) When Sharing (Dividing) in 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 2.26:

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

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

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

Solution: $4!\frac{52!}{(13!)^4} = 1.28 \times 110^{30}$

Example 2.27: 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

$$\frac{18!}{(6!)^3} = 17,153,136 \text{ ways}$$

(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 requires the application of the permutation and combination approaches simultaneously. We shall give a theory which may not be proved.

Theorem: If there are m different things of one kind, 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

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

Example 2.28: How many ways can 5 boys and 4 girls selec.ed 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 2.29:

(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 consc ·nts can be chosen from 8 in $\binom{8}{3}$

the 5 let· s can be arranged among themselves in 5! ways. The requ red 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

$$\begin{aligned}&\binom{5}{3}\binom{8}{2}5! + \binom{5}{2}\binom{8}{2}5!\\ = &\ 33600 + 67200\\ = &\ 100800 \text{ ways}\end{aligned}$$

(F) Combined 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

$$\begin{aligned} nHr &= \binom{n+r-1}{r} = \frac{(n+r-1)!}{(n+r-1-r)!r!}\\ &= \frac{)n+r-1)(n+r-2)\cdots(n+r-r-1)(n-1)n}{r!}\\ &= \frac{(n+r-1)(n+r-2)\cdots n}{r!}\end{aligned}$$

Example 2.30: 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

$$\begin{aligned} 6H_4 &= \left(\frac{6+4-1}{4!}\right) = \frac{9!}{4!5!}\\ &= 126\end{aligned}$$

Example 2.31: In an experiment, 2 dice are rolled once. Find the total number of outcomes if

(i) they are distinct

(ii) they are ot 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

$$\binom{6+2-1}{2} = \frac{7!}{2!5!} = 21.$$

Exercises:

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

(2) If ${}^nC_{n-4} = 15$; 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 two 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 he 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 5 books be selected?

Chapter 3

Random Variables and Their Distributrions

3.1 Introduction

In a statistical experiment, the set of all possible outcomes is termed the sample space. Some experiments yields sample spaces whose elements are numbers, but some other experiments do not yield numerically valued elements. For mathematical convenience, it is often desirable to associate one or more numbers (in addition to probabilities) with each possible outcomes of an experiment.

In this chapter we shall study the integer-valued random variables which is known as discrete random variables, continuous random variable and their distribution.

Definition: Random Variable

A random variablex X on a sample space S is a function $X : X \longrightarrow \mathbb{R}$ that assigns a real number $X(s)$ to each sample point $s \in S$.

Capital letters, such as X, Y and Z will be used to denote random variables. The lower case letters $x, y, z, \ldots$ will be used to denote possible values that the corresponding random variables can attain.

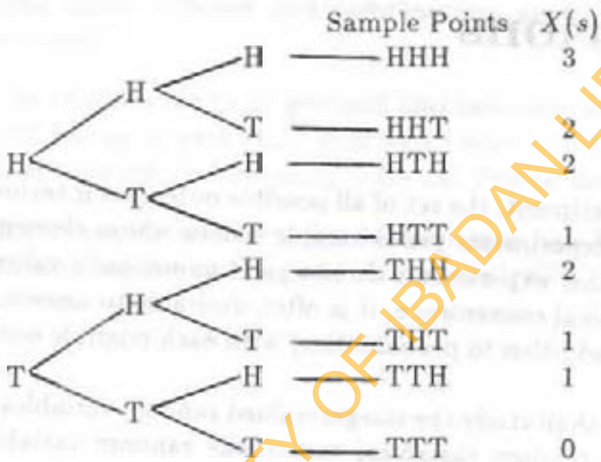
Example 1: A fair coin is tossed three times. The sample space

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

Let X denote the number of heads which appears. Then

$$X(s) = \{0, 1, 2, 3\}.$$

The tree diagram is shown in figure



If the outcome of one performance of the experiment were $S = THT$, then the resulting experimental value of the random variable X be 1 - that is,

$$X(THT) = 1$$

Note that two or more sample points might give the same value for X (i.e., X may not be a one-to-one function), but that two different numbers in the range cannot be assigned to the same sample point (i.e., X is a well-defined function). For example

$$X(HTT) = X(THT) = X(TTH) = 1$$

Discrete Random Variables

Random variables that arise from counting operations, such as the random variable in example 1 are integer-valued. Integer-valued random variables are examples of an important special type known as discrete random variables.

Definition

If the set of all possible values of a random variable, X, is a countable set, $x_1, x_2, \cdots, x_n$, or $x_1, x_2, \cdots$, then X is called a discrete random variable. The function

$$f(x) = P[X = x] \qquad x = x_1, x_2, \cdots \tag{1.1}$$

that assigns the probability to each possible value x will be called the *discrete probability density function* (discrete pdf).

Another common terminology is *probability mass function* (pmf), and the possible values, x_i, are called *mass points* of X.

Sometmes a subscribed notation, $f_X(x)$, is used. The following theorem gives general properties that any discrete pdf must satisfy.

Theorem 1.1

A function $f(x)$ is a discrete pdf if and only if it satisfies both of the following properties for at most a countably infinite set of reals $x_1, x_1, \cdots$

$$f(x_i) \geq 0 \tag{1.2}$$

for all x_i, and

$$\sum_{\text{all } x_i} f(x_i) = 1 \tag{1.3}$$

Proof

Property (1.2) follows from the fact that the value of a discrete pdf is a probability and must be non-negative. In case of property (1.3), since $x_1, x_2, \cdots$ represent all possible values of X, the events $[X = x_1], [X = x_2] \cdots$ constitute an exhaustive partition of the sample space. Thus,

$$\sum_{\text{all } x_i} f(x_i) = \sum_{\text{all } x_i} P[X = x_i] = 1.$$

Example 2

Returning to example 1,

$$f(0) = P[X = 0] = \frac{1}{8}$$

$$f(1) = P[X = 1] = \frac{3}{8}$$

$$f(2) = P[X = 2] = \frac{3}{8}$$

$$f(3) = P[X = 3] = \frac{1}{8}.$$

Here

$$f(x_i) \geq 0 \quad \text{for} \quad x_i = 0, 1, 2, 3$$

and

$$\begin{aligned} \sum_{x_i=0}^{3} f(x_i) &= f(0) + f(1) + f(2) + f(3) \\ &= \frac{1}{8} + \frac{3}{8} + \frac{3}{8} + \frac{1}{8} \\ &= \frac{8}{8} \\ &= 1 \end{aligned}$$

Definition: Cumulative distribution function (CDF)

The *cumulative distribution function* (CDF) of a random variable X is defined for any real x by

$$F(x) = P[X \leq x] \qquad (1.4)$$

The function $F(x)$ is often referred to simply as the *distribution function* of X, and the subscripted notation, $F_X(x)$, is sometimes used.

For brevity, we will often employ a short notation to indicate that a distribution of a particular form is appropriate. If we write $X \sim f(x)$ or $X \sim F(x)$, this will mean that the random variable X has pdf $f(x)$ and CDF $F(x)$.

The general relationship between $F(x)$ and $f(x)$ for a discrete distribution is given by the following theorem.

Theorem 1.2

Let X be a discrete random variable with pdf $f(x)$ and CDF $F(x)$. If the possible values of X are indexed in increasing order, $x_1 < x_2 < x_3 < \cdots$, then

$$f(x_1) = F(x_1), \quad \text{and for any } i > 1$$
$$f(x_i) = F(x_i) - F(x_{i-1}) \tag{1.5}$$

Also, if $x < x_1$ then $F(x) = 0$, and for any other real x

$$F(x) = \sum_{x_i \le x} f(x_i) \tag{1.6}$$

Expectation and Variable of Random Variable Definition

If X is a discrete random variable with pdf $f(x)$, then the *expected value* of X is defined by

$$E(X) = \sum_x x f(x) \tag{1.7}$$

The sum (1.7) is understood to be over all possible values of X. Also, it is an ordinary sum if the range of X is finite, and an infinite series if the range X is infinite. In the latter case, if the infinite series is not convergent, then we will say that $E(X)$ does not exist. Other common notations for $E(X)$ include μ, possibly with a subscript, μ_X. The terms mean and expectation are also often used.

The *variance* of X is given by

$$var(X) = E[(X - \mu)^2] \tag{1.8}$$

Other common notations for the variance are σ^2, σ_X^2, or $V(X)$, and a related quantity, called the *standard deviation* of X, is the positive square root of the variance, $\sigma = \sigma_X = \sqrt{var(X)}$. It is possible to express the variance in terms of $\mu = E(X)$ and $E(X^2)$.
In particular

$$\begin{aligned} var(X) &= E[(X - \mu)^2] = \sum_x (x - \mu)^2 f(x) \\ &= \sum_x (x^2 - 2\mu x + \mu^2) f(x) \end{aligned}$$

$$= \sum_x x^2 f(x) - 2\mu \sum_x x f(x) + \mu^2 \sum_x f(x)$$
$$= E(X^2) - 2\mu E(X) + \mu^2(1) \text{ by eqn. 1.3}$$
$$= E(X^2) - 2\mu \cdot \mu + \mu^2$$
$$= E(X^2) - \mu^2$$
$$= E(X^2) - [E(X)]^2$$

Example

A discrete random variable X has a pdf of the form

$f(x) = k(8 - x)$ for $x = 0, 1, 2, 3, 4, 5$, and zero otherwise.

(a) Find the constant C.

(b) Find the CDF, $F(x)$

(c) Find $P[X > 2]$

(d) Find $E(X)$

(e) Find $var(X)$.

Solution

(a) If $f(x)$ is a pdf, then

$$\sum_{x=0}^{5} f(x) = 1$$
$$\therefore \sum_{x=0}^{5} k(8 - x) = 1$$
$$k \sum_{x=0}^{5} (8 - x) = 1$$
$$k[8 + 7 + 6 + 5 + 4 + 3] = 1$$
$$33k = 1$$
$$k = \frac{1}{33}$$
$$\therefore \; f(x) = \frac{1}{33}(8 - x).$$

(b)
$$F(x) = P(X \le x) = \sum_{0}^{x} \frac{1}{33}(8-x)$$

(c)
$$\begin{aligned} P[X>2] &= 1-F(2) \\ &= 1-P(X\le 2) \\ &= 1-\sum_{0}^{2}\frac{1}{33}(8-x) \\ &= 1-\left[\frac{8}{33}+\frac{7}{33}+\frac{6}{33}\right] \\ &= 1-\frac{21}{33} \\ &= \frac{12}{33} \\ &= 0.636 \end{aligned}$$

(d)
$$\begin{aligned} E(X) &= \sum_{x=0}^{5} x f(x) \\ &= 0\times\frac{8}{33}+1\times\frac{7}{33}+2\times\frac{6}{33}+3\times\frac{5}{33}+4\times\frac{4}{33}+5\times\frac{3}{33} \\ &= \frac{0}{33}+\frac{7}{33}+\frac{12}{33}+\frac{15}{33}+\frac{16}{33}+\frac{15}{33} \\ &= \frac{65}{33} \\ &= 1.9697 \end{aligned}$$

(e)
$$\begin{aligned} E(X^2) &= 0^2\times\frac{8}{33}+1^2\times\frac{7}{33}+2^2\times\frac{6}{33}+3^2\times\frac{5}{33}+4^2\times\frac{4}{33}+5^2\times\frac{3}{33} \\ &= \frac{0}{33}+\frac{7}{33}+\frac{24}{33}+\frac{45}{33}+\frac{64}{33}+\frac{75}{33} \\ &= \frac{215}{33} \\ &= 6.5152 \end{aligned}$$

$$\begin{aligned} \therefore\ var(X) &= E[X^2]-[E(X)]^2 \\ &= 6.5152-[1.9697]^2 \\ &= 2.6355 \end{aligned}$$

Example

A distribution of positive integers has probability function

$$P(x) = \frac{1}{31}\binom{5}{x} \qquad \text{for } x = 1,2,3,4,5$$

$$P(x) = 0 \qquad \text{for } x > 5$$

Prove that the expected mean value is $\frac{80}{31}$ and that the variance is $\frac{1040}{961}$

Solution

$X = x$	1	2	3	4	5
$P(x)$	$\frac{5}{31}$	$\frac{10}{31}$	$\frac{10}{31}$	$\frac{5}{31}$	$\frac{1}{31}$

$$\begin{aligned}
\text{Mean} = E(X) &= \sum_{x=1}^{5} x f(x) \\
&= 1 \times \frac{5}{31} + 2 \times \frac{10}{31} + 3 \times \frac{10}{31} + 4\frac{5}{31} + 5 \times \frac{1}{31} \\
&= \frac{5}{31} + \frac{20}{31} + \frac{30}{31} + \frac{20}{31} + \frac{5}{31} \\
&= \frac{80}{31}
\end{aligned}$$

$$\begin{aligned}
E(X^2) &= \sum_{x=1}^{5} x^2 f(x) \\
&= 1^2 \times \frac{5}{31} + 2^2 \times \frac{10}{31} + 3^2 \times \frac{10}{31} + 4^2\frac{5}{31} + 5^2 \times \frac{1}{31} \\
&= \frac{5}{31} + \frac{40}{31} + \frac{90}{31} + \frac{80}{31} + \frac{25}{31} \\
&= \frac{240}{31}
\end{aligned}$$

$$\begin{aligned}
Var(X) &= E(X^2) - [E(X)]^2 \\
&= \frac{240}{31} - \left(\frac{80}{31}\right)^2 \\
&= \frac{1040}{961}
\end{aligned}$$

3.2 Bernoulli Distribution

Suppose that, on a single trial of an experiment, there are only two events of interest, say E and its complement E'. For example, E and E' could represent the occurrence of a "head" or a "tail" on a single coin toss, or, in general, "success" or "failure" on a particular trial of an experiment. Suppose that E occurs with probability $P = P(E)$ and consequently E' occurs with probability $q = P(E') = 1 - P$.

A random variable, X, that assumes only the values 0 or 1 is known as *Bernoulli variable*, and a performance of an experiment with only two types of outcomes is called a *Bernoulli trial*.

$f(0) = (X = 0) = q$

$f(1) = P(X = 1) = p$

where $p + q = 1$.

The pdf can be expressed as

$$f(x) = p^x q^{1-x} \quad x = 0, 1 \tag{1.9}$$

The corresponding CDF is given by

$$F(x) = \begin{cases} 0 & \text{for } x < 0 \\ q & \text{for } 0 \leq x < 1 \\ 1 & \text{for } x \geq 1 \end{cases} \tag{1.10}$$

Expectation And Variance of Bernoulli Distribution

$$\begin{aligned} E(X) &= 0 \cdot q + 1 \cdot p = p \\ E(X^2) &= 0^2 \cdot q + 1^2 \cdot p = p \end{aligned}$$

so that

$$Var(X) = p - p^2 = p(1 - p) = pq$$

3.3 Binomial Distribution

In a sequence of n independent Bernoulli trials with probability of success p on each trial, let X represent the number of successes. The discrete pdf of X is given by

$$b(x; n, p) = \binom{n}{x} p^x q^{n-x} \quad x = 0, 1, \ldots, n \tag{1.11}$$

The general properties (1.2) and (1.3) are satified by equation (1.11) since $0 \le p \le 1$ and

$$\begin{aligned}\sum_{x=0}^{n} b(x; n, p) &= \sum_{x=0}^{n} \binom{n}{x} p^x q^{n-x} \\ &= (p+q)^n \\ &= 1^n \\ &= 1\end{aligned}$$

The CDF of a binomial distribution is given at integer values by

$$B(x; n, p) = \sum_{k=0}^{x} b(k; n, p) \quad x = 0, 1, \cdots, n \qquad (1.12)$$

A short notation to designate that X has the binomial distribution with parameters n and p is

$$X \sim B(x; n, p)$$

or an alternative notation

$$X \sim BIN(n, p)$$

The Mean and Variance of $BIN(n, p)$

$$\begin{aligned}E[X] &= \sum_{x=0}^{n} x f(x) \\ &= \sum_{x=0}^{n} x \binom{n}{x} (1-p)^{n-x} p^x \\ &= \sum_{x=1}^{n} \frac{x}{x!} \frac{n!}{(n-x)!} (1-p)^{n-x} p^x,\end{aligned}$$

since the term when $x = 0$ contributes zero to the sum,

$$\begin{aligned}&= np \sum_{x=1}^{n} \frac{(n-1)!}{(x-1)![n-1-(x-1)]!} (1-p)^{(n-1)-(x-1)} p^{x-1} \\ &= np \sum_{R=0}^{N} \frac{N!}{R!(N-R)!} (1-P)^{N-R} P^R,\end{aligned}$$

writing $N = n-1$, $R = r-1$ in the summation,

$$\begin{aligned} &= np[(1-p)+p]^N \\ &= np. \end{aligned}$$

We could have shortened this work a little by using the notation $q = 1-p$, and we use this when finding $var(X)$.

$$\begin{aligned} E(X^2) &= \sum_{x=0}^{n} x^2 \binom{n}{x} q^{n-x}p^x \\ &= \sum_{x=1}^{n} [x(x-1)+x] \binom{n}{r} q^{n-x}p^x, \end{aligned}$$

Since the term for which $r = 0$ contributes zero to the summation,

$$\begin{aligned} &= \sum_{r=1}^{n} \frac{x(x-1)\cdot n!}{(n-x)!x!} q^{n-x}p^x + \sum_{x=1}^{n} x\binom{n}{x} q^{n-x}p^x \\ &= n(n-1)p^2 \sum_{x=2}^{n} \frac{(n-2)!}{(x-2)!(n-x)!} q^{n-x}p^{x-2} + np \end{aligned}$$

Since the term for which $x = 1$ contributes zero to the summation, and we have already shown, when finding $E(X)$, that the second summation is equal to np.

If we put $N = n-2$, $R = r-2$ in the first summation, we have

$$\begin{aligned} E[X^2] &= n(n-1)p^2 \sum_{R=0} \frac{N!}{R!(N-R)!} q^{N-R}p^R + np \\ &= n(n-1)p^2[q+p]^N + np \\ &= n(n-1)p^2 + np, \quad \text{since } p+q=1 \\ \therefore Var(X) &= n(n-1)p^2 + np - n^2p^2 \\ &= np(1-p) \\ &= npq \end{aligned}$$

3.4 Useful equality

In a problems on the binomial distribution where we have to calculate more than one probability, a useful equality connects consecutive terms in a binomial expansion. It is

$$p(x+1) = \frac{n-x}{x+1} \cdot \frac{p}{1-p} p(x).$$

For example, the value of $P(2)$ can be obtained from that of $(P(1)$ by using the equality

$$P(2) = \frac{n-1}{2} \cdot \frac{p}{1-p} p(1)$$

where n and p will be given in the problem.

3.5 Hypergeometric Distribution

Suppose a population or collection consists of a finite number of items, say N, and there are M items of type 1 and the remaining $N-M$ items are of type II. Suppose n items are drawn at random without replacement, and denote by X the number of items of type 1 that are drawn. The discrete pdf of X is given by

$$f(x; n, M, N) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} \tag{1.13}$$

The underlying sample space is taken to be the collection of all subsets of size n, of which there are $\binom{N}{n}$, and there are $\binom{M}{x}\binom{N-M}{n-x}$ outcomes that correspond to the event $[X = x]$.

A short notation to designate that X has the hypergeometric distribution with parameters $n, M,$ and N is

$$X \sim \text{HYP}(n, M, N)$$

An Identity

For a hypergeometric distribution we must have $\sum_{x=0}^{M} P(x) = 1$, because $P(x)$ represent all the possibilities that can occur; it follows that

$$\sum_{x=0}^{M} \binom{M}{x}\binom{N-M}{n-x} = \binom{N}{n}$$

Mean and Variance of HYP(n, M, N)

$$E(X) = \sum_{x=0}^{M} \frac{x\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$$

$$\binom{N}{n} E(X) = \sum_{x=0}^{M} x \binom{M}{x}\binom{N-M}{n-x}$$
$$= M \sum_{x=1}^{M} \frac{(M-1)!}{(x-1)!(M-x)!}\binom{N-M}{n-x}$$
$$= M \sum_{t=0}^{M-1} \frac{(M-1)!}{t!(M-t-1)!}\binom{N-M}{n-t-1}$$

(setting $t = x - 1$, that is, $x = t + 1$, so that, as x goes from 1 to M, t goes from 0 to $M - 1$)

$$= M \sum_{t=0}^{M-1} \binom{M-1}{t}\binom{N-1-(M-1)}{n-t-1}$$
$$= M \binom{N-1}{n-1}$$

Using the identity given above, it follows that

$$E(X) = M \binom{N-1}{n-1} \Big/ \binom{N}{n}$$
$$= M \times \frac{(N-1)!}{(n-1)!(N-n)!} \times \frac{n!(N-n)!}{N!}$$
$$= n \cdot \frac{M}{N}$$

Note:
If we were sampling with replacement, $P = \frac{M}{N}$ would be the appropriate binomial parameter and the binomial mean would be $np = n\frac{M}{N}$, the same as for the hypergeometric distribution.

Variance We first find $E[X(X-1)]$

$$\binom{N}{n} E[X(X-1)] = \sum_{x=0}^{M} x(x-1) \binom{M}{x}\binom{N-M}{n-x}$$
$$= M(M-1) \sum_{x=2}^{M} \frac{(M-2)!}{(x-2)!(M-x)!}\binom{N-M}{n-x}$$

Now set $t = x - 2$

$$= M(M-1)\sum_{t=0}^{M-2}\frac{(M-2)!}{t!(M-t-2)!}\binom{N-M}{n-2-t}$$

(use the identity)

$$\begin{aligned} &= M(M-1)\sum_{t=0}^{M-2}\binom{M-2}{t}\binom{N-2-(M-2)}{n-2-t} \\ &= M(M-1)\binom{N-2}{n-2} \end{aligned}$$

Thus

$$\begin{aligned} E[X(X-1)] &= M(M-1)\binom{N-2}{n-2}\Big/\binom{N}{n} \\ &= M(M-1)\frac{n(n-1)}{N(N-1)} \end{aligned}$$

It follows ghat

$$\begin{aligned} V(X) &= E[X(X-1)] + E[X] - [E(X)]^2 \\ &= M(M-1)\frac{n(n-1)}{N(N-1)} + n\frac{M}{N} - n^2\frac{M^2}{N^2} \\ &= n\frac{M}{N}\left(1-\frac{M}{N}\right)\left(\frac{N-n}{N-1}\right) \end{aligned}$$

Note:
If we are sampling with replacement, $p = \frac{M}{N}$ would be the appropriate binomial parameter and the binomial variance would be

$$np(1-p) = n\frac{M}{N}\left(1-\frac{M}{N}\right)$$

is is slightly greater than the hypergeometric variance because of the ... tor $(N-n)/(N-1)$ in the latter. As N becomes very large compared with n (the number of trials), the hypergeometric distribution tends to the binomial distribution.

Theorem

If $X \sim \text{HYP}(n, M, N)$, then for each value $x = 0, 1, \ldots, n$ and as $N \to \infty$ and $M \to \infty$ with $M/N \to P$, as a positive constant,

$$\lim_{N\to\infty} \frac{\binom{M}{n}\binom{N-M}{n-x}}{\binom{N}{n}} = \binom{n}{x} p^x (1-p)^{n-x} \tag{1.14}$$

Note:

The binomial distribution is applicable when we sample with replacement, while the hypergeometric distribution is applicable when we sample without replacement. If the size of the collection sampled from is large, it shoud not make a great deal of difference whether or not a particular item is returned to the collection before the next one is selected.

3.6 The Geometric Distribution

This distribution can arise in an experiment which fulfils the conditions which are required to be satisfied for the binomial distribution except that, instead of counting the number of 'succeses' which occur in the n trials, as we die for binomial, we carry on with the trials only until we get one 'success'.

If we denote by X the number of trials required to obtain the first success, then the discrete pdf of X is given by

$$f(x;p) = pq^{x-1}; \quad x = 1, 2, 3, \ldots \tag{1.15}$$

A special notation to designate that X has geometric distribution is $X \sim \text{Geo}(P)$.

The general properties of (1.2) and (1.3) are satisfied by (1.15), since $0 < p < 1$ and

$$\begin{aligned}\sum_{x=1}^{\infty} f(x;p) = P\sum_{x=1}^{\infty} q^{x-1} &= P(1 + q + q^2 + \cdots) \\ &= P\left(\frac{1}{1-q}\right) \\ &= \frac{p}{q} \\ &= 1\end{aligned}$$

3.7 Mean and Variance of GEO(P)

The mean of $X \sim$GEO(P) is obtained as follows:

$$\begin{aligned} E(X) &= \sum_{x=1}^{\infty} xpq^{x-1} \\ &= \sum_{x=1}^{\infty} p\frac{d}{dq}q^x \\ &= p\frac{d}{dq}\sum_{k=0}^{\infty} q^k \\ &= p\frac{d}{dq}\left(\frac{1}{1-q}\right) \\ &= \frac{p}{(1-q)^2} \\ &= \frac{p}{p^2} \\ &= \frac{1}{p} \end{aligned}$$

In other words, if independent trials, having a common probability p of being successful, are performed until the first success occurs, then the expected number of required trials equals $\frac{1}{p}$.

To determine the $Var(X)$ let us first compute $E[X^2]$. With $q = 1 - p$:

$$\begin{aligned} E[X^2] &= \sum_{x=1}^{\infty} x^2pq^{x-1} \\ &= \sum_{x=1}^{\infty} \frac{d}{dq}(xq^x) \\ &= p\frac{d}{dq}\left(\sum_{x=1}^{\infty} xq^x\right) \\ &= p\frac{d}{dq}\left(\frac{q}{1-q}E[X]\right) \\ &= p\frac{d}{dq}[q(1-q)^{-2}] \\ &= p\left[\frac{1}{p^2} + \frac{2(1-p)}{p^3}\right] \end{aligned}$$

$$= \frac{2}{p^2} - \frac{1}{p}$$

Hence, since $E[X] = \frac{1}{p}$

$$\begin{aligned} Var(X) &= \frac{(1-p}{p^2} \\ &= \frac{q}{p^2} \end{aligned}$$

3.8 Negative Binomial Distribution

In repeated independent Bernoulli trials, let X denote the number of trials requird to obtain r successes. Then the probability distribution of X is the Negative binomial distribution with discrete pdf given by

$$f(x; r, p) = \begin{pmatrix} x-1 \\ r-1 \end{pmatrix} p^r (1-p)^{x-r} \tag{1.16}$$

A special notation, which designates that X has the negative binomial distribution is

$$X \sim NB(r, p).$$

The general properties (1.2) and (1.3) are satisfied by (1.16), since $0 < p < 1$

$$\begin{aligned} \sum_{x=r}^{\infty} \begin{pmatrix} x-1 \\ r-1 \end{pmatrix} p^r q^{x-r} &= p^r \sum_{i=0}^{\infty} \begin{pmatrix} i+r-1 \\ r-1 \end{pmatrix} q^i \\ &= p^r (1-q)^{-r} \\ &= 1 \end{aligned}$$

Note that $\sum_{i=0}^{\infty} \begin{pmatrix} i+r-1 \\ r-1 \end{pmatrix} q^i$ is the series expansion of $(1-q)^{-r}$.

3.9 Mean and Variance of NB(Y, P)

$$\begin{aligned} E[X^k] &= \sum_{x=r}^{\infty} x^k \begin{pmatrix} x-1 \\ r-1 \end{pmatrix} p^r (1-p)^{x-r} \\ &= \frac{r}{p} \sum_{x=r}^{\infty} x^{k-1} \begin{pmatrix} x \\ r \end{pmatrix} p^{r+1} (1-p)^{x-r} \text{ since } x \begin{pmatrix} x-1 \\ r-1 \end{pmatrix} = r \begin{pmatrix} x \\ r \end{pmatrix} \\ &= \frac{r}{p} \sum_{m=r+1}^{\infty} (m-1)^{k-1} \begin{pmatrix} m-1 \\ r \end{pmatrix} p^{r+1} (1-p)^{m-(r+1)} \text{ by setting } m = x+1 \\ &= \frac{r}{p} E[(Y-1)^{k-1}] \end{aligned}$$

where Y is a negative binomial random variable with parameters $r+1$, p. Setting $k=1$ in the preceding equation yields

$$E[X] = \frac{r}{p}.$$

Setting $k=2$ in the preceding equation, and using the above formula for the expected value of a negative binomial random variable, gives that

$$\begin{aligned} E[X^2] &= \frac{r}{p}E[Y-1] \\ &= \frac{r}{p}\left(\frac{r+1}{p}-1\right) \end{aligned}$$

Therefore,

$$\begin{aligned} Var(X) &= \frac{r}{p}\left(\frac{r+1}{p}-1\right) - \left(\frac{r}{p}\right)^2 \\ &= \frac{r(1-p)}{p^2} \end{aligned}$$

3.10 Poisson Distribution

A discrete random variable X with probability function

$$f(x;\mu) = \frac{\mu^x e^{-\mu}}{x!}, \quad x = 0,1,2,\ldots \tag{1.17}$$

where $\mu > 0$ is said to have a Poisson distribution with parameter μ.

A special notation that designates that a random variable has the Poisson distribution with parameter μ is

$$X \sim POI(\mu)$$

The properties given by (1.2) and (1.3) are clearly satisfied, since $\mu > 0$ implies $f(x;\mu) \geq 0$ and

$$\sum_{x=0}^{\infty} f(x;\mu) = e^{-\mu}\sum_{x=0}^{\infty}\frac{\mu^x}{x!} = e^{-\mu}e^{\mu} = 1$$

3.11 Mean and Variance of $POI(\mu)$

$$\begin{aligned}
\text{Mean} = E[X] &= \sum_{x=0}^{\infty} \frac{xe^{-\mu}\mu^x}{x!} \\
&= e^{-\mu}\sum_{x=1}^{\infty} \frac{\mu\mu^{x-1}}{(x-1)!} \\
&= \mu e^{-\mu}\sum_{m=0}^{\infty} \frac{\mu^m}{m!} \quad \text{where } m = x-1 \\
&= \mu e^{-\mu}e^{\mu} \\
&= \mu
\end{aligned}$$

Next,

$$\begin{aligned}
E[X(X-1)] &= \sum_{x=0}^{\infty} x(x-1)e^{-\mu}\frac{\mu^x}{x!} \\
&= \sum_{x=2}^{\infty} x(x-1)e^{-\mu}\frac{\mu^x}{x(x-1)(x-2)!} \\
&= \mu^2 e^{-\mu}\sum_{x=0}^{\infty}\frac{\mu^x}{x!} = \mu^2
\end{aligned}$$

Hence

$$E(X^2) = \mu^2 + \mu$$

so that

$$Var(X) = \mu^2 + \mu - \mu^2 = \mu.$$

3.12 Poisson as a Limiting Form of the Binomial

If $X \sim \text{BIN}(n,p)$, then for each value $x = 0, 1, 2, \ldots$, and as $P \to 0$ with $np \to \mu$ constant,

$$\lim_{n\to\infty} \text{BIN}(n,p) = POI(\mu)$$

or

$$\lim_{n\to\infty} \binom{n}{x} P^x(1-P)^{n-x} = \frac{e^{-\mu}\mu^x}{x!} \tag{1.18}$$

Proof

$$\binom{n}{x} P^x(1-P)^{n-x} = \frac{n!}{x!(n-x)!}\left(\frac{\mu}{n}\right)^x\left(1-\frac{\mu}{n}\right)$$

$$= \frac{\mu^x}{x!}\left(\frac{n}{n}\right)\left(\frac{n-1}{n}\right)\cdots\left(\frac{n-x+1}{n}\right)\left(1-\frac{\mu}{n}\right)^n\left(1-\frac{\mu}{n}\right)^{-x}$$

$$\therefore \lim_{n\to\infty}\binom{n}{x}P^x(1-P)^{n-x} = \frac{\mu^x e^{-\mu}}{x!} \text{ since } \lim_{n\to\infty}\left(1-\frac{\mu}{n}\right)^n = e^{-\mu}$$

$$\text{and, forf fixed } x, \lim_{n\to\infty}\left(1-\frac{\mu}{n}\right)^{-x} = 1$$

Theorem 1.3

A function $f(x)$ is a continuous pdf if and only if it satisfies kboth of the following properties for all $x \in S$, the sample space

(i) $f(x) \geq 0$ (1.19)

(ii) $\int_S f(x)dx = 1$ (1.20)

Eq. (1.20) means integration over the sample space.

For any $x_0 < x_1$ in S,

$$P(x_0 < X \leq x_1) = \int_{x_0}^{x_1} f(x)dx \tag{1.21}$$

Equation (1.21) represents the area of the region under the graph of the probability density function $f(x)$ between the limits $x = x_0$ and $x = x_1$

Definition

The probability (cumulative) distribution function, $F(x)$ is defined by the relation

$$F(x_0) \equiv P(X \leq x_0) = \int_{-\infty}^{x_0} f(x)dx \tag{1.22}$$

This represents the area of the region under the graphy of the pdf, $f(x)$, from $x = -\infty$ to $x = x_0$. The function $F(x)$ obviously increases from zero at the bottom of the range to unity at the top of the range.

From Eq. (1.22)

$$f(x) = \frac{dF(x)}{dx} \tag{1.23}$$

3.13 Expectation and Variance of a Continuous Random Variable

We define

$$\mu = E[X] = \int_{-\infty}^{+\infty} xf(x)dx$$

the expected mean value of X, and

$$\sigma^2 = var(X) = \int_{-\infty}^{+\infty} (x-\mu)^2 f(x)dx = \int_{-\infty}^{+\infty} x^2 f(x)dx - \mu^2$$

the variance of X.

Proof

$$\begin{aligned}\int_{-\infty}^{+\infty} (x-\mu)^2 f(x)dx &= \int_{-\infty}^{+\infty} x^2 f(x)dx - 2\mu \int_{-\infty}^{+\infty} xf(x)dx \\ &\quad + \mu \int_{-\infty}^{+\infty} f(x)dx \\ &= \int_{-\infty}^{+\infty} x^2 f(x)dx - 2\mu^2 + \mu^2 \\ &= \int_{-\infty}^{+\infty} x^2 f(x)dx - \mu^2\end{aligned}$$

Since

$$\int_{-\infty}^{+\infty} xf(x)dx = \mu, \int_{-\infty}^{+\infty} f(x)dx = 1,$$

or

$$Var(X) = \int_{-\infty}^{+\infty} x^2 f(x)dx - \mu^2 = E(X^2) - \mu^2.$$

Example 1

The pdf of the continuous random variable X is given by

$$f(x) = kx^2, \quad 0 \leq x \leq 1$$

$$f(x) = 0, \quad \text{elsewhere}$$

Find (a) the value of k

(b) $P(Xl;eq\frac{1}{2})$

(c) $P(\frac{1}{4} < X < \frac{1}{2})$

Solution

(a) $\int_{-\infty}^{+\infty} f(x)dx = 1$

$\Rightarrow \int_0^1 kx^2 dx = \frac{k}{3} = 1$

$\Rightarrow k = 3$

(b) $P(X \leq \frac{1}{2}) = \int_{-\infty}^{\frac{1}{2}} f(x)dx = \int_0^{\frac{1}{2}} 3x^2 dx = \frac{1}{8}$

(c) $P(\frac{1}{4} < X < \frac{1}{2}) = \int_{\frac{1}{4}}^{\frac{1}{2}} 3x^2 dx = \frac{7}{64}$

Example 2

A random variable X has cumulative distribution function

$$F(x) = 0, \quad x \leq 0$$

$$F(x) = kx^4, \quad 0 < x \leq 2,$$

$$F(x) = 1, \quad x > 2.$$

Solution

(a) $f(x) = \frac{dF(x)}{dx} = 4kx^3$

$F(x)$ must be unity at $x = 2$,

$\Rightarrow 16k = 1, \; k = \frac{1}{16}$

giving

$$f(x) = \frac{x^3}{4}, \quad 0 \leq x \leq 2,$$

$$f(x) = 0, \quad \text{elsewhere.}$$

(b) $E(X) = \int_{-\infty}^{+\infty} xf(x)dx = \int_0^2 \frac{x^3}{4} \cdot x dx = \left[\frac{x^5}{20}\right]_0^2 = \frac{8}{5}$

(c)
$$
\begin{aligned}
Var(X) &= \int_0^2 x^2 \frac{x^3}{4} dx - \left(\frac{8}{5}\right)^2 \\
&= \left[\frac{x^6}{24}\right]_0^2 - \left(\frac{8}{5}\right)^2 = \frac{64}{24} - \frac{64}{25} = \frac{8}{75}
\end{aligned}
$$

3.14 The Continuous Function Distribution

A continuous random variable X whose probability density function (pdf) is given by

$$f(x;a,b) = \frac{1}{b-a} \text{ forf } a \le x \le b, \text{ where } a, b \text{ are finite} \quad (1.24)$$

$$f(x;a,b) = 0 \text{ otherwise,}$$

has a continuous uniform distribution over the interval $[a,b]$. A notation that designates that X has pdf of the form (1.24) is

$$X \sim \text{UNIF}(a,b).$$

We can see that this is a valid density function, since $f(x) \ge 0$, and

$$\int_{-\infty}^{+\infty} f(x)dx = \int_a^b \frac{1}{b-a}dx = 1.$$

Mean and Variance of UNIF(a,b)

The mean of X is obtained as follows:

$$\begin{aligned} E[X] &= \int_a^b x\left(\frac{1}{b-a}\right)dx \\ &= \frac{b^2-a^2}{2(b-a)} \\ &= \frac{(b+a)(b-a)}{2(b-a)} \\ &= \frac{a+b}{2} \end{aligned}$$

Furthermore,

$$\begin{aligned} E(X^2) &= \int_a^b x^2\left(\frac{1}{b-a}\right)dx \\ &= \frac{b^3-a^3}{3(b-a)} \\ &= \frac{(b^2+ab+a^2)(b-a)}{3(b-a)} \\ &= \frac{b^2+ab+a^2}{3} \end{aligned}$$

Thus,

$$\begin{aligned} Var(X) &= E[X^2] - [E(X)]^2 \\ &= \frac{b^2+ab+a^2}{3} - \frac{(a+b)^2}{4} \\ &= \frac{(b-a)^2}{12} \end{aligned}$$

3.15 The Exponential Distribution

A continuous random variable X whose pdf $f()$ is given by

$$f(x) = \lambda e^{-\lambda x} \text{ for } x \geq 0, \text{ where } \lambda \text{ is a positive constant}$$
$$f(x) = 0 \text{ otherwise} \qquad (1.25)$$

has an exponential distribution with parameter λ. Since

$$\begin{aligned} \int_{-\infty}^{+\infty} f(x)dx = \int_0^{\infty} \lambda e^{-\lambda x}dx &= \left[-e^{-\lambda x}\right]_0^{\infty} \\ &= 1, \text{ for } \lambda > 0 \end{aligned}$$

The notation that designate that X has an exponential distribution with parameter λ is $X \sim \text{Exp}(\lambda)$. **Mean and Variance of Exp(λ)**

$$\begin{aligned} E(X) &= \int_0^{\infty} \lambda x e^{-\lambda x}dx \\ &= \left[-xe^{-\lambda x}\right]_0^{\infty} + \int_0^{\infty} e^{-\lambda x}dx, \text{ integrating by parts,} \\ &= \left[\frac{e^{-\lambda x}}{-\lambda}\right]_0^{\infty} = \frac{1}{\lambda}, \text{ for } \lambda > 0 \\ E[X^2] &= \int_0^{\infty} \lambda x^2 e^{-\lambda x}dx \\ &= \left[-x^2e^{-\lambda x}\right]_0^{\infty} + 2\int_0^{\infty} xe^{-\lambda x}dx, \text{ integrating by parts,} \\ &= \frac{2}{\lambda}E(X) \\ &= \frac{2}{\lambda^2} \\ \Rightarrow Var(X) &= \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2} \end{aligned}$$

3.16 Weibull Distribution

A continuous random variable X is said to have the Weibull distribution with parameters $\beta > 0$ and $\theta > 0$ if it has a pdf of the form

$$f(x;\theta,\beta) = \frac{\beta}{\theta^{\beta}} x^{\beta-1} e^{-(x/\theta)^{\beta}},\ x > 0$$

and zero otherwise.
A notation that designates that X has Weibull distribution with parameters β and θ is

$$X \sim WEI(\theta, \beta).$$

It is a distribution that lhas been successful used in reliability theory.

3.17 The Mean and Variance of WEI(θ, β)

$$\begin{aligned} E(X) &= \int_0^{\infty} x \frac{\beta}{\theta^{\beta}} x^{\beta-1} e^{-(x/\theta)^{\beta}} dx \\ &= \frac{\beta}{\theta^{\beta}} \int_0^{\infty} x^{(1+\beta)-1} e^{-(x/\theta)^{\beta}} dx \end{aligned}$$

Following the substitution $t = (x/\theta)^{\beta}$, and some simplification,

$$E(X) = \theta \int_0^{\infty} t^{(1+\frac{1}{\beta})-1} e^{-t} dt = \theta\Gamma\left(1 + \frac{1}{\beta}\right)$$

Similarly,

$$E(X^2) = \theta^2 \Gamma\left(1 + \frac{2}{\beta}\right),$$

and thus

$$Var(X) = \theta^2 \left[\Gamma\left(1 + \frac{2}{\beta}\right) - \Gamma^2\left(1 + \frac{1}{\beta}\right)\right].$$

Chapter 4

Moment-Generating Functions

4.1 Introduction

In this chapter, we shall define the moment-generating function of a random variable. Specifically, we shall consider both discrete and continuous random variables. We shall also show how the mean and variance of a distribution can be determined using the method of moment-generating function.

Definition

Let X be a random variable of the discrete type with p.d.f. $f(x)$. If there is a positive number h such that

$$E[e^{tx}] = \sum_x e^{tx} f(x)$$

exists for $-h < t < h$, the the function of t defined by

$$M(t) = E[e^{tx}]$$

is called the moment-generating function of X.
For any positive integer r,

$$M^{(r)}(t) = \sum_x x^r e^{tx} f(x)$$

Thus, it can be shown that for $-h < t < h$, the derivatives of $M(t)$ of all orders exist at $t = 0$.

$$M'(t) = \sum_x x e^{tx} f(x)$$

$$M''(t) = \sum_x x^2 e^{tx} f(x)$$

Setting $t = 0$, we have

$$\begin{aligned} M'(0) &= \sum_x x f(x) = E[X] \\ M''(t) &= \sum_x x^2 f(x) = E[X^2] \end{aligned}$$

and in general

$$M^{(r)}(0) = \sum_x x^r f(x) = E[X^r]$$

Therefore, if the moment-generating function exists,

$$\mu = M'(0) \text{ and } \sigma^2 = M''(0) - [M'(0)]^2$$

The definition of moment-generating function remains the same for continuous-type random variables except that integrals replace summations. So for a continuous random variable X with p.d.f. $f(x)$, the moment-generating function is

$$M(t) = \int_{-\infty}^{\infty} e^{tx} f(x)dx, \quad -h < t < h.$$

Example 1: Let X be a binomial distribution with p.d.f.

$$f(x) = \begin{cases} \binom{n}{x} p^x(1-p)^{n-x} &, \quad x = 0, 1, 2, \ldots, n \\ 0 &, \text{elsewhere} \end{cases}$$

The moment generating function of X is

$$\begin{aligned} M(t) &= E[e^{tx}] = \sum_x e^{tx} f(x) \\ &= \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x(1-p)^{n-x} \\ &= \sum_{x=0}^{n} \binom{n}{x} (pe^t)^x(1-p)^{n-x} \\ &= [(1-p) + pe^t]^n \end{aligned}$$

4.2 Mean and variance of binomial distribution

$$\begin{aligned} \text{Since } M(t) &= [(1-p) + pe^t]^n \\ M'(t) &= npe^t[(1-p) + pe^t]^{n-1} \end{aligned}$$

and

$$\begin{aligned} E[X] &= M'(0) = np[(1-p)+p]^{n-1} = np \\ M''(t) &= n(n-1)[(1-p)+pe^t]^{n-2}(pe^t)^2 + npe^t[(1-p)+pe^t]^{n-1} \\ M''(0) &= n(n-1)p^2 + np \end{aligned}$$

It follows that

$$\begin{aligned} V(X) &= M''(0) - [M'(0)]^2 \\ &= n(n-1)p^2 + np - (np)^2 \\ &= n^2p^2 - np^2 + np - n^2p^2 \\ &= np - np^2 \\ &= np(1-p) \\ &= npq \text{ where } q = 1-p \end{aligned}$$

Example 2: Let X have a poisson distribution with p.d.f.

$$f(x) = \begin{cases} \dfrac{\lambda^x e^{-\lambda}}{x!} & , x = 0, 1, 2, \ldots \text{ and } \lambda > 0 \\ 0 & , \text{elsewhere} \end{cases}$$

The moment generating function of X is

$$\begin{aligned} M(t) = E[e^{tx}] &= \sum_{x=0} e^{tx} f(x) \\ &= \sum_{x=0}^{\infty} \frac{e^{tx}\lambda^x e^{-\lambda}}{x!} \\ &= e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda e^t)^x}{x!} \\ &= e^{-\lambda} e^{\lambda e^t} \\ &= e^{\lambda(e^t-1)} \text{ for all real values of } t \end{aligned}$$

Mean and variance of Poisson distribution

$$\begin{aligned} \text{Since } M(t) &= e^{\lambda(e^t-1)} \\ M'(t) &= e^{\lambda(e^t-1)}(\lambda e^t) \end{aligned}$$

and

$$M'(0) = \lambda$$

and

$$M''(t) = e^{\lambda(e^t-1)}(\lambda e^t) + e^{\lambda(e^t-1)}(\lambda e^t)^2$$

then

$$M''(0) = \lambda + \lambda^2$$

Therefore

$$\begin{aligned} V(X) &= M''(0) - [M'(0)]^2 \\ &= \lambda + \lambda^2 - (\lambda)^2 \\ &= \lambda \end{aligned}$$

That is, a poisson distribution has

$$\mu = \sigma^2 = \lambda > 0$$

Example 3: Let X have a gamma distribution with p.d.f.

$$f(x) = \begin{cases} \dfrac{x^{\alpha-1}e^{-x/\beta}}{\beta^\alpha \Gamma(\alpha)} & , x > 0; \alpha > 0, \beta > 0 \\ 0 & , \text{otherwise} \end{cases}$$

where α and β are the parameters of the distribution.
The moment-generating function is

$$\begin{aligned} M(t) = E(e^{tx}) &= \int_0^\infty e^{tx} \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} dx \\ &= \int_0^\infty \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{tx - x/\beta} dx \\ &= \int_0^\infty \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{\frac{\beta tx}{\beta} - \frac{x}{\beta}} dx \\ &= \int_0^\infty \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x(1-\beta t)/\beta} dx \end{aligned}$$

If we set

$$y = x(1-\beta t)/\beta, \quad t < \frac{1}{\beta}$$

or

$$x = \beta y/(1-\beta t)$$

We obtain

$$\begin{aligned} M(t) &= \int_0^\infty \frac{\beta/(1-\beta t)}{\Gamma(\alpha)\beta^\alpha}\left(\frac{\beta y}{1-\beta t}\right)^{\alpha-1} e^{-y}dy \\ &= \left(\frac{1}{1-\beta t}\right)^\alpha \int_0^\infty \frac{1}{\Gamma(\alpha)} y^{\alpha-1}e^{-y}dy \\ &= \left(\frac{1}{1-\beta t}\right)^\alpha \frac{1}{\Gamma(\alpha)}\int_0^\infty y^{\alpha-1}e^{-y}dy \\ &= \left(\frac{1}{1-\beta t}\right)^\alpha \frac{1}{\Gamma(\alpha)}\cdot\Gamma(\alpha) \\ &= \left(\frac{1}{1-\beta t}\right)^\alpha \quad \text{provided} \quad t < \frac{1}{\beta}. \end{aligned}$$

4.4 Mean and variance of gamma distribution

$$\text{Since } M(t) = \left(\frac{1}{1-\beta t}\right)^\alpha$$

$$M'(t) = (-\alpha)(1-\beta t)^{-\alpha-1}(-\beta)$$

then

$$M'(0) = \alpha\beta$$

and

$$M''(t) = (-\alpha)(-\alpha-1)(1-\beta t)^{-\alpha-2}(-\beta)^2$$

then

$$M''(0) = \alpha(\alpha+1)\beta^2$$

Therefore,

$$\begin{aligned} V(X) &= M''(0) - [M'(0)]^2 \\ &= \alpha(\alpha+1)\beta^2 - (\alpha\beta)^2 \\ &= \alpha(\alpha+1)\beta^2 - \alpha^2\beta^2 \\ &= \alpha^2\beta^2 + \alpha\beta^2 - \alpha^2\beta^2 \\ &= \alpha\beta^2 \end{aligned}$$

Example 4: Let X have a Normal distribution with p.d.f.

$$f(x) = \begin{cases} \frac{1}{\sigma\sqrt{2\pi}} e^{\{-\frac{(x-\mu)^2}{2\sigma^2}\}} & , -\infty < x < \infty \\ 0 & , \text{ otherwise} \end{cases}$$

We can find the moment-generating function of a normal distribution as follows.

$$\begin{aligned} M(t) &= E(e^{tx}) = \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} dx \\ &= \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{-2\sigma^2 tx + x^2 - 2\mu x + \mu^2}{2\sigma^2}\right\} dx \end{aligned}$$

We complete the square in the exponent. Thus

$$\begin{aligned} M(t) &= \exp\left\{-\frac{1}{2\sigma^2}(\mu^2 - (\mu + \sigma^2 t)^2\right\} \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2}(x - \mu - \sigma^2 t)^2\right\} dx \\ &= \exp\{-\frac{1}{2\sigma^2}(\mu^2 - \mu^2 - 2\mu\sigma^2 t - (\sigma^2)^2 t^2\} dx \end{aligned}$$

because the integrand of the last integral is equal to 1

$$= \exp\{\mu t + \frac{\sigma^2 t^2}{2}\}$$

4.5 The Mean and variance of Normal distribution

Since

$$M(t) = \exp\{\mu t + \frac{\sigma^2 t}{2}\}$$

Now,

$$M'(t) = \exp\{\mu t + \frac{\sigma^2 t^2}{2}\} \times (\mu + \sigma^2 t)$$

so,

$$M'(0) = \mu = \text{mean}$$

and

$$\begin{aligned} M''(t) &= \sigma^2 \exp\{\mu t + \frac{\sigma^2 t^2}{2}\} + \exp\{\mu t + \frac{\sigma^2 t^2}{2}\} \times (\mu + \sigma^2 t^2) \\ M''(0) &= \sigma^2 + \mu^2 \end{aligned}$$

Thus

$$\begin{aligned} \text{Variance} &= M''(0) - (M'(0))^2 \\ &= \sigma^2 + \mu^2 - \mu^2 \\ &= \sigma^2 \end{aligned}$$

Theorem

The moment generating function of a sum of two independent random variables is the product of their moment generating functions.

If X and Y are independent random variables, the moment generating function of the sum $X + Y$

$$\begin{aligned} M_{X+Y}(t) &= E(e^{t(X+Y)}) = E(e^{tX}e^{tY}) \\ &= E(e^{tX})E(e^{tY}) \\ &= M_X(t) \cdot M_Y(t) \end{aligned}$$

Where $M_X(t)$ is the moment generating function of X and $M_Y(t)$ is the moment generating function of Y.

Proof: The joint p.d.f. of X, Y is $f(x, y)$

$$\begin{aligned} \therefore \quad M_{X+Y}(t) &= E[e^{t(X+Y)}] \\ &= \int_Y \int_X e^{t(x+y)} f(x,y)dxdy \\ &= \int_Y \int_Y e^{tx}e^{ty} f(x,y)dxdy \\ &= \int_Y \int_X e^{tx} \cdot e^{ty} f_X(x) \cdot f_Y(y)dxdy \\ & \quad \text{(since } X \text{ and } Y \text{ are independent)} \\ &= \int_X e^{tx} f_X(x)dx \int_Y e^{ty} f_Y(y)dy \\ &= M_X(t) \cdot M_Y(t). \end{aligned}$$

Finite induction extends this theorem to the sum of any finite number of independent random variables: if $X_1, X_2, \ldots, X_n$ are independent, then

$$M_{X_1+X_2+\cdots+X_n}(t) = \prod_{i=1}^{n} M_{X_i}(t).$$

If the random variables have identical distributions, say, with common moment generating function $M_X(t)$, then

$$M_{X_1+X_2+\cdots+X_n}(t) = [M_X(t)]^n.$$

4.6 Moments

Moments has its origin in the study of Mechanics. It is one of the devices for measuring the characteristics of a distribution.

There are three different kinds of moments.

(i) Moment about the origin

(ii) Moment about the mean

(iii) Factorial moments

Each of these shall be treated separately.

(i) Moment about the origin is defined as $U' = E(X^r)$

$= \sum_{i=1}^{\infty} X^r P(X)$ for a discrete random variable

and $\int_{-\infty}^{\infty} X^r f(x) \cdots$ for a continuous random variable.

(ii) The r^{th} moment about the mean is defined as $\mu_r = E(X-\mu)^r$

$\therefore \mu_r = \sum_{i=0}^{\infty}(X-\mu)^r P(X)$ for discrete random variable.

$\mu_r = \int_{-\infty}^{\infty}(X-\mu)^r f(X)$ for continuous random variable.

For a simple series:

r^{th} moment about the origin is given by

$$\boxed{\mu_r = \frac{1}{n}\sum_{i=1}^{n}(X-0)^r}$$

$$\mu_1^1 = \frac{1}{n}\sum(X-0)^1 = \frac{\sum X}{n} = \bar{X} \text{ (mean)}$$

$$\mu_2^1 = \frac{1}{n}\sum_{i=1}^{n}(X-0)^2 = \frac{1}{n}\sum_{i=1}^{n} X^2$$

$$\mu_3^1 = \frac{1}{n}\sum_{i=1}^{n}(X-0)^2 = \frac{1}{n}\sum_{i=1}^{n}X^3$$

$$\mu_r^1 = \frac{1}{n}\sum_{i=1}^{n}(X-0)^3 = \frac{1}{n}\sum_{i=1}^{n}X^r$$

For grouped data $\mu_3^1 = \dfrac{\sum f(X-0)^3}{\sum f}$

The r^{th} moment about the mean is given by

$$\boxed{\mu_r = \frac{1}{n}\sum_{i=1}^{n}(X-\bar{X})^r}$$

Note that the first moment about the mean is zero and the second moment about the mean is the variance.

$$\therefore \quad \mu_1 = \frac{1}{n}\sum_{i=1}^{n}(X-\bar{X})^1 = 0$$

$$\mu_2 = \frac{1}{n}\sum_{i=1}^{n}(X-\bar{X})^2 = \text{(variance)}$$

It is otherwise known as the <u>central moment.</u> However, the central moment can be expressed as a function of the moment about the origin as follows:

$$\mu_2 = \frac{1}{n}\sum X^2 - n\bar{X}^2$$

$$= \frac{1}{n}\sum X^2 - \frac{(\sum X)^2}{n}$$

$$\boxed{\therefore \mu_2 = \mu_2^1 - (\mu_1^1)^2}$$

$$\mu_3 = \frac{1}{n}\sum(X-\bar{X})^3$$

$$= \frac{1}{n}\sum[X^3 - 3X^2\bar{X} + 3X - \bar{X}^2 - \bar{X}^3]$$

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

$$= \frac{1}{n}X^3 + 2\bar{X}^3 + 2\bar{X}^3 - 3\frac{X}{n}\frac{x^2}{n}$$

$$\boxed{\therefore\ \mu_3 = \mu_3^1 + 2\mu_1^2 - 3\mu_1^1\mu_2^1}$$

(iii) **Factorial Moment**
The r^{th} factorial moment is defined as $\mu_{(r)} = E[X(X-1)(X-2)(X-r+1)]$

$$\mu_{(1)} = E(X)$$

The second is factorial moment is $\mu_{(2)} = E[X(X-1)]$

3rd factorial moment $= E[X(X-1)(X-2)]$

4.7 Mean and Variance of a Random Variable Mean of a random variable can be defined as $E(\bar{X})$.

$$\begin{aligned} E(X) &= \sum_{i}^{n} XP(X) \text{ for close variable} \\ &= \int_{-\infty}^{\infty} xf(x)dx \text{ for continuous variable} \end{aligned}$$

The variance of a random variable is defined as

$$\begin{aligned} V(\bar{X}) &= E(X^2) - (E(X))^2 \\ &= \mu_2^1 - (\mu_1^1)^2 \end{aligned}$$

The moment coefficient of skewness is defined as $\beta_1 = \frac{\mu_3}{\sigma^2} = \mu_3/\mu_2$
$\beta_2 = mu_4/\mu_2^2 - 3$
Illustration: The probability distribution function (pdf) of a continuous random variable

$$f(x) = \begin{cases} \frac{2x+1}{6} & 0 < X < 2 \\ 0 & \text{otherwise} \end{cases}$$

Find the 3rd moment about the mean.
Solution:

$$\mu_3 = \int_0^2 (X-\mu)^3 f(x)dx$$

where μ^3 is not given.

$$\boxed{\text{but}\quad \mu_3 = \mu_3^1 + 2(\mu_1^1)^3 - 3\mu_1^1\mu_2^1}$$

$$\mu_1^1 = \int_0^2 xf(x)dx$$
$$= \int_1^2 \frac{1}{6}x(2x+1)dx = \frac{1}{6}\int_1^2 x(2x+1)dx$$

$$\therefore \frac{1}{6}\left[\frac{2x^3}{3}+\frac{x^2}{2}\right]_0^2 = \frac{1}{6}\left[\frac{16}{3}+\frac{4}{2}\right] = \frac{1}{6}[\frac{32+12}{6}]$$
$$= \frac{1}{6}[\frac{44}{6}] = \frac{44}{36} = 1.22$$

And,

$$\mu_2^1 = \int_0^2 (x-0)^2 f(x)dx$$
$$= \frac{1}{6}\int_0^2 x^2(2x+1)dx$$
$$= \frac{1}{6}\left[\frac{2x^4}{4}+\frac{x^3}{3}\right]_0^2 = \frac{1}{6}\left[\frac{32}{4}+\frac{8}{3}\right] = 1.78$$
$$\mu_3^1 = \int_0^2 (x-0)^3 f(x)dx$$
$$= \frac{1}{6}\int_0^2 x^3(2x+1)dx$$
$$= \frac{1}{6}\left[\frac{2x^5}{5}+\frac{x^4}{4}\right]_0^2$$
$$= \frac{1}{6}\left[\frac{64}{5}+\frac{16}{4}\right] = 2.8$$
$$\therefore \mu_3 = \mu_3' + 2\mu_1'^3 - 3\mu_1'\mu_2'$$
$$= 2.8 + 2(1.22)^3 - 3(1.22)(1.78)$$
$$= 2.8 + 2(1.816) - 3(2.172)$$
$$= 2.8 + 3.632 - 6,516 + 6.432 - 6.516$$
$$= -0.084$$

4.8 The Mode and Median of a Distribution

Just like it is possible to find the mean of a probability distribution, it is also possible to obtain the median and mode of a distribution.
Given probability mass function (pmf), the median is defined as the value

of M such that

$$\sum_{x=\infty}^{m} P(X) \geq \frac{1}{2} \text{ or } \sum_{x=m}^{\infty} P_X(x) \geq \frac{1}{2}$$

Mode:

For a discrete distribution the mode is the value of X for which the probability is maximum. For a continuous function, the mode can be obtained by method of calculus.

Illustration 1: The p.m.f. of a random variable X is given by

$$P_X(x) = \begin{cases} \frac{2}{8} & x = -1 \\ \frac{3}{8} & x = 0 \\ \frac{2}{8} & x = 1 \\ \frac{1}{8} & x = 2 \end{cases}$$

Find the (i) mode and (ii) median.

Solution:

(i) The mode of the above distribution is $X = 0$, since $P_X(x)$ is maximum at this point.

(ii) Median $= \sum_{x=\infty}^{m} P_X(x) \geq 0.5 = 0$

Illustration 2: Given the p.m.f. for a random variable Y

$$P_Y(y) = \begin{cases} \frac{2}{9}; & y = -1 \\ \frac{2}{9}; & y = 0 \\ \frac{1}{9}; & y = 1 \\ \frac{2}{9}; & y = 2 \\ \frac{2}{9}; & y = 3 \end{cases}$$

Find the (i) mode and (ii) median.

Solution: (i) The above is a Multimodal distribution.

(ii) Median: $\sum_{y=-\infty}^{m} P_Y(y) \geq 0.5 = 1$ i.e. $y = 1$

The mode and median for a continuous random variable with the pdf. $f(x)$ is

$$\int_{-\infty} f(x)dx \leq \frac{1}{2} \text{ and } \int_{m}^{\infty} f(x)dx \geq \frac{1}{2}$$

Illustration 3: Given a random variable X with pdf.

$$f(x) = \begin{cases} Xe^{-x} & 0 < \infty < \infty \\ 0 & \text{elsewhere} \end{cases}$$

Find (i) the mode, (ii) the median (iii) 1st Quartile (iv) 3rd Quartile and (v) Semi-interquartile range.

Solution:

(i) Mode

$$f(x) = Xe^{-x}$$
$$f'(x) = \frac{d}{dx}(Xe^{-x})$$

$$f'(x) = -xe^x + e^x = 0$$
$$\Rightarrow e^{-x} = Xe^{-x}$$
$$1 = X \text{ mode}$$

(ii) $\int_0^m f(x)dx \geq \frac{1}{2}$

$$\int_0^m Xe^{-x}dx = \frac{1}{2}$$
$$\Rightarrow -Xe^{-x} - e^{-x} \times 1\, dx$$
$$\Rightarrow \left[-Xe^{-x} + e^{-x}\right]_-^m = 0.5$$

Substitute for m and 0

$$\left[-Xe^{-x} - e^{-x}\right]_0^m = 0.5$$
$$-me^{-m} - e^{-m} - (0.1) = 0.5$$
$$-me^{-m} - e^{-m} + 1 = 0.5$$
$$-(m+1)e^{-m} + 1 = 0.5$$
$$-(m+1)e^{-m} = -0.5$$
$$e^{-m}(m+1) = +0.5$$

Solve for m

$$\log_e e^{-m} = \log_e 0.5$$
$$-m = \log_e 0.5$$
$$m = -\log_e 0.5 = \underline{0.693}$$

$$f(x) = \begin{cases} 3e^{-3} & 0 < x < \infty \\ 0 & \text{otherwise} \end{cases}$$

Find the median Q_1 and Q_3 and hence the interquartile range.

Solution:

Median: $\int_0^m f(x)dx = 0.5$

$$\int_0^m 3e^{-3x}dx = 0.5$$
$$\therefore -e^{-3x}\Big|_0^m = 0.5$$

$\Rightarrow \; -e^{-3m} - [-1] = 0.5$
$1 - e^{3m} = 0.5$
$-e^{-3m} = -0.5 \quad \therefore \; e^{-3m} = 0.5$
Solve m

$$\begin{aligned} \log_e e^{-3m} &= \log_e 0.5 \\ 0 - 3m &= \log_e 0.5 \\ m &= -\frac{1}{3}\log_e 0.5 = 0.23 \end{aligned}$$

$$\begin{aligned} Q_1 &= \int_0^{q_1} f(x)dx = 0.25 \\ &= 1 - e^{-3q_1} = 0.25 \\ & \quad e^{-3q_1} = 0.75 \\ & \quad q_1 = -\frac{1}{3}\log_e 0.75 \\ & \quad q_1 = 0.10 \\ Q_3 &= \int_0^{q_3} f(x)dx = 0.75 \\ &= 1 - e^{-3q_3} = 0.75 \\ & \quad e^{-3q_3} = 0.25 \\ & \quad q_3 = -\frac{1}{3}\log_e 0.25 = 0.46 \end{aligned}$$

$$\text{Semi-interquatile Range} = \frac{Q_3 - Q_1}{2} = \frac{0.46 - 0.10}{2} = 0.12$$

$$\text{Quartile Deviation} = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

4.9 Limit Theorems

The important results in probability theory are those that involve the limit theorems. Some of these are:

(i) Chebyshev's inequality

(ii) Central limit theorem

(iii) Law of large numbers.

Before we prove the Chebyshev's inequality let us consider the Markov's inequality.

Markov's inequality: If X is a continuous random variable that takes only non-negative values then for any value $a > 0$

$$P(X \geq a) \leq \frac{E(X)}{a}$$

Proof:
If X is continuous with $f(x)$ as the density function

$$\begin{aligned} E(X) &= \int_0^\infty xf(x)dx \\ &= \int_0^a xf(x)dx + \int_a^\infty xf(x)dx \\ &\geq \int_a^\infty xf(x)dx \\ &\geq \int_a^\infty af(x)dx \\ &= a\int_a^\infty f(x)dx \\ &= aP[X \geq a] \Longrightarrow P(X \geq a) = \frac{E(X)}{a} \end{aligned}$$

Chebyshev's inequality
If X is a random variable with finite mean μ and variance σ^2, then for any value $k > 0$

$$P\{|X - \mu| \geq k\} \leq \frac{\sigma^2}{k^2}.$$

Proof: Since $(X - \mu)^2$ is a non-negative random variable, we can apply Markov's inequality (with $a = k^2$) to obtain

$$P\{(X - \mu)^2 \geq k^2\} \leq \frac{E(X - \mu)^2}{k^2} \qquad (*)$$

but since $E(X - \mu)^2 = \sigma^2$ and

$$(X - \mu)^2 \geq k^2 \text{ iff } |X - \mu| > k$$

then $\star$ is equivalent to

$$P\{|X - \mu| \geq K\} \leq \frac{E(X - \mu)^2}{k^2} = \frac{\sigma^2}{k^2}$$

$$\therefore \quad P\{|X-\mu| \geq k\} \leq \frac{\sigma^2}{k^2}$$

and equivalently

$$P\{|X-\mu| \leq\} \geq 1-\frac{\sigma^2}{k^2}$$

The above theorems enables us to derive bounds on probabilities when only the mean or both the mean and variance of the probability distribution are known.

The central limit theorems: is one of the most remarkable results in probability theory. It states that

"Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variable each having mean μ and variance σ^2. Then the distribution of

$$\frac{X_1+X_2+\cdots+X_n-n\mu}{\sigma/\sqrt{n}}$$

tends to the standard norma as $n \to \infty$ i.e.

$$P\left\{\frac{X_1+X_2+\cdots+X_n-n\mu}{\sigma/\sqrt{n}} \leq a\right\} \to \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-x^2/2}dx$$

as $n \to \infty$

i.e. the distribution of nearly all variables tends to be normally distributed (i.e. $X \sim N(0,1)$] as the sample size becomes large.

Examples (Chebyshev's inequality)

If it is known that bags of pure water produced in a factory during a month is a random variable with mean 50 litres.

(i) What can be said about the probability that this month production will exceed 80 litres?

(ii) If the variance of the month's production is 36, what can be said about the probability that this month's production will be between 35 and 65 litres?

Solution:

If X represent the number of items in a month by Markov's theorem

$$P(X>80) \leq \frac{E(X)}{80}=\frac{50}{80}=\frac{5}{8}=0.625$$

by Chebvshev's theorem.

$$P\{|X-50|\geq 15\}\leq \frac{\sigma^2}{225}=\frac{36}{225}=0.16$$

$$\Longrightarrow P\{|X-50|<15\}\geq 1-0.16=0.84$$

4.10 Chebyshev's inequality

$$\begin{aligned}\sigma^2 &= E(X-\mu)^2=\sum_{x\in\mathbb{R}}(X-\mu)^2 f(x)\\ &= \sum_{x\in A}(X-\mu)^2 f(x)+\sum_{x\in A'}(X-\mu)^2 f(x)\ \forall A\in\mathbb{R}\end{aligned}$$

and

$$A=x:(X-\mu)\geq k\sigma\}$$

but

$$\sum_{x\in A'}(X-\mu)^2 f(x)\geq 0$$

$$\Longrightarrow \sum_{x\in A}(X-\mu)^2 f(x)\geq \sum_{x\in A'}(X-\mu)^2 f(x)$$

$$\therefore\ \sigma^2\geq \sum_{x\in A}(X-\mu)^2 f(x)$$

However in A, $|X-\mu|\geq k\sigma$.

$$\sum_{x\in A}|X-\mu|^2 f(x)\geq \sum_{x\in A}k^2\sigma^2 f(x)\geq k^2\sigma^2\sum_{x\in A}f(x)$$

$$\therefore\ \sigma^2\geq k^2\sigma^2 P\{|X-\mu|\geq k\sigma\}$$

$$\Longrightarrow k^2\sigma^2 P\{|X-\mu|\geq k\sigma\}\leq \sigma^2$$

$$P|X-\mu|\geq k\sigma^2\}\leq \frac{\sigma^2}{k^2\sigma^2}$$

$$\therefore\ P\{|X-\mu|\geq k\sigma\}\leq \frac{1}{k^2}$$

Exercise:

(1) Express the forth moment about the mean as a function of the moment about the origin.

(2) For a continuous random variable X

$$f(x) = \begin{cases} \frac{2}{3} & 0 < x < 4 \\ 0 & \end{cases}$$

(i) Find the expectation and variance of X

(ii) Find the 4th moment about the mean.

(3) The table below is the Probability distribution of random variable table

t	a	b
$f(t)$	$\frac{2}{3}$	$\frac{1}{3}$

Given that $a = 3b$ and $a + b = 8$, find (i) the value of a and b, hence or otherwise find (ii) mean and variance of t. (iii) The 3rd moment about the mean.

Chapter 5

Sampling Theory

5.1 Sampling Theory

Sampling theory deals with the study of relationships between a population and samples drawn from the population. It is important in most applications. For instance, it is used in the estimation of population parameters from a knowledge of the corresponding sample statistics.

In statistics, sample data are observed in order to make inferences or decisions concerning the populations from which the samples are drawn. Hence, statistical inference deals with the question of how inferences can be made about population characteristics from information contained in sample. However, if the statistician knew the population values, there would be no need to make inferences about them! A value such as $\bar{X}$, that is, a number computed from sample data, is referred to as statistic. Thus, the sample mean $\bar{X}$ is a statistic which may be considered as an estimate of the population mean from which the sample was drawn. A statistic may be used as an estimate of an analogous population measure, known as a parameter.

5.1.1 Populations and Samples

Population is the term used to describe a large number set or collection of items that have something in common. That is, universe or population consists of the total collection of items or elements that fall within the scope of a statistical investigation. The purpose of defining a statistical population is to provide very explicit limits for the data collection process and for the inferences and conclusions that may be drawn from the study.

The items or elements which comprise the populationmay be individuals, animals, families, employees, schools, etc.

A sample on the other hand, is a subset of the population, selected in such a way that it is representative of the larger population. The terms population and sample are relative. An aggregate of elements which constitutes a population for one purpose may merely be a sample for another.

5.1.2 Fundamentals of Sampling

Purpose

There are a wide variety of reasons why sampling is important. In most situations, a study of an entire population is impossible; hence sampling may represent the only possible or practicable method to obtain the desired information. For example, in the case of processes, such as manufacturing, where the universe is conceptually infinite including all future as well as current production, it is not possible to accomplish a complete enumeration of the population. Also, in destructive sampling of a finite population, it is possible to effect a complete enumeration of the population but it would not be practical to do so.

Sampling procedures are often employed for overall effectiveess if properly selected. The results from samples are often more accurate than results based on a population. A study of a sample is also less expensive than a study of an entire population, because a smaller number of items or subjects are examined. Samples can be studied more quickly than populations.

5.2 Random and Non-random Selection

Samples can be drawn or selected from statistical universes in a variety of ways. It is therefore, important to distinguish random from non-random methods of selection. In this book, attention is focussed on random sampling or probability sampling, that is, sampling in which the probability of inclusion of every element in the population is known. Non-random sampling methods are referred to as "judgement sampling", that is, selection methods in which judgement is exercised in deciding which elements of a universe to include in the sample.

The basic reason random sampling is preferable to non-random sampling is that in judgement selection, there is no objective method of measuring the precision or reliability of estimates made from the sample. On the other hand, in random sampling, the precision with which estimates of population values can be made is obtained from the sample itself.

5.3 Simple Random Sampling

A random sample or probability sample is a sample drawn such that every element in the population has a known, equal probability of inclusion.

Suppose we have a finite population of N elements, a simple random sample of N elements is a sample selected in a way that every combination of n elements has an equal chance of being included. If the sampling is drawn without replacement, then every element in the population has an equal probability of $\frac{1}{N}$ of being selected on the first draw; each of the remaining $N-1$ elements has an equal probability of $\frac{1}{N-1}$ of being selected on the second draw and so on until the last sample item is drawn. There are $\binom{N}{n}$ possible ways in which the samples of n items can be drawn, then the probability that any sample of size n will be drawn is $\frac{1}{\binom{N}{n}}$.

For example, let a population consist of five letters, A, B, C, D and E. Here $N = 5$, suppose a simple random sample of size $n = 2$ is drawn. The possible number of sample is

$$\binom{5}{2} = \frac{5!}{2!3!} = 10$$

these ten possible samples are

$(A, B), (A, C), (A, D), (A, E), (B, C), (B, D), (B, E), (C, D), (C, E)$ and (D, E).

The probability that any one of these ten samples will be chosen is $\frac{1}{10}$.

For sampling with replacement, since it is possible for the same item to appear more than once in the sample, it is not always used for practical purposes, hence we shall not discuss it here.

A simple random sample may be drawn by the method of "Drawing slips from a Bowl" where the population is usually finite and the elements or items easily identified and numbered.

Another method is the use of "Tables of Random Numbers". This method is preferred especially when the poplation is very large, the preceding method becomes quite unwieldy and time consuming. Bias may also be introduced if the slips are not properly mixed.

5.4 Stratified Random Sampling

In stratified random sampling, the population is first divided into mutually exclusive subgroups called strata and probability samples are then drawn independently from each stratum. Samples may be drawn from each stratum by simple random sampling or cluster sampling or systematic sampling. The objective may be to combine sample statistics from each stratum to obtain an overall estimate of a population parameter or to make comparison between strata to investigate strata differences.

There is an increase in precision i.e., a reduction in sampling error when this method is employed. This reduction in sampling error is achieved by minimizing differences among elements within strata and maximizing differences among strata. This method is most effective when the elements within each stratum are homogeneous (elements within each stratum are alike with respect to the characteristics being studied) and when difference of elements among strata are heterogeneous.

5.5 Cluster Sampling

In cluster sampling the population is divided into groups called clusters and then a probability sample of the clusters is drawn and observed. Cluster sampling is used to achieve a reduction in cost of sample design. Sampling error is also reduced in clustering especially when the elements within each clusters are heterogeneous and the elements among cluster are homogeneous. The advantages of cluster sampling from the point of view of cost arise from the fact that collection of nearby units is easier, faster, cheaper, and more convenient than observing units scattered over a region.

5.6 Systematic Sampling

Systematic sampling is often used in place of simple random sampling as a procedure of obtaining random selection in most practical applications. In a systematic sampling, every *kth* element is taken from the population arranged in some specified order. The first sample which is the starting point is drawn randomly from the first k elements. The value k is obtained by dividing the number of items in the sampling frame (a list of all elements in the population) by the sample size. The results of systematic sampling are similar to those of simple random sampling if the elements in the population occur in random order. A population can be assumed to be random when for instance, items contained in this population are thoroughly mixed. However, the results of this method is not reliable when there is a cyclical variation in the way the elements of the population are arranged.

Chapter 6

Introduction to Statistical Inference

6.1 Introduction

We have been exploring some basic principles and applications of probability in the preceding chapters. A few of the probability distributions that have been found by experience to be particularly useful in solving certain classes of problems have also been carefully examined. We have seen that probability theory is a useful and coherent framework for dealing with the problems of uncertainty. We now introduce one of the key areas to which probability is applied-statistical inference. This field called statistical inference, which is the subject matter of this chapter and the next chapter, uses the theory of probability for making reasonable decisions concerning a population on the basis of the samples drawn from it.

Statistical inference deals with two different classes of problems: (1) Estimation, which is discussed in this chapter and (2) hypothesis testing, which is examined in the next chapter.
In both cases, the problem is structured in such a way that inferences about relevant population values can be made from sample data.

6.2 Estimation

The subject of estimation is concerned with the methods by which population characteristics are estimated from sample information. The objectives are

(i) to present properties for judging how well a given sample statistic estimates the parent population parameter,

(ii) to present several methods for estimating these parameters.

Very often we know or are willing to assume that a random variable X follows a particular probability distribution but we do not know the value(s) of the parameter(s) of the distribution. For example, if X is assumed to follow a normal distribution we may be interested in obtaining the values of its two parameters, namely, the mean and the variance which are unknown. In this problem of parameter estimation, the usual procedure is to assume that we have available a random sample of size n of a random variable X, whose probability distribution is assumed known, and use the sample data to estimate the unknown parameters. Let us take a closer look at this problem of estimation. This estimation problem can be broken into two major categories: point estimation and interval estimation.

6.2.1 Point Estimation

The basic reasons for the need to estimate population parameters from sample information as earlier noted in the previous chapters are that it is ordinarily too expensive or simply infeasible to enumerate complete populations to obtain the required information. The cost of complete censuses may be prohibitive in finite populations while complete enumeration are impossble in the case of infinite populations. Hence estimation procedures are useful in providing the means of obtaining estimates of population parameters with desired degrees of precision.

A point estimation is a single number which is used as an estimate of the unknown population parameter. For instance, let X be a random variable with p.d.f. $f(x, \theta)$, where θ is the parameter of the distribution which is unknown. Suppose also that a random sample of size n is drawn from this distribution, then a function of the sample values such that

$$\hat{\theta} = f(x_1, x_2, \cdots, x_n)$$

provides an estimate of the true θ. $\hat{\theta}$ is called a statistic or an estimator (a function or rule that is used to guess the value of a parameter) and a particular numerical value taken by the estimator is known as an estimate (that is, a particular value calculated from a particular sample of observations).

Note that $\hat{\theta}$ can be treated as a random variable because it is a function of the sample data. $\hat{\theta}$ provides us with a rule or formular, that tells us how to estimate the true θ. Thus, if

$$\hat{\theta} = \frac{1}{n}(x_1 + x_2 + \cdots + x_n) = \bar{X}$$

then $\bar{X}$ (the sample mean) is an estimator of the true mean value, μ. So if $\bar{X} = 25$, then this provides an estimate of μ. The estimator $\hat{\theta}$ obtained is a point estimator because it provides only a single estimate of θ.

6.2.2 Interval Estimation

For most practical purposes, it would not suffice to have merely a single value estimate of a population parameter. Any single point estimate will be either right or wrong. Therefore, instead of obtaining only a single estimate of θ, it would certainly seem to be extremely useful, and perhaps necessary, to obtain two estimates of θ by constructing two estimators $\hat{\theta}_1(x_1, x_2, \cdots, x_n)$ and $\hat{\theta}_2(x_1, x_2, \cdots, x_n)$, and say with some confidence that the interval between $\hat{\theta}_1$ and $\hat{\theta}_2$ includes the true θ. Thus, an interval estimates of a population parameter is a statement of two values between which it is estimated that the parameter lies.

We shall be discussing the construction of confidence intervals as a means of interval estimation in our subsequent discussion.

6.3 Properties of Estimators

Sometimes we are faced with questions such as, how good are some estimates? what makes them good? can we say anything about the closeness of a particular estimate to an unknown parameter? Suppose the arithmetic mean, $\bar{x}$, the median, $\tilde{x}$, and the mid range, r, are calculated from a random sample drawn from a given population. Which method would be the best estimator for obtaining the population mean? Your answer probably would be the sample mean, $\bar{x}$. Why do you think the sample mean represents the best estimator? These and many more questions will be answered in this section.

6.3.1 Unbiasedness: If $\hat{\theta}$ is to be a good estimator of θ, a very desirable property is that its mean be equal to θ, that is, $E(\hat{\theta}) = \theta$.

Definition: Let $X_1, X_2, \ldots, X_n$ be identically independently distributed random variables with p.d.f. $f(x;\theta)$ and $\hat{\theta} = (X_1, X_2, \ldots, X_n)$ be a statistic. Then we shall say that $\hat{\theta}$ is unbiased for θ is $E(\hat{\theta}) = \theta$.

An estimator Γ of an unknown parameter θ is unbiased if $E(\hat{\theta}) = \theta$, for all values of θ. The difference

$$B[\hat{\theta}] = E[\hat{\theta} - \theta] \text{ is called the bias of } \theta$$

If $B[\hat{\theta}] \neq 0$ then $\hat{\theta}$ is said to be a biased estimator. Thus, an unbiased estimator is a random variable whose expected value is the parameter being estimated.

Example 6.1: Let $X_1, X_2, \ldots, X_n$ be a random sample from a nromal distribution $N(\mu, \sigma^2)$. Show that $\bar{X}$ is an unbiased estimator of μ.

Proof: $\bar{X} = \frac{1}{n}\sum X_i$

$$\begin{aligned} E(\bar{X}) &= \frac{1}{n}\sum_{1}^{n} E(X_i) \\ &= \frac{1}{n}\sum_{1}^{n} \mu \\ &= \frac{1}{n} \cdot n\mu = \mu \end{aligned}$$

Example 6.2: Show that $S^2 = \frac{1}{n}\sum_{1}^{n}(X_i - \bar{X})^2$ is not an unbiased estimator of the population variance σ^2.

Proof:

$$\begin{aligned} E(S^2) &= E\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2\right] \\ &= E\left\{\frac{1}{n}\sum_{1}^{n}[(X_i - \mu) - (\bar{X} - \mu)]^2\right\} \\ &= E\left[\frac{1}{n}\sum_{1}^{n}(X_i - \mu)^2 - (\bar{X} - \mu)^2\right] \\ &= \frac{1}{n}\sum_{1}^{n}E(X_i - \mu)^2 - E(\bar{X} - \mu)^2 \\ &= \frac{1}{n}\sum[V(X) - V(\bar{X})] \end{aligned}$$

$$= \frac{1}{n}\sum[\sigma^2 - \sigma^2/n]$$
$$= \sigma^2 - \sigma^2/n$$
$$= \frac{(n-1)\sigma^2}{n}$$

This shows that S^2 is not an unbiased estimator of σ^2. However, an unbiased estimator can be constructed by multiplying the sample variable, S^2, by the factor $n/(n-1)$. Thus

$$E\left[\left(\frac{n}{n-1}\right)S^2\right] = \frac{n}{n-1}E(S^2) = \sigma^2$$

since

$$\frac{n}{n-1}S^2 = \frac{\sum_1^n (X_i - \bar{X})^2}{n-1}$$

Note that it is possible for an estimator not to satisfy some of the desirable statistical properties especially in small sample cases as above. But as the sample size increases indefinitely, the estimator possesses several desirable properties. These are large-sample or asymptotic properties.

An estimator $\hat{\theta}$ is said to be an asymptotically unbiased estimator of θ if

$$\lim_{n\to\infty} E(\hat{\theta}_n) = \theta$$

where $\hat{\theta}_n$ impies that the estimator is based on a sample size n. That is, $\hat{\theta}$ is an asymptotically unbiased estimator of θ if its expected approaches the true value as the sample size increases. For example,

$$E(S^2) = \frac{(n-1)\sigma^2}{n} = \sigma^2\left(1 - \frac{1}{n}\right)$$

then

$$\lim_{n\to\infty} E(S^2) = \lim_{n\to\infty} \sigma(1 - \frac{1}{n}) = \sigma^2$$

6.3.2 Efficiency: This concept refers to the sampling variability of an estimator. The most efficient estimator among a group of unbiased estimators is the one with the smallest variance. Thus, if $\hat{\theta}_1$ and $\hat{\theta}_2$ are two unbiased estimators of θ, and the variance of $\hat{\theta}_1$ is smaller than the variance of $\hat{\theta}_2$, then $\hat{\theta}_1$ is an efficient estimator.

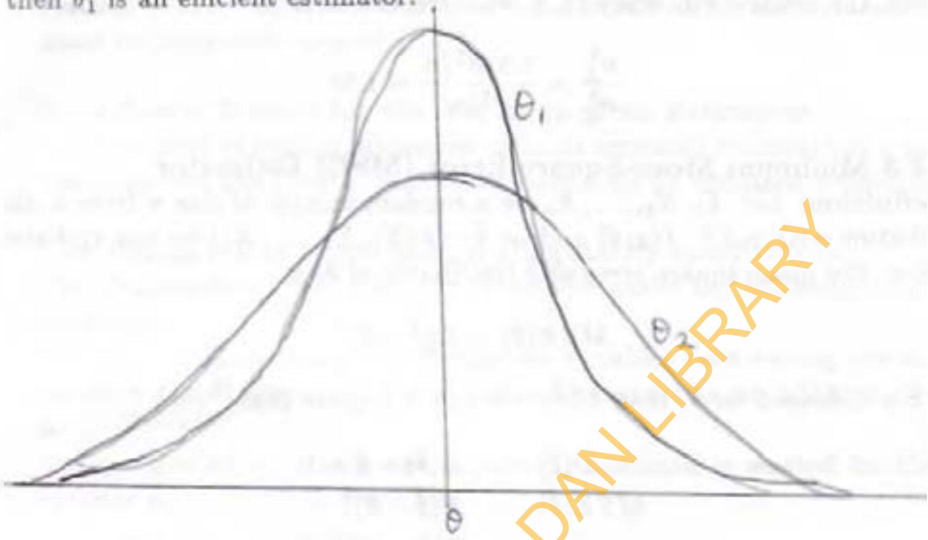


Fig 6.1

Of the two unbiased estimators $\hat{\theta}_1$ and $\hat{\theta}_2$ in fig 3.1, $\hat{\theta}_1$ is best unbiased or efficient. Their relative efficiency is measured by the ratio

$$\frac{V(\hat{\theta}_2)}{V(\hat{\theta}_1)} = \frac{\sigma^2_{\hat{\theta}_2}}{\sigma^2_{\hat{\theta}_1}}$$

where $\sigma^2_{\hat{\theta}_1}$ is the smaller variance.

Example 6.3: A simple random sample of size n is drawn from a normal population with mean μ and variance σ^2. If the sample mean $\bar{X}$ and the sample median $\tilde{X}$ are two estimators of the population mean μ. Obtain the relative efficiency.

Solution

the variance of the sample mean, $\bar{X}$, is $\sigma^2_{\bar{X}} = \dfrac{\sigma^2}{n}$

the variance of the sample median, $\tilde{X}$, is $\sigma^2_{\tilde{X}} = 1.57\dfrac{\sigma^2}{n}$

Thus, the relative efficiency of $\bar{X}$ with respect to $\tilde{X}$ is

$$\frac{\sigma^2_{\tilde{X}}}{\sigma^2_{\bar{X}}} = \frac{1.57\sigma^2/n}{\sigma^2/n} = 1.57$$

6.3.3 Minimum Mean-Square-Error (MSE) Estimator

Definition: Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from a distribution with p.d.f $f(x;\theta)$ and let $\hat{\theta} = h(X_1, X_2, \ldots, X_n)$ be any statistic. Then, the mean square error of $\hat{\theta}$ (estimator of θ) is

$$MSE(\hat{\theta}) = E(\hat{\theta} - \theta)^2$$

If $\hat{\theta}$ is unbiased for θ, then $E(\hat{\theta}) = \theta$ and it implies that

$$\begin{aligned} B(\hat{\theta}) &= E(\hat{\theta}) - \theta = 0 \\ MSE(\hat{\theta}) &= E(\hat{\theta} - \theta)^2 \\ &= E[\hat{\theta} - E(\hat{\theta})]^2 \\ &= Var(\hat{\theta}). \end{aligned}$$

The difference is that, $Var(\hat{\theta})$ measures the dispersion of the distribution of $\hat{\theta}$ around its mean, whereas $MSE(\hat{\theta})$ measures dispersion around the true value of the parameter. This relationships shows the following;

$$\begin{aligned} MSE(\hat{\theta}) &= E(\hat{\theta} - \theta)^2 \\ &= E[\hat{\theta} - E(\hat{\theta}) + E(\hat{\theta}) - \theta]^2 \\ &= E[\hat{\theta} - E(\hat{\theta})] + E[E(\hat{\theta}) - \theta]^2 + 2E[\hat{\theta} - E(\hat{\theta})][E(\hat{\theta}) - \theta] \end{aligned}$$

since the last term is zero.

$$\begin{aligned} MSE(\hat{\theta}) &= E[\hat{\theta} - E(\hat{\theta})]^2 + E[E(\hat{\theta}) - \theta]^2 \\ &= var(\hat{\theta}) + B^2(\hat{\theta}) \\ &= \text{variance of } \hat{\theta} \text{ plus square bias} \end{aligned}$$

If the bias is zero, $MSE(\hat{\theta}) = var(\hat{\theta})$.

Definition: The statistic $\hat{\theta} = h(X_1, X_2, \ldots, X_n)$ that minimizes $E[(\hat{\theta} - \theta)^2]$ is the one with minimum mean square error. If our attention is restr... to unbiased estimators only, then $var(\hat{\theta}) = E(\hat{\theta} - \theta)^2$, and the unbiased statistic $\hat{\theta}$ that minimizes this expression is said to be the unbiased minimum variance estimator of θ.

6.3.4 Lower Bound for the Variance of an Estimator

A method of finding minimum-variance unbiased estimator of a parameter is to find the Cramer-Rao lower bound for an unbiased estimator.

The Cramer-Rao lower bound (Regularity conditions).
The Cramer-Rao lower bound is obtainable under the following regularity conditions.
Let $X_1, \ldots, X_n$ be independent random variables each having the density function $f(x;\theta)$ absolutely continuous with respect to a σ-additive measure μ.
An estimate $\sigma(x_1, \ldots, x_n)$ not necessarily unbiased is wanted for the parameter θ.

(i) $\frac{\partial}{\partial\theta} \log f(x;\theta)$ exists for all θ.

(ii) $$\frac{\partial}{\partial\theta}\int \cdots \int f(x_1;\theta), \cdots f(x_n;\theta)dx_1 \cdots dx_n = \int \cdots \int \frac{\partial}{\partial\theta} f(x_1;\theta) \cdots f(x_n;\theta)dx_1, \cdots, dx_n$$

(iii) $$\frac{\partial}{\partial\theta}\int \cdots \int \sigma(x, \ldots, x_n) f(x_1;\theta) \cdots f(x_n;\theta)dx_1 \cdots dx_n = \int \cdots \int \sigma(x_1, \ldots, x_n)\frac{\partial}{\partial\theta} f(x_1;\theta) \cdots f(x_n;\theta)dx_1, \cdots, dx_n$$

(iv) $$0 < E\left(\left[\frac{\partial}{\partial\theta}\log f(X;\theta)\right]^2\right) < \infty$$

where $\sigma(x_1, \ldots, x_n)$ is a statistic satisfying

$$E\sigma(X_1, \ldots, X_n) = \theta + b(\theta)$$

where $b(\theta)$ is the expected bias.

Theorem: Cramer-Rao inequality

If the regularity conditions (i) to (iv) above are satisfied, then

$$Var(\sigma(X_1,\cdots,X_n)) \geq \frac{[1+b'(\theta)]^2}{nE\left[\frac{\partial}{\partial\theta}\log f(x;\theta)\right]^2}$$

Proof: Let

$$V_i = \frac{\partial}{\partial\theta}\log f(x;\theta)$$

and

$$V = \sum_{i=1}^{n} V_i = \sum \frac{\partial}{\partial\theta}\log f(x_i;\theta)$$

Then

$$\begin{aligned} E(V_i) &= \int \frac{\partial}{\partial\theta}\log f(x_i;\theta)\cdot f(x_i;\theta)dx_i \\ &= \int \frac{1}{f(x_i;\theta)}\frac{\partial}{\partial\theta}f(x_i;\theta)\cdot f(x_i;\theta)dx_i \\ &= \int \frac{\partial}{\partial\theta}f(x_i;\theta)dx_i \\ &= \frac{\partial}{\partial\theta}\int f(x_i;\theta)dx_i = 0 \end{aligned}$$

so that $E(V) = 0$.
Consequently,

$$Var(V) = nE(V^2) = nE(\frac{\partial}{\partial\theta}f(x;\theta))^2$$

note that, $E\sigma(x_1,\ldots,x_n) = \theta + b(\theta)$ that is

$$\int\cdots\int \sigma(x_1,\cdots,x_n)\prod_{i=1}^{n} f(x_i;\theta)dx_i = \theta + b(\theta)$$

Differentiating with respect to θ, we get

$$\frac{\partial}{\partial\theta}\int\cdots\int \sigma(x_1,\cdots,x_n)\prod_{i=1}^{n} f(x_i;\theta)dx_i = 1 + b'(\theta)$$

$$\int\cdots\int \sigma(x_1,\cdots,x_n)\frac{\partial}{\partial\theta}\prod_{i=1}^{n} f(x_i;\theta)dx_i = 1 + b'(\theta)$$

$$\int \cdots \int \sigma \left(\sum \frac{\partial}{\partial \theta} \frac{1}{f(x_i;\theta)} \frac{\partial}{\partial \theta} f(x_i;\theta) \right) \prod_{i=1}^{n} f(x_i;\theta) dx_i = 1 + b'(\theta)$$

$$\int \cdots \int \sigma \left(\sum \frac{\partial}{\partial \theta} \log f(x_i;\theta) \right) \prod_{i=1}^{n} f(x_i;\theta) dx_i = 1 + b'(\theta)$$

$$\int \cdots \int \sigma (\sum V_i) \prod_{i=1}^{n} f(x_i;\theta) dx_i = 1 + b'(\theta)$$

$$E(\sigma V) = 1 + b'(\theta).$$

Note that $\rho_{\sigma v}^2 \leq 1$, where $\rho_{\sigma v}$ is the correlation coefficient of σ and v.
Thus

$$\frac{[cov\sigma; v)^2}{var(\sigma)var(v)} \leq 1$$

i.e. $[E(\sigma v) - E(\sigma)E(v)]^2 \leq var(\sigma) \cdot var(v)$
i.e. $[1 + b'(\theta)]^2 \leq var(\sigma) \cdot var(v)$
since $E(v) = 0$ nd $E(\sigma v) = 1 + b'(\theta)$
i.e. $Var(\sigma) \geq \dfrac{[1 + b'(\theta)]^2}{nE\left(\frac{\partial}{\partial \theta} \log f(x_i;\theta)\right)^2}$

Remark: It can be shown that

$$E\left(\frac{\partial}{\partial \theta} f(x_i;\theta)\right)^2 = -E\left(\frac{\partial^2}{\partial \theta^2} f(x_i;\theta)\right)$$

which gives a Cramer-Rao lower bound as

$$\frac{(1 + b'(\theta))^2}{-nE\left(\frac{\partial^2}{\partial \theta^2} f(x_i;\theta)\right)}$$

Example 1: Let $X_1, \ldots, X_n$ be a random sample from $f(x;\theta) = \dfrac{e^{-\theta}\theta^x}{x!}$ for $x = 0, 1, 2, \ldots$
Obtain the Cramer-Rao lower bound for the unbiased estimator of θ.

Solution: $f(x;\theta) = \dfrac{e^{-\theta}\theta^x}{x!}, \quad x = 0, 1, 2, \ldots$

$$\log f(x;\theta) = -\theta + x \ln \theta - \ln x!$$

$$\frac{\partial}{\partial\theta}\log f(x;\theta) = \frac{x}{\theta} - 1$$

$$E\left[\frac{\partial}{\partial\theta}\log f(x_i;\theta)\right]^2 = \frac{1}{\theta^2}E[x-\theta]^2 = \frac{Var(x)}{\theta^2} = \frac{\theta}{\theta^2} = \frac{1}{\theta}$$

thus,

$$\frac{1}{nE\left[\frac{\partial}{\partial\theta}\log f(x_i;\theta)\right]^2} = \frac{1}{n/\theta} = \frac{\theta}{n}$$

Hence, $\bar{X}$ is the UMVUE of θ.

Example 2: Let $X_1, X_2, \ldots, X_n$ be a random sample from the exponential distribution having a p.d.f. of $f(x;\theta) = \frac{1}{\theta}e^{-x/\theta}$, $\theta > 0$, $0 < x < \infty$. Obtain the Cramer-Rao lower bound for the variance of an unbiased estimator for θ.

Solution:

$$f(x;\theta) = \frac{1}{\theta}e^{-x/\theta}$$

$$\log f(x;\theta) = -\ln\theta - x/\theta$$

$$\frac{\partial\log f(x;\theta)}{\partial\theta} = -\frac{1}{\theta} + \frac{x}{\theta^2}$$

hence,

$$E\left[\frac{\partial\log f(x;\theta)}{\partial\theta}\right]^2 = E\left[-\frac{1}{\theta} + \frac{x}{\theta^2}\right]^2$$

$$= \frac{1}{\theta^4}e[x-\theta]^2$$

$$= \frac{\theta^2}{\theta^4} = \frac{1}{\theta^2}$$

The Cramer-Rao lower bound is

$$Var(\hat{\theta}) \geq \frac{1}{n\cdot E\left[\frac{\partial\log f(x;\theta)}{\partial\theta}\right]^2}$$

$$= \frac{1}{n/\theta^2}$$

$$= \frac{\theta^2}{n} = \frac{var(X)}{n} = var(\bar{X})$$

$$F_D(r_i),\ \ r_i < t$$

6.3.5 Consistency

Knowing that an estimator is unbiased gives little information as to the goodness of the method of estimation. It would seem that closeness of the estimator to the parameter is of importance. The concept of consistency is slightly sharper than the variance of an estimator which gives a better idea of how close the estimate of the parameter is to the actual parameter. Roughly speaking, if an estimator, $\hat{\theta}$, approaches the parameter θ closer and closer as the sample size n increases, $\hat{\theta}$ is said to be a consistent estimator of θ.

Definition: Let $\hat{\theta}$ be an estimator for θ, based on a random sample of size n. If

$$\lim_{n\to\infty} P(|\hat{\theta} - \theta| \geq \epsilon) = 0, \text{ for any } \epsilon > 0,$$

then $\hat{\theta}$ is a consistent estimator for θ.

Thus, an estimator is consistent if as the sample size becomes larger, the probability increases that the estimates will approach the true value of the population parameter. Alternatively, $\hat{\theta}$ is consistent if it satisfies

(i) $V(\hat{\theta}) \to 0$ as $n \to \infty$

(ii) $\hat{\theta}$ becomes unbiased as $n \to \infty$

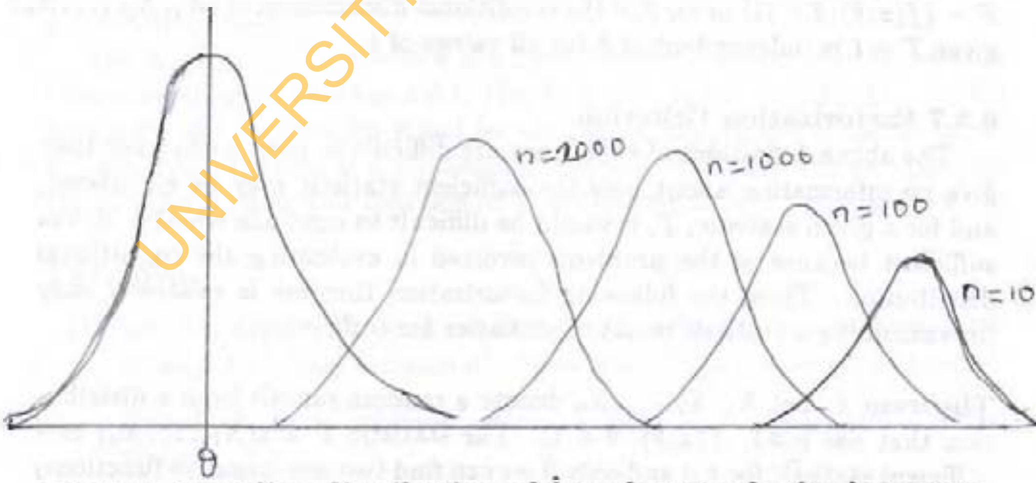


Fig 6.2: Sampling distribution of $\hat{\theta}$ as the sample size increases

From Chebyshev inequality (chapter?) we know that

$$\begin{aligned} P(|\hat{\theta} - \theta| \geq \epsilon) &\leq \frac{1}{\epsilon^2} E[(\hat{\theta} - \theta)^2] \\ &= \frac{1}{\epsilon^2} MSE(\hat{\theta}) \end{aligned}$$

It follows that if $MSE(\hat{\theta}) \to 0$ as $n \to \infty$, then $\hat{\theta}$ is consistent.

6.3.6 Sufficiency

A sufficient statistic is an estimator that summarizes from the sample data all information contained in these data, and no other estimator can provide additional information.

Definition 1: A statistic, $T(X_1, X_2, \ldots, X_n)$ is said to be sufficient for θ if the conditional distribution of $X_1, X_2, \ldots, X_n$, $T = t$, does not depend on θ for any values of t.

Definition 2: Let X_j, $j = 1, 2, \ldots, n$ be iid random variables with p.d.f. $f(x; \theta)$, and let $\theta = (\theta_1, \theta_2, \ldots, \theta_r) \in \Omega \subseteq R^r$ be a vector, let $T = (T_1, T_2, \ldots, T_m)$ where $T_j = t_j(X_1, X_2, \ldots, X_n)$, $j = 1, 2, \ldots, m$ are statistics. We say that T is an m-dimensional sufficient statistic for the family $F = \{f(x; \theta); \theta \in \Omega\}$ or for θ, if the conditional distribution of $(X_1, X_2, \ldots, X_n)$ given $T = t$ is independent of θ for all values of t.

6.3.7 Factorization Criterion

The above definitions of sufficiency are difficult to work with, since they give no information about how the sufficient statistic may be calculated, and for a given statistic, T, it would be difficult to conclude whether it was sufficient because of the problems involved in evaluating the conditional distribution. Thus, the following factorization theorem is relatively easy for examining a statistic or set of statistics for sufficiency.

Theorem 1: Let $X_1, X_2, \ldots, X_n$ denote a random sample from a distribution that has p.d.f. $f(x; \theta)$, $\theta \in \Omega$. The statistic $T = t(X_1, \ldots, X_n)$ is a sufficient statistic for θ if and only if we can find two non-negative functions, k_1 and k_2 such that

$$f(x_1; \theta) f(x_2; \theta) \cdots f(x_n; \theta) = k_1[t(x_1, \cdots, x_n); \theta] k_2(x_1, \cdots, x_n]$$

where $k_2(x_1, \dots, x_n)$ does not depend upon θ.

Theorem 2: Let $X_1, X_2, \cdots, X_n$ be iid random variable with p.d.f. $f(x;\theta), \theta = (\theta_1, \dots, \theta_r) \in \Omega$. An m-dimensional statistic

$$T = t(X_1, X_2, \dots, X_n) = (t_1(X_1, \dots, X_n), \dots, t_m(X_1, \dots, X_n))$$

is sufficient for θ if and only if the joint p.d.f. of $X_1, \dots, X_n$ factors as follows:

$$f(x_1, x_2, \dots, x_n; \theta) = g[t(x_1, x_2, \dots, x_n); \theta] h(x_1, x_2, \dots, x_n)$$

where g depends on $x_1, x_2, \dots, x_n$ and h is entirely independent of θ.

6.3.8 Completeness

Let X be a t-dimensional random variable with p.d.f. $f(x;\theta)$, $\theta \in \Omega \subseteq R^r$ and let $g : R^k \to R$ be measurable, so that $g(x)$ is a random variable. We assume that $E[g(x)]$ exists for all $\theta \in \Omega$ and set $F = \{f(x;\theta); \theta \in \Omega\}$.

Definition: We say that the family F (or the random variable X) is complete if for every g, $E[g(x)] = 0$ for all $\theta \in \Omega$ implies that $g(x) = 0$ except possibly on a set N of X's such that $P_\theta(X \in N) = 0$ for all θ.

6.3.9 Uniqueness

Let $X_1, X_2, \dots, X_n$, where n is a fixed integer, denote a random sample from a distribution that has p.d.f. $f(x;\theta), \theta \in \Omega$. Let $Y = y(X_1, X_2, \dots, X_n)$ be a sufficient statistic for θ and let the family $[g(y;\theta); \theta \in \Omega]$ of the p.d.f. be complete. If there is a conditional function of Y which is an unbiased function for θ, then this function of Y is the unique best statistic for θ.

Examples:

(1) Let $X_1, X_2, \dots, X_n$ be a random sample of a random variable with mean μ and finite variance σ^2. Show that $\bar{X}$ is a consistent estimator

Proof: $\bar{X}$ is an unbiased estimator for μ and its variance is σ^2/n
Since

$$MSE(\bar{X}) = \frac{\sigma^2}{n} \to 0 \text{ as } n \to \infty$$

$\therefore$ $\bar{X}$ is a consistent estimator for μ.

(2) Let $X_1, X_2, \ldots, X_n$ denote a random sample from a distribution with p.d.f.

$$f(x;\theta) = \begin{cases} \theta x^{\theta-1} & , 0 < x < 1 \\ 0 & , \text{elsewhere} \end{cases}$$

where $0 < \theta$. Prove that the product $u(X_1, X_2, \ldots, X_n) = X_1 \cdot X_2 \cdots X_n$ is a sufficient statistic for θ.

Solution

The joint p.d.f. of $X_1, X_2, \ldots, X_n$ is

$$\begin{aligned} f(x;\theta) &= \prod_{i=1}^{n} f(x_i;\theta) \\ &= \prod_{i=1}^{n} \theta x_i^{\theta-1} \\ &= \theta^n (x_1 x_2 \cdots x_n)^{\theta-1} \\ &= [\theta^n (x_1 x_2 \cdots x_n)^{\theta}] \left(\frac{1}{x_1, x_2, \ldots, x_n}\right) \end{aligned}$$

where $0 < x_i < 1$, $i = 1, 2, \ldots, n$. In the factorization theorem let

$$k_1[u(x_1, x_2, \ldots, x_n); \theta] = \theta^n (x_1, x_2, \ldots, x_n)^{\theta}$$

and

$$k_2(x_1, x_2, \ldots, x_n) = \left(\frac{1}{x_1, x_2, \ldots, x_n}\right)$$

since $k_2(x_1, x_2, \ldots, x_n)$ does not depend upon θ, the product $X_1 \cdot X_2 \ldots X_n$ is a sufficient statistic for θ.

(3) Let $X_1, X_2, \ldots X_n$ be a random sample from a distribution with p.d.f.

$$f(x;\theta) = \theta^x (1-\theta)^{1-x}, \; x = 0, 1, \; 0 < \theta < 1$$

Show that $\sum_{1}^{n} X_i$ is a sufficient statistics for θ.

Solution

$$
\begin{aligned}
f(x;\theta) &= \theta^{x}(1-\theta)^{1-x} \\
L(x;\theta) &= \prod_{i=1}^{n} f(x_i;\theta) \\
&= \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i} \\
&= \theta^{\Sigma x_i}(1-\theta)^{n-\Sigma x_i} \\
&= \left[\left(\frac{\theta}{1-\theta}\right)^{\Sigma x_i}(1-\theta)^{n}\right]\times 1
\end{aligned}
$$

This function is of the form $k_1[u(x_1, x_2, \ldots, x_n);\theta]k_2(x_1, x_2, \ldots, x_n)$ where

$$k_1[u(x_1, x_2, \ldots, x_n)] = \left(\frac{\theta}{1-\theta}\right)^{\Sigma x_i}(1-\theta)^{n}$$

and $k_2(x_1, x_2, \ldots, x_n) = 1$.

$\therefore \sum_{1}^{n} X_i$ is a sufficient statistic for θ

(4) Assume X is uniform on the interval $(0,\gamma)$. Based on a random sample of n observations, the maximum likelihood estimator $\hat{\Gamma} = \max(x_1, x_2, \ldots, x_n)$ whereas the method of moments estimator is $\tilde{\Gamma} = 2\bar{X}$. Compute

(i) $E[\hat{\Gamma}]$ and $E[\tilde{\Gamma}]$

(ii) the mean square errors of the estimators

(iii) Are they consistent?

Solution: Let $X \sim u(a,b)$, $f(x) = \frac{1}{b-a}, a < x < b$
If X is uniformly distributed over $(0,\gamma)$, the p.d.f. of

$$Y = \max(X_1, X_2, \ldots, X_n) \text{ is } n[F(y)]^{n-1}f(y)$$

or

$$g(y) = n\left(\frac{y}{\gamma}\right)^{n-1}\frac{1}{\gamma},\ 0 \le y \le \gamma$$

The expected value on this density and hence the estimates of Γ and $\tilde{\Gamma}$ are

$$\text{(i)}\quad E[\hat{\Gamma}] = \int_0^\gamma y g(y) dy = \int_0^\gamma yn\left(\frac{y^{n-1}}{\gamma^{n-1}}\right)\frac{1}{\gamma}dy$$

$$= \frac{1}{\gamma^n}\int_0^\gamma ny^n dy = \frac{n\gamma}{n+1}$$

$$E[\tilde{\Gamma}] = E(2\bar{X}) = 2E\left[\frac{\sum X_i}{n}\right]$$

$$= \frac{2}{n}\sum E(X_i)$$

$$= \gamma$$

Note that in the limit $\lim_{n\to\infty} E(\hat{\Gamma}) = \gamma$, the estimate of $\hat{\Gamma}$ is asymptotically unbiased.
(ii) $MSE[\hat{\Gamma}] = Var(\hat{\Gamma}) + (B(\hat{\Gamma}))^2$
but $Var(\hat{\Gamma}) = E[\hat{\Gamma}^2] - (E[\hat{\Gamma}])^2$
so

$$E[\hat{\Gamma}^2] = \int_0^\gamma y^2 n(\frac{y}{\gamma})^{n-1}\frac{1}{\gamma}dy = \frac{n}{\gamma^n}\int_0^\gamma y^{n+1}dy$$

$$= \frac{n\gamma^2}{n+2}$$

$$\therefore\quad Var(\hat{\Gamma}) = \frac{n\gamma^2}{n+2} - \frac{n^2\gamma^2}{(n+1)^2}$$

$$= \frac{\gamma^2 n}{(n+2)(n+1)^2}$$

$$B(\hat{\Gamma}) = E[\hat{\Gamma}] - \gamma$$

$$= \frac{n\gamma}{n+1} - \gamma = \frac{-\gamma}{n+1}$$

$$\therefore\quad MSE(\hat{\Gamma}) = Var(\hat{\Gamma}) + [B(\hat{\Gamma})]^2$$

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

$$= \frac{\gamma^2 n + (n+2)\gamma^2}{(n+2)(n+1)^2}$$

$$= \frac{2\gamma^2}{(n+1)(n+2)}$$

$$MSE(\tilde{\Gamma}) = Var(\tilde{\Gamma}) + (B(\tilde{\Gamma}))^2$$

But

$$Var(\tilde{\Gamma}) = Var(2\bar{X}) = 4E(\bar{X}^2) - (E(\tilde{\Gamma}))^2$$
$$= \frac{4\gamma^2}{12n}$$

$\therefore\ MSE(\tilde{\Gamma}) = \dfrac{4\partial^2}{12n}$ since $B(\tilde{\Gamma}) = E(\tilde{\Gamma}) - \gamma = 0$

(iii) As $n \to \infty$, $MSE(\hat{\Gamma}) = \dfrac{2\gamma^2}{(n+1)(n+2)} \to 0$

$\therefore\ \hat{\Gamma}$ is a consistent estimator.

As $n \to \infty$, $MSE(\tilde{\Gamma}) = \dfrac{4\gamma^2}{12n} \to 0$

$\therefore\ \tilde{\Gamma}$ is a consistent estimator.

(5) Let $P^* = \{b(\theta, n),\ 0 \le \theta \le 1\}$ be the family of binomial distribution corresponding to n independent trials with constant probability θ with

$$P(X = x) = \binom{n}{x} \theta^x (1-\theta)^{n-x}, \quad x = 0, 1, \ldots, n$$

Then P^* is complete.

Proof: We have to show that

$$(*) \quad \sum_{x=0}^{n} \binom{n}{x} \theta^x (1-\theta)^{n-x} g(x) = 0$$

implies $g(x) = 0$ for $x = 0, 1, \ldots, n$.

We can write $*$ as

$$(**) \quad \sum_{x=0}^{n} a(x)\varphi^x = 0,$$

where

$$a(x) = \binom{n}{x} g(x) \text{ and } \varphi = \frac{\theta}{1-\theta}$$

Equation (**) is a polynomial of degree n in φ and vanishes for every value of φ which is true only when
$a(x) = 0$ for $x = 0, 1, \ldots, n$.
Hence $g(x) = 0$ for $x = 0, 1, \ldots, n$.

6.4 Methods of Finding Estimators

6.4.1 Method of Moments

The oldest method of determining estimators is the method of moments. This is derived by equating sample moments to population moments.

Let $X_1, X_2, \ldots, X_n$ be a random sample from a population which depends on unknown parameters $\theta_1, \ldots, \theta_k$. Assume that the first k moments about the origin exist as function $\phi_r(\theta_1, \ldots, \theta_k)$ of the parameters, (where $r = 1, 2, \cdots, k$). The expectation $E(X^r)$ is frequently called the rth moment of the distribution, $r = 1, 2, 3, \ldots$. The sum $M_r = \frac{\sum_1^n X_i^r}{n}$ is the rth moment of the sample about the origin, $r = 1, 2, 3, \ldots$. The method of moments can be described as follows.

Equate $E(X^r)$ to M_r to obtain K equations in k unknown parameters. Thus one can get an estimate of a parameter.

6.4.2 Method of Least Squares

Generally the method of least squares is used to estimate parameters in a linear model.

Let $Y_1, \ldots, Y_n$ be independent random variables such that $E(Y_i) = \alpha + \beta x_i$, where x_i is a known constant and α and β are unknown parameters. Consider the quadratic function

$$Q(x_1, \ldots, x_n; \alpha, \beta) = \sum (y_i - \alpha - \beta x_i)^2$$

Let $\hat{\alpha}$ and $\hat{\beta}$ be the values of α and β that minimizes Q.
Then

$$\frac{\partial Q}{\partial \alpha} = -2\sum (y_i - \hat{\alpha} - \hat{\beta} x_i) = 0$$

$$\frac{\partial Q}{\partial \beta} = -2\sum (y_i - \hat{\alpha} - \hat{\beta} x_i) x_i = 0$$

Solving the two equations simultaneusly, we get

$$\hat{\beta} = \frac{\sum(y_i - \bar{y})(x_i - \bar{x})}{\sum(x_i - \bar{x})^2}$$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$$

6.4.3 Minimum Chi-Square Method

This method is applicable where observed values of the random variables themselves are frequencies of a finite number of mutually exclusive events with probabilities $P_1, \ldots, P_m$ which are functions of K unknown parameters to be estimated. The data consists of $n_1, \ldots, n_k$ frequencies for the m mutually exclusive events.

Let $\sum n_i = n$.

Obviously, $nP_1, nP_2, \ldots, nP_m$ are the expected frequencies of the m events. Then, a measure of discrepancy

$$\chi^2 = \sum_{i=1}^{m} \frac{(x_i - nP_i)^2}{nP_i} \qquad (*)$$

may be used.

The values of the unknown parameters that minimize $*$ are referred to as minimum chi-square estimates.

Usually non-linear equations result in minimizing $*$ therefore, computational methods are usually employed. However, an approximation to the chi-square expression was obtained by Berkson where

$$P_i = [1 + e^{-(\alpha + \beta x_i)}]^{-1}$$

is the logistic curve used often for analysing bioassay data.

The resulting approximate chi-square expression which he called logit chi-square is given as

$$Q = \sum_{i=1}^{m} (l_i - \alpha - \beta x_i)^2$$

where $l_i = \log P_i/q_i$, P_i is the proportion responding in a dose response experiment.

It is important to note that, it is usually very difficult to obtain minimum chi-square estimators thus, most statisticians use maximum likelihood for estimating the parameters.

6.4.4 Method of Maximum Likelihood

Let $x_1, x_2, \ldots, x_n$ be the observed values of a random sample $X_1, X_2, \ldots, X_n$. Then if $X_1, \ldots, X_n$ are discrete, random variables, the likelihood function of the sample denoted by L is defined as the joint probability of $x_1, x_2, \ldots, x_n$.

Such that

$$\begin{aligned} L(\theta) &= P(x_1, x_2, \ldots, x_n; \theta) \\ &= P(x_1;\theta)P(x_2;\theta), \ldots, P(x_n;\theta). \end{aligned}$$

If $X_1, \ldots, X_n$ are continuous random variables, the likelihood L, is defined to be the joint density evaluated at $x_1, x_2, \ldots, x_n$.

$$\begin{aligned} L(\theta) &= f(x_1, \ldots, x_n; \theta) \\ &= f(x_1;\theta)f(x_2;\theta)\ldots f(x_n;\theta) \end{aligned}$$

The value of θ that maximizes $L(\theta)$ will be taken as the estimate $\hat{\theta}$.

In general, let $X_1 \ldots X_n$ be independent random variables with density function $f(x_i; \theta_1, \ldots, \theta_k)$. The likelihood function is defined as

$$L(x_1, \ldots, x_n; \theta_1, \ldots, \theta_k) = \prod_{i=1}^{n} f(x_i; \theta_1, \cdots, \theta_k)$$

This gives the likelihood of obtaining the particular sample values. $(\hat{\theta}_1, \ldots, \hat{\theta}_k)$ is said to be the maximum likelihood estimate of $(\theta_1, \ldots, \theta_k)$ if it maximizes the likelihood function.

Examples

1. Let $X_1, X_2, \ldots, X_n$ be a random sample from each of the distributions having the following probability density functions:

   (a) $f(x;\theta) = \theta x^{\theta-1}, 0 < \theta < \infty, 0 < x < 1$

   (b) $f(x;\theta) = \frac{1}{\theta} e^{-x/\theta}, 0 < x < \infty, 0- < \theta < \infty$

   (c) $f(x, \alpha, \beta) = \frac{\beta^{\alpha} X^{\alpha-1} e^{-\beta X}}{\Gamma(\alpha)}, x > 0, \alpha < 0, \beta > 0$

   (i) Find the moment estimates for the unknown parameter(s) in (a-c).

(ii) Find the maximum likelihood estimator for the unknown parameters in (a) and (b).

Solution:

(ia) $E(X) = \int_0^1 x \cdot \theta x^{\theta-1} dx = \int_0^1 \theta x^{\theta} dx$

$$= \left.\frac{\theta x^{\theta+1}}{\theta+1}\right|_0^1 = \frac{\theta}{\theta+1}$$

The corresponding 1st sample moment is

$$M_1 = \frac{1}{n}\sum x_i = \bar{x}$$

Equating the corresponding moments and solving for the unknown parameter θ, we get

$$\mu_1' = \bar{x} \Rightarrow \frac{\theta}{\theta+1} = \bar{x}$$

$$\hat{\theta} = \frac{\bar{X}}{1-\bar{X}}$$

(b) $f(x;\theta) = \frac{1}{\theta}e^{-x/\theta}$, $0 < x < \theta$, $0 < \theta < \infty$

Solution: $\mu_1' = E(X) = \theta$

The first moment is

$$m_1 = \frac{1}{n}\sum_1^n x_i = \bar{x}$$

Setting $m_1 = E(X)$ we have

$$\hat{\theta} = \bar{X}$$

(c) $f(x,\alpha,\beta) = \dfrac{\beta^{\alpha} X^{\alpha-1} e^{-\beta X}}{\Gamma(\alpha)}$, $x > 0$, $\alpha > 0, \beta > 0$

Solution: Since we seek estimators for two parameters α and β, we equate two pairs of population and sample moments.

Then

$$E(X) = \mu_1' = \frac{\alpha}{\beta} \text{ and } E(X^2) = \mu_2' = \frac{\alpha(\alpha+1)}{\beta^2}$$

Given a random sample of size n, the first two moments are given by

$$m_1 = \frac{\sum_1^n x_i}{n} = \bar{x} \text{ and } m_2 = \frac{\sum_1^n x_i^2}{n}$$

We set $m_1 = E(X)$ and $m_2 = E(X^2)$ and solve for α and β. That is,

$$M_1 = m_1 \equiv \frac{\alpha}{\beta} = \bar{x} = M_1 \qquad \text{(i)}$$

and

$$\mu_2 = m_2 \equiv \frac{\alpha(\alpha+1)}{\beta^2} = \frac{1}{n}\sum x_i^2 = M_2 \qquad \text{(ii)}$$

From equation (i)

$$\beta = \frac{\alpha}{\bar{x}} = \frac{\alpha}{M_1}$$

Substituting this into equation (ii) and solving for α, we have

$$\begin{aligned}
\frac{\alpha(\alpha+1)}{\frac{\alpha^2}{M_1^2}} &= M_2 \\
\frac{M_1^2(\alpha+1)}{\alpha} &= M_2 \\
\alpha M_1^2 + m_1^2 &= \alpha M_2 \\
M_1^2 &= \alpha M_2 - \alpha M_1^2 \\
\hat{\alpha} &= \frac{M_1^2}{M_2 - M_1^2} = \frac{\bar{X}^2}{\frac{1}{n}\sum X_i^2 - \bar{X}^2}
\end{aligned}$$

Substituting α into equation (i)

$$\begin{aligned}
\hat{\beta} &= \frac{M_1^2}{\frac{M_2 - M_1^2}{m_1}} = \frac{m_1}{m_2 - m_1^2} \\
&= \frac{\bar{X}}{\frac{1}{n}\sum X_i^2 - \bar{X}^2}
\end{aligned}$$

ii(a) $f(x;\theta) = \theta x^{\theta-1}, 0 < \theta < \infty, 0 < x < 1$

$$L(\theta) = \prod_{i=1}^{n} f(x_i;\theta)$$

$$
\begin{aligned}
&= \prod_{i=1}^{n} \theta x_i^{\theta-1} \\
&= \theta^n (x_1 \dots x_n)^{\theta-1} \\
\ln L(\theta) &= n \ln \theta + (\theta - 1) \ln x_i \\
&= n \ln \theta + \theta \ln x_i - \ln x_i \\
\frac{\partial \ln L(\theta)}{\partial \theta} &= \frac{n}{\hat{\theta}} + \ln x_i = 0 \\
\hat{\theta} &= \frac{-n}{\ln X_i}
\end{aligned}
$$

(b) $f(x;\theta) = \frac{1}{\theta} e^{-x/\theta}$, $0 < x < \infty$, $0 < \theta < \infty$

$$
\begin{aligned}
L(\theta) &= \prod_{i=1}^{n} f(x_i;\theta) \\
&= \prod_{i=1}^{n} \frac{1}{\theta} e^{-x_i/\theta} \\
&= \frac{1}{\theta^n} e^{-\Sigma x_i/\theta} \\
\ln L(\theta) &= n \ln \theta - \frac{\Sigma x_i}{\theta} \\
\frac{\partial \ln L(\theta)}{\partial \theta} &= \frac{-n}{\hat{\theta}} + \frac{\Sigma x_i}{\hat{\theta}^2} = 0 \\
\hat{\theta} &= \frac{\Sigma X_i}{n} = \bar{X}
\end{aligned}
$$

Chapter 7

Tests of Hypotheses

7.1 Tests of Hypotheses

We now discuss the subject of hypothesis testing, which as earlier noted is one of the two basic classes of statistical inference. Testing of hypothesis involves using statistical inference to test the validity of postulated values for population parameter. If the hypothesis specifies the distribution completely it is called simple, otherwise it is called composite. For example, a demographer interested in the mean age of residents in a certain local government area might pose a simple hypothesis such as $\mu = 42$ or he might specify a composite hypothesis such as $\mu \neq 42$ or $\mu > 42$.

A statistical test is usually structured in terms of two mutually exclusive hypotheses referred to as the null hypothesis and the alternative hypothesis denoted by H_0 and H_1 respectively.

Two types of error occur in hypothesis testing; these are type I error and type II error. Type I error occurs if H_0 is rejected when it is true. The probability of a type I error is the conditional probability, $P(\text{reject } H_0 | H_0$ is true) and is denoted by α.

Hence,

$$\begin{aligned} \alpha &= P(\text{reject } H_0 | H_0 \text{ is true}) \text{ and} \\ 1 - \alpha &= P(\text{accept } H_0 | H_0 \text{ is true}) \end{aligned}$$

Type II error occurs if H_0 is accepted when it is false. It's probability is denoted by the symbol β, where

$$\begin{aligned} \beta &= P(\text{accept } H_0 | H_0 \text{ is false}) \\ 1 - \beta &= P(\text{reject } H_0 | H_0 \text{ is false}) \text{ called power of the test} \end{aligned}$$

Types I and II error can be explained as follows:

	H_0 is true	H_0 is false
Accept H_0	$1-\alpha$ correct devision	β Type II errors
Reject H_0	α Type I error	$1-\beta$ coirrect decision

Standard format of hypothesis testing: This format involves 5 steps.

Step 1: State the null and alternative hypotheses.

Step 2: Determine the suitable test statistics. This involves choosing the appropriate random variable to use in deciding to accept or reject the null hypothesis.

Test Statistics

Unknown Parameter	Appropriate Test Statistic
"μ" σ known, population normal	$Z = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$
σ known, population normal	$Z = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$, if n is "large" usually $n \geq 30$
"μ" σ known, population normal	$t = \frac{\bar{X}-\mu_0}{s/\sqrt{n}}$, with $(n-1)$df
"σ^2" population normal	$X^2 = \frac{(n-1)s^2}{\sigma^2}$, $(n-1)$df
"P" population normal	$Z = \frac{(X/n) - P_0}{\sqrt{P\frac{(1-P)}{n}}}$

Step 3: Determine the critical region using the cumulative distribution table for the test statistic. The set of values that lead to the rejection of the null hypothesis is called the critical region. A statistical test may be one one-tail or two-tail test. Whether one uses a one- or two-tail test of significance depend upon how the alternative hypothesis is formulated.

Types of hypothesis	H_0	H_1	Decision Rule Rejected H_0 if
Two-tail	$\mu = \mu_0$	$\mu \neq \mu_0$	$z < z_{\alpha/2}$ or $z > z_{\alpha/2}$
Right-tail	$\mu \leq \mu_0$	$\mu > \mu_0$	$z > z_{\alpha/2}$
Left-tail	$\mu \geq \mu_0$	$\mu < \mu_0$	$z < z_{\alpha/2}$

Step 4: Compute the value of the test statistic based on the sample information, e.g. Z_c, t_c, χ^2_c.
Step 5: Make a statistical decision and interpretation. H_0 is rejected if the computed value of the test statistic falls in the critical region otherwise it is accepted.

Example 1. An Emir from the north belives that the mean monthly income of messengers in the north is ₦1,000. Suppose a random sample of 144 workers is taken and a mean income of ₦1,200 found. If the population standard deviation is known to be ₦300. Check the claim of the Emir based on the sample information at $\alpha = 0.10$ level of significance.

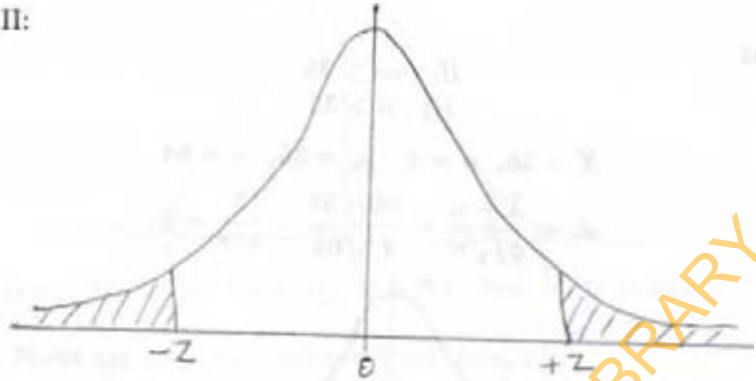
Solution:

Step I: $H_0 : \mu_0 =$ ₦1,000
$H_1 : \mu_0 \neq$ ₦1,000

Step II: The test statistics to use is $Z = \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$

since μ_0 and σ are known.

Step III:



Step IV: $\bar{X} = 1,200$, $\mu_0 = 1000$, $n = 144$, $\sigma = 300$

$$Z_c = \frac{1200 - 1000}{300/\sqrt{144}}$$
$$= 8$$

Step V: Since the computed value $Z_c = 8$ lies in the critical region, we reject the null hypothesis, H_0, that $\mu = 1000$.

7.1.1 One-sample test about μ (σ known)

The standardized normal random variable, Z, is the test statistic to use when the population variance is known. The two cases are:

1. For a sampling with replacement or an infinite population, the appropriate test statistic is

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$

2. For a sampling without replacement or a finite population, use

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}\sqrt{\frac{N-n}{N-1}}}$$

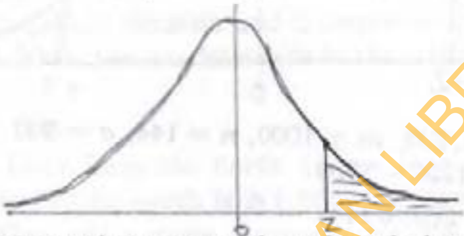
Example: Suppose a counsellor in a local government area believes that the teachers in that area are working at most 35 hours a week. A random sample of 64 teachers yields an average of 38 hrs of work per week. The population standard deviation is known to be 4 hrs. Determine whether the claim is correct at $\alpha = 0.05$.

Solution

$$H_0 : \mu_0 \leq 35$$
$$H_1 : \mu > 35$$

$$\bar{X} = 38, \quad \sigma = 4, \quad \mu_0 = 35, \quad n = 64$$

$$Z_c = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{38 - 35}{4/\sqrt{64}} = \frac{3}{4/8} = 6$$



Since Z_c (calculated value) is greater than the tabulated value ($Z_t = 1.645$) reject H_0 and conclude that the counsellor's claim is not true.

7.1.2 One-sample tests about μ (σ unknown)

If sampling is drawn from a normal population and the population variance is not known, then the test statistic to use is

$$t_c = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$$

with $n - 1$ degrees of freedom.

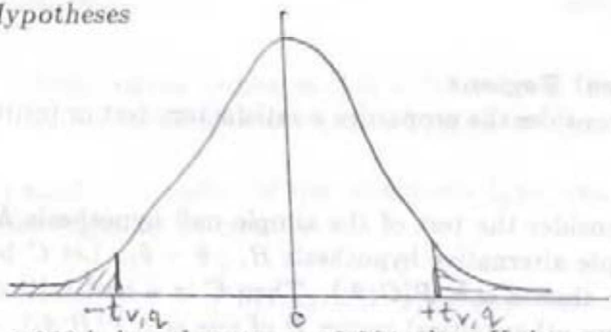
Example: A production company decided to examine the weights of their products. The manager believes the mean weight is 3.0 pounds. Suppose the company took a sample of 25 items from their production and found the mean weight was 3.6 pounds, with a sample standard deviation of 2.2 pounds. Using $\alpha = 0.05$, can this company's claim be regarded as correct?

Solution:

$$H_0 : \mu_0 = 3.0$$
$$H_1 : \mu \neq 3.0$$

$$\bar{X} = 3.6, \quad s = 2.2, \quad n = 25$$

$$t_c = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} = \frac{3.6 - 3.0}{2.2/5} = 1.363$$



Since $t_c = 1.363$ is less than $t_{\text{tab}} = 2.064$, then H_0 is accepted.

7.1.3 Tests on the Population Variance, σ^2

The following test statistic is used to test hypotheses about an unknown population variance

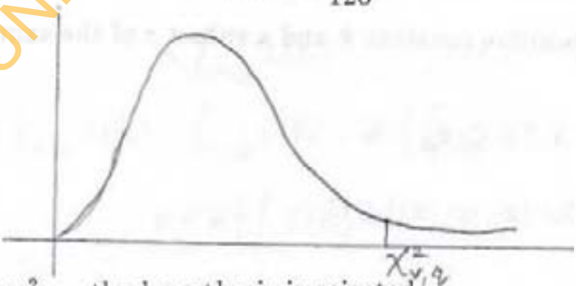
$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$$

where σ_0^2 is the hypothesized value of the population vairance, and $n-1$ is the degrees of freedom for this test statistic.

Example: A government wage review board claims that the wages of some cooks across the country show a variance of at least ₦120 per year. A random sample of 25 workers revealed a sample variable of ₦40/year. What is your conclusion about the review board's claim using $\alpha = 0.01$?

Solution: $H_0 : \sigma_0^2 \geq 120$ against $H_1 : \sigma_0^2 < 120$
$n = 25$, $s^2 = 40$ $\quad \chi^2_{0.01,24} = 10.86$

$$\chi_c^2 = \frac{24 \times 40}{120} = 8$$



Since $\chi_c^2 < \chi^2_{\text{tab}}$, the hypothesis is rejected.

7.2 Best Critical Regions

We want to consider the properties a satisfactory test or (critical region) should possess.

Definition: Consider the test of the simple null hypothesis $H_0 : \theta = \theta_0$ against the simple alternative hypothesis $H_1 : \theta = \theta_1$. Let C be a critical region of size α; that is $\alpha = P(C;\theta_0)$. Then C is a best critical region of size α if, for every other critical region B of size $\alpha = P(B;\theta_0)$, we have

$$P(C;\theta_1) \geq P(B;\theta_1)$$

That is, when $H_1 : \theta = \theta_1$ is true, the probability of rejection $H_0 : \theta = \theta_0$ using the critical region C is at least as great as the corresponding probability using any other critical region B of size α.

Thus a best critical region of size α is the critical reigon that has the greatest power among all critical regions of size α. The Neyman-Pearson lemma provides sufficient conditions for a best critical region of size α.

Theorem (Neyman-Pearson Lemma): Let $X_1, X_2, \cdots, X_n$ be a random sample of size n from a distribution having p.d.f. $f(x;\theta)$, where θ_0 and θ_1 are two possible values of θ. Denote the joint p.d.f. of $X_1, X_2, \cdots, X_n$ by the likelihood function

$$\begin{aligned} L(\theta) &= L(\theta; x_1, x_2, \ldots, x_n) \\ &= f(x_1;\theta) f(x_2;\theta) \cdots f(x_n;\theta) \\ &= \prod_{i=1}^{n} f(x_i;\theta) \end{aligned}$$

If there exists a positive constant k and a subset c of the sample space $\ni$

(a) $P[(X_1, \cdots, X_n) \in C; \theta_0] = \alpha$.

(b) $\dfrac{L(\theta_0)}{L(\theta_0)} \leq k$ for $(x_1, \cdots, x_n) \in C$

(c) $\dfrac{L(\theta_0)}{L(\theta_0)} \geq k$ for $(x_1, \cdots, x_n) \in C'$

then C is a best critical reigon of size α for testing the null hypothesis $H_0 : \theta = \theta_0$ against the simple alternative hypothesis $H_1 : \theta = \theta_1$

Proof: For random variables of the continuous type; replace the integrals by summation signs for discrete type.

$$\int L(\theta) = \int \cdots \int L(\theta; x_1, x_2, \cdots, x_n) dx_1, dx_2, \cdots, dx_n$$

If there exists anothei critical region of size α, say $B, \ni$

$$\begin{aligned} \alpha &= \int_C L(\theta_0) = \int_B L(\theta_0) \\ 0 &= \int_C L(\theta_0) - \int_B L(\theta_0) \end{aligned}$$

Since C is the union of the disjoint sets $C \cap B$ and $C \cap B'$ and B is the union of the disjoints sets $B \cap C$ and $B \cap C'$

$$= \int_{C \cap B'} L(\theta_0) + \int_{C \cap B} L(\theta_0) - \int_{C \cap B} L(\theta_0) + \int_{C' \cap B} L(\theta_0)$$

hence,

$$0 = \int_{C \cap B'} L(\theta_0) + \int_{C' \cap B} L(\theta_0)$$

By hypothesis (b), $KL(\theta_1) \geq L(\theta_0)$ at each point in C, and $C \cap B'$ in particular, thus,

$$k \int_{C \cap B'} L(\theta_1) \geq \int_{C \cap B'} L(\theta_0)$$

By hypothesis (c), $KL(\theta_1) \leq L(\theta_0)$ at each point in C' and $C' \cap B$ in particular, thus

$$K \int_{C' \cap B} L(\theta_1) \leq \int_{C' \cap B} L(\theta_0)$$

$$\therefore \; 0 = \int_{C' \cap B} L(\theta_0) - \int_{C' \cap B} L(\theta_0) \leq K \left\{ \int_{C \cap B'} L(\theta_1) - \int_{C' \cap B} L(\theta_1) \right\}$$

or

$$0 \leq k \left\{ \int_C L(\theta_1) - \int_B L(\theta_1) \right\}$$

thus,

$$\int_C L(\theta_1) \geq \int_B L(\theta_1);$$

that is, $P(C;\theta_1) \geq P(B;\theta)$. C is a best critical region of size α, since the result is true for every critical region B of size α.

Example 1: Let $X_1, X_2, \cdots, X_n$ be a random sample of size n from a Poisson distribution with mean λ. A best critical region for testing $H_0 : \lambda = 3$ against $H_1 : \lambda = 5$ is given by

$$\begin{aligned}
\frac{L(3)}{L(5)} &= \frac{3^{\Sigma x_i} e^{-3n}}{5^{\Sigma x_i} e^{-5n}} \times \frac{x_i!}{x_i!} \leq k \\
&= \left(\frac{3}{5}\right)^{\sum x_i} e^{2n} \leq k \\
&\sum x_i \ln \frac{3}{5} + 2n \leq \ln k \\
&\sum x_i \geq \frac{\ln k - 2n}{\ln(3/5)} = C
\end{aligned}$$

The best critical region is $\{(x_1, \cdots, x_n) : \sum x_i \geq C\}$

Example 2: Let $x_1, \cdots, X_n$ be a random sample from a normal distribution $N(\mu, 64)$.

(a) Show that $C = \{(x_1, x_2, \cdots, x_n) : \bar{x} \leq c\}$ is a best critical reigon for testing $H_0 : \mu = 80$ against $H_1 : \mu = 76$

(b) Find n and c so that $\alpha = 0.05$ and $\beta = 0.05$ approximately.

Solution:

(a) $H_0 : \mu = 80$ against $H_1 : \mu = 76$

$$\begin{aligned}
\frac{L(80)}{L(76)} &= \frac{(128)^{-n/2} \exp[-\frac{1}{128} \sum (x_i - 80)^2]}{(128)^{-n/2} \exp[-\frac{1}{128} \sum (x_i - 76)^2]} \\
&= \exp -\frac{1}{128} [\sum (x_i - 80)^2 - \sum (x_i - 76)^2] \\
&\Rightarrow \exp[-\frac{1}{128}(-8 \sum x_i + n80^2 - n76^2)] \leq k \\
&\Rightarrow 8 \sum x_i - 624n \leq 128 \ln k \\
&\sum x_i - 78n \leq 16 \ln k \\
&\bar{x} \leq 78 + \frac{16}{n} \ln k \\
&\bar{x} \leq c
\end{aligned}$$

where $C = 78 + \frac{16}{n} \ln k$.
According to N.P. lemma, the best critical region is

$$\therefore C = \{(x_1, \cdots, x_n) : \bar{x} \le c\}.$$

(b) Find n and c so that $\alpha = 0.05$ and $\beta = 0.05$

$$\begin{aligned} \alpha &= P(\bar{X} \le c; \mu = 80) \\ &= P\left(\frac{\bar{X} - 80}{8/\sqrt{n}} \le \frac{c - 80}{8/\sqrt{n}}; \mu = 80\right) \\ &= \phi\left(\frac{c - 80}{8/\sqrt{n}}\right) \\ \beta &= (\bar{X} > c; \mu = 76) \\ &= P\left(\frac{\bar{X} - 76}{8/\sqrt{n}} > \frac{c - 76}{8/\sqrt{n}}; \mu = 76\right) \\ &= 1 - \phi\left(\frac{c - 76}{8/\sqrt{n}}\right) \end{aligned}$$

$$\phi\left(\tfrac{c-76}{8/\sqrt{n}}\right) = 0.05 \quad \ldots (i) \;\Rightarrow\; \tfrac{c-80}{8/\sqrt{n}} = -1.645$$
$$\phi\left(\tfrac{c-76}{8/\sqrt{n}}\right) = 0.95 \quad \ldots (ii) \;\Rightarrow\; \tfrac{c-80}{8/\sqrt{n}} = -1.645$$

$$\begin{aligned} \frac{c - 80}{c - 76} &= \frac{-1.645}{1.645} = -1 \\ c - 80 &= -c + 76 \\ c &= 78 \\ \text{and } n &= 43 \end{aligned}$$

7.2.1 Uniformly Most Powerful (UMP) Tests

We now consider testing a simple hypothesis H_0 against a composite hypothesis H_1

Definition: The critical region C is a uniformly most powerful critical region of size α for testing a simple hypothesis H_0 against an alternative composite hypothesis H_1 if C is the best critical region of size α for testing H_0 against each simple hypothesis in H_1. A test defined by this critical

region C is called uniformly most powerful test with significance level α for testing a simple hypothesis H_0 against a composite hypothesis H_1.
UMP tests do not always exist, however when they exist the $N-P$ lemma can be used to find them.

Example 1. Let $X_1, \cdots, X_n$ be a random sample from a distribution with density function

$$f(x, \lambda) = \lambda e^{-\lambda x}, \quad x \geq 0, \ \lambda > 0$$

Show that there exists a uniformly most powerful test for testing the simple hypothesis $H_0 : \lambda = \lambda_0$ against an alternative composite hypothesis $H_1 : \lambda > \lambda_0$.

Solution: Let $\lambda_1 > \lambda_0$.
Consider testing $H_0 : \lambda = \lambda_0$ against $H_1 : \lambda = \lambda_1$.
By $N-P$ lemma,

$$\frac{L(H_0)}{L(H_1)} = \left(\frac{\lambda_0}{\lambda_1}\right)^n e^{(\lambda_0 - \lambda_1)\sum x_i} \leq k$$

that is,

$$n \log\left(\frac{\lambda_0}{\lambda_1}\right) + (\lambda_1 - \lambda_0)\sum x_i \leq \log k$$

$$(\lambda_1 - \lambda_0)\sum x_i \leq \log k - \log 0 \left(\frac{\lambda_0}{\lambda_1}\right)$$

$$\sum x_i \leq [\log k - \log\left(\frac{\lambda_0}{\lambda_1}\right) / (\lambda_1 - \lambda_0)]$$

This inequality holds for all $\lambda_1 > \lambda_0$. Thus UMP critical reigon exists and is

$$C = \{(x_1, \ldots, x_n) : \sum x_i \leq c\}$$

where

$$c = [\log k - \log\left(\frac{\lambda_0}{\lambda_1}\right) / (\lambda_1 - \lambda_0)]$$

Example 2. Let $X_1, \ldots, X_n$ be a random sample from $N(\mu, 25)$. Show that there exists a UMP for testing $H_0 : \mu = 40$ against the composite hypothesis $H_1 : \mu > 40$ $(H_1 : \mu = \mu_1 > 40)$.

Solution:

$$\begin{aligned}
\frac{L(H_0)}{L(H_1)} = \frac{L(40)}{L(\mu_1)} &= \frac{(50\pi)^{\frac{n}{2}} \exp[-\frac{1}{50}\sum(x_i - 40)^2]}{(50\pi)^{\frac{n}{2}} \exp[-\frac{1}{50}\sum(x_i - \mu_1)^2]} \\
&= \exp\left[-\frac{1}{50}[\sum(x_i - 40)^2 - \sum(x_i - \mu_1)^2\right] \\
&= \exp\{-\frac{1}{50}[2(\mu_1 - 40)\sum x_i + n(40^2 - \mu_1^2)]\} \leq k \\
&-\frac{1}{50}[2(\mu_1 - 40)\sum x_i + n(40^2 - \mu_1^2)] \leq \ln k \\
&2(\mu_1 - 40)\sum x_i + n(40^2 - \mu_1^2) \geq -50\ln k \\
&2(\mu_1 - 40)\sum x_i \geq -50\ln k - n(40^2 - \mu_1^2) \\
&\bar{x} \geq \left[\frac{-50\ln k}{2n(\mu_1 - 40)} + (40 + \mu_1)\right] = c
\end{aligned}$$

The best critical region of size α for testing $H_0 : \mu = 40$ against $H_1 : \mu = \mu_1$ where $\mu_1 > 40$ is given by

$$C = \{(x_1 \cdots x_n) : \bar{x} \geq c\}$$

C is selected such that $P(\bar{X} \geq c : H_0 : \mu = 40) = \alpha$.

Exercises:

1. Let $X_1 \cdots X_n$ be a random sample from $N(\theta, 1)$. Show that there exists no UMP test for testing the simple hypothesis $H_0 : \theta = \theta_0$ against an alternative composite hypothesis $H_1 : \theta \neq \theta_0$.

2. Let X have an exponential distribution with a mean of θ; that is, p.d.f. of X is $f(x;\theta) = \frac{1}{\theta}e^{-x/\theta}$; $0 < x < \infty$.

   (a) Show that a best critical region for testing $H_0 : \theta = 3$ against $H_1 : \theta = 5$ can be based on the statistic $\sum_1^n X_i$

   (b) If $n = 12$, find a best critical region of size $\alpha = 0.10$ for testing $H_0 : \theta = 3$ against $H_1 : \theta = 7$.

3. Let $X_1, X_2, \cdots, X_n$ denote a random sample from a normal distribution $N(\theta, 100)$.

   (i) Show that $c = \{c : c \leq \bar{x})\}$ is a best critical region for testing $H_0 : \theta = 75$ against $H_1 : \theta = 78$

   (ii) Find n and c so that

$$P[(X_1, X_2, \cdots, X_n) \in c : H_0] = P(\bar{x} \geq c; H_0) = 0.05$$

   and

$$P[(X_1, X_2, \cdots, X_n) \in c : H_1] = P(\bar{x} \geq C : H_1) = 0.09$$

   approximately.

7.3 Likelihood Ratio Tests

A general test-construction method that is applicable when both the null and alternative hypotheses, say H_0 and H_1, are composite shall be considered here. The functional form of the p.d.f. is assumed known but depends on an unknown parameter or parameters. That is, assume the p.d.f. of X is $f(X;\theta)$, where θ represents one or more unknown parameters.

Let Ω denote the total parameter space i.e. the set of all possible values of the parameter θ given by either H_0 or H_1.
Consider the following hypotheses

$$H_0 : \theta \in w \text{ against } H_1 : \theta \in w'$$

where w is a subset of Ω and w' is the complement of w with respect to Ω.

Definition 1: The likelihood ratio is the quotient

$$\lambda = \frac{L(\hat{w})}{L(\hat{\Omega})}$$

$L(\hat{w})$ is the maximum likelihood function with respect to θ when $\theta \in w$ and $L(\hat{\Omega})$ is the maximum likelihood functin with respect to θ when $\theta \in \Omega$.
Since λ is the quotient of non-negative functions, it implies that $\lambda \geq 0$.
And since $w \subset \Omega$, then $L(\hat{w}) \leq L(\hat{\Omega})$ and hence $\lambda \leq 1$. Thus $0 \leq \lambda \leq 1$.
If the maximum of L in w is much smaller than that in Ω, then the data $x_1, \ldots, x_n$ do not support the hypothesis.

$H_0 : \theta \in w$. That is, a small value of $\lambda = \dfrac{L(\hat{w})}{L(\hat{\Omega})}$ leads to the rejection of H_0. However, a value of the ratio λ that is close to 1 supports the null hypothesis H_0.

Definition 2: To test $H_0 : \theta \in w$ against $H_1 : \theta \in w'$, the critical region for the likelihood ratio test is the set of points in the sample space for which

$$\lambda = \frac{L(\hat{w})}{L(\hat{\Omega})} \leq k,$$

where $0 < k < 1$ and k is selected so that the test has a desired significance level α.

Example. Let $X_1, \cdots, X_n$ be a random sample from $N(\theta_1, \theta_2)$. Consider testing the hypotheses

$$H_0 : \theta_1 = 0,\ \theta_2 > 0,\quad -\infty < \theta_1 < \infty$$

$$H_1 : \theta_1 \neq 0,\ \theta_2 > 0,\ 0 < \theta_2 < \infty \quad \Omega = \{\theta_1, \theta_2\}$$

Solution.

$$\begin{aligned}
f(x; \theta_1, \theta_2) &= \left(\frac{1}{2\pi\theta_2}\right)^{\frac{1}{2}} e^{-\frac{1}{2\theta_2}(x-\theta_1)^2} \\
L(w) &= \left(\frac{1}{2\pi\theta_2}\right)^{\frac{n}{2}} e^{-\frac{1}{2\theta_2}\Sigma x_i^2} \\
\ln L(w) &= -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\theta_2 - \frac{\Sigma x_i^2}{2\theta_2} \\
\frac{\partial \ln L(w)}{\partial \theta_2} &= \frac{-n}{2\theta_2} + \frac{\Sigma x_i^2}{2\theta_2^2} = 0 \\
n &= \frac{\Sigma x_i^2}{\theta_2} \\
\hat{\theta}_2 &= \frac{\Sigma x_i^2}{n} \\
L(\hat{w}) &= \left(\frac{1}{(2\pi\Sigma x_i^2/n)}\right)^{\frac{n}{2}} e^{-\frac{1}{2}\Sigma x_i^2/\Sigma x_i^2/n}
\end{aligned}$$

$$= \left(\frac{n}{2\pi\Sigma x_i^2}\right)^{\frac{n}{2}} e^{-\frac{n}{2}}$$

$$L(\Omega) = \left(\frac{1}{2\pi\theta_2}\right)^{\frac{n}{2}} e^{-\frac{1}{2\theta_2}\Sigma(x_i-\theta_1)^2}$$

$$\ln L(\Omega) = \frac{-n}{2}\ln 2\pi - \frac{n}{2}\ln\theta_2 - \frac{1}{2\theta_2}\Sigma(x_i-\theta_1)^2$$

$$\frac{\partial \ln L(\Omega)}{\partial\theta_1} = \frac{1}{\hat{\theta}_2}\Sigma(x_i-\theta_1) = 0$$

$$= \Sigma x_i - n\hat{\theta}_1 = 0$$

$$\hat{\theta}_1 = \bar{x}$$

$$\frac{\partial \ln L(\Omega)}{\partial\theta_2} = \frac{-n}{2\hat{\theta}_2} + \frac{1}{2\hat{\theta}_2^2}\Sigma(x_i-\bar{x})^2 = 0$$

$$= \frac{n\hat{\theta}_2 + \Sigma(x_i-\bar{x})^2}{2\hat{\theta}_2^2} = 0$$

$$\hat{\theta}_2 = \frac{\Sigma(x_i-\bar{x})^2}{n}$$

$$L(\hat{\Omega}) = \left(\frac{1}{\left(2\pi\frac{\Sigma(x_i-\bar{x})^2}{n}\right)}\right)^{\frac{n}{2}} e^{-\frac{1}{2}\frac{\Sigma(x_i-\bar{x})^2}{\frac{\Sigma(x_i-\bar{x})^2}{n}}}$$

$$= \left(\frac{1}{\left(2\pi\frac{\Sigma(x_i-\bar{x})^2}{n}\right)}\right)^{\frac{n}{2}} e^{-\frac{n}{2}}$$

Taking the ratio,

$$\frac{L(\hat{w})}{L(\hat{\Omega})} = \frac{\left(\frac{n}{2\pi(\Sigma x_i^2}\right)^{\frac{n}{2}} + e^{-\frac{n}{2}}}{\left(\frac{n}{2\pi\Sigma(x_i-\bar{x})^2}\right)^{\frac{n}{2}} e^{-\frac{n}{2}}} = \left(\frac{\Sigma(x_i-\bar{x})^2}{\Sigma x_i}\right)^{\frac{n}{2}}$$

Note that $\Sigma x_i^2 = \Sigma(x_i-\bar{x})^2 + n\bar{x}^2$

$$\frac{L(\hat{w})}{L(\hat{\Omega})} = \left(\frac{\Sigma(x_i+\bar{x})^2}{\Sigma(x_i-\bar{x})^2 + n\bar{x}^2}\right)^{\frac{n}{2}}$$

$$= \left(\frac{1}{1+\frac{n\bar{x}^2}{\Sigma(x_i-\bar{x})^2}} \right)^{\frac{n}{2}} \leq \lambda_0$$

$$= \frac{1}{1+\frac{n\bar{x}^2}{\sum(x_i-\bar{x})^2}} \leq \lambda_0^{2/n}$$

Converting, we have

$$1 + \frac{n\bar{x}^2}{\sum(x_i - \bar{x})^2} \geq \lambda'$$

$$\frac{n\bar{x}^2}{\sum(x_i - \bar{x})^2} \geq \lambda' - 1 = \lambda''$$

Taking the square root,

$$\frac{\sqrt{n}\bar{x}}{\sqrt{\sum(x_i - \bar{x})^2}} \geq \lambda'''$$

Multiply through by $\lambda n - 1$

$$\frac{\bar{x}\sqrt{n}}{\sqrt{\frac{\sum(x_i-\bar{x})^2}{n-1}}} \geq C$$

Divide through by $\sqrt{n}$

$$\frac{\bar{x}}{s/\sqrt{n}} \sim t_{(n-1)}$$

Thus, the test can be based on the t distribution with $n - 1$ degrees of freedom.

Exercise: Let $X_1, \ldots, X_n$ be a random sample from $N(\mu, 5)$. Find the likelihood ratio for testing the hypothesis. $H_0 : \mu = 72$ against $H_1 : u \neq 72$

7.4 Sequential Probability Ratio Test

The two assumptions made on the previous methods of hypothesis testing are:

(i) that a sample of fixed size is taken

(ii) choice has to be made in favour of one or two possible outcomes

However, in sequential sampling, samples are taken one at a time and a decision is made at any point in time. At each stage of the sampling one of three decisions is made:

(i) accept

(ii) reject

(iii) continue sampling.

Sequential analysis refers to techniques for testing hypothesis or estimating parameters when the sample size is not fixed in advance but is determined during the course of the experiment by criteria which depend on the observations as they occur.

Consider testing a simple null hypothesis against a simple alternative hypothesis. That is, suppose a sample can be drawn from one of two distributions and it is desired to test that the sample came from one distribution against the possibility that it came from the other. If $X_1, X_2, \ldots$ denote the random variables, we want to test

H_0: $X_i \sim f_0(\cdot)$ against $H_1 : X_i \sim f_1(\cdot)$
the simple likelihood-ratio-test was of the following form;
Reject H_0 if $\lambda = \dfrac{L_0}{L_1} \leq k$ for some constant $k > 0$.

the sequential test employs the likelihood-ratio sequentially.

Define

$$\lambda_m = \lambda_m(x_1, \cdots, x_m) = \frac{L_0(x_1, \cdots, x_m)}{L_1(x_1, \cdots, x_m)}$$
$$= \frac{L_0(m)}{L_1(m)} = \frac{\prod_{i=1}^m f_0(x_i)}{\prod_{i=1}^m f_1(x_i)}$$

for $m = 1, 2, \ldots$ and compute sequentially $\lambda_1, \lambda_2, \ldots$. For fixed k_0 and k_1 satisfying $0 < k_0 < k_1$, adopt the following procedure:
Take observation x_1 and compute λ_1: if $\lambda_1 \leq k_0$ reject H_0;
If $\lambda_1 \geq k_1$ accept H_0: but if $k_0 < \lambda_1 < k_1$ take observation x_2 and compute λ_2.
If $\lambda_2 \leq k_0$ reject H_0; if $\lambda_2 \geq k_1$ accept H_0 and if $k_0 < \lambda_2 < k_1$ take x_3, etc.
The idea is to continue sampling as long as $k_0 < \lambda_j < k_1$ and stop as soon as $\lambda_m \leq k_0$ or $\lambda_m \geq k_1$ where you reject or accept H_0.

The critical region is defined as

$$C = \bigcup_{n=1}^{\infty} C_n,$$

where

$$C_n = \{(x_1 \cdots x_n); k_0 < \lambda_j(x_1 \cdots x_j) < k_1, j = 1, \ldots, n-1, \ \lambda_n(x_1 \cdots x_n) \leq k_0\}$$

i.e. a point in C_n indicates that H_0 is to be rejected for a sample of size n. Similarly, the acceptance region is defined as

$$A = \bigcup_{n=1}^{\infty} A_n$$

$$A_n = \{(x_1 \cdots x_n); k_0 < \lambda_j(x_1 \cdots x_j) < k_1; j = 1, \ldots, n-1, \ \lambda_n(x_1 \cdots x_n) \geq k_1\}$$

Definition: For fixed $0 < k_0 < k_1$, a test as described above is defined to be a sequential probability ratio test.

Determining k_0 and k_1 so that the sequential probability ratio test will have pre-assigned α and β for its respective sizes of Type I and Type II errors.

$$\alpha = P[\text{reject } H_0 | H_0 \text{ is true}] = \sum_{n=1}^{\infty} \int_{C_n} L_0(n) \tag{1}$$

and

$$\beta = P[\text{accept } H_0 | H_0 \text{ is false}] = \sum_{n=1}^{\infty} \int_{A_n} L_1(n) \tag{2}$$

For fixed α and β, (1) and (2) are two equations in the two unknowns k_0 and k_1. Solutions to equations (1) and (2) would give the SPRT having the desired preassigned error sizes α and β.

Theorem I: Let k_0 and k_1 be defined so that the SPRT corresponding to k_0 and k_1 has error sizes α and β; then k_0 and k_1 can be approximated by k_0' and k_1' where

$$k_0' \approx \frac{\alpha}{1-\beta} \quad \text{and} \quad k_1' \approx \frac{1-\alpha}{\beta}$$

Proof:

$$
\begin{aligned}
\alpha &= P[\text{reject } H_0 | H_0 \text{ is true}] \\
&= \sum_{n=1}^{\infty} \int_{C_n} L_0(n) \\
&\leq \sum_{1}^{\infty} \int_{C_n} k_0 L_1(n) \leq k_0 \sum_{1}^{\infty} \int_{C_n} L_1(n) \\
&\leq k_0 P[\text{reject } H_0 | H_1 \text{ is true}] \\
&\leq k_0 [1-\beta] \\
&\text{and hence } k_0 \geq \frac{\alpha}{1-\beta} \\
&\therefore \; k_0' \approx \alpha/1-\beta
\end{aligned}
$$

Also,

$$
\begin{aligned}
1-\alpha &= Pr[\text{accept } H_0 | H_0 \text{ is true}] \\
&= \sum_{n=1}^{\infty} \int_{A_n} L_0(n) \\
&\geq k_1 \sum_{1}^{\infty} \int_{A_n} L_1(n) \\
&\geq k_1 Pr[\text{accept } H_0 | H_1 \text{ is true}] \\
&\geq k_1 \beta \\
\text{hence, } \frac{1-\alpha}{\beta} &\geq k_1
\end{aligned}
$$

That is $k_1' \approx \dfrac{1-\alpha}{\beta}$

Theorem II: Let α' and β' be the error sizes of the SPRT defined by k_0' and k_1'.
Then $\alpha' + \beta' \leq \alpha + \beta$.

Proof: Let A' and C' denote the acceptance and critical regions of the SPRT deifned by k_0' and k_1'. Then

$$
\alpha' = \sum_{1}^{\infty} \int_{C_n'} L_0(n) \leq k_0' \sum_{1}^{\infty} \int_{C_n'} L_1(n) = \frac{\alpha}{1-\beta}(1-\beta')
$$

and

$$1-\alpha' = \sum_1^\infty \int_{A'_n} L_0(n) \geq k'_1 \sum_1^\infty \int_{A'_n} L_1(n) = \frac{1-\alpha(\beta')}{\beta}$$

Combining the two results, we have:

$$\alpha' \leq \frac{\alpha}{1-\beta}(1-\beta') \text{ and } 1-\alpha' \geq \frac{1-\alpha}{\beta}(\beta')$$

or $\quad \alpha'(1-\beta) \leq \alpha(1-\beta') \text{ and } \beta(1-\alpha') \geq (1-\alpha)\beta'$

$\quad \alpha'(1-\beta) \leq \alpha(1-\beta') \text{ and } \beta'(1-\alpha) \leq \beta(1-\alpha')$

or $\quad \alpha'(1-\beta) + \beta'(1-\alpha) \leq \alpha(1-\beta') + \beta(1-\alpha')$

$\quad \alpha' + \beta' \leq \alpha + \beta$

7.4.1 Approximate Expected Sample Size of SPRT

Select two numbers k_0 and k_1 and continue sampling as long as $k_0 < \lambda_m < k_1$ and stop as soon as $\lambda_m \leq k_0$ or $\lambda_m \geq k_1$. If

$$Z_i = \log_e \frac{f_0(x_i)}{f_1(x_i)}$$

an equivalent test is described as follows:

Continue samping as long as $\log_e k_0 < \sum_{i=1}^m z_i < \log_e k_1$ and stop as soon as

$$\sum_1^m z_i \leq \log_e k_0$$

when you reject H_0

or

$$\sum_1^m z_i \geq \log_e k_1$$

when you accept H_0.

Let N be the random variable denoting the sample size of the SPRT, and let

$$Z_i = \log_e \frac{f_0(X_i)}{f_1(X_i)}$$

Theorem: Wald's equation: Let $Z_1, Z_2, \cdots, Z_n \cdots$ be independent identically distributed random variables satisfying $E[|Z_i|] < \infty$. Let N be an integer-valued random variable whose value n depends only on the values of the first n Z_i's.

Suppose $E(N) < \infty$. Then $E[z_1 + \cdots + z_n] = E(N) \cdot E(Z_i)$.

Proof.

$$
\begin{aligned}
E[N|H_0 \text{ is true}] &\simeq \frac{\alpha \log_e k_0 + (1-\alpha)\log_e k_1}{E[Z_i|H_0 \text{ is true}]} \\
&\simeq \frac{\alpha \log_e[\alpha|(1-\beta)] + (1-\alpha)\log_e[(1-\alpha)|\beta]}{E[Z_i|H_0 \text{ is true}]}
\end{aligned}
$$

and

$$
\begin{aligned}
E[N|H_0 \text{ is false}] &\simeq \frac{(1-\beta)\log_e k_0 + \beta \log_e k_1}{E[Z_i|H_0 \text{ is false}]} \\
&\simeq \frac{(1-\beta)\log_e[\alpha/(1-\beta)] + \beta\log_e[(1-\beta)/\beta]}{E[Z_i|H_0 \text{ is false}]}
\end{aligned}
$$

Example 1: Let X have a Poisson distribution with mean θ. Find the sequential probability ratio test for testing $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1$. Show that this test can be based upon the statistic $\sum_{i=1}^{n} x_i$. If $\theta_0 = 0.02$, $\theta_1 = 0.07$, $\alpha = 0.2$ and $\beta = 0.1$, find k_0 and k_1.

Solution:

$$
\begin{aligned}
L &= \frac{\prod_{i=1}^{n} \theta^{x_i} e^{-\theta}}{x_i!} \\
&= \frac{\theta^{\Sigma x_i} e^{-n\theta}}{\sum x_i!}
\end{aligned}
$$

$$
\frac{L(\theta_0)}{L(\theta_1)} = \frac{\frac{\theta_0^{\sum x_i} e^{-n\theta_0}}{\sum x_i!}}{\frac{\theta_1^{\sum x_i} e^{-n\theta_1}}{\sum x_i!}} = \left(\frac{\theta_0}{\theta_1}\right)^{\sum x_i} e^{-n(\theta_0 - \theta_1)}
$$

$$k_0 < \left(\frac{\theta_0}{\theta_1}\right)^{\sum x_i} e^{-n(\theta_0-\theta_1)} < k_1$$

$$\ln\left(\frac{\alpha}{1-\beta}\right) < \sum x_i \ln\left(\frac{\theta_0}{\theta_1}\right) - n(\theta_0-\theta_1) < \ln\left(\frac{1-\alpha}{\beta}\right)$$

$$\ln\left(\frac{\alpha}{1-\beta}\right) + n(\theta_0-\theta_1) < \sum x_i \ln\left(\frac{\theta_0}{\theta_1}\right) < \ln\left(\frac{1-\alpha}{\beta}\right) + n(\theta_0-\theta_1)$$

$$\frac{\ln\left(\frac{\alpha}{1-\beta}\right) + n(\theta_0-\theta_1)}{\ln\left(\frac{\theta_0}{\theta_1}\right)} > \sum x_i > \frac{\ln\left(\frac{1-\alpha}{\beta}\right) + n(\theta_0-\theta_1)}{\ln\left(\frac{\theta_0}{\theta_1}\right)}$$

i.e.

$$\frac{\ln\left(\frac{1-\alpha}{\beta}\right) + n(\theta_0-\theta_1)}{\ln\left(\frac{\theta_0}{\theta_1}\right)} < \sum x_i < \frac{\ln\left(\frac{\alpha}{1-\beta}\right) + n(\theta_0-\theta_1)}{\ln\left(\frac{\theta_0}{\theta_1}\right)}$$

$$\frac{\ln(\left(\frac{0.8}{0.1}\right) - 0.05n}{\ln\left(\frac{0.02}{0.07}\right)} < \sum x_i < \frac{\ln\left(\frac{0.2}{0.9}\right) - 0.05n}{\ln\left(\frac{0.02}{0.07}\right)}$$

$$\frac{\ln 8 - 0.05n}{-1.25} < \sum x_i < \frac{\ln 0.22 - 0.05n}{-1.25}$$

$$k_0 = -1.66 + 0.04n$$

$$k_1 = 1.20 + 0.04n$$

Example 2: Consider testing the hypothesis $H_0 : \mu < \mu_0$ Vs $H_1 : \mu = \mu_1 > \mu_0$, where μ is the mean of a normal density with known variance σ^2. Here the likelihood ratio

$$\begin{aligned}
\frac{L_1}{L_0} &= \frac{\exp[-\frac{1}{2\sigma^2}\sum(x_i-\mu_1)^2]}{\exp[-\frac{1}{2\sigma^2}\sum(x_i-\mu_0)^2]} \\
&= \exp\left[-\frac{1}{2\sigma^2}[\sum(x_i-\mu_1)^2 - \sum(x_i-\mu_0)^2]\right] \\
&\quad \exp\left[-\frac{1}{2\sigma^2}[2(\mu_0-\mu_1)\sum x_i + n(\mu_1^2-\mu_0^2)]\right] \\
&= \exp\left[\frac{\mu_1-\mu_0}{\sigma^2}\sum x_i + \frac{n(\mu_0^2-\mu_1^2)}{2\sigma^2}\right]
\end{aligned}$$

The critical reigon C will be determined by the inequality

$$\exp\left[\frac{\mu_1 - \mu_0}{\sigma^2}\sum x_i + \frac{n(\mu_1^2 - \mu_0^2)}{2\sigma^2}\right] \geq k$$

k is non-negative constant.
Taking logarithms

$$\frac{\mu_1 - \mu_0}{\sigma^2}\sum x_i \geq \ln k + \frac{n(\mu_1^2 - \mu_0^2)}{2\sigma^2}$$

$$\sum x_i \geq \frac{\sigma^2 \ln k}{\mu_1 - \mu_0} + \frac{n(\mu_1 + \mu_0)}{2}$$

$$\frac{L_{1n}}{L_{0n}} = \exp\left[\frac{\mu_1 - \mu_0}{\sigma^2}\sum x_i + \frac{n(\mu_0^2 - \mu_1^2)}{2\sigma^2}\right]$$

$$\log\frac{\beta}{1-\alpha} < \frac{\mu_1 - \mu_0}{\sigma^2}\sum x_i + \frac{n(\mu_0^2 - \mu_1^2)}{2\sigma^2} < \log\frac{1-\beta}{\alpha}$$

Since $\mu_1 > \mu_0$, then

$$\log\frac{\beta}{1-\alpha} + \frac{n(\mu_0^2 - \mu_1^2)}{2\sigma^2} < \frac{\mu_1 - \mu_0}{\sigma^2}\sum x_i < \log\frac{1-\beta}{\alpha} + \frac{n(\mu_1^2 - \mu_0^2)}{2\sigma^2}$$

$$\frac{\sigma^2}{\mu_1 - \mu_0}\left[\log\frac{\beta}{1-\alpha} + \frac{n(\mu_0^2 - \mu_1^2)}{2\sigma^2}\right] < \sum x_i < \frac{\sigma^2}{\mu_1 - \mu_0}\left[\log\frac{1-\beta}{\alpha} + \frac{n(\mu_1^2 - \mu_0^2)}{2\sigma^2}\right]$$

$$\frac{\sigma^2}{\mu_1 - \mu_0}\log\frac{\beta}{1-\alpha} + \frac{n(\mu_0^2 + \mu_1^2)}{2} < \sum x_i < \frac{\sigma^2}{\mu_1 - \mu_0}\log\frac{1-\beta}{\alpha} + \frac{n(\mu_1 + \mu_0)}{2}$$

Suppose we have chosen $\alpha = 0.05$, $\beta = 0.10$, and that we are testing $\mu_0 = 10$ against $\mu_1 = 10.5$ with $\sigma = 1$, then

$$-4.50 + 10.25n < \sum x_i < 5.78 + 10.25n$$

and the sequence test proceeds as follows:

(i) $\sum x_i \leq -4.50 + 10.25n$ accept $\mu = 10$

(ii) If $\sum x_i \geq 5.78 + 10.25n$ accept $\mu = 10.5$

(iii) If neither inequality is satisfied, take another observation.

Suppose 15 values of X_i obtained in sampling are as follows:
10.91, 8.88, 10.35, 9.84, 10.42, 10.05, 9.94, 11.15, 9.92, 10.23
10.53, 8.61, 9.70, 8.83, 8.96

n	1	2	3	4	5	6	7	8	9	10
b_0	5.75	16.00	26.25	36.50	46.75	57.00	67.25	77.50	87.77	98.00
$\sum x_i$	10.91	19.79	30.14	39.98	50.40	60.54	70.39	81.54	91.46	101.69
b_1	16.03	26.28	36.53	46.78	57.03	67.28	77.53	87.78	98.03	108.28

n	11	12	13	14	15
b_0	109.25	118.50	128.75	139.00	149.25
$\sum x_i$	112.22	120.83	130.53	139.36	148.32
b_1	118.53	128.78	139.03	149.28	159.53

The decision to accept $\mu = 10$ which is the correc decision occurred at the 15th sample.

Chapter 8

Probability Calculus and Limit Theorems

8. The Probability Calculus

8.1 Space of Events and Class of Subsets

Events arising out of an experiment in nature may not be predicted with certainty; but certain events occur more often than others, so that there is a natural enquiry as to whether such phenomenon can be described by attaching precise measures or indices (to be called probabilities) to events. These quantities, to be useful in practice, should satisfy some consistency conditions based on intuitive notions.

The calculus of probability is devoted to the construction and study of such quantities. Their applicability to problems of the real world, that is, as examination of the relevance of hypothetical probabilities in a contigent situation or the estimation of appropriate probabilities and prediction of events based on them constitute the subject matter of statistical theory and methods to which this part of the textbook is all about. As a first step, we must specify the set of all outcomes of an experiment which are distinguishable in some sense. We shall call them elementary events for some practical reasons demanding the definition of an event as a more general statement governing an elementary event.

Consider the heights of individuals drawn from a population, the observations can theoretically assume continuously values on the entire real line, or if height is measured in intervals of 0.5cm, the elementary events will be confined to a countable set of points on the real line. Events of interest in such cases may be the occurence of individuals with heights in specified intervals.

Let the set of all elementary events be **S**, also called the sample space with elements or points **A** representing elementary events. If an event A_1 is defined by a property π possessed by an element, it is natural to consider the event where π does not hold. Some basic fundaments of calculus have been discussed earlier. Our class C of sets on which probabilities are to be defined should at least have the property that if A_1 and $A_2 \in C$, then A_1^c, $A_1 \cup A_2$ and $A_1 \cap A_2 \in C$. Such a class of subsets is called Boolean field and is represented as $\mathcal{F}$, we shall discuss more on this later in this part.

8.2 Probability Set Function

For each set $A \in \mathcal{F}$ we have to assign a value $P(A)$ (called probability of A) or we define a set function P over the members of $\mathcal{F}$. The probability function has to satisfy some intuitive requirements. First, its range shoud be $[0,1]$, the value 0 for impossibility and 1 for certainty. Second, let $A_1, \ldots, A_k$ $(A_i \in \mathcal{F})$ be disjoint sets whose union is S, which means that any elementary event that occur has one and only one of K possible descriptions $A_1, \ldots, A_k$. The relative frequencies of the events $A_1, \ldots, A_k$ must then add up to unity which suggests, the following requirements stated in form of two fundamental axioms governing the set function P on the chosen field $\mathcal{F}$;

First Axiom: $P(A) \geq 0$, $A \in \mathcal{F}$

Second Axiom:

$$\bigcup_{i=1}^{\infty} A_i = S, \quad A_i \cap A_j = \phi \quad \forall\, i \neq j$$

This implies

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

A set function P defined for all sets in $\mathcal{F}$ and satisfying the first and second axioms is called a probability measure. The consequencies of these axioms are:

(i) $0 \leq P(A) \leq 1, \quad A \in \mathcal{F}$

(ii) $P(\phi) = 0$ [by observing $\phi \cup (\bigcup_{i=1}^{\infty} A_i) = S = \bigcup_{1}^{\infty} A_i$ and the second axiom holds for a finite decomposition of S

(iii) $P(S) = 1$ [since $S \cup \phi \cup \phi = \cdots S$]

(iv) $P(\cup A_i) = \Sigma P(A_i)$, for any countable union of disjoint sets in $\mathcal{F}$, whose union also belong to $\mathcal{F}$. $[(\cup A_i) \cup (\cup A_i)] = S = (\cup A_i)^c \cup A_1 \cup A_2 \ldots]$ then $(P(\cup A_i) + P[(\cup A_i)^c]) = 1 = P[(\cup A_i)^c + P(A_1) + P(A_2) + \cdots]$

(v) Let A_i be a non-increasing sequence of sets in $\mathcal{F}$ s.t. $\lim_{i\to\infty} A_i = \cap A_i \in \mathcal{F}$. Then $\lim_{i\to\infty} P(A_i) = P(\lim_{i\to\infty} A_i)$. The compliments A_i^c form a non-decreasing sequence and

$$\cup A_i^c = (\cap A_i)^c = A_1^c \cup (A_2^c - A_1^c) \cup (A_3^c - A_2^c) \cdots$$

using the consequence (iv) of the second axiom, we see that

$$\begin{aligned} P[(\cap A_i)^c] &= P[A_1^c] + P(A_2^c - A_1^c) + \cdots \\ &= P(A_1^c) + [P(A_2^c) - P(A_1^c)] + \ldots \\ &= \lim P(A_i^c) \quad \text{as } i \to \infty. \end{aligned}$$

i.e. $1 - P(\cap A_i) = \lim[1 - P(A_i)]$ as $i \to \infty$
or $P(\lim A_i) = P(\cap A_i) = \lim P(A_i)$ as $i \to \infty$. This establishes the non-increasing sequnce.

Similarly, if A_i, $i = 1, 2, \ldots$ is a non-decreasing sequence of sets in $\mathcal{F}$ such that $\lim A_i = \cup A_i \in \mathcal{F}$, then

$$\lim_{i\to\infty} P(A_i) = P(\cup A_i) = P(\lim_{i\to\infty} A_i)$$

establishes the non-decreasing sequence.

Example 8.1: Let $A_i \in \mathcal{F}$, $i = 1, 2. \ldots$ be a countable number of disjoint sets such that $\cup A_i = S$. Then what is the behaviour of $\sum_1^\infty P(A_i)$? From the second intuitive requirement that $\sum_1^\infty P(A_i) = 1$, for any finite decomposition of S it follows that $\sum_1^\infty P(A_i) \leq 1$.

Example 8.2: What are the consequencies of $\sum_1^\infty P(A_i) < 1$? Let us consider the sequence of events $B_k = \bigcup_k^\infty A_i$, the sets $B_1, B_2, \ldots$ form a decreasing sequence tending in the limit (we shall discuss this shortly) to the empty set ϕ. We may then expect $P(B_k)$ to decrease to zero as k increases. This does not happen if $\sum_1^\infty P(A_i) < 1$, it appears then that we need as a conveniest condition $\sum_1^\infty P(A_i)$ for a finite or countable decomposition of S.

8.3 Borel Field and Its Extension to Probability Measure

A field which contains all countable unions and intersections of sequences is called a Borel field or a sigma field (σ-field). Given a field $\mathcal{F}$ there exists a minimal Borel field containing $\mathcal{F}$ which we denote by $\mathcal{B}(\mathcal{F})$, this we can show as follows:

There is at least one Borel field which contains $\mathcal{F}$. All arbitrary intersections of Borel fields $\mathcal{B}$ are also Borel fields. Hence the intersection of all Borel fields containing $\mathcal{F}$ is precisely the minimal Borel field, $\mathcal{B}(\mathcal{F})$ #.

A set function defined n $\mathcal{F}$ and satisfying the first and second axioms can be uniquely extended to all sets in $\mathcal{B}(\mathcal{F})$, that is, there exists a unique function P^* such that:

(i) $0 \le P^*(A) \le 1, \quad A \in \mathcal{B}(\mathcal{F})$

(ii) $P^*(\cup A_n) = \Sigma P^*(A_n)$ for a countable sequence A_n of disjoint sets in $\mathcal{B}(\mathcal{F})$ and

(iii) $P^*(A) = P(A)$, if $A \in \mathcal{F}$.

We define a function P^* as follows:

Consider a set A in $\mathcal{B}(\mathcal{F})$ and a collection of sets A_i in $\mathcal{F}$ s.t. $A \subset \bigcup_{i=1}^\infty A_i$, then

$$P^*(A) = \inf_{A_i} \sum P(A_i)$$

It may then be of some advantage to consider the wider field $\mathcal{B}(\mathcal{F})$ as our basic class of sets for defining the probability function, we are then in a position to build up a calculus of probability based on the basic space S of element A, a Borel field or σ-field $\mathcal{B}$ of sets (events) in S and a probability measure P on $\mathcal{B}$. The triplet $(S, \mathcal{B}, P)$ is called a probability space, while we define S or $(S, \mathcal{B})$ as the sample space.

Let us consider the following Lemma to substantiate our discussion on probability measure.

Lemma 8.1: If $P'(S)$ is a probability measure on $\mathcal{B}_1$ with $P(S) = F(S)$, if $S = (-\infty, y)$, then $P(s) = P'(s)\ \forall\, s \in \mathcal{B}_1$.

Proof: Let $S = [a, b]$; since $(-\infty, b) = (-\infty, a) + [a, b)$
$F(b) = F(a) + P([a, b)) = F(a) + P'([a, b))$
Hence P and P' agree on intervals of the form $[a, b)$.

Let $\mathcal{F}_n$ be the class of finite unions of sets of the form $[a, b)$ in $[-n, n)$. Then $\mathcal{F}_n$ is a field and a set $S \in \mathcal{F}_n$ can always be expressed as union of disjoint sets of the form $S_i = [a_i, b_i)$. Therefore

$$P(s) = \sum P([a_i b_i)) = \sum P'([a_i b_i]) = P'(s)$$

so that P and P' agree on $\mathcal{F}_n$ and therefore on $\mathcal{B}(\mathcal{F})$. Now let $s \in \mathcal{B}_1$, then $S \cap [-n, n) \in \mathcal{B}(\mathcal{F}_n)$ so $P[S \cap [-n, n]] = P'[S \cap [-n, n]]$ and letting $n \to \infty$, $P(s) = P'(s)$ #.

8.4 Limit Theorems

The distributions of many statistics of interest are too complicated to derive in useful forms. In many cases, however, limiting distributins can be obtained; these may be used as approximations to the exact distributions when the number of observations N is large. In this section we study some limits theorems of general applicability. A conventional method of obtaining the limiting distribution of a sequence of random variables is to find the limiting distribution of an approximating sequence discuss as follows:

8.4.1 Convergence of Random Variable

A sequence of random variables $\{X_n\}$, $n = 1, 2, \ldots$ is said to converge to a random variable X (applicable to constant c):

(a) Weakly or in Probability (written as $X_n \xrightarrow{P} X$) if for every given $\epsilon > 0$,

$$\lim_{n\to\infty} P(|X_n - X| > \epsilon) = 0 \tag{8.1}$$

(b) Strongly or almost surely (written as $\lim_{n\to\infty} X_n = X$ with prob 1 or $X_n \xrightarrow{a.s} X$) if

$$P\left(\lim_{n\to\infty} X_n = X\right) = 1 \tag{8.2}$$

or equivalently

$$\lim_{N\to\infty} P\left(\sup_{n\geq N}|X_n - X| > \epsilon\right) = 0 \tag{8.3}$$

for every ϵ.

(c) In quadratic mean (written as $X_n \xrightarrow{\text{q.m}} X$) if

$$\lim_{n\to\infty} E(X_n - X)^2 = 0 \tag{8.4}$$

or in r-th mean if

$$\lim_{n\to\infty} E(|X_n - X|^r) = 0, \ \forall\ r > 0$$

A sequence of random variables $\{X_n(\cdot)\}$, $n = 1, 2, \ldots$ is said to converge to a constant c is sense of equation (8.1) through (8.4) according as the sequence $\{X_n(\cdot) - t\}$, $n = 1, 2, \ldots$ converges to zero in the sense of (8.1) through (8.4).

8.4.2 Relationship Among Various Types of Convergence

(A) **Claim 1**

Convergence in quadratic mean (when $r = 2$) implies convergence in Probability.

Proof:

$$P(|X_n - X| > \xi) \leq \frac{1}{\xi^2}E[|X_n - X|^2]$$

$$\lim_{n\to\infty} P[|X_n - X| > \xi] \leq \frac{1}{\xi^2}\lim_{n\to\infty}(|X_n - X|^2) = 0$$

and by virtue of $P(\cdot) \geq 0$, then L.H.S. takes zero only as its value. #

(B) **Claim 2**

Convergence almost surely implies convergence in probability.

Proof:

Recall that in almost surely

$$\lim_{N\to\infty} P\left[\sup_{n\geq N}|X_n - X| > \xi\right] = 0$$

Then

$$|X_n - X| > \xi \Rightarrow \sup_{n \geq N} |X_n - X| > \xi$$

$$\therefore \quad P(|X_n - X| > \xi) \leq P\left(\sup_{n \geq N} |X_n - X| > \xi\right)$$

$$\lim_{n \to \infty} P(|X_n - X| > \xi) \leq \lim_{n \to \infty} P\left[\sup_{n \geq N} |X_n - X| > \xi\right] = 0 \qquad \#$$

(C) If $X_n \xrightarrow{q.m} X$ in such a way that $\Sigma E(X_n - X)^2 < \infty$ then $X_n \xrightarrow{a.s} X$
Proof: Since $\Sigma E(X_n - X)^2 < \infty$, the infinite sum

$$\sum_1^{\infty}(X_n - X)^2 \tag{8.5}$$

must converge except for a set of sequence of measure zero in R^{∞}. Suppose that (1.5) diverges on a set of measure P, let A_N be the set of sequences such that

$$\sum_1^{\infty}(X_i - X)^2 > \lambda$$

The sequence $\{A_N\}$ is non-decreasing and hence by

$$\lim P(A_N) = P(\lim A_N) \geq P,$$

since a divergent series eventually exceeds any given value λ. Consequently

$$\sum_1^{N} E(X_i - X)^2 = E\left[\sum_1^{N}(X_i - X)^2\right] \geq \lambda P$$

which contradicts the assumption,

$$\sum_1^{\infty}(X_i - X)^2 < \infty \quad \text{for sufficiently large } \lambda$$

Hence $P[\Sigma(X_n - X)^2 \text{ converges}] = 1$, but if an infinite series converges the n-th term tends to zero as n increases to infinity. Hence

$$P(|X_n - X| \to 0 \text{ as } n \to \infty) \geq P\left[\sum_1^{\infty}(X_n - X)^2 \text{ converges}\right) = 1$$

This is almost surely, but incidentally we have proven that if X_n; $n = 1, 2, \ldots$ is a sequence of random variables then

$$E\left(\sum_{1}^{\infty} X_n\right) = \sum_{1}^{\infty} E(X_n)$$

provided $\sum E|Y_n| < \infty$, which ensures the convergence of $\sum_{1}^{\infty} X_n$ with probability one.

(D) **Claim 4**

The convergence in r-th mean implies convergence in Probability.

Proof: Let $g(\cdot)$ be a positive function of X and select $\xi > 0$, then

$$\begin{aligned} E(g(x)) &= \int_{-\infty}^{\infty} g(x)dF(x) = \int_{-\infty}^{\infty} g(x)f(x)dx \\ &\geq \int_{g(x)>\xi} g(x)dF(x) \geq \int_{g(x)>\xi} \xi dF(x) = \xi \int_{g(x)>\xi} dF(x) \\ &= \xi P[g(x) > \xi] \\ \frac{E(g(x))}{\xi} &\geq P[g(x) > \xi] \end{aligned}$$

Set $g(x) = |X_n - X|^r$, then

$$P(|X_n - X|^r > \xi) \leq \frac{1}{\xi^r} E(|X_n - X|^r) \to 0$$

If there is convergence in r-th mean.

We shall endeavour to study some useful theorems on limits of random sequence of variables:

Theorem 8.1:

Let $\{X_n\}$ be a sequence of independent random variables having zero means and variances $\{\sigma_n^2\}$ and suppose $C_n^2 = \sum \sigma_i^2$ such that

$$\lim_{n\to\infty} \frac{C_n^2}{n^2} = 0, \text{ set } \bar{X}_n = \frac{1}{n}\sum X_i$$

The $\bar{X}_n$ converges in probability to zero, that is,

$$\bar{X}_n \xrightarrow{P} 0$$

Proof:

$$E(\bar{X}_n) = 0$$

$$V(\bar{X}_n) = E(\bar{X}_n^2) = \frac{1}{n^2}\Sigma\sigma_i^2 = \frac{C_n^2}{n^2}$$

For every $\xi > 0$, by Chebychev's inequality we have

$$P[|\bar{X}_n| > \xi] \leq \frac{1}{\xi^2}\text{Var}(\bar{X}_n)$$

$$\lim_{n\to\infty} P(|\bar{X}_n| > \xi) \leq \frac{1}{\xi^2}\lim_{n\to\infty} V(\bar{X}_n)$$

$$= \frac{1}{\xi^2}\lim\frac{C_n^2}{n^2} \longrightarrow 0 \qquad \#$$

Theorem 8.2
If $\{R_n\}$ is a sequence of random variables which converges in Probability to (i) S and (ii) T, then S and T are equivalent random variables.

Proof: Suppose we have

$$P(\theta : |S(\theta) - T(\theta)| > \xi)$$
$$= P(\theta : |(S_n(\theta) - T(\theta) - (S_n(\theta) - S(\theta))| > \xi)$$

Let K, K_1 and K_2 respectively denote the set of points in R^∞ for which

$$|(S_n(\theta) - T(\theta) - (S_n(\theta) - S(\theta))|$$
$$= |S - T| > \frac{1}{N} = K,$$
$$|S_n - S| > \frac{1}{2N} = K_1 \text{ and}$$
$$|S_n - T| > \frac{1}{2N} = K_2$$

Searching through we see that $K \subset K_1 \cup K_2$ implies $P(K) \subseteq P(K_1)+P(K_2)$, then

$$P(|S-T|) > \frac{1}{N} \leq P(|S_n - S|) > \frac{1}{2N} + P(|S_n - T|) > \frac{1}{2^N} \qquad (8.6)$$

For n sufficiently large, equation (8.6) tends to zero.

Let H_N denote the set in R^∞ for which $|S - T| > \frac{1}{N}$ is true for $N = 1, 2, \ldots$, so that $H_1 \subset H_2 \subset H_3 \ldots$ is also true. The set of points in R^∞ for which S is not equal to T is therefore given by

$$H = \lim_{n \to \infty} H_n = \cup H_i$$

and so

$$P(S \neq T) = P(\cup H_i) \leq \Sigma P(H_i)$$

But

$$P(H_i) = P(|S - T| > \frac{1}{i}) = 0 \quad \text{(by 8.6)}$$

Therefore,

$$P(S \neq T) = 0$$ #

Theorem 8.3

A function $f : s \to R$ is said to be uniformly continuous if for every $\xi > 0$, there exist a δ such that for every $T \in S$, $P \in S$, if $|T - P| < S$ then $|f(T) - f(P)| < \xi$. Then given $\{T_n\}$ sequence of random variables such that $T_n \xrightarrow{P} T$, let $g(T)$ be a continuous function of T on $\mathbb{R}$, then

$$g(T_n) \xrightarrow{P} g(T).$$

Proof:

$g(T)$ is uniformly continuous on a closed interval $(-a, a)$, for $\xi > 0$, choose a such that

$$P[|T| \geq a] < \frac{\xi}{2} \tag{8.7}$$

for such a and ξ there exist a $\delta(\xi, a)$ such that if:

(i) $|T| \leq a$ (ii) $|T_n - T| \leq \delta$, then

(iii) $|g(T_n) - g(T)| < \xi$.

Let K_1, K_2 and K_3 respectively denote the set of points in R^∞ for which (i) = K_1, (ii) = K_2 and (iii) = K_3 holds then $K_3 \supset K_1 \cap K_2$. From $K_2 \subset K_3$ but $K_1 \cap K_2 \subset K_3$. Therefore

$$\bar{K}_3 \subset \overline{K_1 \cap K_2}$$
$$\bar{K}_3 \subset \bar{K}_1 \cup \bar{K}_2 \quad \text{(By De Morgan's Law)}$$
$$P(\bar{K}_3) \subseteq P(\bar{K}_1) + P(\bar{K}_2)$$

From (iii), we have

$$P(|g(T_n) - g(T) > \xi \leq P(|T| \geq a) + P(|T_n - T| < \delta)$$

But there exist $S(\delta, \xi, a)$ such that for every $S > s(\delta, \xi, a)$ sufficiently large, then $P(|T_n - T| \leq \delta)$ tends to zero, that is

$$P(|T_n - T| \leq \delta) > \xi/2 \qquad (8.8)$$

Using (8.7) and (8.8) we have

$$P(|g(T_n) - g(T)| > \xi) \leq \frac{\xi}{2} + \frac{\xi}{2} = \xi \qquad \#$$

8.4.3 Convergence of a Sequence of Distribution Functions

Introduction

We shall denote the sequence of distribution functions of the random variables $\{T_n\}$, $n = 1, 2, \ldots$ by $\{F_n\}_{n=1}^{\infty}$. The sequence of random variables $\{T_n\}$ is said to converge in distribution or in law to a random variable T with distribution function F_T; if $F_n \to F$ as $n \to \infty$ at all continuity parts of F, such convergence is denoted as $T_n \xrightarrow{L} T$.

The approximating distribution F is called the limiting or asymptotic distribution of T_n. In statistical applications, limiting distribution plays an important role. The random variable T_n stands for a statistic computed from a sample of size n, whose actual distribution is difficult to find. In such a case it may be approximated by the limiting distribution, at least for large values of n. We shall examine some results that are important in studying limit distributions.

Theorem 8.4

Let $\{T_n, R_n\}_{n=1}^{\infty}$ be a sequence of pairs of variables. Then $|T_n - R_n| \xrightarrow{P} 0$, $R_n \xrightarrow{L} R$, this implies $T_n \xrightarrow{L} T$, that is the limiting distribution of T_n exists and is the same as that of R.

Proof:

Let $F_{T_n} = F$ be the distribution function of T_n and that of R_n be F_R. Also let $S_n = T_n - R_n$ and t be a continuity point of F_R, then

$$\begin{aligned} F_{T_n}(t) &= P(T_n \leq t) = P(S_n + R_n \leq t) = P(R_n \leq y - S_n) \\ &= P[R_n \leq t - S_n/S_n > -\xi] + P[R_n \leq t - S_n/S_n \leq -\xi] \\ &\leq P[R_n \leq t + \xi] + P[T_n \leq -\xi] \end{aligned}$$

as $n \to \infty$, $P(S_n \le \xi \to 0$ since

$$T_n - R_n \xrightarrow{P} 0.$$

$$\liminf F_{T_n}(t) \le \limsup F_{T_n}(t) \le F_R(r + \xi),$$

also

$$\begin{aligned} F_{T_n}(t) &= P[R_n \le t - S_n; S_n < \xi] + P[R_n \le t - S_n; S_n \ge \xi] \\ &\ge P[R_n \le t - S_n; S_n < \xi] - P[R_n \le t - \xi] \\ &\ge P[R_n \le t - \xi] - P[R_n \le t - \xi; S_n \ge \xi] \\ &\ge P[R_n \le t - \xi] - P[S_n \ge \xi] \end{aligned}$$

as $n \to \infty$, $\quad P[S_n \ge \xi) \to 0$

$\therefore \ F_{T_n}(t) \ge F_{R_n}(t - \xi)$, since $R_n \xrightarrow{L} R$ and collating these results we have

$$F_R(t - \xi) \le \liminf F_{T_n}(t) \le \limsup F_{T_n}(t) < F_T(t + \xi)$$

$$\liminf F_{T_n}(t) = \limsup F_{T_n}(t) = F_R(t)$$

i.e.

$$F_{T_n} \longrightarrow F_R \Longrightarrow Y_n \longrightarrow R$$

\#

Theorem 8.5

Let $\{T_n, R_n\}_{n=1}^{\infty}$ be a sequence of pairs of random variables, then:

(a) $T_n \xrightarrow{L} T$, $R_n \xrightarrow{P} 0 \Rightarrow T_n R_n \xrightarrow{P} 0$

Proof: Consider

$$\begin{aligned} P(|T_n R_n| > \xi) &= P(|T_n R_n| > \xi,\ |R_n| \le \xi/k) \\ &\quad + P(|T_n R_n| > \xi,\ |R_n| > \xi/k) \\ \Rightarrow \limsup_{n\to\infty} P(|T_n R_n| > \xi) &= P(|T| > k) \end{aligned}$$

for any fixed k and any δ positive, we can choose k sufficiently large such that

$$P(|T| > K) < \delta, \quad \text{is true}$$

$$\therefore \lim_{n\to\infty} P(|T_n R_n| > \xi) = 0$$

\#

(b) $T_n \xrightarrow{L} T,\ R_n \xrightarrow{P} C \Rightarrow T_n + R_n \xrightarrow{L} T + C$

Proof: We have that $T_n \xrightarrow{L} T$, $R_n \xrightarrow{P} C$ hence $T_n + C \xrightarrow{L} T + C$, now

$$(T_n + R_n) - (T_n + C) = T_n - C \xrightarrow{P} 0$$

and by Theorem (8.4)

$$T_n + R_n \xrightarrow{L} T + C \text{ since } T_n + R_n \xrightarrow{P} 0$$

and

$$T_n + C \xrightarrow{L} Y + C$$ #

(c) $T_n \xrightarrow{L} T,\ R_n \xrightarrow{P} C \Rightarrow T_n R_n \xrightarrow{L} CT$

Proof: $T_n R_n - CT_n = T_n(R_n - C)$ but $T_n \xrightarrow{L} T$ and $R_n \xrightarrow{P} C$, then $T_n(R_n - C) \xrightarrow{P} 0 \Rightarrow T_n R_n - CT_n \xrightarrow{P} 0$, but $CT_n \xrightarrow{L} CT$. Hence $T_n R_n \xrightarrow{L} CT$ #

Theorem 8.6

A sequence $\{F_n\}_{n=1}^{\infty}$ of distribution converges weakly to F iff $F_n \to F$ at every continuity point of F.

Proof:

A seequence A is said to be dense in B if for every $b \in B$, we can find a subsequence in A which converges to B. Also a sequence $\{F_n\}_{n=1}^{\infty}$ of functions is said to converge weakly to a function F if $F_n \to F$ at every $t \in C(F)$.

(i) The *necessary* condition is true because the continuity intervals of F are dense in the real line.

(ii) For the sufficiency condition, let D be a dense set, take T' and T'' in D, then $T' < T < T'' \Rightarrow F(T') < F(T) < F(T'')$, so

$$F_n(T') \leq \liminf F_n(T) \leq \limsup F_n(T) \leq F_n(T'')$$

then

$$T' \in D, F_n \to F \Rightarrow F_n(T') \to F(T')$$

and since $T'' \in D$,

$$F_n \to F \Rightarrow F_n(T'') \to F(T'')$$

Let $T' \to T$ and $T'' \to T$, then

$$F_n(T') \to F(T), \;\; F_n(T'') \to F(T)$$

Then

$$\liminf F_n(T) \longrightarrow F(T), \;\; \limsup F_n(T) \to F(T)$$

$\Rightarrow F_n(T) \to F(T)$ which is weakly convergence. #

Theorem 8.7 (Helly's Sequential Extraction)

Every sequence of distribution is weakly compact, that is, there is a subsequence which tends to a function (not necessarily a distribution function) at all continuity points of the latter.

Proof:

Let $D = \{r_k\}$ be the set of all rationals. Since $F_n(T_r)$ is bounded, there exists a convergent subsequence. Consider the sequence $F_{n_1}(t)\}$ which converges for the particular value of $t = r_1$. From the sequence $\{F_{n_1}(t)\}$ we can extract another subsequence $\{F_{n_2}(t)\}$ in a similar way, which converges at $t = r_2$, and of course such a sequence will converge at $t = r_1$ and so on.

Let $F_{n_{jj}}$ be the j-th member of $F_{nj}(t)$ then the sequence $\{F_s\} = \{F_{n_{11}}, F_{n_{22}}, \cdots, F_{n_{jj}} \ldots\}$ necessarily converges for all $t \in D$. F_D is bounded and non-decreasing for all any t.

Let $F(y)$ = upper bound

$$F_D(r_i), \;\; r_i < t \tag{8.9}$$

F is continuous from the left, bounded and non-decreasing if $t \in C(f)$, there exists a sequence of rational values $\{t_i', t_i''\}$ such that $t_i' < t < t_i''$ and $F(t_i'') - F(t_i') \to 0$ as $i \to \infty$.

Also

$$F_s(t_i') \leq F_s(t) \leq F_s(t_i'')) \tag{8.10}$$

For each s, where $F_s(t)$ is the sequence that converges to F_D.

Taking limits in (4.10) we obtain

$$F_D(t_i') \leq \liminf F_s(t) \leq \limsup F_s(t) \leq F_D(t_i'')$$

and is true for each i. Then $F_D(t_i'') - F_D(t_i') \to 0$ so that the limit $F_s(t)$ exists and is equal to $F(t)$ defined in (8.9) #

Theorem 8.8 (Helly Bray)

A sequence $\{F_n\}$ is said to converge completely (c) to F, that is, $F_n \xrightarrow{c} F$, if $F_n(-\infty) = F(-\infty)$ and $F_n(\infty) = F(\infty)$

The theorem states that if $F_n \xrightarrow{c} F$, then

$$\int g dF_n \longrightarrow \int g dF \text{ for bounded continuous function } g.$$

Proof:

Take a_0 and a_1 with $a_0 < a_1$ as two continuous points and consider

$$\begin{aligned} R &= \int_{-\infty}^{\infty} g dF_n - \int_{-\infty}^{\infty} g dF = \int_{-\infty}^{\infty} g(dF_n - dF) \\ &= \int_{-\infty}^{a_0} g(dF_n - dF) + \int_{a_0}^{a_1} g(dF_n - dF) + \int_{a_1}^{\infty} g(dF_n - dF) \\ |R| &\le |J_1| + |J_2| + |J_3| \end{aligned}$$

where

$$\begin{aligned} J_1 &= \int_{-\infty}^{a_0} g(dF_n - dF); \\ J_2 &= \int_{a_0}^{a_1} g(dF_n - dF); \\ J_3 &= \int_{a_1}^{\infty} g(dF_n - dF) \end{aligned}$$

Since g is bounded there exist C such that $|g| < C$

$$\begin{aligned} |J_1| &= \left|\int_{-\infty}^{a_0} g(dF_n - dF)\right| \le \int_{-\infty}^{a_0} |g(dF_n - dF)| \\ &\le \int_{-\infty}^{a_0} (dF_n - dF) = C\,[F_n(a) + F(a)] = \frac{\xi}{s} \end{aligned}$$

For $n > n_0$ and a_0 sufficiently small, similarly $|J_3| \le \xi/5$ for $n > n_0$ and a_1 sufficiently large. g is uniformly continuous in the interval (a_0, a_1), suppose we divide the interval (a_0, a_1) into K parts, that is,

$$a_0 = t_0 < t_1 < \ldots < t_k = a_1$$

with $t_i \in 5(F)$; $i = 1, \ldots, K$. Such that $|g(t_i) - g(t_i + 1)| < \frac{\xi}{5}$ for $t_i \leq t \leq t_{i+1}$, uniformly for all i.

Let us define $g_m(t) = g(t_i)$, $t_i \leq t \leq t_{i+1}$, then

$$\begin{aligned}\int_{a_0}^{a_1} g_m(t)dF_n &= \sum_{i=0}^{m-1} \int_{t_i}^{t_{i+1}} g(t_i)dF_n(t) \\ &= \sum_{i=0}^{m-1} g)t_i) \int_{t_i}^{t_{i+1}} dF_n = \sum_{i=0}^{m-1} g(t_i)\left[F_n(t_{i+1}) - F_n(t_i)\right]\end{aligned}$$

This tends to

$$\sum_{i=0}^{m-1} g(t_i)\left[F(t_{i+1}) - F(t_i)\right] = \int_{a_0}^{a_1} g_m(t)dF$$

as $n \to \infty$ we have

$$\int_{a_0}^{a_1} g_m(t)(dF_n - dF) < \frac{\xi}{5}$$

Now,

$$\begin{aligned}J_2 &= \int_{a_0}^{a_1} g(dF_n - dF) = \int_{a_0}^{a_1}(g - g_m)dF_n - dF + \int_{a_0}^{a_1} g_m(dF_n - dF) \\ &= \int_{a_0}^{a_1}(g_m - g)dF\end{aligned}$$

$$|J_2| \leq \frac{\xi}{5}\left[\int_{a_0}^{a_1} dF_n + 1 + \int_{a_0}^{a_1} dF\right] \leq \frac{3\xi}{5}$$

$$\begin{aligned}\therefore |R| &\leq |J_1| + |J_2| + |J_3| \\ &\leq \frac{\xi}{5} + \frac{\xi}{5} + \frac{3\xi}{5} = \xi\end{aligned}$$

#

Problems 8.1

1(a) Define each of the following concepts:

(i) a σ-algebra

(ii) a probability measure

(iii) a probability space

(iv) a random variable

(b) Show that if $A_i \in \mathcal{B}$, then $\bigcap_{i=1}^{\infty} A_i \in \mathcal{B}$, where $\mathcal{B}$ is an algebra of sets.

2. Prove that the convergence of a sequence of random variables in quadratic mean to a constant C in a way that $\Sigma E(X_i - C)^2 < \infty$ implies convergence almost surely to C.

3. If $Q'(s)$ is a probability measure on $\mathcal{B}_1$ with $Q(s) = F(s)$, for $s = (-\infty, x)$, show that

$$Q(s) = Q'(s) \quad \forall\, s \in \mathcal{B}_1$$

4. If $\{A_n\}_{n=1}^{\infty}$ be a sequence in $\{$, show that if $\Sigma P(A_n) < \infty$ then $P(A) = 0$, where A is the set of elements common to an infinity of these sets in $\{$.

5. Show that if $\Sigma P(A_n) = \infty$ and the events A_n are independent, then $P(A) = 1$.

Solutions 8.1

1(a) (i) A field which contains unions of all countable sequences of set (and therefore countable intersections) is called a Borel field or σ-field.

(ii) A set function K defined for all sets in s and satisfyng the axioms:

(a) $P(A) \geq 0$, $A \in K$

(b) $\bigcup_{i=1}^{\infty} A_i = \Omega$, $A_i \cap A_j = \phi$ for all $j \neq j$ this implies $\sum_{1}^{\infty} P(A_i) = 1$ is called probability measure.

(iii) Suppose we define the space Ω of elements ω, a Borel or σ-field of sets in Ω, and a probability measure P on $\mathcal{B}$. The triplet $(\Omega, \mathcal{B}, P)$ is called a probability space.

(iv) A real valued point function $X(\cdot)$ defined on the space $(\Omega, \mathcal{B}, P)$ is called a random variable of the set $\{\omega : X(\omega) < x\} \in \mathcal{B}$ for every x in $\mathbb{R}$.

2. Since $\sum_{1}^{\infty} E(X_n - C)^2 < \infty$, then the infinite sum

$$\sum_{1}^{\infty}(X_n - C)^2 \tag{i}$$

must converge except for a set of sequences of measure zero in $\mathbb{R}^{\infty}$. Suppose that (i) diverges on a set of measure P. Let A_N be the set in R^{∞} such that $\sum_{1}^{N}(X_i - C)^2 > \lambda$.

The sequence $\{A_n\}$ is non-decreasing and hence by the consequence of the axioms of probability, that is,

$$P(\lim A_i) = \lim_{i\to\infty} P(A_i),$$

we have

$$\lim P(A_n) = P(\lim A_N) \geq P$$

Consequently,

$$\sum_{1}^{N} E(X_i - C)^2 = e[\sum (X_i - C)^2] \geq \lambda P$$

which contradicts the assumption, $\sum_{1}^{\infty} E(X_i - C)^2 < \infty$, if λ is chosen sufficiently large.

Hence, $\sum_{1}^{\infty}(X_n - C)^2$ converges) = 1, but if an infinite series converges the n-th term $\to 0$ as $\to \infty$. Hence

$$P(|X_n - C| \to 0 \text{ as } n \to \infty) \geq P(\sum_{1}^{\infty}(X_n - C)^2 \text{ converges}) = 1$$

3. see Lemma 4.1

4. A is the set defined as

$$A = \bigcap_{r=1}^{\infty} \bigcup_{n=r}^{\infty} A_n$$

it follows that $A \subset \bigcup_{n=r}^{\infty} A_n$, for each r and

$$P(A) \leq P\left(\bigcup_{r}^{\infty} A_n\right) \leq \sum_{r}^{\infty} P(A_n) < \xi,$$

for $r \to \infty$ since $\Sigma P(A_n)$ converges, that $P(A) = 0$

5. A^c is defined as

$$\begin{aligned} A^C &= \bigcup_{r=1}^{\infty} \bigcap_{n=r}^{\infty} A_n^c \\ 1 - P(A) &= P(A^c) = P\left(\bigcup_{r=1}^{\infty} \bigcap_{n=r}^{\infty} A_n^c\right) \leq \sum_{r=1}^{\infty} P\left(\bigcap_{n=r}^{\infty} A_n^c\right) \\ &= \sum_{r=1}^{\infty} \prod_{n=r}^{\infty} [(1 - P(A_n))], \end{aligned}$$

using independence of events, since $\Sigma P(A_n) = \infty$, the infinite product diverges to zero for each r., Hence $P(A) = 1$.

Chapter 9

Law of Large Numbers

9.0 Introduction to Large Number Concepts

In practical terms, estimates are usually made of an unknown parameter by considering the average of the number of replicated observations of the quantity, each of which may be in error. The properties of the estimate is of interest, as an initial step, we observe its behaviours as the number of repeated observations increases and the nature of its convergence.

Let $\{Y_n\}_{n=1}^{\infty}$ be a sequence of observations and $\bar{Y}_n$ the average of the first n observations. Then we ask the question on what conditions can any of these convergence forms exist:

$$\bar{Y}_n \longrightarrow \xi \tag{9.1}$$

and also

$$\bar{Y}_n - \xi_n \longrightarrow 0 \tag{9.2}$$

where $\{\xi_n\}$ is a sequence of constants measured by the sequence of observations $\{Y_n\}_{n=1}^{\infty}$.

We say that the large law of numbers holds if the convergence is either in form of (9.1) or (9.2). When the convergence is in probability, we say that the weak law of large numbers (WLLN) holds, and when the convergence is almost surely, we say the strong law of large numbers (SLLN) holds. In the next subsections we shall consider some important theorems on both weak law and strong law of large numbers.

9.1 Weak Law of Large Numbers

Let us study some situations when the weak law of large number hold by Theorems and Lemmas:

Theorem 9.1 (Chebychev's Theorem)

Let $E)Y_i) = \mu_i$, $Var(Y_i) = \sigma_i^2$ and $Cov(Y_iY_j) = 0$; $i \neq j$. Then

$$\lim \frac{1}{n^2}\sum \sigma_i^2 = 0 \Rightarrow \bar{Y}_n - \bar{\mu}_n \xrightarrow{p} 0$$

Proof:

Define $\bar{\mu}_n = \frac{1}{n}\sum_1^n \mu_i$ and by Chebychev's inequality, we have

$$\begin{aligned} P(|\bar{Y}_n - \bar{\mu}_n| > \xi) &\leq \frac{1}{\xi^2} Var(\bar{Y}_n) \\ &= \frac{1}{\xi^2 n^2}(\sum \sigma_i^2) \\ \lim P(|\bar{P}_n - \bar{\mu}_n| > \xi) &\leq \lim_{n\to\infty} \frac{1}{\xi^2 n^2|} \sum_1^n \sigma_i^2 \to 0 \end{aligned}$$

this implies that $\bar{Y}_n - \bar{\mu}_n \xrightarrow{p} 0$ #

Theorem 9.2 (Khintchine's Theorem)

Let $\{Y_i\}$, $i = 1, 2, \ldots$ be independent and identically distributed $(i.i.d)$ and $E(Y_i)$ exists, then

$$E(Y_i) = \mu < \sigma \Rightarrow \bar{Y} \xrightarrow{p} \mu$$

Proof

Define a pair of new random variables for $i = 1, 2, \ldots, n$ and fror fixed θ, we have

$$w_i = Y_i,\ V_i = 0, \text{ if } |Y_i| < \theta n$$

$$w_i = 0,\ V_i = Y_i \text{ if } |Y_i| \geq \theta n$$

so that $Y_i = W_i + V_i$. Let $E(W_i) = \mu_n$ for $i = 1, \ldots, n$ since $(Y_i) = \mu$ then

$$|\mu_n - \mu| < \xi \qquad (9.3)$$

For any given ξ, if n is chosen sufficiently large. Now

$$V(w_i) = \int_{-\theta n}^{\theta n} y^0 dF(y) - \mu_n^2 \leq \int_{-\theta n}^{\theta n} |y| dF(y) \leq b\theta n$$

where $b = E(|Y|)$ exists. Using theorem (2.1), we have

$$P\left(\frac{1}{n}\sum^{n} w_i - \mu_n| \geq \xi\right) \leq \frac{b\theta}{\xi^2}$$

or by equation (2.3), we have

$$P\left(\left|\frac{1}{n}\sum_{1}^{p} w_i - \mu\right| \geq \eta\xi\right) \leq \frac{b\theta}{\xi^2}$$

since

$$P(V_i \neq 0) = \int_{|y|\geq\theta n} dF(y) \leq \frac{1}{\theta n}\int_{|y|\geq\theta n} |y| dF(y) \leq \frac{\theta}{n}$$

by chosen n sufficiently large, we have

$$P\left(\sum_{1}^{A} V_i \neq 0\right) \leq \sum_{1}^{n} P(V_i \neq 0) \leq \theta$$

Let us consider

$$\begin{aligned} P(|\frac{1}{n}\sum Y_i - \mu| \geq 4\xi] &= P[|(\frac{1}{n}\sum w_i - \mu) + \frac{1}{n}\sum V_i| \geq 4\xi] \\ &\leq P[|\frac{1}{n}\sum w_i - \mu| \geq 2\xi + P[|\frac{1}{n}\sum V_i| \geq 2\xi] \\ &\leq \frac{b\theta}{\xi^2} + P[sumV_i \neq 0] \leq \frac{b\theta}{\xi^2} + \theta \end{aligned}$$

The quantity $\frac{b\theta}{\xi} + \theta \to 0$, by correct choice of θ.

$$\therefore \bar{Y}_n \xrightarrow{p} \mu \qquad \#$$

Theorem 9.3 (Hajeck Renyi Inequality)

Let $\{Y_n\}_{n=1}^{\infty}$ be independent random variables such that $E(Y_n) = 0$, $V(Y_n) = \sigma_n^2 < \infty$. If $c_1, c_2, \ldots$ be a non-increasing sequence of positive constant, then for any positive integers m and n such that $m < n$, and any arbitrary $\xi > 0$, we have

$$P\left[\max_{m \le k \le n} C_k |Y_1 + Y_2 + \cdots + Y_k| > \xi\right] \le \frac{1}{\xi^2}\left[C_m^2 \sum_1^m \sigma_i^2 + \sum_{m+1}^{n} C_k^2 \sigma_k^2\right]$$

Proof

Set

$$Y = \sum_{k=m}^{n-1} (C_k^2 - C_{k+1}^2) S_k^2 + C_n^2 S_n^2$$

where $S_k = \sum_{i=1}^{k} Y_i$. Then

$$\begin{aligned} E(Y) &= \sum_{k=m}^{n-1} (C_k^2 - C_{k+1}^2) E(S_k^2) + C_n^2 E(S_n^2) \\ &= \sum_{k=m}^{n-1} \left\{ (C_k^2 - C_{k+1}^2) \sum_1^k \sigma_i^2 \right\} + C_n^2 \sum_1^k \sigma_i^2 \\ &= C_m^2 \sum_1^m \sigma_i^2 + C_{m+1}^2 \left[\sum_1^{m+1} \sigma_i^2 - \sum_1^n \sigma_i^2\right] + \cdots + C_n^2 \left[\sum \sigma_i^2 - \sum_1^{n-1} \sigma_i^2\right] \\ &= C_m^2 \sum_1^m \sigma_i^2 + \sum_{m+1}^{n} C_k^2 \sigma_k^2 \qquad (9.4) \end{aligned}$$

Let $\epsilon_i; i = m, m+1, \cdots, n$ be the event $C_j|S_j| < \xi$ for $m \le j < i$ and $C_i|S_i| \ge \xi$. Then

$$P\left[\max_{m \le k \le n} C_k |S_k| \ge \xi\right] = \sum_m^n P(\epsilon_i) \qquad (9.5)$$

and because ϵ_i are mutually exclusive. Let ϵ_0 denote the event $C_j|S_j| < \xi$; $m \le j \le n$ in this sense we shall have from (2.4)

$$E(Y) \ge \sum_{i=m}^{n} E(Y|\epsilon_i) P(\epsilon_i) \text{ if } F(Y|\epsilon_0) P(\epsilon_0) > 0 \qquad (9.6)$$

We shall consider the case when $k \ge i$ and $j > i$ as follows:

Case A

Consider $k \ge i$, we have

$$\begin{aligned} E(S_{kj}^2|\epsilon_i) &= E[\{S_i^2 + (Y_{i+1} + \cdots + Y_k)^2 + 2S_i(Y_{i+1} + \cdots + Y_k)\}|\epsilon_i] \\ &\ge E(S_i^2/\epsilon_i) + 2E[S_i(Y_{i+1} + \cdots + Y_k)/\epsilon_i] \end{aligned}$$

Case B

Consider $j > i$, we have $E(S_iY_j/\epsilon_i) = 0$, then $E(S_k^2/\epsilon_i) \ge E(S_i^2/\epsilon_i)$, given ϵ_i (event $C_j|S_{jk}| \ge \xi$) then

$$|S_i| \ge \xi|C_i$$

$$\therefore \; E(S_i^2|\xi_i) \ge \xi^2|C_i^2$$

Now,

$$\begin{aligned} E(Y/\epsilon_i) &= \sum_{k=m}^{n-1} E(S_k^2/\epsilon_i)(C_k^2 - C_{k+1}^2) + C_n^2 E(S_n^2/\epsilon_i) \\ &\ge \sum_{k=1}^{n-1} E(S_k^2|\epsilon_i)(C_k^2 - C_{k+1}^2) + C_n^2 E(S_n^2|\epsilon_i) \\ &\ge \frac{\xi^2}{C_i^2}\sum_{k=i}^{n-1}[(C_k^2 - C_{k+1}^2) + C_n^2] \\ &\ge \frac{\xi^2}{C_i^2}(C_i^2 - C_{i+1}^2 + C_{i+1}^2 + C_{i+2}^2 - C_{i+3}^2 + \cdots + C_n^2 + C_n^2) \\ &\ge \frac{\xi^2}{C_i^2} \cdot C_i^2 = \xi^2 \end{aligned} \qquad (9.7)$$

By using (9.6) in (9.7) gives

$$E(Y) \ge \xi^2 \sum_m^n P(\epsilon_i)$$

and by (9.5), we have

$$\begin{aligned} \sum P(\epsilon_i) &= P\left(\max_{m \le k \le n} C_k|S_k| \ge \xi\right) \\ &\le \frac{1}{\xi^2}E(Y) = \frac{1}{\xi^2}\left[C_m^2\sum_1^m \sigma_i^2 + \sum_{m+1}^n C_k^2\sigma_k^2\right] \end{aligned}$$

9.2 Strong Law of Large Numbers

When convergence is almost surely we say we have a strong law of large numbers being satisfied. Let us consider the following Theorems and Lemma to study the strong law of large numbers.

Theorem 9.4 (Kolmogorov I)

Let $\{Y_i\}; i = 1, 2, \ldots$ be a sequence of independent random variables such that $E(y_i) = \mu_i$, and $Var(Y_i) = \sigma_i^2$. Then

$$\sum_{i=1}^{\infty} \sigma_i^2 \Big/ C_i^2 < \infty \Rightarrow \bar{Y}_n - \bar{\mu}_n \xrightarrow{a.s} 0$$

That is, the sequence $Y_1, Y_2, \ldots$ obeys the strong law of large numbers.

Proof

Let us consider the random variables $X_i = Y_i - \mu_i$ and apply the Hajek-Rexji's inequality, we have

$$P\left\{\max_{m \le i \le n} C_i |X_1 + \cdots + X_i| \ge \xi\right\} \le \frac{1}{\xi^2}\left(C_m^2 \sum_1^m \sigma_i^2 + \sum_{m+1}^{n} C_i^2 \sigma_i^2\right)$$

Choosing $C_i = 1/i$, we have

$$P\left\{\max_{m \le i \le n} |\bar{X}_i| \ge \xi\right\} \le \frac{1}{\xi^2}\left(\frac{1}{m^2}\sum_1^m \sigma_i^2 + \sum_{m+1}^{n} \sigma_i^2 / i^2\right)$$

By letting $n \to \infty$, we see that

$$P\left\{\max_{m \le i} |\bar{X}_i| \ge \xi\right\} \le \frac{1}{\xi^2}\left(\frac{1}{m^2}\sum_1^m \sigma_i^2 + \sum_{m+1}^{\infty} \sigma_i^2 / i^2\right)$$

(since $\sum \sigma_i^2 / i^2$ converges) it follows that

$$\lim_{m \to \infty} P\left\{\max_{m \le i} |\bar{X}_i \ge \xi\right\} = 0$$

It implies

$$P\left\{\lim_{m \to \infty} \bar{X}_m = 0\right\} = 1$$

#

Theorem 9.5 (Kolmogorov II)

Let $y_1, y_2, \ldots$ be a sequence of independent, identically distributed variables. Then a necessary and sufficient condition that $\bar{Y}_n \xrightarrow{a.s} \mu$ is that $E(Y_i)$ exists and is equal to μ.

Proof

(A) **Necessary Condition**

Let E_n be the event that $|Y_n| \geq n$. Then

$$\frac{Y_n}{n} = \bar{Y}_n - \frac{(n-1)(\bar{Y}_{n-1}}{n} \xrightarrow{a.s} 0 \tag{9.8}$$

since $\bar{Y}_n \xrightarrow{a.s} \mu$. The result (2.8) implies that the probability of infinitely many events ϵ_n occurring is zero. Also the independence of Y_n implies the independence of ϵ_n, and by the following lemma credited to Borel Cantelli:

If $\{A_n\}$ is a sequence in $\mathcal{F}$, then

(a) $\sum P(A_n)" < \infty \Rightarrow P(A_n i.o) = 0$

(b) $\sum P(A_n) = \infty$ and A_i are independent then $\Rightarrow \quad P(A_n i.o) = 1$

By this lemma we see that

$$\sum P(|Y_n| \geq n) = \sum_{n=1}^{\infty} P(\epsilon_n) < \infty \tag{9.9}$$

Let $P_j = P(|Y| \geq j)$, then

$$\begin{aligned} E(|Y|) &\leq |(1 - P_1) + 2(P_1 - P_2) + \cdots \\ &= 1 + P_1 + P_2 + \cdots = 1 + \sum_{n=1}^{\infty} P(\epsilon_n) \end{aligned} \tag{9.10}$$

From (9.9), the last expression in (2.10) is less than ∞. Hence $E(|Y|)$ exists and from the sufficiency condition it follows that $E(Y) = \mu$.

(B) **Sufficiency Condition**

Consider the sequence of truncated variables

$$Y_n^* = \begin{cases} Y_n & \text{for } |Y_n| < n \\ 0 & \text{for } |Y_n| \geq n \end{cases}$$

We then obtain

$$V(Y_n^*) \leq E(Y_n^*)^2 = \int_{-n}^{+n} y^2 dF(y) \leq \sum_{k=0}^{n-1}(k+1)^2 P(k \leq |Y| < k+1)$$

and

$$\begin{aligned}
\sum_{n=1}^{\infty} \frac{V(Y_n^*)}{n^2} &\leq \sum_{n=1}^{\infty}\sum_{k=0}^{n-1} \frac{(k+1)^2}{n^2} P(k \leq |Y| < k+1) \\
&\leq \sum_{k=1}^{\infty} P(k-1 \leq |Y| <, k) K^2 \sum_{n=k}^{\infty} \frac{1}{n^2} \\
&\leq 2\sum_{k=1}^{\infty} KP(k-1 \leq |Y| < K) \\
&\leq 2[1 + E(|Y|)] < \infty
\end{aligned}$$

By Kolmogorov theorem 1, the sequence Y_n^* obeys the law of large numbers, that is,

$$Y_n^* - \frac{1}{n}\sum_{1}^{n} E(Y_i^*) \xrightarrow{\text{a.s}} 0$$

Now as $n \to \infty$, $E(Y_n^*) \to E(Y_n) = \mu$, hence $\frac{1}{n}\sum E(Y_i^*) \to \mu$ as $n \to \infty$

$$\therefore \; Y_n^* \xrightarrow{\text{a.s}} \mu$$

Also, we have to establish that Y_n^* and Y_n are equivalent sequences, that is,

$$P(Y_n \neq Y_n^*;\, n \geq N) \to 0 \text{ as } N \to \infty;$$

which implies that Y_n obeys the S.L.L.N. If Y_n^* does and that the limits are the same.

Consider

$$\begin{aligned}
P(Y_n \neq Y_n^*, n \geq N) &\leq \sum_{n \geq N} P(Y_n \neq Y_n^*) = \sum_{n=N}^{\infty} P(|Y_n| \geq n) \\
&\geq \sum_{n=N}^{\infty} (n - N + 1)P(n \geq |Y_n| < (n+1)) \\
&\leq \sum_{n=N}^{\infty} nP(n \leq N < (n+1)),
\end{aligned}$$

since all Y_n have the same distribution function.

$$\therefore \quad P(Y_n \neq Y_n^*, n \geq N) \leq \int_{|Y| \geq N} |Y| dF(Y) \to 0$$

as $N \to \infty$. #

Problems 9.2

1. Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of observations and for which $E(Y_i) = \mu_i$, $Var(Y_i) = \sigma_i^2$ and $Cov(Y_i, y_j) = \sigma_{ij}, i \neq j$. Define $\bar{Y}_n = \frac{1}{n}\sum Y_i$ and $\bar{\mu}_n = \frac{1}{n}\sum_1^n \mu_i$. Show that

$$\lim_{n \to \infty} \frac{1}{n^2} \sum_{i=1}^{n} \sigma_i^2 = 0,$$

implies that $\bar{Y}_n$ converges to $\bar{\mu}_n$.

2. Does the strong law of large numbers hold for the following sequences?

   (i) $P(Y_n = \pm 2^n) = \frac{1}{2}$

   (ii) $P(Y_n = \mp n) = \frac{1}{2}\sqrt{n}, P(Y_n = 0) = 1 - \frac{1}{\sqrt{n}]}$

   (iii) $P(Y_n = \mp 2^n) = \frac{1}{2}\rho^{n+1}, P(Y_n = 0) = 1 - \frac{1}{2^{2n}}$

3. Show that if $\{Y_n\}_{n=1}^{\infty}$ is a sequence of independent random variables with $E(Y_i) = \mu_i$, $V(Y_i) = \sigma_i^2; i = 1, 2, \ldots$ then

$$\sum \sigma_i^2 / i < \infty,$$

it implies that the sequence of random variables obeys the strong law of large numbers.

4. If Y_k has the binomial distribution with parameters k and p, does the sequence $Y_1, Y_2, Y_3, \ldots$ obey the strong law of large numbers?

5. Compare the assumptions and results of Khinchine's theorem (WLLN) and Kolmogorov II theorem (SLLN).

Solution 9.2

1. Suppose we have

$$\text{Prob}(|\bar{Y}_n - \bar{\mu}_n| > \xi) \leq \frac{V(\bar{Y}_n)}{\xi_n^2} = \frac{1}{\xi_n^2}\sum_1^n \sigma^2 \to 0$$

as $n \to \infty$. Hence

$$\lim_{n\to\infty} \text{Prob}\{(|\bar{Y}_n - \bar{\mu}_n| > \xi)\} = 0$$

and so $\bar{Y}_n \to \bar{\mu}_n$.

2(i) $E(Y_n) = 0,\ V(Y_n) = \frac{1}{2}(2^{2n}) + \frac{1}{2}(2^{2n}) = 2^{2n}$

Now,

$$\sum_{n=1}^{\infty} V(Y_n) = \sum_{n=1}^{\infty} 2^{2n} = \infty$$

Hence, Y_n does not obey the S.L.L.N.

(ii) $E(Y_n) = 0,\ V(Y_n) = \frac{1}{2\sqrt{n}}(n^2) + \frac{1}{2\sqrt{n}}(n^2) = \frac{n^2}{\sqrt{n}}$

Also

$$\sum_{n=1}^{\infty} Var(Y_n) = \sum_{n=1}^{\infty} n^{3/2} = \infty$$

Y_n does not obey the S.L.L.N.

(iii) $E(Y_n) = 0,\ Var(Y_n) = \frac{1}{2}\frac{2^{2n}}{2^{2n}} + \frac{1}{2}\frac{2^{2n}}{2^{2n}} = 1$

which is bounded, since the variances are uniformly bounded, $\sum \frac{Var(Y_n)}{n^2} < \infty$ and hence $Y_n \xrightarrow{\text{a.s}} E(Y_m) = 0$. Hence the sequence Y_n obeys the S.L.L.N.

3. Let $\{Y_n\}_{n=1}^{\infty}$ be a sequence of independent random variables, such that $E(Y_n) = 0$, $Var(Y_n) = \sigma_n^2 < \infty$. If $C_1, C_2, \ldots$ is a non-increasing sequence of positive constant, then for any positive intefgers m and n, with $m < n$ and arbitrary $\xi > 0$ we have

$$P\left[\max_{m\leq k\leq n} C_k\left|\sum_{i=1}^{k} Y_i\right| > \xi\right] \leq \frac{1}{\xi^2}\left[C_m^2\sum_{i=1}^{m}\sigma_i^2 + \sum_{m+1}^{n} C_k^2\sigma_k^2\right] \qquad (i)$$

Set $Y_i = X_i - \mu_i$ in (i), we have

$$P\left[\max_{m\le i\le n} C_k \left|\sum_1^k Y_i\right|\right] > \xi \le \frac{1}{\xi^2}\left[C_m^2\sum_{i=1}^m \sigma_i^2 + \sum_{m+1}^n C_k^2\sigma_k^2\right]$$

Select $C_i = 1/i$ as a sequence of non-increasing constants, therefore

$$P\left[\max_{m\le i\le n} |\bar{Y}| > \xi\right] \le \frac{1}{\xi^2}\left[\frac{1}{m^2}\sum_1^m \sigma_i^2 + \sum_{m+1}^\infty \sigma_k^2/m^2\right] \qquad (ii)$$

If m is as large as possible $\sum \sigma^2/m^2 \to 0$, and so for sufficiently large m the expression on the r.h.s. of (ii) tends to zero. Thus makes the probability on the L.H.S. to converge in almost surely to zero.

4. Let $\{Y_k\}$ be the sequence of I.I.D. binomial random variables, a sufficient condition that $\bar{Y} \xrightarrow{a.s} \mu$ is that $E(Y_i)$ exist and is equal to $\mu = KP$. Now assume $E(Y_i) < \infty$ and $E(Y_i) = KP$, set

$$\begin{aligned} Y_k^* &= Y_k \text{ if } |Y_k| < K \\ &= 0, \text{ otherwise} \end{aligned}$$

But

$$\begin{aligned}
Var(Y_k^*) &\le E[(Y_k^*)^2] = \int_{-k}^{k} y^2 dF \\
&= \sum_{-k}^{-1}\int_n^{n+1} y^2 dF + \sum_0^{k-1}\int_n^{n+1} y^2 dF \\
&\le \sum_0^{k-1}(n+1)^2 P(n \le |Y| < n+1) \\
\sum \frac{V(Y_n^*)}{K^2} &\le \sum_{k=1}^\infty\sum_{n=0}^{k-1}\frac{(n+1)^2}{k^2}P(n \le |Y| \le n+1 \\
&= \sum_{n=0}^\infty \{(n+1)^2 P(n \le |Y| < n+1)\sum_{k=n+1}^\infty \frac{1}{k^2} \\
&= \sum_1^\infty n^2 P(n-1 \le |Y| < n)\sum_{k=n}^\infty \frac{1}{k^2} \\
&\le 2\sum_1^\infty nP(n-1 \le |Y| < n) \\
&\le 2(1 + E(|Y|))
\end{aligned}$$

That is $\sum V(Y_n^*)/K^2 < \infty$ hence the sequence obeys the S.L.L.N.

5. Both assume i.i.d. random variables and existence of the mean. The result in Khinchine's theorem is that the assumptions are necessary for $\bar{Y}_n \xrightarrow{p} \mu$ but not the reverse. That is, consequence in probability does not necessarily imply finite expectation. However, the result in Kolmogorov II theorem is that the assumptions are necessary and sufficient for $\bar{Y}_n \xrightarrow{\text{a.s}} \mu$.

Chapter 10

Generating Functions and Inversion Theorem

10.1 Introduction

The first moment about an arbitrary point α by the Stieltjes integral is defined as

$$\mu_1' = \int_{-\infty}^{\infty} (y - \alpha) dF \tag{10.1}$$

and the second moment

$$\mu_2' = \int_{-\infty}^{\infty} (y - \alpha)^2 dF \tag{10.2}$$

The generalization of these equations defined by a series of coefficients $\mu_k'; k = 1, 2, \ldots$ by the relation

$$\mu_k' = \int_{\infty}^{\infty} (y - \alpha)^k dF \tag{10.3}$$

μ_k' is called the moment of order k about the point α. When α is the mean μ_1', we write the moment without the prime as

$$\mu_k = \int_{-\infty}^{\infty} (y - \mu_1')^k dF \tag{10.4}$$

called the central moment or moment about the mean. In the specific, when $\mu_1 = 0$, we may define a moment of order zero as

$$\mu_0' = \mu_0 = \int_{-\infty}^{\infty} dF = 1$$

Because of some theoretical reasons, μ'_k exists only when

$$\gamma_k = E(|Y-\alpha|^k) = \int |Y-\alpha|^k dF$$

exists, this is true when the integral defining μ'_k is of Lebesque-Stieltjes type.

The relationships between the central moments and moment about an arbitrary origin may be verified as follows:

If α and θ are two variate-values, let $\theta - \alpha = \tau$ and denote the moments about α and θ by $\mu'(\alpha)$ and $\mu'(\theta)$ respectively. Then we have, by binomial theorem

$$\begin{aligned}(y-\alpha)^k &= (y-\theta+\theta-\alpha)^k \\ &= (y-\theta+\tau)^k \\ &= \sum_{j=0}^{k}\binom{k}{j}(y-\theta)^{k-j}\tau^j\end{aligned}$$

Hence

$$\begin{aligned}\mu'_k(\theta) &= \int_{-\infty}^{\infty}(y-\alpha)^k dF \\ &= \int_{-\infty}^{\infty}\sum_{j=0}^{k}\binom{k}{j}(y-\theta)^{k-j}\tau^j dF \\ &= \sum_{j=0}^{k}\binom{k}{j}\tau^j\int_{-\infty}^{\infty}(y-\theta)^{k-j}\tau^j dF \\ &= \sum_{j=0}^{k}\binom{k}{j}\mu'_{k-j}(\theta)\tau^j \\ &= \{\mu'(\theta)+\tau\}^k \qquad (10.5)\end{aligned}$$

This equation is of particular importance if one of the values α and θ is the mean of the distribution. In this case we shall have

$$\mu'_k = \sum_{j=0}^{k}\binom{k}{j}\mu_{k-j}\mu_1'^{(j)} \qquad (10.6)$$

$$\mu_k = \sum_{j=0}^{k}\binom{k}{j}\mu'_{k-j}(-\mu'_1)^j \qquad (10.7)$$

In particular,

$$\mu_2' = \mu_2 + \mu_1'^{(2)} \tag{10.8}$$

$$\mu_3' = \mu_3 + 3\mu_1'\mu_2 + \mu_1'^{(3)} \tag{10.9}$$

$$\mu_4' = \mu_4 + 4\mu_1'\mu_3 + 6\mu_1'^{(2)}\mu_2 + \mu_1'^{(4)} \tag{3.10}$$

Let us discuss a few results concerning the mean and variance of a given distribution in terms of moments;

Result (A): The second moment $\mu_2'(\alpha)$ is a min when takes about the mean μ. That is,

$$\begin{aligned} E(Y-\alpha)^2 &= E(Y-\mu+\mu-\alpha)^2 \\ &= E(Y-\mu)^2 + (\mu-\alpha)^2 + 2(\mu-\alpha)E(Y-\mu) \\ \therefore\ E(Y-\mu)^2 + (\mu-\alpha)^2 &\geq E(Y-\alpha)^2. \end{aligned}$$

Result (B): The Chebychev's inequality discussed in Chapter two is an important inequality as it is independent of the exact nature of the distribution of the variable, say, Y. By definition, let

$$\begin{aligned} \sigma^2 &= \int_R (Y-\mu)^2 dF \\ &= \int_{-\infty}^{Y-\xi\sigma} (Y-\mu)^2 dF + \int_{\mu-\xi\sigma}^{\mu+\xi\sigma} (Y-\mu)^2 dF \\ &+ \int_{\mu+\xi\sigma}^{\infty} (Y-\mu)^2 dF \end{aligned}$$

By dropping the middle term and replacing $(Y-\mu)^2$ by the smallest value in the first and third terms, we have

$$\begin{aligned} \sigma^2 &\geq \xi^2\sigma^2 \int_{-\infty}^{\mu-\xi\sigma} dF + \xi^2\sigma^2 \int_{\mu+\xi\sigma}^{\infty} dF \\ &\geq \xi^2\sigma^2 \rho[|Y-\mu| \geq \xi\sigma] \end{aligned}$$

Result (C): Suppose Y_1 and Y_2 are two random variables with means μ_1, μ_2 variables σ_1^2, σ_2^2 and distribution functions F_1, F_2, then

$$F_1(y+\mu_1) - F_1(-y+\mu_1) \geq F_2(y+\mu_2) - F_2(-y+\mu_2) \tag{10.11}$$

Because of some theoretical reasons, μ'_k exists only when

$$\gamma_k = E(|Y-\alpha|^k) = \int |Y-\alpha|^k dF$$

exists, this is true when the integral defining μ'_k is of Lebesque-Stieltjes type.

The relationships between the central moments and moment about an arbitrary origin may be verified as follows:

If α and θ are two variate-values, let $\theta - \alpha = \tau$ and denote the moments about α and θ by $\mu'(\alpha)$ and $\mu'(\theta)$ respectively. Then we have, by binomial theorem

$$\begin{aligned}(y-\alpha)^k &= (y-\theta+\theta-\alpha)^k \\ &= (y-\theta+\tau)^k \\ &= \sum_{j=0}^{k}\binom{k}{j}(y-\theta)^{k-j}\tau^j\end{aligned}$$

Hence

$$\begin{aligned}\mu'_k(\theta) &= \int_{-\infty}^{\infty}(y-\alpha)^k dF \\ &= \int_{-\infty}^{\infty}\sum_{j=0}^{k}\binom{k}{j}(y-\theta)^{k-j}\tau^j dF \\ &= \sum_{j=0}^{k}\binom{k}{j}\tau^j\int_{-\infty}^{\infty}(y-\theta)^{k-j}\tau^j dF \\ &= \sum_{j=0}^{k}\binom{k}{j}\mu'_{k-j}(\theta)\tau^j \\ &= \{\mu'(\theta)+\tau\}^k \qquad (10.5)\end{aligned}$$

This equation is of particular importance if one of the values α and θ is the mean of the distribution. In this case we shall have

$$\mu'_k = \sum_{j=0}^{k}\binom{k}{j}\mu_{k-j}\mu_1^{(j)} \qquad (10.6)$$

$$\mu_k = \sum_{j=0}^{k}\binom{k}{j}\mu'_{k-j}(-\mu'_1)^j \qquad (10.7)$$

In particular,

$$\mu_2' = \mu_2 + \mu_1'^{(2)} \tag{10.8}$$

$$\mu_3' = \mu_3 + 3\mu_1'\mu_2 + \mu_1'^{(3)} \tag{10.9}$$

$$\mu_4' = \mu_4 + 4\mu_1'\mu_3 + 6\mu_1'^{(2)}\mu_2 + \mu_1'^{(4)} \tag{3.10}$$

Let us discuss a few results concerning the mean and variance of a given distribution in terms of moments;

Result (A): The second moment $\mu_2'(\alpha)$ is a min when takes about the mean μ. That is,

$$\begin{aligned} E(Y-\alpha)^2 &= E(Y-\mu+\mu-\alpha)^2 \\ &= E(Y-\mu)^2 + (\mu-\alpha)^2 + 2(\mu-\alpha)E(Y-\mu) \end{aligned}$$

$$\therefore \; E(Y-\mu)^2 + (\mu-\alpha)^2 \geq E(Y-\alpha)^2.$$

Result (B): The Chebychev's inequality discussed in Chapter two is an important inequality as it is independent of the exact nature of the distribution of the variable, say, Y. By definition, let

$$\begin{aligned} \sigma^2 &= \int_R (Y-\mu)^2 dF \\ &= \int_{-\infty}^{Y-\xi\sigma} (Y-\mu)^2 dF + \int_{\mu-\xi\sigma}^{\mu+\xi\sigma} (Y-\mu)^2 dF \\ &\quad + \int_{\mu+\xi\sigma}^{\infty} (Y-\mu)^2 dF \end{aligned}$$

By dropping the middle term and replacing $(Y-\mu)^2$ by the smallest value in the first and third terms, we have

$$\begin{aligned} \sigma^2 &\geq \xi^2\sigma^2 \int_{-\infty}^{\mu-\xi\sigma} dF + \xi^2\sigma^2 \int_{\mu+\xi\sigma}^{\infty} dF \\ &\geq \xi^2\sigma^2 \rho[|Y-\mu| \geq \xi\sigma] \end{aligned}$$

Result (C): Suppose Y_1 and Y_2 are two random variables with means μ_1, μ_2 variables σ_1^2, σ_2^2 and distribution functions F_1, F_2, then

$$F_1(y+\mu_1) - F_1(-y+\mu_1) \geq F_2(y+\mu_2) - F_2(-y+\mu_2) \tag{10.11}$$

for each y implies that $\sigma_1^2 \leq \sigma_2^2$. To prove this, let G_1 and G_2 be the distribution functions of $|y_1 - \mu|$ and $|y_2 - \mu_2|$ respectively. Then integrating by parts, we have

$$\begin{aligned}\sigma_1^2 - \sigma_2^2 &= \lim_{T\to\infty}\int_0^T y^2 d(G_1 - G_2)\\ &= \lim_{T\to\infty} T^2[G_1(T) - G_2(T)]\\ &\quad -2\lim_{T\to\infty}\int_0^T y(G_1 - G_2)dy\\ &= 0 - 2\int_0^\infty y(G_1 - G_2)dy \leq 0\end{aligned}$$

The condition in (10.11) implies that $G_1 \geq G_2$, the converse proposition is not true, however, if $\sigma_1^2 \leq, \sigma_2^2$, then it implies that for at least one value of y the inequality (3.11) is true but not necessarily true for all values of y.

10.2 Moment Generating Functions (M.G.F)

The results (A) and (C) in subsection 3.1 show that for some cases we can derive from the distribution function a function $M(t)$ which, when expanded in powers of t, will yield the moments of the distribution as the coefficients of those powers. This function is accordingly be referred to as a moment-generating function (m.g.f).

The m.g.f of the distribution of a random variable Y is formally defined as

$$\phi(t)E(e^{tY}) = \int_{-\infty}^{\infty} e^{ty}dF(y) = \sum e^{ty_i}f(y_i)$$

But for many frequency functions the integral $\int_{-\infty}^{\infty} e^{ty}dF$ or the sum $\{\sum e^{ty_i}f(y_i)\}$ does not exist for some or all real values of t. When $\phi(t)$ exists, this expectation depends on the choice of t, and so defines a function of t.

Depending on the distribution; $\phi(t) \neq 1$ at $t = 0$ and for other values of t, $\phi(t)$ may and may not exists. To generate moments, suppose the exponential function in the integrand of $\phi(t)$ is replaced by its power series expansion, then

$$\phi(t) = \int_{-\infty}^{\infty}\sum_0^\infty \frac{(ty)^k}{k!}dF(y) = \sum_0^\infty \frac{t^k}{k!}\left(\int_{-\infty}^{\infty} y^k dF(y)\right)$$

$$= \sum_{0}^{\infty} E(Y^k)t^k/k! \qquad (10.12)$$

under commutation functions, the *k-th* moment of a distribution is simply be coefficiencies of $t^k/k!$ in the power series expansion of the m.g.f. The coefficient $t^t/k!$ is the Maclaurin power series expansion, that is, the $k-th$ derivative at 0 is

$$E(X^k) = \phi^{(k)}(0),$$

it is equivalent to

$$\frac{d^k}{dt^k}E(e^{ty}) = E(Y^k e^{tY})\bigg|_{t=0}$$

gives the required coefficients.

Example 10.1

Let us consider a simple random variable Y having just two possible values, 1 with probability P and 0 with probability $1 - P$. The moment generating function is

$$\begin{aligned}
\phi(t) &= E(e^{tY}) = e^t \cdot P + e^0 \cdot (1 - P) \\
&= Pe^t + (1 - P) \\
&= P(1 + \frac{t}{1!} + \frac{t^2}{2!} + \cdots) + (1 - P) \\
&= 1 + P(t/1!) + P(t^2//2!) + \cdots
\end{aligned}$$

Clearly the coefficients of $t^k/k!$ are equal to p (for $k = 1, 2, \ldots$) and so $E(Y^k) = P$, except for $k = 0$.

The moment generating function of two independent random variables whose term is $(Y + Z)$, is a particular simple combination of the moment generating functions of the summands, namely, their product:

$$\begin{aligned}
\phi_{Y+Z}(t) &= E(e^{t(Y+Z)}) = E(e^{tY}e^{tZ}) \\
&= E(e^{tY})E(e^{tZ}) = \phi_Y(t)\phi_Z(t)
\end{aligned}$$

The finite induction extends this result to the sum of any finite number of independent random variables $Y_1, Y_2, \cdot, \ldots, s, Y_n$, thus

$$\phi_{\Sigma Y_i}(t) = \prod_{i=1}^{n} \phi_Y(t)$$

If the summands have identical distributions, with common m.g.f. $\phi_Y(t)$, then

$$\phi_{\Sigma Y_i}(t) = [\phi_Y(t)]^n$$

In conclusion, if one obtaines the moment generating function of a distribution indirectly, he can then calculate the moments of that distribution; but one puzzle is on the precise density function or probability function of the distribution. There is a uniqueness theorem (treated later in this chapter) for moment generating functions, which says that there can only be one distribution leading to a given m.g.f (under certain conditions). Thus, if themoment generating function of a random variable Y is obtained directly- but is recognised as themoment generating function of a known distribution, then that distribution is the distribution of Y.

10.3 The Factorial Moment Generating Function (fmgf)

Another function that is closely associated to m.g.f. is a function that generates factorial moments. This function is well defined as follows:

$$\mu_Y(t) = E(t^Y) = E[e^{Y \log t}] = \phi_Y(\log t)$$

Because $\log 1 = 0$, it is the point at $t = 1$ which might interest us and this produces the factorial moments from the derivatives. On a formal note we have

$$\mu'(t) = E(Yt^{Y-1})\Big|_{t=1} = E(Y) \qquad (10.13)$$

also

$$\mu''(t) = E(Y(Y-1)t^{Y-2})\Big|_{t=1} = E(Y(Y-1))$$

$$\mu^{[k]} = E[Y(Y-1)(Y-2)\cdots(Y-K+1)] \qquad (10.14)$$

Equation (10.14) is called the $k-th$ factorial moment.

For a discrete random sequence Y, the FMGF is called probability generating function

$$\mu_Y(t) = E(t^Y) = \sum_{k=0}^{\infty} t^k P(Y = K) \qquad (10.15)$$

Because of the awkwardness in writing out k factors starting with n, we may write the factorial expression as

$$y(y-h)(y-2h)\cdots\{y(k-1)h\}$$

which is conveniently written as $y^{(k)}$, a notation which brings out an analogy with the power y^k. Taking first differences w.r.t. y and with unit h, we have

$$\begin{aligned}
\Delta y^{[k]} &= (y+h)^{(k)} - y^{(k)} \\
&= (y+h)y(y-h)\cdots\{y-(k-2)h\} - y(y-h)\cdots\{y-(k-1)h\} \\
&= ky^{(k-1)}h = \partial y^k/dx
\end{aligned}$$

conversely,

$$\sum_{y=0}^{\infty} y(k) = \frac{1}{(k+1)h}(y+h)^{(k+1)}$$

and corresponding to

$$\int_0^y y^k dy = \frac{1}{k+1}y^{k+1}$$

Thus the $k-th$ factorial moment about an arbitrary origin may then be defined by the equation

$$\mu'_{(k)} = \sum_{j=-\infty}^{\infty} (y_j - a)^{(k)} f(y_j) \tag{10.16}$$

In statistical theory the f.m.g. functions are not very prominent, but they provide very concise formulae for the moments of certain discontinuous distributions of the binomial type. When it is necessary to distinguish between factorial moments about the mean and those about an arbitrary point we may write the former without the prime.

Factorial moments obey the laws of binomial transformation governing ordinary moments. The expressions are:

$$(a+b)^{[k]} = \sum_{j=0}^{k} \binom{k}{j} a^{[k-j]} b^{[j]}$$

and so

$$\begin{aligned}
(y-a)^{[k]} &= (y-b+c)^{[r]}, \text{ where } c = b-a \\
&= \sum_{j=0}^{k} \binom{k}{j} (y-b)^{[k-j]} c^{[j]}
\end{aligned}$$

and hence

$$\mu'_{[k]}[a] = \sum_{j=0}^{k} \binom{k}{j} \mu'_{[k-j]}(b)c^{[j]}$$
$$= \{\mu'(b) + c\}^{[k]} \quad (10.17)$$

Example 10.2

Let $Y_1, Y_2, \cdots, Y_n$ be independent sequence of random variables, each with the distribution $P(Y = 1) = p$, $P(Y = 0) = q = 1 - p$ the m.s.f. of this distribution is

$$\phi_Y(t) = E(e^{tY}) = e^t P + e^0 q$$

and so the moment generating function of the sum is

$$\phi_{\Sigma Y_i}(t) = [\phi_Y(t)]^n = (Pe^t + q)^n$$

The factorial moment generating function is then

$$\mu_{(t)} = \phi(\log t) = (pe^{\log t} + q)^n$$
$$= (pt + q)^n$$

also

$$\mu''_{(t)} = n(n-1)(pt+q)^{n-2}p^2$$

whence

$$E(Y(Y-1)) = \mu''(1) = n(n-1)p^2.$$

10.4 Cumulant Generating Function (c.g.f.)

The moment functions are a set of descriptive constants of a distribution which are useful for measuring its properties and in certain circumstances, for specifying it. They are not the ony set of constants for this purpose, cumulants have properties which are more useful from the theoretical viewpoint.

By definition, the cumulants $K_1, K_2, \cdots, K_r$ are defined by the identity in t

$$\exp\{K_1 t + \frac{K_2 t^2}{2!} + \cdots + \frac{k_r t^r}{r!} + \cdots\} = 1 + \mu'_1 t + \frac{\mu'_2 t^2}{2!} + \cdots + \frac{\mu'_r t^r}{r!} + \cdots$$
$$= \sum_{r=0}^{\infty} (Y^r) \frac{t^r}{k!} = \phi(t) \quad (10.18)$$

Log $\phi(t)$ may then be called a cumulant-generating function and denoted by c.g.f.

Example 10.3

Consider the discrete Poisson distribution whose frequencies at $0, 1, \ldots, y$ are

$$e^{-\lambda}(1, \frac{\lambda}{1!}, \cdots, \frac{\lambda^j}{j!}, \cdots).$$

The moment generating lfunction if exists, is given by

$$\begin{aligned}\phi(t) &= e^{-\lambda}\sum_{j=0}^{\infty}\frac{\lambda^j}{j!}e^j \\ &= e^{-\lambda}\exp(\lambda e^t) \\ &= \exp\{\lambda(e^t-1)\}\end{aligned}$$

Since the variate is non-negative, for any r the absolute moment is the same as the oridnary moment, and we have

$$\mu_r' = e^{-\lambda}\sum_{j=0}^{\infty}\frac{\lambda^j j^r}{j!}$$

and since this converges commulants of all orders exist. Then

$$\begin{aligned}\log\phi(t) &= \lambda(e^t-1) \\ &= \lambda\sum_{j=1}^{\infty}t^j/j!\end{aligned}$$

and hence $k_r = \lambda$ for all r. Thus the cumulants of this distribution are equal to λ.

10.5 Characteristic Functions (c.f.)

The characteristic function of a random variable has many useful and important properties which gives it a central role in statistical theory, we shall give an account of some in this subsection.

Formally, the characteristic function $\Phi(t)$ defined on t of a random variable Y whose distribution function is F is defined as

$$\Phi_Y(t) = E(e^{(tY)}) = \int_{R_Y} e^{(tY)}dF(y)$$

$\Phi(t)$ always exists, since

$$|\Phi(t)| = \left|\int_{-\infty}^{\infty} e^{ity} dF\right| \le \int_{-\infty}^{\infty} \left|e^{ity}\right| dF = \int_{-\infty}^{\infty} dF - 1$$

so that the defining integral converges absolutely. Also $\Phi(t)$, is uniformly continuous in$t$ and differentiable k times under the integral of the resulting expressions exist and are uniformly converges for which it is sufficient that V_j exists. Then

$$\begin{aligned} \left|\Phi^{[k]}(t)\right| &= \left|\int_{-\infty}^{\infty} y^k e^{ity} dF\right| \\ &\le \int_{-\infty}^{\infty} \left|y^k\right| dF = v \end{aligned}$$

Example 10.4

Consider the distribution defined by the density e^{-y} for $y > 0$, the characteristic function of the distribution is

$$\Phi(t) = E(e^{ity}) = \int_{-\infty}^{\infty} e^{ity-y} dy = \frac{1}{1-it}$$

The characteristic function genertes moments in such a way as the moment generating function, except that each differentiation introduces factor of i:

$$E(Y^k) = i^{-k}\phi^{[k]}(0)$$

This is possible only if the $k-th$ moment of the distribution exists. If moments up to a certain order, say r, exists, then it is possible to express $\Phi(t)$ as a Maclaurin series. Although the c.f. does generate moments, its principal use is as a too. in deriving distributions, as such it is necessary to know several facts about characteristic functions by theorems that establishits facts.

Theorem 10.1

Let $\Phi(t)$ be the characteristic function of a random variable Y_n, whose distribution function is F_n. Suppose $\Phi(t)$ is the characteristic function of a random variable Y whose distribution function is F. Then

$$Y_n \xrightarrow{L} Y \text{ iff } \Phi_n(t) \longrightarrow \Phi(t).$$

To prove, we shall consider both the necessary and sufficient conditions.

Proof

(A) **Necessary**

Assume $Y_n \xrightarrow{L} Y$, then

$$\int e^{ity} dF_n \longrightarrow \int e^{ity} dF \text{ by Helly Bray's theorem.}$$

(B) **Sufficiency**

Assume $\Phi_n(t) \longrightarrow \Phi(t)$, this implies that

$$\int e^{ity} dF_n \longrightarrow \int e^{ity} dF$$

choose a subsequence $\{F_m\}$ which tends to a non-decreasing bounded function G, then

$$\Phi_m = \int_{-\infty}^{\infty} e^{ity} dF_m$$

Taking the limits we obtain

$$\Phi_m(t) \longrightarrow \phi(t) = \int_{-\infty}^{\infty} e^{ity} dF \qquad (10.19)$$

from this we have,

$$\Phi(0) = G(\infty) - G(-\infty) = 1,$$

because $\Phi(0) = 1$. This means that G is necessarily a distribution and $G = F$. Since the distribution function satisfies equation (10.19), it is unique. All subsequence necessarily lead to the same distribution function for some reasons. #

Theorem 10.2

If in theorem (10.1), it is only supposed that $\Phi_n(t) \rightarrow \Phi(t)$ for all t, the limit function is a characteristic function provided only that it is continuous at $t = 0$ or if the limit is uniform in as interval containing zero.

Proof

Choose a subsequence F_m tending to a non-decreasing bounded function G and consider

$$\int_0^V \Phi_m(t)dt = \int_0^V \left(\int_{-\infty}^{\infty} e^{ity} dF_m\right) dt$$
$$= \int_{-\infty}^{\infty} \left(dF_m \int_0^V e^{ity}\right) dt$$
$$= \int_{-\infty}^{\infty} \frac{e^{ivy}-1}{iy} dF_m$$

Taking limit as m tends to infinity

$$\int_0^V \Phi(t)dt = \int \frac{e^{ivy}-1}{iy} dG$$

$$\int_0^V \frac{\Phi(t)dt}{V} = \int \frac{e^{ivy}-1}{ivy} dG$$

$$\lim_{V\to 0} \int_0^V \frac{\Phi(t)}{V} dt = \lim_{V\to 0} \int \frac{e^{ivy}-1}{ivy} dG$$

$$\Rightarrow \Phi(0) = \int dG$$

Then

$$\lim_{m\to\infty} \Phi(t) = \Phi(0)$$

$$\Phi(0) = G(\infty) - G(-\infty) = 1$$

and by theorem 10.1, $\Phi_m(t) \to \Phi(t)$, the characteristic function of G. #

10.6 The Inversion Theorem

Theorem 10.3

We now state and prove the fundamental theorem of the theory of characteristic functions called Inversion theorem or Uniqueness theorem: The characteristic function uniquely determines the distribution function more precisely if

$$\Phi(t) = \int_{-\infty}^{\infty} e^{ity} dF,$$

then

$$F(y) - F(0) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\left(\frac{1-e^{ity}}{it}\right)\Phi(t)dt$$

Proof: Let

$$F(Y_2) - F(Y_1) = \frac{1}{2\pi}\lim_{c\to 0}\int_{-c}^{c}\left(\frac{e^{-ity_1} - e^{ity_2}}{it}\right)\Phi(t)dt$$

Set $Y_1 = Z$ and $Y_2 = Y$ and let $z \to 0$ then

$$F(Y) - F(0) = \frac{1}{2\pi}\lim_{z\to 0}\lim_{c\to\infty}\int_{-c}^{c}\left(\frac{e^{-it(0)} - e^{ity}}{it}\right)\Phi(t)dt$$

The last results hold at the continuity point of F and the limit evaluated with respect to any set of points in continuity point of F.

Theorem 10.4

Let Y be a random variable having characteristic function $\Phi(t)$ and distribution function $F(\cdot)$. If Y_1 and Y_2 are points in $c(F)$, then

$$F(Y_1) - F(Y_2) = \frac{1}{2\pi}\lim_{c\to\infty}\int_{-c}^{c}\left(\frac{e^{-ity_1} - e^{ity_2}}{it}\right)\Phi(t)dt$$

Proof: Set

$$J_c = \frac{1}{2\pi}\int_{-c}^{c}\left(\frac{e^{-ity_1} - e^{ity_2}}{it}\right)\Phi(t)dt$$

Then

$$J_c = \frac{1}{2\pi}\int_{-c}^{c}\left[\left(\frac{e^{-ity_1} - e^{ity_2}}{it}\right)\int_{-\infty}^{\infty} e^{ity}dF(y)\right]dt$$

$$= \frac{1}{2\pi}\int_{-c}^{c}\left[\int_{-\infty}^{\infty}\left(\frac{e^{it(y-y_1)} - e^{it(y-y_2)}}{it}\right)dF(y)\right]dt$$

The integral w.r.t. F or y converges absolutely and integral w.r.t. t is within finite limit $(-c, c)$ and so we may need to change the order of integration to obtain

$$J_c = \frac{1}{2\pi}\int_{-\infty}^{\infty}\left[\int_{-c}^{c}\left(\frac{e^{-ity_1} - e^{ity_2}}{it}\right)dt\right]dF(y)$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty}\left[\int_{0}^{\infty}\left(e^{it(y-y_1)} - e^{-t(y-y_1)} - e^{it(y-y_2)}\right.\right.$$

$$= \frac{1}{\pi}\int_{-\infty}^{\infty}\left[\int_{0}^{c}\left(\frac{\sin t(y-y_1)}{t} - \frac{\sin t(y-y_2)}{t}\right.\right.$$

$e^{it} = \cos x + \sin x$, then as $c \to \infty$

$$\frac{1}{\pi}\int_0^c \frac{\sin \alpha t}{t}dt \quad \to \quad \tfrac{1}{2};\text{ if } \alpha > 0 \quad \text{or} \quad \to \quad -1/2; \text{if } \alpha < 0 \tag{10.20}$$

The convergence is uniform w.r.t. α in every region $\alpha > \delta > 0$ or $\alpha < -\delta$ and for $|\alpha| \leq \delta$, and for all values of C.

Also

$$\left|\frac{1}{\pi}\int_0^t \frac{\sin \alpha t}{t}dt\right| < 1 \tag{10.21}$$

Take $y_2 > y_1 \epsilon\ c(f)$ and set

$$\psi(c, y, y_1, y_2) = \frac{1}{\pi}\int_0^t \left[\frac{\sin t(y-y_1)}{t} - \frac{\sin t(y-y_2)}{t}\right] dt$$

Then

$$\begin{aligned} J_t &= \int_{-\infty}^{\infty} \psi(t, y, y_1, y_2)dF(y) \\ &= \int_{-\infty}^{y_1-\delta} \psi(c, y, y_1, y_2)dF(y) + \int_{y_1-\delta}^{y_1+\delta} \psi(\cdot)dF(y) + \int_{y_1+\delta}^{y_2-\delta} \psi(\cdot)dF(y) \\ &+ \int_{y_2-\delta}^{y_2+\delta} \psi(\cdot)dF(t) + \int_{y_2+\delta}^{\infty} \psi(\cdot)dF(y) \end{aligned}$$

with δ chosen such that $y_1 + \delta < y_2 - \delta$ and gives the interval $-\infty < y < y_1 - \delta$ and $y - y_1 < -\delta$, for the first integral $-\infty < y < y_1 - \delta$ and so $y - y_1 < -\delta$, $y - y_2 < -\delta$ since $y_2 > y_1$ then

$$\int_{-\infty}^{y_1-\delta} \psi(\cdot)dF = 0$$

the same way the fifth integral gives

$$\int_{y_2+\delta}^{\infty} \psi(\cdot)dF = 0$$

$-\delta < y < y_2 - \delta$, we have $y - y_1 > \delta$ and $y - y_2 < -\delta$, this gives

$$\int_{y_1+\delta}^{y_2-\delta} \psi(\cdot)dF = 1$$

and so

$$\int_{y_1+\delta}^{y_2-\delta} \psi(\cdot)dF = F(y_2 - \delta) - F(y_1 + \delta)$$

as $c \to \infty$.
Sismilarly,

$$\left|\int_{y_2-\delta}^{y_2+\delta} \psi(\cdot)dF\right| < 2\int_{y_1-\delta}^{y_1+\delta} dF(y) = 2[F(y_2+\delta) - F(y_2-\delta)]$$

By collating results, we have for $\delta > 0$,

$$\lim_{c\to\infty} J_c = F(y_2-\delta) - F(y_1+\delta) + 2[F(y_1+\delta) - F(y_1-\delta) + F(y_2+\delta) - F(y_2-\delta)]$$

so

$$\lim_{c\to\infty} J_c = \lim_{c\to\infty} J_c = F(y_2) - F(y_1) \qquad \#$$

This is so because F is continuous and J_c does not depend on δ.

Theorem 10.5

If the characteristic function $\phi(t)$ is Lebesque integrable over the entire line $(-\infty, \infty)$, then the distribution function $F(y)$ that corresponds to it is continuous and $F(y)$ is also continuous with

$$F'(y) = f(y) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-ity}\Phi(t)dt$$

Proof

If $\Phi(t)$ is Lebesque integrable, then

$$\left(\frac{e^{-ity_1} - e^{-ity_2}}{it}\right)\Phi(t)$$

is Lebesque integrable and the inversion formula may be written in the form

$$F(y_2 - F(y_1) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\left(\frac{e^{-ity_1} - e^{-ity_2}}{it}\right)\Phi(t)dt$$

select h such that

$$\begin{aligned} y_1 &= y - h \\ y_2 &= y + h, \end{aligned}$$

Select A so large such that

$$\int_{|t|>A} |e^{-ith} - 1|\Phi(t)dt < \xi/2$$

and also select h as small as possible, such that

$$\int_{|t|\leq A} |e^{-ith} - 1|\Phi(t)dt < \xi/2$$

Then

$$|F(y+h) - F(y)| \leq \frac{\xi}{2} + \frac{\xi}{2} = \xi$$ #

Problem 10.3

1. The distribution function for Y with cumulative distribution function is

$$F(y) = \begin{cases} 0, & \text{if } y < 0 \\ 1 - 0.6e^{-y}, & \text{if } y \geq 0 \end{cases}$$

obtain the moment generating function at point t and the $k-th$ moment of the distribution.

2. Let Y denote the number of points on a die. Suppose the die is tossed three times, obtain the probability generating function and its probability if the total number of points is 7.

3. By considering the characteristic function of thebilateral exponential distribution

$$f(Y) = \frac{1}{2}e^{-|y|}, \quad -\infty < y < \infty$$

and using the inversion theorem or otherwise, show that the characteristic function of the Cauchy distribution is $\Phi_Y(t) = e^{-it\theta - |t|}$

4. Compute the moment generating function of the distribution defined by the density

$$f(y) = e^{-y}, \quad y > 0$$

Expand it in a power series.

then

$$F(y+h) - F(y-h) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\left(\frac{e^{-it(y-h)} - e^{-it(y+h)}}{it}\right)\Phi(t)dt$$

noting that $e^{ity} = \cos ty + i\sin ty$, we obtain

$$\begin{aligned} F(y+h) - F(y-h) &= \frac{1}{\pi}\int_{-\infty}^{\infty}\sin th(\cos ty - i\sin ty)\Phi(t)dt \\ &= \frac{h}{\pi}\int_{-\infty}^{\infty}\frac{\sin th}{th}e^{-ity}\Phi(t)dt \\ |F(y+h) - F(y-h)| &\le \frac{h}{\pi}\int_{-\infty}^{\infty}|\phi(t)|dt \end{aligned}$$

Let $h \to 0$, we obtain

$$F(y+0) - F(y-0) \text{ and } F(y^+) - F(y^-) = 0$$

which shows that F is continuous, but

$$\frac{F(y+h) - F(y-h)}{2h} = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{\sin th}{th}e^{-ity}\Phi(t)dt$$

When $h \to 0$, then we ootain

$$\begin{aligned} F'(y) &= f(y) = \lim_{h\to 0}\frac{F(y+h) - F(y-h)}{2h} \\ &= \frac{1}{2\pi}\int_{\infty}^{\infty}e^{-ity}\Phi(t)dt \end{aligned}$$

We need to show that $f(y)$ is continuous:

$$\begin{aligned} F(y+h) - F(y) &= \frac{1}{2\pi}\int_{-\infty}^{\infty}(e^{-it(y+h)} - e^{ity})\Phi(t)dt \\ |F(y+h) - F(y)| &\le \frac{1}{2\pi}\int_{-\infty}^{\infty}|e^{-it(y+h)} - e^{-ity}|\Phi(t)dt \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty}|e^{-ity}|\,|e^{-ity} - 1|\Phi(t)dt \\ |F(y+h) - F(y)| &\le \frac{1}{2\pi}\int_{-\infty}^{\infty}|e^{-ith} - 1|\Phi(t)dt \\ &= \frac{1}{2\pi}\left[\int_{|t|\le A}|e^{-ith} - 1|\Phi(t)dt \right. \\ &\quad \left. + \int_{|t|>A}|e^{-ith} - 1|\Phi(t)dt\right] \end{aligned}$$

5. Compute the factorial moment generating function of the discrete distribution defined by the probability function.

$$f(k) = \frac{1}{2^{k+1}},\ k = 0, 1, 2, \ldots$$

Determine the mean and variance of the distribution.

Solution 10.3

1. In this distribution there is a discrete lump to probability, in the amount 0.2, at $y = 0$. The moment generating function is

$$\begin{aligned}\Phi(t)\int_{\infty}^{\infty} e^{ty}dF(y) &= e^{0}(0.2) + \int_{0}^{\infty} e^{ty}(0.8e^{-y})dy \\ &= 0.2 + 0.8(1-t)^{-1} = 1 + 0.8t + 0.8t^2 + \ldots \\ &= (0.8)k!(t^k/k!);\ k = 0, 1, 2, \ldots\end{aligned}$$

Hence the $k-th$ moment is

$$E(Y^k) = 0.8K!$$

2. Let Z = total number of points thrown is a random variable Z whose factorial moment generating function is

$$\begin{aligned}E(t^z) &= E(t^{z_1})E(t^{z_2})E(t^{z_3}) \\ &= \left[\frac{t(1-t^6)}{6(1-t)}\right] \\ &= \frac{t^3}{216}\sum_{k=0}^{3}\binom{3}{k}(-t)^{6k}\sum_{j=0}^{\infty}\binom{-3}{j}(-t)^j\end{aligned}$$

By writing the expansion for $E(t^z)$ as a double deem, as

$$\begin{aligned}E(t^z) &= \frac{1}{216}\sum_{k=0}^{3}\sum_{j=0}^{\infty}(-1)^{j+k}\binom{3}{k}\binom{-3}{j}t^{6k+j+3} \\ P(Y=7) &= \frac{1}{216}\binom{3}{0}\binom{-3}{4} = \frac{(-3)(-4)(-5)(-6)}{216\times 24} \\ &= \frac{15}{216} = 0.069\end{aligned}$$

3. The characteristic function of Y at t is

$$\phi(t) = \frac{1}{\lambda\pi}\int_{\infty}^{\infty} e^{itz}\frac{1}{[1+(\frac{y-\theta}{\wedge})^2}dy \qquad (i)$$

Set $z = \frac{y-\theta}{\lambda}$, then $y = \theta + \lambda z$ and $dy = \lambda dz$,

$$\therefore \quad \phi(t) = \frac{1}{\lambda\pi}\int_{-\infty}^{\infty} e^{it(\theta+\lambda z)}\frac{1}{(1+z^2)}\lambda dz$$

$$= \frac{1}{\pi}\int_{-\infty}^{\infty} e^{it\theta}\cdot e^{it\lambda z}(\frac{1}{1+z^2})dz \qquad (ii)$$

but

$$\frac{1}{2}e^{-|p|} = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{itp}\frac{1}{1+t^2}dt$$

then

$$\int_{-\infty}^{\infty} e^{itp}\frac{1}{1+t^2}dt = \Pi e^{-|p|} \qquad (iii)$$

In (ii), set $V = -z$, then $r = -dv$ and

$$\phi(t) = \frac{1}{\pi}\int_{\infty}^{-\infty} e^{it\theta}\cdot e^{-it\lambda v}\frac{1}{1+\nu^2}dV$$

$$= e^{it\theta}\left(\frac{1}{\pi}\int_{-\infty}^{\infty} e^{-it\lambda v}\frac{1}{1+v^2}dv\right)$$

Let $6 = tv$, then

$$\phi(t) = e^{it\theta}\cdot\frac{1}{\pi}\Pi e^{-|b|}$$

$$= e^{it\theta}e^{-\lambda|t|}, \quad \text{since } \lambda > 0$$

4. $$M_Y(t) = \int_0^{\infty} e^{-y}e^{ty}dy = \int_0^{\infty} e^{-y(1-t)}dy$$

$$= \frac{-1}{1-t}\left[e^{-y(1-t)}\right|_0^{\infty}$$

$$= \frac{1}{1-t} = (1-t)^{-1}$$

$$= 1 + x + x^2 + \cdots + x^k + \cdots$$

5. $$\eta_X(t) = E(t^x) = \sum_0^{\infty} t^k P(X=K) = \sum t^k/2^{k+1}$$

$$= \frac{t^0}{2} + \frac{t}{2^2} + \frac{t^2}{2^3} + \cdots$$

From $S_\infty = \dfrac{a}{1-r}$ (sum to infinity), we have

$$\begin{aligned}
\eta_X(t) &= \frac{1}{2-t} = (2-t)^{-1} \\
E(X) &= \eta_X^{[1]}(t)\Big|_{t=1} = \frac{1}{(2-1)^2}\Big|_{t=1} = 1 \\
E(X(X-1)) &= \eta_X^{[2]}(t)\Big|_{t=1} = \frac{2}{(2-t)^3}\Big|_{t=1} = 2 \\
Var(X) &= \eta^{[2]} + E(X) - (E(X))^2 \\
&= \eta^{[2]} + \eta^{[1]} - [\eta^{[1]}]^2 \\
&= 2 + 1 - 1^2 = 2
\end{aligned}$$