U.P. Rajarshi Tandon Open
University, Prayagraj

UGSTAT – 102

Probability, Distribution and Statistical Inference

Block: 1 Probability Theory

Unit – 1 : Random Experiments and Probability

Unit – 2 : Conditional Probability

Block: 2 Probability Distributions and Expectations

Unit – 3 : Random Variables and Probability Distributions

Unit – 4 : Mathematical Expectation

Unit – 5 : Inequalities for Moments

Block: 3 Concepts of Probability Distribution

Unit – 6 : Univariate distributions

Unit – 7 : Discreet Distributions

Unit – 8 : Normal Distribution

Unit – 9 : Continuous Distributions

Unit – 10 : Sampling Distributions

Block: 4 Basic Principles of Statistical Inference

Unit – 11 : Estimation

Unit – 12 : Methods of Estimation

Unit – 13 : Testing of Hypothesis

Block: 5 Tests of Significance

Unit – 14 : Exact Tests and Fisher Z-Transformations

Unit – 15 : Large Sample Tests

Unit – 16 : Non-Parametric Tests

Course Design Committee

Dr. Ashutosh Gupta **Chairman**
Director, School of Sciences
U. P. Rajarshi Tandon Open University, Prayagraj

Prof. Anup Chaturvedi **Member**
Ex. Head, Department of Statistics
University of Allahabad, Prayagraj

Prof. S. Lalitha **Member**
Head, Department of Statistics
University of Allahabad, Prayagraj

Prof. Himanshu Pandey **Member**
Department of Statistics, D. D. U. Gorakhpur University,
Gorakhpur.

Dr. Shruti **Member-Secretary**
Sr. Assistant Professor, School of Sciences
U.P. Rajarshi Tandon Open University, Prayagraj

Course Preparation Committee

Dr. Shruti **Course Coordinator**
Sr. Assistant Professor, School of Sciences
U. P. Rajarshi Tandon Open University, Prayagraj

Block: 1 ***Probability Theory***

Prof. Anoop Chaturvedi **Writer**
Department of Statistics, University of Allahabad
Prayagraj

Prof. V. P. Ojha **Reviewer**
Department of Statistics and Mathematics,
D. D. U., Gorakhpur University, Gorakhpur

Prof. Ram Lal **Reviewer**
Department of Mathematics and Statistics, Allahabad
Agricultural Institute (Deemed University), Naini, Prayagraj

Prof. A. K. Saxena **Editor**
Department of Statistics, Lucknow University, Lucknow

Prof. S. K. Pandey **Editor**
Department of Statistics, Lucknow University, Lucknow

Block: 2 Probability Distributions and Expectations

Prof. S. K. Pandey **Writer**
Department of Statistics, Lucknow University, Lucknow

Prof. Anoop Chaturvedi **Writer**
Department of Statistics, University of Allahabad
Prayagraj

Dr. Shruti **Writer**
School of Sciences, U. P. Rajarshi Tandon Open University,
Prayagraj

Prof. V. P. Ojha **Reviewer**
Department of Statistics and Mathematics,
D. D. U., Gorakhpur University, Gorakhpur

Prof. G. S. Pandey **Reviewer**
Department of Statistics, University of Allahabad,
Prayagraj

Dr. Alok Kumar **Reviewer**
Department of Community Medicine, Institute of
Medical Sciences, Banaras Hindu University, Varanasi

Prof. S. K. Pandey **Editor**
Department of Statistics, Lucknow University, Lucknow

Block: 3 Concept of Probability Distributions

Prof. B. P. Singh **Writer**
Department of Statistics, Banaras Hindu University, Varanasi

Dr. Padmakar Singh **Writer**
Department of Statistics, U.P.P.G. College, Varanasi

Dr. Alok Kumar **Writer**
Department of Community Medicine, Institute of
Medical Sciences, Banaras Hindu University, Varanasi

Dr. Shruti **Writer**
School of Sciences, U. P. Rajarshi Tandon Open University,
Prayagraj

Prof. V. P. Ojha **Reviewer**
Department of Statistics and Mathematics,
D. D. U., Gorakhpur University, Gorakhpur

Dr. Padmakar Singh **Reviewer**
Department of Statistics, U.P.P.G. College, Varanasi

Dr. D. C. Singh **Reviewer**
Department of Statistics, Harish Chandra P.G. College,
Varanasi

Prof. R. C. Yadava **Editor**
Department of Statistics, Banaras Hindu University, Varanasi

Prof. S. K. Pandey **Editor**
Department of Statistics, Lucknow University, Lucknow

Prof. G. S. Pandey **Editor**
Department of Statistics, University of Allahabad,
Prayagraj

Block: 4 *Basic Principles of Statistical Inference*

Prof. B. N. Pandey **Writer**
Department of Statistics, Banaras Hindu University, Varanasi

Prof. Ram Ji Tiwari **Writer**
Department of Statistics, University of Jammu, Jammu

Dr. Sanjay Singh **Writer**
Department of Statistics, Banaras Hindu University, Varanasi

Prof. V. P. Ojha **Reviewer**
Department of Statistics and Mathematics,
D. D. U., Gorakhpur University, Gorakhpur

Prof. S. K. Upadhyay **Reviewer**
Department of Statistics, Banaras Hindu University, Varanasi

Dr. Shruti **Reviewer**
School of Sciences, U.P. Rajarshi Tandon Open University,
Prayagraj

Prof. V. P. Ojha **Editor**
Department of Statistics and Mathematics, D. D. U.,
Gorakhpur University, Gorakhpur

Block: 5 *Tests of Significance*

Prof. Ram Lal **Writer**
Department of Mathematics and Statistics, Allahabad
Agricultural Institute (Deemed University), Naini, Prayagraj

Prof. Umesh Singh **Writer**
Department of Statistics, Banaras Hindu University, Varanasi

Prof. S. K. Pandey **Writer**
Department of Statistics, Lucknow University, Lucknow

Prof. Anoop Chaturvedi **Reviewer**
Department of Statistics, University of Allahabad
Prayagraj

Prof. B. P. Singh **Reviewer**
Department of Statistics, Banaras Hindu University, Varanasi

Prof. A. K. Saxena **Reviewer**
Department of Statistics, Lucknow University, Lucknow

Prof. S. K. Pandey **Editor**
Department of Statistics, Lucknow University, Lucknow

Prof. S. K. Upadhyay **Editor**
Department of Statistics, Banaras Hindu University, Varanasi

First Edition: *March 2008* (Published with the support of the Distance Education Council, New Delhi)

Second Edition: ***July 2020***

ISBN :

Printed and Published by Dr. Arun Kumar Gupta Registrar, Uttar Pradesh Rajarshi Tandon Open University, 2020.

Printed By: Chandrakala Universal Pvt. 42/7 Jawahar Lal Neharu Road, Prayagraj.

Blocks Introduction

The present SLM on ***Probability Distribution and Statistical Inference*** consists of five Blocks. *Block - 1 – Probability Theory* has two units; *Block - 2 – Probability Distributions and Expectations* has three units; *Block - 3 – Concepts of Probability Distributions* ha*s* five units; *Block - 4 – Basic Principles of Statistical Inference* ha*s* three units and at the last *Block - 5 Tests of Significance* has three units.

The ***Block - 1 – Probability Theory*** consist two units. The *first unit* of this block describes the elementary principles and concepts of probability, along with their usefulness, and explains how tw mathematical definition of probability can be used to calculate. The probabilities of occurrence of events, certain basic theorems and results on probability are also given in this unit. The *second unit*, discusses the concept of conditional probability, independence of events and Baye's theorem with applications..

The ***Block - 2 – Probability Distributions and Expectations*** is the second block having three units. The *first unit* of this block gives the concepts of random variable, distribution function, discrete and continuous random variables, probability mass function, probability density function and their properties. The *second unit* of this block discusses the concept of expectation along with additive and multiplicative theorems on expectation. The concept of moments in terms of expectation has also been defined in this unit. The *third unit* of this block provides some important inequalities concerning moments, Chebychew's inequality and its applications are also given.

The ***Block - 3 – Concepts of Probability Distributions*** consists of five units. The *first unit* of this block introduces the concept of probability distribution, discrete probability distributions namely Bernoulli, Binomial and Poisson have been discussed in this unit along with their properties, applications and importance. The *second unit* of this block describes geometric, negative binomial and hyper geometric distributions with their applications, properties and importance. The *third unit* of this block defines normal distribution with its properties, applications and importance. The *fourth unit* of this block provides uniform and exponential distributions with their properties, applications and importance. The *fifth unit* of this block deals with the sampling distributions like $\chi2$, t, F, z distributions, Beta, Gamma, Chauchy densities.

The ***Block - 4 Basic Principles of Statistical Inference*** concentrated on the study of inferential statistics which deals with taking judgement, drawing conclusion or inferences about a population on the basis of information available in a sample/s drawn from the population. Traditionally, the problems in statistical inference are classified into 'problems of estimation' and 'tests of hypothesis'. Although, in fact, they are decision problems and can be tackled by unified approach- decision problems and can be tackled by unified approach – decision theoretic approach. The estimations can be done in either of two ways: (i) Point estimation and (ii) confidence interval estimation. It consists of three units. The *first unit* of this block deals with the estimators of the parameters and properties of a good estimator, it also discussed briefly the confidence interval estimation. The *second unit* of this block discusses with the two methods of estimation of the parameters-viz. method of moment's estimation (MME) and method of maximum likelihood estimation (MLE). The *third unit* of this block gives a brief account of the concepts of the tests of hypotheses.

The ***Block - 5 – Tests of Significance*** consists of three units. The *first unit* of this contains exact tests with examples based on Chi-square, t and F-distributions. It also provides Fisher’s z-transformation and its uses. The *second unit* of this block describes large sample tests with illustrations. The *third unit* of this block discusses one-sample, two samples, and non-parametric tests for location and scale along with run test for randomness. Illustrations and examples on these topics have also been given.

At the end of every block/unit the summary, self-assessment questions and further readings are given.

U.P. Rajarshi Tandon Open University, Prayagraj

UGSTAT – 102

Probability, Distribution and Statistical Inference

Block

1

Probability Theory

Unit – 1
Random Experiments and Probability

Unit – 2
Conditional Probability

Course Design Committee

Dr. Ashutosh Gupta **Chairman**
Director, School of Sciences
U. P. Rajarshi Tandon Open University, Prayagraj

Prof. Anup Chaturvedi **Member**
Ex. Head, Department of Statistics
University of Allahabad, Prayagraj

Prof. S. Lalitha **Member**
Head, Department of Statistics
University of Allahabad, Prayagraj

Prof. Himanshu Pandey **Member**
Department of Statistics, D. D. U. Gorakhpur University,
Gorakhpur.

Dr. Shruti **Member-Secretary**
Sr. Assistant Professor, School of Sciences
U.P. Rajarshi Tandon Open University, Prayagraj

Course Preparation Committee

Dr. Shruti **Course Coordinator**
Sr. Assistant Professor, School of Sciences
U. P. Rajarshi Tandon Open University, Prayagraj

Prof. Anoop Chaturvedi **Writer**
Department of Statistics, University of Allahabad
Prayagraj

Prof. V. P. Ojha **Reviewer**
Department of Statistics and Mathematics,
D. D. U., Gorakhpur University, Gorakhpur

Prof. Ram Lal **Reviewer**
Department of Mathematics and Statistics, Allahabad
Agricultural Institute (Deemed University), Naini, Prayagraj

Prof. A. K. Saxena **Editor**
Department of Statistics, Lucknow University, Lucknow

Prof. S. K. Pandey **Editor**
Department of Statistics, Lucknow University, Lucknow

First Edition: *March 2008* (Published with the support of the Distance Education Council, New Delhi)

Second Edition: ***July 2020***

ISBN :

Printed and Published by Dr. Arun Kumar Gupta Registrar, Uttar Pradesh Rajarshi Tandon Open University, 2020.

Printed By: Chandrakala Universal Pvt. 42/7 Jawahar Lal Neharu Road, Prayagraj.

Unit-1: Random Experiments and Probability

Structure

1.1 Introduction

In various fields of social, biological physical sciences etc., we come across with experiments and phenomenons in which some kind of uncertainty is involved. This unit introduces the basic idea of such random experiments, sample space and events associated with the help of several examples. Different algebraic operations like union, intersection, complement, Cartesian product etc. will also be explained to you. The probability associated with an event has been defined using axiomatic approach. The classical and empirical approach to probability will be explained and their shortcomings will be discussed. Different results associated with the probability will be derived.

1.2 Objectives

After going through this unit you shall be able to:

- Understand the difference between the deterministic and random experiments.
- Define sample space and events associated with a random experiment.
- Use different algebraic operations for the events.
- Define probability using axiomatic approach.
- Apply classical probability to solve problems with finite sample space and equally likely elementary events.

1.3 Deterministic and Random Experiments

The experiments in various fields are fields are usually performed to derived adequate description on the basis of the measurements obtained. The major steps involved in any experiment are:

- ***Input:*** The input of an experiment may consist of equipments, material, input data etc.
- ***Action:*** The experiment is conducted using the input.
- ***Output:*** The action leads to result or results, called output of the experiment.

The experiments can be divided into following two major classes on the basis of performance of their outputs.

Deterministic Experiments: In deterministic experiments, a precisely deterministic input and action yields a precisely deterministic output. If input and action are fixed, we get the same output. We can even predict the output of the experiment.

Examples of Deterministic Experiments: *Some examples of deterministic experiments are the distance covered by a car travelling at a constant speed; Verifying Ohm's law: determining gravitational constant at a place etc.*

In all the above mentioned experiments, for a fixed input and action, we get a fixed output. For instance if a car is traveling with a speed of 60 km./hr., it will cover a fixed distance of 120 km. in two hours journey. In other words, a fixed input of speed v = 60 km/hr. and time t = 2 hr. will yield fixed output of distance $d = v \times t = 120$ km. Similarly, if obtaining gravitational constant at a fixed place, if the experiment is conducted without any errors, each time it will lead to some output, i.e., the value of gravitational constant "g".

Non-deterministic or Random Experiments: In non deterministic or random experiments, fixed input and action does not always yield fixed output. The exact knowledge of input and action does not allow exact prediction of outcome of the experiment.

Examples of Random Experiments: *Some examples of random experiments are, tossing a coin; throwing a dice; Life of an electric bulb; Number of road accidents in a day at Allahabad; Queue size at a railway reservation counter; time taken to download a particular website etc.*

In all these examples of random experiments, fixed input and action does not yield fixed output. For instance, if life of 10 electric bulbs manufactured by the same company under similar conditions is measured, each bulb will have a different lifetime. Some kind of uncertainly is involved in output of these experiments.

In probability theory we are mainly concerned with the random experiments, i.e. the experiments in which some kind of uncertainty is involved.

1.4 Sample Space

In random experiments we cannot predict the exact outcome of the experiment but we know the set of all possible outcomes.

Definition: Sample space is the set of all possible outcomes of a random experiment. Usually we denote it by Ω and a point belonging to Ω by $\omega \in \Omega$.

Some examples of sample space associated with different random experiments:

- In tossing a coin the sample space is given by $\Omega = \{H, T\}$.
- In throwing a dice the sample space is $\Omega = \{1, 2, 3, 4, 5, 6\}$.
- Suppose we toss a coin until we obtain a Head. Then the sample space is given by $\Omega = \{H, TH, TTH, TTTH........\}$

  Here the sample space has countably infinite number of points.
- In observing the life of an electric bulb (in hrs), the sample space is given by

  $\Omega = \{t: 0 \le t < \infty\}$.

  In this example the sample space has uncountable number of points.
- Suppose we toss a coin three times. Then the sample space is

  $\Omega = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$.
- If we toss a coin three times and count the number of H's obtained, the sample space is $\Omega = \{0, 1, 2, 3\}$.
- If we toss a coin until we obtain two H's in succession or two T's (not necessarily in succession). Then the sample space becomes

  $\Omega = \{HH, TT, THH, THT, HTT, HTHH, HTHT\}$
- In observing the queue size at a railway reservation counter, the required sample space is $\Omega = \{0, 1, 2, ...\}$
- Let us consider the time taken (in minutes) to download a website. The sample space is observed as

  $\Omega = \{t : 0 < t < \infty\}$.

1.5 Events

Definition: An event is a set of possible outcomes of a random experiment. Obviously an event is a subset of the sample space. If we denote an event associated with a random experiment by A and Ω is the corresponding sample space then A $\subseteq$ Ω.

Some Examples of Events:

- In tossing a coin, let A be the event that outcome is "H". Then A can be written as A = {H}. Obviously, A is a subset of the corresponding sample space Ω = {H, T}.
- In throwing a dice the event A that outcome is an even number is given by A= {2, 4, 6}.
- Let A be the event that lifetime of an electric bulb is more than 1000 hrs, then A = {t: 1000 < 1 < ∞}.
- In tossing a coin until we obtain a "H", let A be the event that number of tosses is more than 3. Then A = {TTTH, TTTTH}.
- The event that there is no customer in the queue at the railway reservation counter is given by A= {0}.

An event A is said to occur if outcome of the random experiment under consideration has a description that is a member of A.

Thus in throwing a dice, if A= {2, 4, 6}, and outcome of the throw is 4, then A occurs whereas if outcome is 3 then A does not occur.

The sample space Ω of a random experiment is the sure event as the outcome of the experiment is bound to be a member of Ω. In other words Ω is sure to occur. Further, the null or impossible event is denoted by $\emptyset$.

1.6 Algebra of Events

Since an event is subset of the sample space, it is also a set and we can perform different algebraic operations like union, intersection, complement etc. on events.

- ***Union:*** For two events A and B, A∪B is an event which occurs when either A or B (or both) occur, Let A_1, A_2,…….,A_k events, Then $\cup_{i=1}^{k} A_i$ be an event which occurs when at least one of the events A_i occurs.

- ***Intersection:*** For two events A and B, A∩B is an event which occurs when both A or B occur. In general for k events A_1, A_2,…….,A_n events, Then $\bigcap_{i=1}^{k} A_i$ be an event which occurs when all of the events A_i occurs.
- ***Complement:*** The complement of an event A, denoted by A^c, is an event, which occurs whenever A does not occur.
- **Difference:** For two events A and B, A~B = A∩B^c be an event which occurs when A occurs and B does not occur. Obviously $A^c = \Omega \sim A$.
- **Cartesian product:** Let Ω be the sample space associated with a random experiment. If the experiment is repeated twice, The sample space for two repetitions of the random experiment is the Cartesian product.

  $\Omega\times \Omega = \{(\omega_1,\ \omega_2) : \omega_1,\ \omega_2\ \epsilon\ \Omega\ \}$. In general the sample space for n repetitions of the random experiment is $\Omega\times \Omega \ldots\ldots\ldots \Omega$ (times).

  For example, in tossing a coin the sample space is Ω = {H,T}. However, the sample space in tossing a coin two times is the Cartesian product $\Omega\times \Omega$= {(H,H), (H,T), (T,H), (T,T)}.
- If occurrence of event A implies the occurrence of event B, i.e., whenever event A occurs event B also occurs, we denote it by A⇒B. In notation of set theory, If A⇒ B then A⊂B. For example, in throwing a dice let A= {2, 4} be the event that outcome is either 2 or 4 and B = {2, 4, 6} is the event the outcome is an even number then whenever A occurs, B also occurs, i.e., A⇒ B So that A⊆B.

Mutually Exclusive Events: Two events A and B are said to be mutually exclusive or disjoint events, if A ∩ B = ∅ .

Example 1.1: In throwing a dice, if A is the event that outcome is an even number and B is the event that outcome is an odd number, then A and B are mutually exclusive events as A ∩ B = ∅ .

Example 1.2: Let A be the event that life of an electric bulb is more than 1000 hours and B be the event that life of the bulb is more than 200 hours but less than 800 hours, then A and B are mutually exclusive as A ∩ B = ∅ .

1.7 Axiomatic Definition of Probability

Let Ω be the sample space associated with a random experiment. With each event A⊆ Ω, we associated a real number P(A), called the probability of A, satisfying the following four axioms:

Axiom 1: $0 \leq P(A) \leq 1$

Axiom 2: $P(\Omega)=1$

Axiom 3: For two mutually exclusive events A and B, $(A \cap B = \emptyset)$.

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

Axiom 4: For pair wise mutually exclusive events $A_1, A_2, \ldots\ldots\ldots$ $(A_i \cap A_j = \emptyset\ \forall_i \neq j)$,

$$P\left(\bigcup_i A\right) = \sum_i (A_i)$$

A function P satisfying above axioms is called a probability Measure.

Note: Axiom 4 of the probability measure is also called countable additivity.

Result I: The probability of null event is zero, i.e., $P(\emptyset)=0$.

Proof: We observe that $\emptyset \cup \Omega = \emptyset$, i.e., $\emptyset$ and Ω are mutually exclusive. Further

$$\emptyset \cup \Omega = \Omega.$$

Hence,

$P(\emptyset) + P(\Omega) = P(\Omega)$,

which gives $P((\emptyset)= 0$.

P(A) is a measure of how confident we are that the outcome will be in A.

1.8 Classical or Mathematical Definition of Probability

Equally Likely Events: Events $A_1, A_2 \ldots\ldots\ldots A_n$ are said to be equally likely if

$$P(A_1) = P(A_2) = \ldots\ldots\ldots = P(A_n).$$

Elementary Events: Let $\Omega = \{\omega_1, \omega_2, \ldots\ldots \omega_n\}$ be the sample space associated with a random experiment. Then each single point event $\{\omega_1\}$; j= 1, 2,…….n; is called an elementary event.

Let $\Omega = \{\omega_1, \omega_2, \ldots\ldots \omega_n\}$ be the sample space associated with a random experiment and each elementary event $\{\omega_j\}$; j = 1, 2, …….n: be equally likely. Let event A consists of n_A points of Ω . Then

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

Thus probability of event A is equal to the number of favorable cases to event A, n_A divided by the total number of all possible cases n.

Example 1.3: An unbiased dice is thrown. (Unbiased dice means all faces of the dice are equally likely). Let A be the event that we obtain a number less than 4.

Then

$$\Omega = \{1, 2, 3, 4, 5, 6\} \text{ ; } A = \{1, 2, 3\}$$

$$\text{So that, } P(A) = 3/6 = 1/2.$$

Drawbacks of Classical Probability:

(i) Circular definition – The term equally likely is defined is terms of probability and the term equally likely in used in defining probability.
(ii) Fails when all the elementary events are not equally likely.
(iii) Classical probability cannot be used when the sample space has countably infinite or uncountable number of points.

Still classical probability can be used to solve many problems when the sample space is finite and all elementary events are equally likely.

Result II: The classical Probability satisfies all the axioms of a probability measure.

Proof: Let $\Omega = \{\omega_1, \omega_2, \ldots\ldots \omega_n\}$ be the sample space.

1. Since $0 \le n_A \le n$, if follows that $0 \le \frac{n_A}{n} \le 1$ $0r$ $0 \le P(A) \le 1$
2. $P(\Omega) = \frac{n}{n} = 1$
3. Let A and B be two mutually exclusive events. Without loss of generality, we assume that event A consists of n_A points of Ω and B of next n_B points of Ω , i.e.,

$$A = A\{\omega_1, \ldots\ldots\ldots \omega_{n_A}\} \; and \; B \; \left\{\omega_{n_{A+1}} \ldots\ldots\ldots . \omega_{n_{A+n}} \right\},$$

Then

$$A \cup B = \left\{\omega_1, \ldots\ldots\ldots \omega_{n_A}, \omega_{n_{A+1}} \ldots\ldots\ldots . \omega_{n_{A+nB}}\right\}$$

Further,

$$P(A) = \frac{n_A}{n}, \qquad P(B) = \frac{n_B}{n}$$

and

$$P(A \cup B) = \frac{n_A + n_B}{n} = P(A) + P(B).$$

Hence, Axiom 3 is satisfied.

We can easily generalize the proof for the case of more than two events.

1.9 Empirical or Statistical Definition of Probability

Relative Frequency Ratio: Let an event A occurs n_A times in n repetitions of a random experiment, Relative frequency ratio of event A is defined as

$$f_A = \frac{n_A}{n}$$

Obviously.

(i) $0 \leq f_A \leq 1$

(ii) $f_A = 1$, iff, A occurs every time in n repetitions of the experiment;

(iii) $f_A = 0$ iff A never occurs in n repetitions;

(iv) For two mutually exclusive events A and B

$$f_{A \cup B} = f_A + f_B$$

As $n \rightarrow \infty$, f_A converges to P(A) in certain probabilities sense. This property gives the empirical or statistical definitions of probability.

Empirical or Statistical Probability:

The statistical probability approach can be used for estimating probability of an event empirically but it also does not serve as a general definition of probability.

It is only the axiomatic approach, which leads to proper definition of probability. Though the Mathematical and Statistical probability have their own importance, they do not provide a proper definition of probability.

1.10 Some Important Results on Probability

We have some important results as below.

(i) $P(A^c) = 1 - P(A)$

(ii) $P(A \cap B^c) + P(A \cap B) = P(A)$

(iii) For two events A and B, not necessarily mutually exclusive.

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

(iv) If $A \Rightarrow B$ i.e. $A \subset B$ the $P(A) \leq P(B)$.

(v) Probability of occurrence of exactly one of the two events A and B is given by P $[(A \cap B^c) \cup (A^c \cap B)] = P(A) + P(B) - 2P(A \cap B)$.

Proof of the results:

(i) We have

$A \cap A^c = \emptyset$

Thus A and A^c are mutually exclusive. Further

$A \cap A^c = \Omega$

Hence ,

$P(A \cap A^c) = P(\Omega)$

$P(A) + P(A^c) = 1$, (by axioms 2 and 3)

Or,

$P(A) = 1 - P(A^c)$

(ii) We observe that

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

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

$= A \cap \emptyset$

$= \emptyset$

so that $(A \cap B)$ *and* $(A \cap B^c)$ *are mutually exclusive. Further*

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

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

$= A \cap \Omega$

$= A$

Hence

$P[(A \cap B) \cup (A \cap B^c)] = P(A)$

Or,

$P(A \cap B) + P(A \cap B^c) = P(A)$ (using axiom 3).

(iii) We can write

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

Again $A \cup (A^c \cap B) = \emptyset$, so that A and $A^c \cap B$ are mutually exclusive.

Hence,

$P(A \cup B)$

$= P[A \cup (A^c \cap B)]$

$= P(A) + P(A^c \cap B)$

Further, using result (ii) we have

$P(A^c \cap B) = P(A) + P(B) - P(A \cap B)$

Hence we obtain

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

(iv) If $A \Rightarrow B$ i.e. $A \subset B$ then we can write

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

Since A and $A^c \cap B$ are mutually exclusive, we have

$P(B) = P(A) + P(A^c \cap B)$

Further, utilizing $P(A^c \cap B) \geq 0$, the result immediately follows:

(v) The occurrence of exactly one of the two events A and B can be written as $(A \cap B^c) \cup (A^c \cap B)$. Since $(A \cap B^c)$ *and* $(A^c \cap B)$ are mutually exclusive,

We have,

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

$= P\ (A \cap B^c) + P\ (A^c \cap B)$

Further, utilizing the result (ii), we have

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

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

This leads to the required result.

1.11 Self Assessment Exercises

1. An unbiased dice is thrown two times. What is the probability that (i) the two throws leads to the same number, (ii) the number obtained in the first throw is one more than the number obtained in the second throw, (iii) the sum of two number obtained is 5?
2. Out of twenty lottery tickets, three tickets have a cash prize. If three tickets are selected at random, what is the probability that at least one ticket has a cash prize?
3. From all two digit numbers (numbers between 10 and 99), a number is selected at random. What is the probability that (i) number is divisible by 5, (ii) the number is divisible by at least one of the numbers 5 and 7?
4. A person can have birthday on any of the seven days of a week with equal probability. If two people are selected, what is the probability that they both have birthday on the same day of the week?
5. An unbiased coin is tossed six times. What is the probability that (i) two heads are obtained,

(ii) even number of heads are obtained.

6. A person can go from his home to this office using one of the four possible routes. If he selects the route randomly, what is the probability that there are different routes on four consecutive days

7. Verify if the following assignment of probabilities is possible. Give the reason:
   i. $P(A) = 0.4$, $P(B) = 0.8$, $P(A\cap B) = 0.7$
   ii. $P(A) = 0.6$, $P(B) = 0.4$, $P(A\cup B) = 0.5$
   iii. $P(A) = 0.45$, $P(B) = 0.30$, $P(A\cap B^c) = 0.50$

8. If $P(A) = P(B) = 1/4$ and $P(A\cap B)= 1/8$, then obtain the probability of occurrence of
   (a) exactly one of the two events A and B,
   (b) $P(A\cap B^c)$.

9. If $A\subset B$, B and C are mutually exclusive and $P(A) = 1/3$, $P(B) = 1/2$, $P(C) = 1/4$, then obtain $P(A\cup B\cup C)$.

10. For two events A and B, if $P(A) = 0.4$, $P(B) = 0.6$ and $A\Rightarrow B$. Find $P(B/A)$.

11. IF $P(A) = 1/2$, $P(A\cup B) = 5/6$, and $P(A\cap B) = 1/3$, then obtain $P(B)$.

12. For three events A, B and C prove that
    $P(A\cup B\cup C) = P(A) + P(B) + P(C) - P(A\cap B) - P(A\cap C) - P(B\cap C) + P(A\cap B\cap C)$.

13. For two events A and B, prove that
    $P(A) + P(B) -1\leq P(A\cap B) \leq$ minimum of $\{ P(A), P(B)\} \leq$ maximum of $\{P(A), P(B)\} \leq P(A\cup B) \leq P(A) + P(B)$.

14. Let $P(A) = P(B) = P(C) = 1/2$, $P(A\cap B) = P(B\cap C) = P(A\cap C) = 1/4$, and $P(A\cap B\cap C) = 1/6$.
    Find (i) P [exactly one of the events occur]
    (ii) P (at least two of the events occur),
    (iii) P (at most two of the events occur).

15. Let $P(A) = x$, $P(B) = y$, and $P(A\cap B) = z$. In terms of x, y and z find the probability that (i) exactly one of the two events will occur, (ii) none of the two events A, B occur.

16. A fair dice is thrown twice. Let A be the event that first throw shows an even number and B is the event that sum of the numbers obtained in two throws is an number then obtain $P(A\cap B)$.

17. For k events $A_1,\ldots\ldots A_k$, show that.

$$(i)\ \ P\left(\bigcup_{i=1}^{k} A_i\right) \leq \sum_{i=1}^{k} (A_i)$$

$$(ii)\quad P\left(\bigcap_{i=1}^{k} A_i\right) \geq \sum_{i=1}^{k} P(A_i) - \ (k-1)$$

1.12 Answers/ Suggestions

1. (i) 1/6 (ii) 5/12 (iii) 1/12

2. The number of ways in which exactly one ticket one ticket with case prize can be selected is $\binom{3}{1}$. Similarly you can obtain number of ways in which exactly two or three tickets can be selected. Total number of possible ways in which is 3 tickets can be selected out of 20 tickets is $\binom{20}{3}$. The required probability is 7/1140. Since 7= $\binom{3}{1}+\binom{3}{2}+\binom{3}{3}$.

3. (i) $\frac{18+13-2}{90} = \frac{29}{90}$
4. 1/7
5. $\frac{\binom{6}{2}}{2^6}$ $\quad (ii)\ \frac{\binom{6}{4}-\binom{6}{5}-\binom{6}{6}}{2^6}$
6. $\frac{4^1}{4^4} = \frac{3}{32}$
7. (i) No because P(A∩B) can not be greater than either P(A) or P(B).
   (ii) No (iii) No
8. (i) 1/4 (ii) 1/8
9. P(A∩B) = P(A), P(B∩C) = 0, P(A∩C) is also 0 as A⊂B, P(A∩B∩C) = 0. Then P(A∪B∪C) = 3/4.
10. P(B-A) = 0.2
11. 2/3
14. (i) P [exactly one of the three events occur] = P(A)- P(A∩B)- P(A∩C)+ P(A∩B∩C)+ [P (B)- P(A∩B)- P(B∩C)+ P(A∩B∩C)] + [P (C)- P(B∩C)- P(A∩C)+ P(A∩B∩C)] = 7/8.

(ii) 3/8. (iii) 1-P (all the three events occur) = 7/8

15. (i) x+y-2z, (ii) 1- (x + y-z)

16. 1/6 (Notice that A⊂B)

17. Use method of indication to prove the result

1.13 Summary

Those experiment in which outcome depends on chance are called random or non-deterministic experiments. The set of all possible outcomes of a random experiment is known as sample space. A sub set of sample space is defined as an event. Various definitions of probability of occurrence of an event are given. In classical or mathematical definition of probability, it is defined as the ratio of favorable outcomes to the total number of outcomes provided outcomes in the sample space are mutually exclusive, exhaustive and equally likely to occur. In statistical definition, we calculate the probability by relative frequency or occurrence of event provided experiment has been repeated essentially under the similar condition for sufficiently large number of times. In axiomatic definition, it is defined as a real number lying between 0 and 1 (inclusive of both) provided following three conditions are satisfied (i) P(S) =1 (ii) P(∅)= 0 (iii) $P\left(\bigcup_{i=1}^{k} A_i\right) = \sum_{i=1}^{k} P(A_i)$ where A_i^5 are mutually exclusive events.

1.15 Further Readings

1. Cramer H, Mathematical Methods of Statistics, Princeton University Press, 1946 and Asia Publishing House, 1962.
2. Hogg R.V. and Craig A.T., Introduction to Mathematical Statistics, Macmillan, 1978.
3. Prazen E. Modern Probability Theory and its applications, John Wiley, 1960 and Wiley Eastern 1972.
4. Rao C.R., Linear Statistical Inference and Its Applications, John Wiley, 1960 and Wiley Eastern 1974.
5. Rahtagi V.K. (1984), An Introduction to Probability Theory and Mathematical Statistics, John Wiley, 1976 and Wiley Eastern 1985.
6. Vikas S.S., Mathematical Statistics, John Wiley, 1962 Toppan.

Unit-2: Conditional Probability

Structure

2.1 Introduction

In events associated with a random experiment, the information about the occurrence (or non occurrence) of an event may influence the probability of occurrence of some other events. The objective of present unit is to introduction the concept of conditional probability of an event subject to the information that some other event has already occurred. Various results regarding the conditional probability have been given. The basic concepts of independence of two events and mutual independence of several events have been introduced.

2.2 Objectives

After going through this unit you should be also to:

- Understand the basic concept of conditional probability and able to obtain conditional probabilities for events.
- Define and differentiate between pair wise independence and mutual independence of events.
- Prove and apply Bay's theorem for solving problems.

2.3 Conditional Probability

Example 2.1: Consider a box containing 100 computer chips out of which 40 are defective. Two chips are selected randomly (i) with replacement; (ii) without replacement. Let

A= {first item is defective}

B= {second item is defective}

In case of with replacement, P(A) = P(B) = 40/100 = 2/5.

In case of without replacement,. P(A) = 2/5. If A has already occurred, probability of occurrence of B is 39/99. If A has not occurred, probability of occurrence of event B is 40/99. Hence the information about the occurrence of event A affects the probability of occurrence of event B.

Definition: Let A, B be two events defined on the same sample space Ω. Then conditional probability of B given A denoted by P(B|A), is defined as

$$P(B|A)=\frac{P\,(A\cap B)}{P(A)}, \quad provided\ P(A) \neq 0.$$

For further motivation, consider mathematical probability. Let n be the total number of points in Ω, n_A be the number of cases favorable to A, n_B be the number of cases favorable to B and $n_{A\cap B}$ be the number of cases favorable to $A \cap B$. Then

$$P\,(B|A) = \frac{n_{A\cap B}}{n_A} = \frac{n_{A\cap B}/n}{n_A/n} = \frac{P(A \cap B)}{P(A)}$$

Result: The conditional probability satisfies all the axioms of a probability measures, that is,

(i) $0 \leq P(B/A) \leq 1$;

(ii) P(A|A)= 1;

(iii) $P\,(B_1 \cup B_2/A) = P(B_2/A) + P(B_1/A); \quad if\ B_1 \cap B_2 = \Phi;$

(iv) $P(\cup_{i=1}^{n} B_i\,/A) = \sum_{i=1}^{n} P(B_i/A);$

If $B_i's$ provided are pair wise mutually exclusive.

Poof:

(i) Since $A \cap B \subset A$, we have
$P(A\cap B) \leq P(A)$
Or,
$0 \leq P(B|A) = P(A\cap B) / P(A) \leq 1$

(ii) We have
P(A|A)
$$= \frac{P(A\cap A)}{P(A)} = \frac{P(A)}{P(A)} = 1$$

(iii) Since B_1 and B_2 are mutually exclusive, $A \cap B_1$ and $A \cap B_2$ are also mutually exclusive. Hence

$$P(B_1 \cup B_2|A)$$

$$= \frac{P[A \cap (B_1 \cup B_2)]}{P(A)}$$

$$= \frac{P[(A \cap B_1) \cup (A \cap B_2)]}{P(A)}$$

$$= P(B_1|A) + P(B_2|A)$$

(iv) We can prove it along the same lines as (iii)

Note:

(i) For the conditional probability P(B|A), A behaves as a new sample space.

(ii) The unconditional probability of an event A, P(A) may be viewed as a conditional probability of event A given Ω.

2.4 Multiplicative Theorem of Probability

From the definition of conditional probability, we observe that

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

This is called ***Multiplicative Theorem of Probability.***

Extension of Multiplicative Theorem: For n events $A_1, A_2 \ldots . A_n$.

$$P(A_1 \cap A_2 \cap \ldots \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \ldots \ldots P(A_n|A_1 \cap \ldots \cap A_{n-1})$$

Proof: Considering the right hand side, we observe that

$$P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \ldots \ldots P(A_n|A_1 \cap \ldots \cap A_{n-1})$$

$$= P(A_1).\frac{P(A_1 \cap A_2)}{P(A_1)}.\frac{P(A_1 \cap A_2 \cap A_3)}{P(A_1 \cap A_2)} \ldots .\frac{P(A_1 \cap A_2 \cap \ldots \cap A_n)}{P(A_1 \cap A_2 \cap \ldots \cap A_{n-1})} = P(A_1 \cap A_2 \cap \ldots \cap A_n)$$

= Left had side

Example 2.2: In the example 2.1 for the without replacement case, P(A) = 40/100, P(B|A) = 39/99, $P(B|A^C)$ = 40/99. Then the probability that both the select chips are defective is given by

$$P(A \cap B) = P(B|A)P(A) = 39 \times 40/99\ 100. = \frac{40 \times 39}{100 \times 99}$$

In fact, P(A∩B) is probability of simultaneous occurrence of events A and B.

2.5 Independent Events

If A and B are two mutually exclusive events then P(B∩A)= 0 and P(B|A)= 0. If A⇒B then A⊆B and (A∩B) =A. So that P(B∩A) = P(A) and P(B|A) = 1. In both the cases, knowledge about the occurrence of event A gives some definite information about the occurrence of event B. However, if P(B|A) = P(B), then we can say that the information about the occurrence of event A does not have any effect on the probability of event B and, in this sense A and B are independent

Definition: Two events A and B are said to be independent whenever

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

Note: If A and B are independent events then P(B|A) = P(B) and P(A|B) = P(A), provided

P(A) > 0, P(B) > 0.

Definition: The n events A_1, A_2,……..A_n, are said to be mutually independent, if for all r = 2, 3,…….. n.

$$P\left(A \cap A_{i_k}\right) = P\left(A_{i_1}\right) \ldots\ldots P\left(A_{i_k}\right);\ i_1 \neq i_2 \neq \cdots \ldots\ldots i_k$$

We can write the above condition for mutual independence in detail as

$$P\left(A_i \cap A_j\right) = P(A_i)\, P\left(A_j\right) for\ i \neq j.$$

$$P\left(A_i \cap A_j \cap A_k\right) = P(A_i)\, P\left(A_j\right) P(A_k) \text{ for } i \neq j \neq k.$$

$$P(A_1 \cap A_2 \cap A_3 \ldots\ldots\ldots. \cap A_n) = \mathrm{P}(A_1) P(A_2). \ldots P(A_n)$$

In all we have 2^n-n-1 conditions for mutual independence.

Note: ***Obviously, if A_1, A_2 ……,A_n are mutually independent, then they are pariwise independent also.*** However its converse is not always true.

Example 2.3: ***(Counter example in which the events are pair wise independent but not mutually independent).***

A fair dice is thrown two times. Let A be the event that an even number comes up in the first throw. B be the event that an even number turns up in the second throw and C be the event that both the throws lead to the same number. Then total number of points in the sample space is 6 × 6 = 36. Now

A= {(2,1) ,(2,2), (2,3), (2,4), (2,5), (2,6), (4,1), (4,2), (4,3), (4,4), (4,5), (4,6), (6,1) (6,2), (6,3) (6,4), (6,5), (6,6)}

B= {(1,2), (1,4), (1,6), (2,2) (2,4), (2,6), (3,2), (3,4), (3,6), (4,2), (4,4), (4,6), (5,2), (5,4), (5,6), (6,2) (6,4), (6,6)}

C= {(1,1), (2,2), (3,3), (4,4), (5,5), (6,6)}

Then

P(A) = 18/36 = 12

P(B) = 18/36 = 1/2

P(C) = 6/36 = 1/6

A∩B is the event that both the throws lead to the even number, i.e.

A∩B = {(2,2), (4,4), (6,6)}

So that

P(A∩B)= 3/36= 1/12.

Further A∩C be the event that outcome of first throw is an even number and outcome of both the throws is same number. Thus A∩C = {(2,2), (4,4), (6,6)} and P(A∩C) = 1/12

Similarly B∩C be the event that outcome of second throw is an even number and outcome of both the throws is same number. Thus B∩C = {(2,2), (4,4), (6,6)} and P(B∩C) = 1/12

A∩B∩C is the event that both the throws lead to same event number and A∩B∩C = {(2,2), (4,4), (6,6)} thus A∩B∩C = 1/12.

Hence we observe that P(A∩B) = P(A). P(B), P(B∩C) = P(B). P(C), and P(A∩C) = P(A). P(C) but P(A∩B∩C) = P(A). P(B). P(C), i.e. A, B and C are pair wise independent but not mutually independent.

2.6 Partition of the Sample Space and Baye's Theorem

Theorem of total probability:

Partition of the Sample Space: We say that the B_1, ………. B_k define a partition of the sample space Ω if.

$$(a) \qquad B_i \cap B_l = \emptyset \qquad \forall \quad i \neq j$$

$$(b) \qquad \bigcup_{i=1}^{k} B_i = \quad \Omega$$

$$(c) \qquad P(B_i) > 0 \qquad\qquad \forall \;\; i$$

Result: Let $B_1, \ldots\ldots\ldots, B_k$ be a partition of the sample space Ω and A be an event (A⊂ Ω), Then

$$P(A) = \sum_{i=1}^{k} P(A|B_i)P(B_i)$$

Proof: Since $B_1, \ldots\ldots\ldots, B_k$ define a partition of the sample space Ω , we have

$$B_i \cap B_j = \emptyset, \qquad\qquad \forall \;\; i \neq j$$

And

$$\bigcup_{j=1}^{k} B_j = \quad \Omega$$

Further, A can be written as

$A = A \cap \Omega$

$= A \cap \left(\cup_{j=1}^{k} B_j\right)$

$= \cup_{j=1}^{k} \left(A \cap B_j\right)$

Hence using axiom 4 of probability measure, for pair wise mutually exclusive events A_1, A_2, $A_3 \ldots\ldots\ldots A_n$, we have

$$P\left(\bigcup_i A_i\right) = \sum_i P(A_i),$$

We observe that

$$P(A) = \sum_{j=1}^{k} P(A \cap B_j)$$

$= \sum_{j=1}^{i} P\left(A \middle| B_j\right) P(B_j)$ (by multiplicative law of probability)

Baye's Theorem:

Let $B_1, \ldots\ldots\ldots, B_k$ be a partition of the sample space Ω and A be an event (A⊂ Ω), Then

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_{j=1}^{k} P(A|B_j)P(B_j)}$$

Proof: We have

$$P(B_i|A) = \frac{P(A \cap B_i)}{P(A)}$$

$$= \frac{P(A|B_i)P(B_i)}{P(A)}$$

Where the last equality follows form the multiplicative law of probability. Further, from the previous result, we have,

$$P(A) = \sum_{j=1}^{k} P(A|B_j)P(B_j)$$

Substituting the value of P(A), we obtain

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_{j=1}^{k} P(A|B_j)P(B_j)}$$

Hence the result

Remark:

Here the probabilities $P(B_j)$'s are known as **"a priori (or prior) probabilities"**. They exist before we gain any information about A (the result of an experiment)

The probabilities $P(A|B_j)$'s are known as "**likelihoods".** They indicate how likely event A to occur under the information that B_j occurs.

The probabilities $P(B_j|A)$'s are known as '**a posteriori (or posterior) probabilities".** They are determined after the results of the experiment are known.

Example 2.4 A bag contains 10 coins out of which five coins of type 1 are unbiased and remaining five coins are biased. Among the biased coins, four coins of type II have probabilities of head 1/3 are remaining one coins of type III have probability of head 9/10. A coin is selected at random and tossed three times. If we get heads in a row, what is the probability that the coin is of type III.

Solution: Let

B_1 : Event the coin is of type I

B_2 : Event the coin is of type II

B_3 : Event the coin is of type III

A: Event that we get three heads in a row.

Given that $P(B_1) = 0.5$ $P(B_2) = 0.4$, $P(B_3) = 0.1$

$P(A|B_1) = (1/2)^3$, $P(A|B_2) = (1/3)^3$, $P(A|B_3) = (9/10)^3$

Then the probabilities that the coin is of type III given that three heads have been obtained in a row, i.e., $P(B_3|A)$, is given by

$$P(B_3|A) = \frac{P(A|B_3)\,P(B_3)}{P(A|B_1)\,P(B_1) + P(A|B_2)\,P(B_2) + P(A|B_3)\,P(B_3)}$$

$$= \frac{\left(\frac{9}{10}\right)^3 \times 0.1}{\left(\frac{1}{2}\right)^3 \times 0.5 + \left(\frac{1}{3}\right)^3 \times 0.4 + \left(\frac{9}{10}\right)^3 \times 0.1}$$

$$= 0.46$$

2.7 Self Assessment Exercises

1. Let A, B and C be three events such that A and C are mutually exclusive, B and c are independent and $P(A) = 1/12$, $P(B) = 1/3$, $P(C) = 1/4$ and $P(A \cap B^C) = 0$, Find
   - (i) $P(A \cap B \cap C)$
   - (ii) $P(A \cup B \cup C)$
   - (iii) P(exactly one of the event A,B,C occur)
   - (iv) $P(C|A)$, (v) $P(B|A)$.
2. For two events A and B if $A \Rightarrow B$, $P(A) = 1/3$, $P(B) = 1/2$ then find

   (i) $P(B|A)$, (ii) $P(A|B)$
3. If $P(A) = P(B) = 1/3$ and $P(A \cap B) = 1/6$. Then obtain (a) $P(B|A)$, (b) $P(A|B^C)$.
4. Prove that for three events A,B,C.
   - (i) $P(A \cup B|C) = P(A|C) + P(B|C) - P(A \cap B|C)$
   - (ii) $P(A \cap B|C) + P(A \cap B^C|C) = P(A|C)$
   - (iii) If $B \Rightarrow C$ and $P(A)>0$ then $P(B|A) \leq P(C|A)$
   - (iv) If $B \Rightarrow C$ then $P(C|B) = 1$.
5. Prove that for two events A and B, if $P(B|A) > P(A|B) > P(A)$.
6. In throwing two fair dice the events A, B and C are defined as

A = “First toss results in a 1, 2, or 3.”

B = “Second toss results in 4, 5, or 6.”

C = “ The sum of the two face is 7.”

Show that A, B and C are pair view independent but not mutually independent.

7. Suppose that factory has to complete a project, which, which may get delayed because of a strike. The probability that there will be a strike is 0.20. The probability that the project will be completed in time if there is no strike is 0.90 and the probability that the project will be completed on time if there is a strike is 0.45. What is the probability that the project will be completed on time? If the project is completed on time, what is the probability that there was no strike?

8. A machine hired by a company could have failed as the result of one of the four possible reasons. (i) due to voltage fluctuation in electricity supply, (ii) due to malfunctioning of equipment (iii) due to carelessness in handling the machine, (iv)due to sabotage. Interviews with manager analyzing the risks involved led to the conclusion that failure would occur with probabilities 0.15 as result of voltage fluctuation in electricity supply, with a probability 0.20 due to malfunctioning of equipment, with probability 0.60 due to carelessness and with probability 0.25 due to sabotage. The prior probabilities of the four causes of machine failure are, respectively, 0.15, 0.30, 0.35 and 0.20 based on this information.
   (a) Find the probability of a failure at the construction site.
   (b) If a failure has occurred at the construction site, what is the most likely cause of failure?

9. An unbiased coin is tossed four times. Let A be the event that first, two tosses lead to the same outcome (both heads or both tails) and B is the event that last three tosses lead to the same outcome (all heads or all tails), then find P(B|A).

10. A finite discrete sample space is consist of the four points denoted by {(100), (001), (010), (111)}.

    Each point has been assigned probability V4 and a point is selected randomly. Let A_i (1=1,2,3) be the event, which occurs if there is a 1 at the i-th place. Thus A_1 = {(100), (111)}. Show that A_1, A_2, A_3 are pair wise independent but not mutually independent.

2.8 Answers/ Suggestions

1. (i) 0, (ii) 7/12 (iii) 1/3 (iv) 0, (v) 1.
2. (i) 1, (ii) 2/3.
3. (a) 12 (b) V4
7. P (Project is completed on time) = 0.81, P(No strike | project is completed on time)= 8/9
8. Let A be the event that machine has failed B_1 is the event that cause of failure is voltage fluctuation, B_2, is the event that cause of failure is malfunctioning, B_3 is the event that cause of failure is carelessness and B_4 is the event that cause of failure is sabotage. Then $P(A|B_1) = 0.15$, $P(A|B_2) = 0.2$, $P(A|B_4) = 0.25$, $P(B_1) = 0.15$, $P(B_2) = 0.3$, $P(B_3) = 0.35$, $P(B_4) = 0.2$ Then
   (a) $P(A) = 0.3425$
   (b) $P(B_1|A) = 0.0657$, $P(B_2|A) = 0.175$, $P(B_3|A) = 0.613$, $P(B_4|A) = 0.146$

   Hence most likely cause of failure is Carelessness
9. A = {HHHH, HHHT, HHTH, HHTT, TTHH, TTHT, TTTH, TTTT}

   B= {HHHH, HTTT, THHH, TTTT}

   A∩B= {HHHH, TTTT}

   P(B|A)= ¼.

2.9 Summary

When occurrence of an event is affected by certain conditions and the numeric value of probability of occurrence of an event varies as per these conditions, then it is conditional probability. In this case, events are not mutually independent and occurrence of one event may affect the occurrence of others.

Bays's theorem revises the initial probabilities of occurrence of events (which partition the sample space) when as a particular event occurs.

2.10 Further Readings

1. Cramer H, Mathematical Methods of Statistics, Princeton University Press, 1946 and Asia Publishing House, 1962.
2. Hogg R.V. and Craig A.T., Introduction to Mathematical Statistics, Macmillan, 1978.
3. Prazen E., Modern Probability Theory and its applications, John Wiley, 1960 and Wiley Eastern 1972.

4. Rao C.R., Linear Statistical Inference and Its Applications, John Wiley, 1960 and Wiley Eastern 1974.
5. Rohtagi V.K. (1984): An Introduction to Probability Theory and Mathematical Statistics, John Wiley, 1976 and Wiley Eastern 1985.
6. Vikas S.S., Mathematical Statistics, John Wiley, 1962 Toppan.

U.P. Rajarshi Tandon Open
University, Prayagraj

UGSTAT – 102

Probability, Distribution and Statistical Inference

Block

2

Probability Distributions and Expectations

Unit – 3

Random Variables and Probability Distributions

Unit – 4

Mathematical Expectation

Unit - 5

Inequalities for Moments

Course Design Committee

Dr. Ashutosh Gupta **Chairman**
Director, School of Sciences
U. P. Rajarshi Tandon Open University, Prayagraj

Prof. Anup Chaturvedi **Member**
Ex. Head, Department of Statistics
University of Allahabad, Prayagraj

Prof. S. Lalitha **Member**
Head, Department of Statistics
University of Allahabad, Prayagraj

Prof. Himanshu Pandey **Member**
Department of Statistics, D. D. U. Gorakhpur University,
Gorakhpur.

Dr. Shruti **Member-Secretary**
Sr. Assistant Professor, School of Sciences
U.P. Rajarshi Tandon Open University, Prayagraj

Course Preparation Committee

Dr. Shruti **Course Coordinator**
Sr. Assistant Professor, School of Sciences
U. P. Rajarshi Tandon Open University, Prayagraj

Prof. S. K. Pandey **Writer**
Department of Statistics, Lucknow University, Lucknow

Prof. Anoop Chaturvedi **Writer**
Department of Statistics, University of Allahabad
Prayagraj

Dr. Shruti **Writer**
School of Sciences, U. P. Rajarshi Tandon Open University,
Prayagraj

Prof. V. P. Ojha **Reviewer**
Department of Statistics and Mathematics,
D. D. U., Gorakhpur University, Gorakhpur

Prof. G. S. Pandey **Reviewer**
Department of Statistics, University of Allahabad,
Prayagraj

Dr. Alok Kumar **Reviewer**
Department of Community Medicine, Institute of
Medical Sciences, Banaras Hindu University, Varanasi

Prof. S. K. Pandey **Editor**
Department of Statistics, Lucknow University, Lucknow

First Edition: *March 2008* (Published with the support of the Distance Education Council, New Delhi)

Second Edition: ***July 2020***

ISBN :

Printed and Published by Dr. Arun Kumar Gupta Registrar, Uttar Pradesh Rajarshi Tandon Open University, 2020.

Printed By: Chandrakala Universal Pvt. 42/7 Jawahar Lal Neharu Road, Prayagraj.

Unit-3: Random Variables and Probability Distributions

Structure

3.1 Introduction

In the previous two units of block 1 you have noticed that in many examples of the random experiments, the outcomes are not numbers. In such situations, it is more convenient for further analysis to assign a real number to each ω point of the sample Ω. The main objective of defining a random variable is to assign numerical values to different points of the sample space.

3.2 Objectives

After going through this unit the students should be able to

- Define random variable and understand the difference between discrete and continuous random variables.
- Define and obtain probability mass function and cumulative distribution function of a continuous random variable.
- Define and obtain probability density function and cumulative distribution function of a continuous random variable

- Obtain different probabilities for discrete and continuous random variables using probabilities mass function/ probability density function.

3.3 Random Variables

Example 1: In tossing a coin the sample space is

$$\Omega = \{H, T\}.$$

We may assign number 1 to 'H' and number 0 to 'T'. In other words, we define a function X(.), such that

X(H)=1 and X(T)=0.

Definition: A random variable X is a real valued function defined on Ω satisfying the condition that for each real number x.

$$\{\omega : X(\omega) \leq x\} \subseteq \Omega.$$

The domain of the random variable X is Ω and its range is entire real line or a subset of real line. We denote the range of X by Rx.

Example 2: Suppose we toss an unbiased coin two times. Then the sample space is

$$\Omega = \{HH, HT, TH, TT\}.$$

Let us write ω_1 = HH, ω_2 = HT, ω_3= TH, ω_4= TH. If we define X(ω) = number of H in ω, then X(ω_1) = 42, $X(\omega_2)$= X (ω_3) = and X(ω_4) = 0. Thus X assigns number 0 to TT, 1 to HT and TH and number 2 to HH. Then for each real x you observe that,

$$\{\omega : X(\omega) \leq x\} = \begin{Bmatrix} \emptyset & if\ x < 0 \\ \{TH\} & if\ 0 \leq x < 1 \\ \{HT, TH, TT\} & if\ 1 \leq x < 2 \\ \Omega & if\ x > 2 \end{Bmatrix}$$

The range of X is R_x = {0, 1, 2}

3.4 Cumulative Distribution Function and its Properties

For a random variable X, the function

$$F(x) = P[\{\omega : X(\omega) \leq x\}] = P[X \leq x]$$

is called the ***Cumulative Distribution Function (cdf)*** of the random variable X.

Example 3: Let X be the number obtained in throwing an unbiased dice. The cdf of X is given by

$$F(x) = \begin{cases} 0 & if\ x < 1 \\ 1/6 & if\ 1 \leq x < 2 \\ 2/6 & if\ 2 \leq x < 3 \\ 3/6 & if\ 3 \leq x < 4 \\ 4/6 & if\ 4 \leq x < 5 \\ 5/6 & if\ 4 \leq x > 6 \\ 1 & if\ x > 6 \end{cases}$$

The graph of F(x) against x is given by

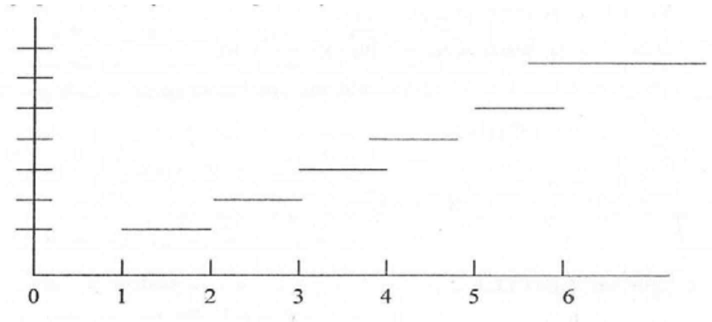


Figure 1

From the above graph, you observe that the cdf is a step function with jumps at the points of discontinuity of F(x), i.e., 1, 2, 3, 4, 5 and 6. The amount of jump at a point is equal to the probability of that point 1/6.

Some properties of Cumulative Distribution Function are

(i) $0 \leq F(x) \leq 1$.

(ii) For a < b, $P(a < X \leq b) = F(b) - F(a)$.

(iii) If a < b, then $F(a) \leq F(b)$, i.e., cdf F(x) is monotonic non-decreasing function of x.

(iv) F(x) is right continuous.

(v) We have $F(\infty) = 0$ and $F(\infty) = 1$.

Lim $F(x) = F(\infty) = 1$; $\lim_{x\to-\infty} F(x) = F(-\infty) = 0$.

Proof:

(i) Obvious as probability lies between 0 and 1.

(ii) You can write $(-\infty, b] = (-\infty, a\,] \cup (a, b]$.

Since $(-\infty, b]$ and (a,b] are mutually exclusive you observe that

$$P\,\{X\epsilon(-\infty, b]\} = P\{X\epsilon(-\infty, a]\} + P\{X\epsilon(a, b]\}$$

Or

$$F(b) = F(a) + P[\, a < X \leq b\}$$

Which leads to the required result.

(iii) In the previous result, since $0 \leq P[a < X \leq b] \leq F(b)$.

(iv) For showing that F(x) is right continuous you have to prove that for h > 0.

$$\lim_{x \to -\infty} F[(x+h) - F(x)] = 0.$$

Since

$$F(x+h) - F(x) = P(x < X \leq x+h)$$

We have

$$\lim_{x \to -\infty} F[(x+h) - F(x)] = 0.$$

(v) Notice that

$$\{\omega \colon X(\omega) < \infty\} = \Omega \text{ so that } P\{\omega \colon X(\omega) < \infty\} = P(\Omega) = 1, \text{which implies that } F(\infty) = 1$$

Further

$$\{\omega \colon X(\omega) < -\infty\} = \emptyset \text{ so that } P\{\omega \colon X(\omega) < \infty\} = P(\emptyset) = 0, \text{which gives } F(-\infty) = 0$$

Result: The cdf F(x) is left continuous at point x if and only if P(X = x) = 0.

Proof: For left continuity of F(x) we have to show that for h > 0.

$$\lim_{x \to -\infty} F[(x) - F(x-h)] = 0$$

Since

$$F(x) - F(x-h) = P(x-h < X \leq x)$$

Taking limit h→0, you obtain,

$$\lim_{x \to -\infty} [F(x) - F(x-h) = P(X = x).$$

Note: For the cdf of a random variable to be continuous at a point X = x, P(X = x) must be equal to 0.

3.5 Discrete Random Variable and Probability Mass Function

A random variable X is called a discrete random variable if it takes finite or countable infinite number of values. Thus the range R_x of a discrete random variable has countable number of points.

Let X be a discrete random variable with $R_x = \{x_1, x_2, \ldots\ldots x_i, x_j \ldots\ldots\}$, with each possible outcome x_i we associate a number $P(x_i) = P(X = x_i)$; i= 1,2,…. Satisfying the following conditions:

(i) $p(x_i) \geq 0$ for all i = 1, 2,…..

(ii) $\sum_{i=1}^{n} p(x_i) = 1$

The function p(x) satisfying the above conditions is called the probability mass function (pmf) of the random variable X.

The collection of pairs (xi, p(xi)); i = 1, 2…. is called the probability distribution of X. the cdf of X is given by

$$F(x) = i \sum_{i, x_i \le x} p(x_i)$$

The values x_1, x_2….of X with which we associated positive probabilities are called mass points and $p(x_i)$ is the probability mass associated with mass point x_i.

Example 4: Let X takes values 0, 1, 2, 3, 4, with corresponding probabilities 1/4, 1/3, 1/8, 1/6, 1/8 respectively, obviously X is a discrete random variable with R_x = {0, 1, 2, 3, 4}.

The pmf of X can be written as

$$p(x) = \begin{cases} \frac{1}{4} & if\ x = 0 \\ \frac{1}{3} & if\ x = 1 \\ \frac{1}{8} & if\ x = 2, 4 \\ \frac{1}{6} & if\ x = 3 \\ 0 & otherwise \end{cases}$$

The cdf of X is given by

$$f(x) = \begin{cases} 0 & if\ x < 0 \\ \frac{1}{4} & if\ 0 \le x < 1 \\ \frac{1}{4} + \frac{1}{3} = \frac{7}{12} & if\ 1 \le x < 2 \\ \frac{1}{4} + \frac{1}{3} + \frac{1}{8} = \frac{17}{24} & if\ 2 \le x < 3 \\ \frac{1}{4} + \frac{1}{3} + \frac{1}{8} + \frac{1}{6} = \frac{21}{24} & if\ 3 \le x < 4 \\ \frac{1}{4} + \frac{1}{3} + \frac{1}{8} + \frac{1}{6} + \frac{1}{8} = 1 & if\ x \ge 4 \end{cases}$$

If the possible values of X are arranged in ascending order so that $x_1 = x_2$, then the cdf of X can be written as

$$F(x) = \begin{cases} 0 & if\ x < x_1 \\ p(x_1) & if\ x_1 \le x < x_3 \\ p(x_1) + p(x_2) & if\ x_2 \le x < x_3 \end{cases}$$

Obviously F(x) is step function with jumps at mass points x_1, x_2……..The magnitude of jump at mass point xi is p(xi).

The domain of the pmf p(x) is R_x and the range is interval [0, 1].

Example 5: The range of r.v. X has a finite number of possible values, say, $R_x = \{x_1, \ldots\ldots, x_n\}$ with $x_1 < \ldots\ldots < x_n$ and equal probabilities associated with each mass point then find the pmf of x is given by

$$p(x) = 1\begin{cases} \frac{1}{n}, & if\ x = x_1, x_2 \ldots\ldots x_n \\ 0 & otherwise \end{cases}$$

Further the cdf of X is

$$F(x) = \begin{cases} 0 & if\ x < x_1 \\ \frac{1}{n}, & if\ x_1 \le x < x_2 \\ \frac{2}{n}, & if\ x_2 \le x < x, \\ \cdot & \\ \cdot & \\ 1 & if\ x \ge x_n \end{cases}$$

Example 6: The pmf of a r.v. X is given by

$$p(x) = \begin{cases} \frac{k}{4^x}, & if\ x = 0, 1, 2, \ldots. \\ 0, & otherwise \end{cases}$$

For obtaining the constant k, we use the property of the pmf which gives k = 3/4. Suppose we want to obtain the probabilities (i) P(x > 3.5) (ii) P(x < 6 | X > 3). Then

$$P(X > 3.5) = \sum_{x=4}^{\infty} \frac{3}{4}\left(\frac{1}{4}\right)^x = \frac{1}{64}$$

Further,

$$P(x < 6 \mid X > 3) = \frac{P(3 < X < 6)}{P(X > 3)} = \frac{p(4) + p(5)}{\sum_{x=4}^{\infty} p(x)} = \frac{\frac{3}{4}\left(\frac{1}{4}\right)^4 + \frac{3}{4}\left(\frac{1}{4}\right)^5}{\sum_{x=4}^{\infty} \frac{3}{4}\left(\frac{1}{4}\right)^x} = \frac{15}{16}$$

3.6 Continuous Random Variable and Probability Density Function

Let X be a random variable with the range R_x entire real line or an interval of the real line and the cdf F(x) of X is a continuous function of x over R_x. Then X is a continuous random variable. Some examples of continuous random variables are (i) time taken to download a particular website, (ii) survival time of a cancer patient after diagnosis.

Notice that if F(x) is continuous over R_x, then probability of a particular point is 0 and, like discrete random variables, we can not associated positive probabilities with different possible values of X. However, with each x, we define a function f(x), called the probability density function of X as follows:

Definition: For continuous random variables, there exists a function f(x) called the probability density function (pdf) of X, such that

(i) $f(x) \geq 0$ for all x.

(ii) $\int_{\infty}^{\infty} f(x)dx = 1$

(iii) For real number a, b with $-\infty < a < b < \infty$, the probability that X lies in interval (a,b] is given by

$$P(a < X \leq b) = \int_{a}^{b} f(x)dx$$

Since for a continuous random variable X, the probability of a single point P(X = x) = 0 for all x, we have

$$P(a < X \leq b) = P(a \leq X \leq b) = P(a < X < b)$$

The cdf of X is given by

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(x')dx'$$

Conversely, the pdf of X, in terms of cdf, is given by

$$f(x) = \frac{d}{dx} F(x)$$

Example 7: Let us verify if the following function is the pdf of a random variable X.

$$f(x) = 4x - 2x^2 - 1, \quad if\ 0 < x < 2$$

$$= 0\,, otherwise$$

The above function is not a proper pdf as f(x) is negative for some values of X. For instance, if we take x = 1/4. Then f(x) = -1/8.

Example 8: Suppose f(x) defined below is the pdf of a random variable. We obtain (i) the constant k (ii) the cdf of X, (iii) P(X > 5), (iv) P(X > 5 | X > 3).

$$f(x) = \begin{cases} ke^{-x/10}, 0 < x < \infty \\ 0, \quad\quad otherwise \end{cases}$$

Solution:

(i) For obtaining the constant k, we use the property of the pdf that the integral of f(x) over the entire range of X is 1. Using the transformation t = 10.x, we have

$$1 = \int_{0}^{\infty} ke^{-x/10}dx$$

$$= k\int_0^\infty e^{-x/10}dx$$

$$= k.10\int_0^\infty e^{-1}dt \;=\; k.10$$

$$Hence, k = 1/10.$$

(ii) The cdf of X is given by

$P(X \leq x) = F(x)$

$$= \int_0^x \frac{1}{10}e^{-x/10}dx = \int_0^{x/10} e^{-1}dt \;\; (t = 10x')$$

Therefore the cdf of X is given by

$$F(x) = \begin{cases} 0, & x \leq \infty \\ 1 - e^{-x/10}, & 0 < x < \infty \end{cases}$$

(iii) We have

$$P(X > 5) = 1 - F(5) = e^{-5/10} = e^{-1/2}$$

and

$$P(X > 5 \mid X > 3) = \frac{P(3 \leq X < 5)}{P(X \leq 3)}$$

$$= \frac{F(5) - F(3)}{1 - F(3)} = \frac{e^{-3/10} - e^{-1/2}}{e^{-3/10}}$$

3.7 Self Assessment Exercises

1. The pmf of a discrete random variable X is given by

$$p(x) = \begin{cases} c.|x|, & if\ x = -2, -1, 1, 2 \\ 0, & otherwise \end{cases}$$

Obtain the value of c. Also find the cdf of X. What is the probability that $|X| \leq 1$.

2. The pmf of a discrete random variable X is given by

$$p(x) = \begin{cases} k\dfrac{4^x}{x!}, & x = 0, 1, 2, \ldots. \\ 0, & otherwise \end{cases}$$

Find the value of k. Also obtain the probability that X is greater than or equal to 3.

3. A fair coin is tossed until a head is obtained. Let X be the number of tosses. Write the pmf of X. Also obtain the probability that number of tosses is greater than or equal to two.

4. For each of the following determine the value of c so that the functions can serve as the probability distribution of a random variable:

(i) $$f(x) = \frac{c}{3^x} \qquad if\ x = 1, 2, 3, \ldots \ldots$$

$$=0, \qquad otherwise$$

(ii) $$f(x) = cxe^{-x} \qquad if\ x > 0$$

$$= 0, \qquad elsewhere$$

5. The pdf of a random variable X is given by

$$f(x) = \begin{cases} 1|1 - x|, & 0 < x < 2 \\ 0, & otherwise \end{cases}$$

Find the cdf of X. Also obtain (i) P (X > 1), (ii) P (X > 1/2 | 0 < x <1).

6. The pdf of a random variable X is given by

$$f(x) = \begin{cases} k.x^2, & if\ 0 < x < 1 \\ 0, & otherwise \end{cases}$$

Find the constant k. Also obtain (i) the cdf of X. (ii) P(1/4 < x <1/2

(iii) P(1/3 < x < 1/2 | x > 1/4).

7. Suppose the pdf of a random variable X, if (x), is given by

$$f(x) = \begin{cases} kx, & if\ 0 < x < 1 \\ kx^2 & if\ 1 \leq x < 2 \\ 0, & otherwise \end{cases}$$

Find the constant k and the cdf of X. Also find (a) P(X > 1/3) (b) P(1 < X < 2 | X > 1/2).

3.8 Solutions/Suggestions

1. c = 1/6.
The cdf of X is

$$F(x) = \begin{cases} 0 & if\ x < -2 \\ \frac{1}{3} & if -2 \le x < -1 \\ \frac{1}{2} & if -1 \le x < 1 \\ \frac{2}{3} & if\ 1 \le x < 2 \\ 1, & if\ x \ge 2 \end{cases}$$

P ($|X| \le 1$) = 1/3

2. k = e^{-4}, P($X \ge 3$) = 1 – 13 e^{-4}

3. p(x) = $1/2^x$, x = 1, 2,……; P($x \ge 2$) = 1/2

4. (i) c = 2 (ii) c = 1

5. The cdf of X is

$$F(x) = \begin{cases} 0, & if x < 0 \\ \frac{x^2}{2}, & if\ 0 \le x > 1 \\ 2x - \frac{x^2}{2} - 1 & if\ 1 \le x < 2 \\ 1 & if\ x \ge 2 \end{cases}$$

(i) 1/2 (ii) 3/4

6. k = 3.

(i) $$F(x) = \begin{cases} 0, & if\ x \le 0 \\ x^3, & if\ \ 0 < x < 1 \\ 1, & if\ x \ge 1 \end{cases}$$

(ii) 7/64 (iii) 152/1701

7. k = 6/17.

$$F(x) = \begin{cases} 0, & if\ x \le 0 \\ \frac{3}{17}x^2, & if\ \ 0 < x < 1 \\ \frac{1 - 2x^3}{17} & if\ \ 1 \le x < 2 \\ 1, & if\ x \ge 2 \end{cases}$$

(a) 50 / 51 (b) 56 / 65.

3.9 Summary

A random variable is defined as a real valued function defined over the sample space. Each value of the random variable represents an event of a random experiment. The distribution of probability over the different values of the random variable is known as probability distribution.

3.10 Further Readings

1. Cramer H, Mathematical Methods of Statistics, Princeton University Press, 1946 and Asia Publishing House, 1962.
2. Hogg R.V. and Craig A.T., Introduction to Mathematical Statistics, Macmillan, 1978.
3. Prazen E., Modern Probability Theory and its applications, John Wiley, 1960 and Wiley Eastern 1972.
4. Rao C.R., Linear Statistical Inference and Its Applications, John Wiley, 1973 and Wiley Eastern 1974.
5. Rohtagi V.K., An Introduction to Probability theory and Mathematical Statistics John Wiley, 1976 and Wiley Eastern 1985.
6. Vikas S.S., Mathematical Statistics, John Wiley, 1962 and Toppan.

Unit-4: Mathematical Expectation

Structure

4.1 Introduction

The notion of expectation or mean has its origin in the theory of games. The concept was first dealth by Huygens (1629-1695). A gambler may be interested to know the average winnings at a game while a businessman wishes to know his average profits on a product. The 'average' value of a random phenomenon is called its mathematical expectation or expected value. The statisticians prefer it due its property of statistical stability. If we know the probability distribution of a random variable/s then we can calculate the expected value of the random variable or the distribution.

4.2 Objectives

After going through this unit you shall be able to,

- Describe the expectation of random variables. X or mean of the distribution.
- Identify the raw moments of the distribution,
- Indentify the central moments of the distribution,
- Analyze expectation of sum of random variables and the product of independent r.v.'s.
- State covariance between two variables.
- Explain coefficient of correlation between two linearly dependent random variables.

4.3 Mathematical Expectation and Types

Definition: A (real-valued) function defined on the sample space is called a random variable (r. v.) or a stochastic variable. Obviously to each value of a random variable x there corresponds a definite probability.

Let $x_1, x_2, \ldots\ldots, x_k$ be the possible values of x, and let $p_1, p_2\ldots\ldots\ldots p_k$ be the corresponding probabilities. A statement of the possible values together with the probabilities gives the probabilities distribution of x.

Types of Random Variables:

There are two types of random variables:

- Discrete Random Variable
- Continuous Random Variable.

4.3.1 Discrete Random Variable

If a random variable X assumes only a finite number or contutably infinite number of values then X is called discrete random variable. The possible values that X may take are $x_1, x_2, \ldots\ldots., x_n$ in finite case and $x_1, x_2\ldots\ldots$, in countably infinite case or $\{x_i$, i= 0,1,2,…..}

4.3.2 Continuous Random Variable

If a random variable X assumes any value in some interval or intervals it is called a continuous random variable. In other words, if a variate X can take an infinite set of values in a given interval, say $a \leq x \leq b$, it is a continuous random variable.

4.3.3 Expectation of a Function of a Random Variable

Let X be a random variable which takes the values $x_1, x_2\ldots..$, then the probability that X = x, is denoted by $P(X = x_i)$. The function $P(X= x_i)$ denoted by $P(x_i)$ or $\boldsymbol{p_i}$ is called probability function of X, we write $P(x_i) = P[X = x_i]$,I = 1, 2……or $p(x) = p[X = x]$, for $x = x_1, x_2, x_3,\ldots\ldots.$

The probability distribution of a discrete random variable is the set of order pairs $[x_i, p(x_i)]$ which must satisfy the conditions.

(1) $p(x_i) \geq 0$, i = 1, 2, 3,……. .... (2.1)

(2) $\sum_{i=1}^{\infty} p(x_i) = 1$. .... (2.2)

Example:

(1) Let X be the number of points appearing in a toss of a die. Then probability distribution of the discrete random variable X is given by the probability function.

$$p(x) = \begin{cases} 1/6, & x = 1, 2, 3, 4, 5, 6 \\ 0, & othrwise \ldots.. \end{cases} \qquad \ldots\ldots\ldots\ldots\ldots(2.3)$$

This discrete probability distribution can be expressed as.

X= x	1	2	3	4	5	6
P(X = x) = p(x)	1/6	1/6	1/6	1/6	1/6	1/6

(2) Let a coin be tossed two times. Let r. v. X = Number of heads Possible values of X are x = 0, 1, 2. The distribution of r.v. X is

X = x	0	1	2
P(X = x) = p(x)	1/4	1/2	1/4

(3) If a coin be tossed 4 times, the distribution of getting heads, 0, 1, 2, 3, 4, times is

X = x	0	1	2	3	4
P(X = x) = p(x)	1/16	4/16	6/16	4/16	1/16

Expected Value or Mathematical Expectation:

Discrete Distribution: If X denotes a discrete random variable which can assume the values x_1, x_2, ….x_n with respective probabilities p_1, p_2,….p_n-----.

Where $p_i = p(x_i) = p[X = x_i]$ *and* $\sum_{i=1}^{\infty} p(x_i) = 1$, then the mathematical expectation of X or expected value of X, denoted by E(X), is

$$E(X) = p_1x_1 + p_2x_2 + \cdots \ldots . p_nx_n = \sum_{i=1}^{\infty} p_ix_i \quad or\ E(X) = \sum_{i=1}^{\infty} xp(x) \qquad \ldots \ldots \ldots (2.4)$$

The expectation exists if the series $\sum_{i=1}^{\infty} |x_i|\, p(x_i)$ is convergent, that is the condition is satisfied.

$$\sum_{i=1}^{\infty} |x_i|\, p(x_i) < \infty \qquad \ldots \ldots \ldots \ldots (2.4.1)$$

The expectation of X may not always exist. You should check the condition (2.4.1)

Please look at the sum (2.4). The summond x_i $p(x_i)$ is the i^{th} value of the r. v. X multiplied by the probability that [X = x_i] and then the summation is taken over all values. Thus, E(X) is also an 'average' of the values that the r. v. assumes and each value is weighted by the probability with which it probable will receive correspondingly higher weight.

Further, E(X) is the centre of gravity (or centraid) of the unit mass that is determined by the probability mass function *(p.m.f.)* of X. Hence, the mean E(X) of X is the measure of which in the values of r. v. X are "centred".

Continuous random variable or distribution:

If X is a continuous r. v. with pdf f(x), then expectation E(X) of r. v. X or expected value of the distribution is given by the integral

$$E(X) = \int_{-\infty}^{\infty} x \quad f(x)dx$$

provided the condition

$$\int_{-\infty}^{\infty} |x|f(x)dx < \infty$$

is satisfied.

4.4 Moments

Let X be a r.v. and ∅(X) be a function of X. Then expected value of ∅(X) denoted by E[∅(X)], is defined as

$$E[\emptyset(\mathrm{X})] = \begin{cases} \sum \emptyset(x)p(x), & if\ X\ is\ a\ discreter\ r.v. with\ p.m.f., \quad p(x) \\ \int_{-\infty}^{\infty} \emptyset(x)f(x)dx, & if\ X\ is\ a\ continuous\ r.v. with\ p.d.f., \quad f(x) \end{cases}$$

……(2.5)

$$E[\emptyset(\mathrm{X})]\ \ exists\ if \sum_{x} \emptyset(x)p(x)\ \ or \int_{-\infty}^{\infty} \emptyset(r)f(x)dx\ is\ conergent\ as\ the\ case\ may\ be.$$

If X is a r.v. which takes the value x_1, x_2, …….x_n with corresponding probabilities p_1, p_2……p_n respectively, then we may write

$$E[\emptyset(\mathrm{X})] = \sum_{i=1}^{n} \emptyset(x)pi$$

In case $\emptyset(X) = X^r$, we have

$$E(X^r) = p_1x_1^r + p_2x_2^r + \cdots \ldots \ldots +p_nx_n^r = \sum p_ix_i^r$$

This is defined as the r[th] moment of the discrete random variable X or of the prob. distribution about x = 0, or origin it is denoted by μ_r for r = 0, 1, 2, 3,…., This μ_r is the expected value of the r[th] power of the variate. In fact, $\mu_0 = 1$ (always) expectation of r.v. X or $\mu_i = \sum_{i=1}^{\infty} x_ip(x_i)$ is the mean of the r.v. X or the prob. distribution is X.

Central Moments

Also r[th] central moment or moment about mean

$$\mu_r = E[X - E(x)]^r = \sum p_i[x_i - E(X)]^r$$

$$\mu_1 = E(X) \qquad\qquad if\ X\ is\ discrete\ r.v.$$

$$\int_{\infty}^{\infty} [X - E(x)]^r\ f(x)dx \qquad\qquad if\ X\ is\ continuous\ r.v.$$

In particular,

$$\mu_1 = E(X) = p_1x_1 + p_2x_2 + \cdots \ldots . = \sum_i p_ix_i = mean$$

$$\mu_2 = E[X - E(X)]^2 = Var\ (X) = \sigma^2$$

$$\mu_3 = E[X - E(X)]^3 = \sum p_i[X - E(X)]^3$$

$$\mu_4 = E[X - E(X)]^4 = \sum p_i[X - E(X)]^4$$

If p_i is replaced $\frac{f_i}{N}$ by where $\sum f_i$= N then

$E(X) = \frac{\sum fx}{N} = \bar{X},$ $\qquad$ $which\ is\ the\ mean.$

$\therefore E(X)$ $\qquad$ $represents\ the\ mean.$

Theorem 2.1 If x = a, a constant then E(x) = a.

Proof: Since X = a, we have x_i = a and P[X = x_i] = P[X = a] = 1 for all i. Hence it gives

for discrete random variable

E(X) = p[X = a] = a.1= a ……………(2.6)

or

for continuous random variable

E(X) = $\int_{-\infty}^{\infty} x\ f(x)dx = cf(x)dx = c\int_{-\infty}^{\infty} f(x)dx = C.$

Example 1: What is the expected value of the number of points, X, that will be obtained in a single throw with an ordinary die? Find its variance also.

Solution: Here the variance X is the number of pointed that comes up on a die. It assumes the values X_i = 1, 2, 3, 4, 5, 6 with probability $p(x_i)$ = P[X = x_i] = 1/6 in each case.

Hence $E(X) = p_1x_1 + p_2x_2 + \cdots \ldots \ldots . + p_6x_6$ = (1/6)*1 + (1/6)*2 + …. + (1/6)*6 = 3.5

Also Var (X) = E(X^2) - $[E(X)]^2$ = 1/6 ($1^2 + 2^2 + \ldots + 6^2$) - $(7/2)^2$ = 35/12

Example 2: Thirteen cards are drawn simultaneously from a deck of 52. If aces count 1, face cards 10 and others according to denomination, find the expectation of the total score on the 13 cards.

Solution: Let Xi be the number on the cord corresponding to the i^{th} draw, then X_i takes the values 1, 2, 3, 4, 5, 6, 7, 8, 9, 10; 10; 10; 10; each having the selection probability = 4/52 = 1/13

Hence Let X be the total score on the 13 cards. Then X = $X_1 + X_2 + X_3 + \ldots\ldots X_{13}$.

E(Xi) = (1/13)*1 + (1/13)*2 + (1/13)*3 ……………+ (1/13)*9 + (1/13)*10 + (1/13)*10 + (1/13)*10 + (1/13)*10

=1/13 (1 + 2 + 3 +…..+ 9 + 10 + 10 + 10 + 10) = 85/13 and

$$E(X) = \sum_{i=1}^{13} E(x_i) = \sum_{i=1}^{13} \left(\frac{85}{13}\right) = 85$$

***Theorem*:** Show that (a) |E(X)| ≤ E(|X|) and

(b) E(X) = $\sum_{x=1}^{\infty} P(X \geq x)$, $x = 0, 1, 2, \ldots \ldots$

Proof:

(a) E(X) = $\sum_{i=1}^{n} x_i p_i = \sum xp$, say

$\Rightarrow \ |E(X)| \leq |\sum xp|$

$\Rightarrow \ |E(X)| \leq |\sum p|x|$ $[\because p \geq 0 \ always]$

$\Rightarrow \ |E(X)| \leq E\ (|X|).$ $[\because \ p \leq 1]$

Hence proof of r. v. X is

P_x = P[X = x] = p(x), for X = 0, 1, 2, 3, ……………

(b) Let

P(X = x) = p(x): p_0, p_1, p_2………

By def.

E(X) = $\sum_{x=0}^{\infty} 0.\mathrm{p}0 + 1.\mathrm{p}_1 + 2.\mathrm{p}_2 + 3.\mathrm{p}_3 + \cdots \ldots \ldots$ ……(2.7)

$= \mathrm{p}_1 + 2\mathrm{p}_2 + 3\mathrm{p}_3 + \cdots \ldots \ldots \ldots \ldots$

But

$P(X \geq 1) = p[X = 1] + p[X = 2] + p[X = 3] + \ldots\ldots + \ldots\ldots$

$P(X \geq 2) = p[X = 2] + p[X = 3] + \ldots\ldots$

$P(X \geq 3) = p[X = 3] + p[X = 4] + \ldots\ldots$ and so on

Adding all, we have

$$\sum_{x=1}^{\infty} P(X \geq x) = \mathrm{p}_1 + 2\mathrm{p}_2 + 3\mathrm{p}_3 + \cdots \qquad (2.8)$$

From (2.7) and (2.8), we have

$$E(X) = \sum_{x=1}^{\infty} P(X \geq x) \ldots\ldots\ldots\ldots\ldots (2.9)$$

4.5 Theorems on Expectation

Theorem 2.2 Consider two r.v.s X and Y defined on the same sample space.

The expectation of the sum of two random variables is equal to the sum of their expectation, mathematically,

$$E(X + Y) = E(X) + E(Y). \qquad \ldots\ldots(2.10)$$

Let $p_1, p_2, \ldots, p_m$ be the probabilities of m values $x_1, x_2, \ldots x_m$ of the variate X and $p_1', p_2, \ldots\ldots, p_m'$ be the probabilities of n values $y_1, y_2 \ldots y_n$ of the variate Y respectively. Then X + Y is a variate which can take *mn* values $x_i + y_j$ ($i = 1, 2, \ldots, m$; $j = 1, 2, \ldots\ldots, n$), since any of the m values of X may be associated with any of the n values of Y.

Let p_{ij} be the probability corresponding to that variate when X assumes the value x_i and Y assumes the value y_j. That is, $p[X = x_i, Y = y_j] = p_{ij}$, for $i = 1, 2, \ldots, m$, $j = 1, 2, \ldots\ldots, n$.

Mathematically the marginal proof of X is

$$P_i = p_i = p[X = x_i] = p[X = x_i, Y = y_1] + p[X = x_i, Y = y_2] + \ldots\ldots + p[X = x_i, Y = y_n]$$

$$= p_{i1} + p_{i2} + \ldots\ldots + p_{in}$$

Similarly the marginal proof of

$$\sum_{j=1}^{n} p_{ij} \ \ proof\ of\ Y\, p_j^n = p_{ij} = \sum_{i=1}^{m} p_{ij}$$

As X assumes a define value x_i, Y assumes one of the values y_1, y_2, ….., y_n so that the sum $\sum_{j=1}^{n} p_{ij} = \ p_i$ represents the probability p_i of X assuming the value x_i, i.e.

$$\sum_{j=1}^{n} p_{ij} = \ p_i \qquad Similarly \sum_{i=1}^{m} p_{ij} = p_{j,} \qquad Then$$

$$E(X+Y) = \sum_{i=1}^{m}\sum_{j=1}^{n} p_{ij}(x_i + y_i) \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots.(2.11)$$

$$= \sum_{i=1}^{m} x_i \ \left[\sum_{i=1}^{n} p_{ij}\right] + \sum_{i=1}^{m} y_i \left[\sum_{j=1}^{n} p_{ij}.y_i\right]$$

$$= \sum_{j=1}^{m} x_i \ \sum_{j=1}^{n} p_{ij} + \sum_{j=1}^{n} y_i \sum_{i=1}^{m} p_{ij}$$

$$= \sum_{j=1}^{m} x_i p_i + \sum_{i=1}^{n} y_j p_j$$

$$\therefore E(X+Y) = E(X) + E(Y).$$

Thus the expected value of the sum of two variates is equal to the sum of their expected values.

Similarly: E(X + Y + Z +….) = E (X) + E(Y) + E(Z) +…… ……. (2.12)

Theorem 2.3: Expected Value of a Constant is a constant.

i.e., if m is a constant, then E(m) = m. ………. (2.13)

Proof: ∵ P(m = m) = 1

And P(m≠ m) = P(m = n) = 0, where n≠ m

Then E(m) = 1.m + 0.n = m.

Mathematical Expectation for Multiplication of two Discrete Random Variables:

Theorem: If X and Y are independent random variables, then

E(XY) = E(X) E(Y). ………………. (2.14)

Proof: Using the notation already introduced the expectation of the product XY may be written as

$$E(XY) = \sum_{i=1}^{k}\sum_{j=1}^{l} x_i y_j p_{ij} \qquad \ldots\ldots\ldots\ldots.\ (2.15)$$

Where we are taking formula for expectation. (The values x_i y_i may not be all different, but they arise from an exhaustive set of mutually exclusive cases.) Now, since x and y are supposed to be independent, $p_{ij} = p_{io}\, p_{oj}$ for all i, j. Hence

$$E(XY) = \sum_{i=1}^{k}\sum_{j=1}^{l} x_i\, y_i\, p_{io}\, p_{oj} \qquad \ldots\ldots\ldots\ldots\ldots \quad (2.16)$$

Since x_i y_i p_{io} p_{oj} = (x_i p_{io}) (y_i p_{oj}), where the first factor depends on i alone and the second depends on j alone, we may write the above double sum as the product of two sums:

$$E(XY) = \left(\sum_{i=1}^{k} x_i p_{io}\right)\left(\sum_{j=1}^{l} y_i p_{oj}\right) \qquad \ldots\ldots\ldots\ldots\ldots.. \quad (2.17)$$

But the first factor on the right hand side is by definition (7.16), the expectation of x and the second factor is similarly the expectation of y. As such,

$$E(XY) = E(x)\, E(y).$$

Thus the expected value of the multiplication of two independent variates is equal to the product of their expected values.

In general if X, Y, Z, …………. are independent r.v.s …… (2.18)

$$E(XYZ\ldots\ldots) = E(X)\, E(Y)\, E(Z)\ldots\ldots\ldots..$$

Theorem: If Y = bX, then var (Y) = b^2var (X).

Proof: From theorem E(Y) = b E(X).

Hence $y - E(Y) = b[x - E(X)]$ ………………….. (2.19)

and $[y - E(Y)]^2 = b^2\ [X - E(X)]^2$

on applying theorem again, we have

$E[y - E(Y)]^2 = b^2\ [X - E(X)]^2$,

i.e., var (Y) = b^2 var(X). ……………… (2.20)

Theorem: If y = a + bx, then var (y) = b^2 var(x).

Proof: From theorem gives E(Y) = a + b E(X)

Hence $y - E(Y) = b[x - E(X)]$.

Next, proceeding as in the proof of theorem, we have the stated result.

Theorem: First moment about mean is zero.

Proof: Let X have the probability distribution

x:	x_1	x_2	-------	x_n
p:	p_1	p_2	-------	p_n

Then $E(X) = \sum xp = \sum_{i=1}^{n} x_i p_i$

and ∵ E(X) = mean = constant

$$\mu_1 = E(X - E(X)] \qquad \dots\dots\dots\dots (2.21)$$

$$= E(X + \{-E(X)\}] \qquad = E(X) + [-E(X)]$$

$[\because E(X+Y) = E(X) + E(Y) \quad and\ E(m) = m]$ (∵m is a constant)

= E(X) - E(X) =0.

Similarly, for variance

$$\mu_2 = Var(X) = E(X - EX)^2 \qquad \dots\dots\dots\dots\dots (2.22)$$

$$= E[X^2 - 2XE(X) + \{E(X)\}^2]$$

$$= E(X^2) - 2E(X)E(X) + [E(X)]^2$$

$$= E(X^2) - 2[E(X)]^2 + [E(X)]^2$$

$$= E(X^2) - [E(X)]^2$$

Since Var (X) ≥ 0

$$\therefore E(X^2 - [E(X)]^2 \geq 0 \Rightarrow E(X^2) \geq [E(X)]^2 \qquad \dots\dots\dots\dots (2.23)$$

Example 4: Find for the following probability distribution:

x	8	12	16	20	24
p(x)	1/8	1/6	3/8	1/4	1/2

Solution: By definition

E(X) = $\sum x.p(x)$

$$= 8.\left(\frac{1}{8}\right) + 12.\left(\frac{1}{6}\right) + 16.\left(\frac{3}{8}\right) + 20.\left(\frac{1}{4}\right) + 24.\left(\frac{1}{12}\right)$$

$$= 16 = \bar{X} = \mu_1$$

This represent the mean of the distribution i.e., E(x) =16.

$$E(X^2) = \sum x^2 . p(x)$$

$$= 8^2.\left(\frac{1}{8}\right) + 12^2.\left(\frac{1}{6}\right) + 16^2.\left(\frac{3}{8}\right) + 20^2.\left(\frac{1}{4}\right) + 24^2.\left(\frac{1}{12}\right)$$

$$= 276 = \bar{X} = \mu_2$$

This is represents the second moment about the origin zero.

$$\mu_2 = Var(X) = E\{(X - \bar{X})^2\} = \sum (x - \bar{x})^2 \ p(x)$$

$$= (8-16)^2.\frac{1}{8} + (12-16)^2.\frac{1}{6} + \ (16-46)^2.\frac{3}{8} + (20-16)^2.\frac{1}{4} + (24-16)^2.\frac{1}{12}$$

$$= 20.$$

$$Alternatively, \quad E(X - \bar{X})^2 = Ex^2 - \bar{x}^2 = 276 - 256 = 20.$$

This is represents the variance of the distribution.

Example 5: Four balls are drawn from a bag containing 5 black, 6 white and 7 red balls. Let X denote the number of white balls drawn find E(X).

Solution:

Total number of balls in the bag = 5 + 6 + 7 = 18

The number of white balls is drawn may be 0, 1, 2, 3, 4.

$$\therefore P(X - X) = \frac{6_{C_0} \times 12_{C_{4-x}}}{18_{C_4}}, \qquad where\ x = 0, 1, 2, 3, 4.$$

Then we have

$$P(X = 0) = \frac{6_{C_0} \times 12_{C_4}}{18_{C_4}} = \frac{495}{3060}$$

$$P(X = 1) = \frac{6_{C_1} \times 12_{C_3}}{18_{C_4}} = \frac{1320}{3060}$$

$$P(X = 2) = \frac{6_{C_2} \times 12_{C_2}}{18_{C_4}} = \frac{990}{3060}$$

$$P(X = 3) = \frac{6_{C_3} \times 12_{C_1}}{18_{C_4}} = \frac{240}{3060}$$

$$P(X = 4) = \frac{6_{C_4} \times 12_{C_0}}{18_{C_4}} = \frac{15}{3060}$$

$$\therefore E(X) = 0.\frac{495}{3060} + \frac{1.1320}{3060} + \frac{2.990}{3060} + \frac{3.240}{3060} + \frac{4.15}{3060}$$

$$= \frac{4080}{3060} = \frac{136}{102} = \frac{4}{3}$$

Example 6. The probability function of a discrete random variable is as follows:

X= x	0	1	2	3	4	5	6	7
p(x)	0	k	2k	2k	3k	k^2	$2k^2$	k^2+k

Find (i) k (ii) P(X < 6) and P(X ≥ 6) (iii) P(0 < X < 5) and (iv) Distribution function of X.

$$E(X) = \sum xp(x) = 0.0 + 1(k) + 2(2k) + 3(2k) + 4(3k) + 5(k^2) + 6\ (2k^2) + 7(k^2 + k)$$

$$= 24\,k^2 + 30\,k = 24\left(\frac{1}{10}\right)^2 + 30\ \left(\frac{1}{10}\right) = 3.24$$

Solution:

(i) Since

$$\sum_{x=0}^{7} p(x) = 1 \qquad \text{… … … … ….} \quad (2.23)$$

$$\Rightarrow \quad 0 + k + 2k + 2k + 3k + k^2 + 2k^2 + 7k^2 + k = 1$$

$$\Rightarrow \quad 10k^2 + 9k - 1 = 0$$

$$\Rightarrow \quad 10k^2 + 10k - k - 1 = 0$$

$$\Rightarrow \quad 10k(k + 1) - (k + 1) = 0$$

$$\Rightarrow \quad (10k - 1) - (k + 1) = 0$$

$$\therefore \quad k = \frac{1}{10}$$

[k = -1 is not possible]

(ii) P(X < 6) = P(X= 0) + P(X=1) +……+ P(X = 5)

$= 0 + \frac{1}{10} + \frac{2}{10} + \frac{2}{10} + \frac{3}{10} + \frac{1}{100} = \frac{81}{100} = 0.81$

P(X ≥ 6) = P(X = 6) + P(X =7)

$= \frac{2}{100} + \frac{7}{100} + \frac{1}{10} = \frac{19}{100} = 0.19$

(iii) P(0 < X < 5) = P(X=1) + P(X = 2) + P(X= 3) + P(X = 4)

$$= \frac{1}{10} + \frac{2}{10} + \frac{2}{10} + \frac{3}{10} = \frac{8}{10} = 0.8$$

(iv) Distribution $F(x) = P(X \leq x)$

$F(0) = 0$, $F(1) = 1/10$, $F(2) = 3/10$, $F(3) = 5/10$.

$F(4) = 8/10$, $F(5) = 81/100$, $F(6) = 83/100$, $F(7) = 1$

Example 7:

A bag contains 9 balls which are numbered from 1 to 9. Three balls are drawn without replacement from this find the expectation of the sum of the numbers on these balls.

Solution:

Let X_i, denote the number on the i^{th} ball drawn,

S = Sum of the numbers on the three balls

$= X_1 + X_2 + X_3$

$E(S) = E(X_1 + X_2 + X_3)$ ………………………… (2.24)

Now $X_i = 1, 2, 3, 4, 5, 6, 7, 8, 9$ with probability 1/9 each.

$E(X_i) = 1/9\ (1 + 2 + 3 + 4 + \ldots.. + 8 + 9)$

$= 1/9.9\text{x}\ 10/2 = 5$ for $i = 1, 2, 3$

$E(S) = 5 + 5 + 5 = 15$, from (1).

Covariance in Term of Expectation:

If X and Y are two variates with the respective expected values (or means) E(X) and E(Y), the covariance between X and Y is defined as

$\text{Cov}\ (X,Y) = E[\{X - E(X)\}\ \{(Y - E(Y)\}]$.

Thus the expected value of product of the deviations of the two variates from their mean is called their covariance.

$\text{Cov.}\ (X,Y) = E[\{(X – E(X)\}\ \{Y-E(Y)\}]$

$= E[(XY - E(X).Y- E(Y)\ X + E(X)\ E(Y)]$

$= E(XY) - E(X)E(Y) - E(Y)\ E(X) + E(X)E(Y)$

$= E(XY) - E(X)\ E(Y)$. …………… (2.25)

Theorem: The covariance of two independent variates is equal to zero.

Proof: If X and Y are independent variates, then their functions are also independent.

$E\{X - E(X)\} = E(X) - E(X) = 0,$

$E\{Y - E(Y)\} = E(Y) - E(Y) = 0.$

$\therefore Cov(X, Y) = E[\{X - E(X)\}\{ Y - E(Y)] = 0.$

Alternative Method: We know that

$Cov(X,Y) = E[\{X - E(X)\}\{Y - E(Y)\}]$

$= E(XY) - E(X) E(Y)$

$= E(X) E(Y) - E(X) E(Y)$ [if X and Y are independent then $E(XY) = E(X) E(Y)$
$= 0.$ ………………… (2.26)

Cor. 3: If X_1 and X_2 are independent then

$Var(x_1 - x_2) = var(x_1 + x_2) = var(x_1) + var(x_2)$ ………………… (2.27)

Cor. 4: If $x_1, x_2, \ldots\ldots x_n$ are independent and $U = a_0 \pm a_1X_1 \pm a_2X_2 \pm \cdots \ldots \ldots. \pm a_nX_n$

We have $var(u) = a_1^2\, var\,(x_1) + a_2^2 var\,(x_2) + \ldots\ldots.. + a_n^2 var(x_n)$ ………..(2.28)

$$r(X,Y) \;\; or \;\; p_{xy} = \frac{Cov\,(X,Y)}{\sqrt{var\,(X)var\,(Y)}} \quad is\ called\ the\ correlation\ coefficient$$

Between X and Y. Thus

$$\delta_{xy} = \frac{E[\{(X - E(X)\}\{(Y - E(Y)\}]}{\sqrt{E\{(X - E(X)\}^2\, E\{(Y - E(Y)\}^2]}}$$

$$= \left.\frac{E(XY) - \; E(X)E(Y)}{\sqrt{[\{E(X)^2 - \; \{(E(X)^2\}\{E(Y)^2 - \; \{(E(Y)^2\}}}\right\} \quad \ldots\ldots.. \; (2.29)$$

We have, if X and Y are independent. Then $\delta_{xy} = 0.$

Also $var\,(X \pm Y) = var\,(X) + \; Var((Y) \pm 2cov\,(X,Y)$

$$= var\,(X) + \; Var((Y) \pm 2\rho_{xy}\sqrt{var(X)var\,(Y)}$$
$$= \; \sigma_x^2 + \sigma_y^2 + 2\rho_{xy}\sigma_x\sigma_{y.} \qquad \ldots\ldots\ldots\ldots. \; (2.30)$$

Example 8:

A box contains a white and b black balls; (< a + b) c balls are drawn. Show that the expectation of the number of white balls drawn is ca/a + b.

Solution:

Let a variate x_i be defined as the colour of the ball at the 1^{th} drawn and as follows:

Xi = 1 if the i^{th} ball drawn is white = 0 if the i^{th} ball drawn is black.

Then the number of white balls, X among the c drawn balls is given by

$X = x_1 + x_2 + \ldots.. + x_c$.

Now

$$P(x_i = 1)\frac{a}{a+b}, P(x_i = 0) = \frac{b}{a+b}$$

$$E(x_i) = 1.\frac{a}{a+b} + 0.\frac{b}{a+b} = \frac{1}{a+b}$$

Therefore,

$E(X) = E(x_1) + E(x_2) + \ldots..+ E(x_c)$

$= \frac{a}{a+b} + \frac{a}{a+b} + \cdots \ldots.c\ times\ = \frac{ca}{a+b}$

Example 9: A coin tossed until the head appears. What is expectation of the number of tosses?

Solution: Let X denote the number of tosses until the first head appears. The values of X with their probabilities are tabulated as follows:

Favorable Events : H, TH, TTH, ………

[X = x] : 1,2,3,…..

$$P(X = x) = P(x):\quad \frac{1}{2}, \left(\frac{1}{2}\right)^2, \left(\frac{1}{2}\right)^3, \ldots \ldots$$

$$\therefore E(X) = \sum xp\,(x) = 1.\frac{1}{2} + 2.\left(\frac{1}{2}\right)^2 + 3.\left(\frac{1}{2}\right)^3 + \cdots \ldots \ldots$$

$$(Aritmetics\ geometric\ seris)$$

Also,

$$\frac{1}{2}\,E(X) = \left(\frac{1}{2}\right)^2 .1 + \left(\frac{1}{2}\right)^3 .2 + \cdots ..$$

Subtracting:

$$E(X) - \frac{1}{2}E(X) = \frac{1}{2}E(X) = \frac{1}{2} + \frac{1}{2^2} + \frac{1}{2^3} + \cdots \ldots.$$

$$= \frac{1/2}{1-(1/2)} = 1$$

$$\Rightarrow E(X) = 2$$

Example 10: What is the mathematical expectation of the sum of points, if n dice are tossed?

Solution: Denoting by x_i the number of points on the i^{th} dice, the sum of points on n dice is

$$X = x_1 + x_2 + \ldots + x_n.$$

$$\therefore \quad E(x_1 + x_2 + \cdots \ldots \ldots + x_n) = E(x_1) + E(x_2) + \ldots + E(x_n)$$

But for every single dice

$$E(x_i) = \frac{7}{2}, \qquad i = 1, 2, 3, \ldots . n$$

$$\therefore \quad E(x_1 + x_2 + \ldots \ldots \ldots + x_n) = (7/2) + (7/2) + \ldots + (7/2) \text{ n times}$$

$$= \frac{7}{2} n.$$

Example 11: Find the expected value of the product of points on n dice tossed.

Solution: Denoting by x_i the number of points on the i^{th} dice the product of points on n dice is $x_1, x_2, x_3, \ldots \ldots, x_n$

For every single dice.

$$E(x_i) = \frac{7}{2}, \qquad i = 1, 2, 3, \ldots, n$$

$$\therefore E(x_1 x_2 \ldots x_n) = E(x_1).E(x_2).E(x_3) \ldots \ldots E(x_n)$$

$$= \frac{7}{2} * \frac{7}{2} * \frac{7}{2} * \ldots \ldots \ldots n \text{ } factors.$$

$$= \left(\frac{7}{2}\right)^n$$

4.6 Self Assessment Exercises

1. A coin is tossed three times. If X is a random variable giving the number of heads which arise, obtain the probability distribution of X. Hence or otherwise, determine the means of the distribution.

2. Define a random variable and its Mathematical expectation. Prove that

   (i) E(X + Y + …..T) = E(X) + E(Y) +…..+ E(T). and

   (ii) E(XY….T) = E(X) E(Y)……..; X, Y, …..T are independent.

3. A random variable X is defined by

$$X = \begin{cases} -2 & with\ probability\ 1/3 \\ 3 & with\ probability\ 1/2 \\ 1 & with\ probability\ 1/6 \end{cases}$$

Find (i) E(X); (ii) E (2X + 5) and (iii) Var (X).

4. If X is a random variable, is any function of X and a, b, c are constants. Show that

(i) E (a) = a

(ii) E(a X) = a E (X)

(iii) E[$a\,\Psi\,(X)$] = a E[Ψ (X)]

(iv) Var (aX + b) = a^2 var (X)

(v) Var $\left(\frac{aX+b}{c}\right) = \frac{a^2}{c^2}\ var\ (X)$

(vi) Cov (X, X) = var (X)

(vii) E[Ψ(X) + a] = E [Ψ(X)] + a

5. If X and Y are two random variables and a, b, c are constant, show that

(i) Cov (aX , b Y) = a b Cov (X,Y)

(ii) Cov (X + a, Y + b) = Cov (X, Y)

(iii) Cov (aX + bY, c X + d Y) = ac var (X) + bd Var (Y) + (ad + bc) Cov (X,Y)

6. If x is a random variable for which E (x) = 10 and var (x) = 25. Find the positive values of a and b such that y = ax - b has expectation 0 and variance 1.

7. The probability function of a random variable X is given by

$$P(x) = \begin{cases} k & if\ x = 0 \\ 2k & if\ x = 1 \\ 3k & if\ x = 2 \\ 0 & otherwise \end{cases}$$

Find (a) the value of k (b) P(X < 2), P(X ≤ 2), P(0 < X < 2)

(c) Distribution functions of X.

8. Define random variable and its expectation. The probability function of a random variable X is as follows:

X = x	-2	-1	0	1	2	3	
P(X=x)		0.1	k	0.2	2k	0.3	3k

Find the value of the constant k and E(X).

4.7 Answers

1.

X=x	0	1	2	3
p(x=X)	1/8	3/8	3/8	1/8

and E(X) = 3/2.

2. (i) e(x) = 1, (ii) E (2X+5)= 7 (iii) Var (X) = 5.

6. a= 1/5, b =2

7. (a) k=1/6 (b) (i) 1/2 (ii) 1 (iii) 1/3

(c)
$$F(x) = \begin{cases} o, & if\ x < 0 \\ \frac{1}{6} & if\ x = 0\ or\ 0\ \leq x < 1 \\ \frac{1}{2}, & if\ x = 1\ or\ 0 \leq x < 2 \\ 1, & if\ x \geq 2 \end{cases}$$

8. k = 0.1, E(X) = 0.8

4.8 Summary

This unit gives an overview of theory of expectation, moments, their application and importance.

4.9 Further Readings

1. Cramer H, Mathematical Methods of Statistics, Princeton University Press, 1946 and Asia Publishing House, 1962.
2. Hogg R.V. and Craig A.T., Introduction to Mathematical Statistics, Macmillan, 1978.
3. Prazen E., Modern Probability Theory and its applications, John Wiley, 1960 and Wiley Eastern 1972.
4. Rao C.R. Linear Statistical Inference and Its Applications, John Wiley, 1960 and Wiley Eastern 1974.
5. Rohtagi V.K., An Introduction to Probability theory and Mathematical Statistics, John Wiley, 1976 and Wiley Eastern 1985.
6. Vikas S.S., Mathematical Statistics, John Wiley, 1962 Toppan.

Unit-5: Inequalities for Moments

Structure

5.1 Introduction

In the last two units you studied about r.v., its d.f. (or c.d.f.) its natures in discrete and continuous r.v.'s p.m.f, p.d.f. and their properties, theory of expectation and moments. This unit describes different inequalities concerning moment along with Chebychev's inequality and its applications.

5.2 Objectives

After going through this unit you shall be able to understand,

- Cauchy- Schwarz Inequality
- Markov Inequality
- Chebychev's Inequality
- Applications and importance of above mentioned inequalities.

5.3 Cauchy –Schwarz Inequality

Statement: For any two random variables X and Y having finite variances, we have

$$[E(XY)]^2 \leq E(X)^2 . E(Y)^2$$

Proof:

For any non negative real valued constant a, we have

$$(aX - Y)^2 \geq 0.$$

Or

$$E(aX - Y)^2 \geq 0$$

Or

$$E\{a^2 X^2 + Y^2 - 2aXY] \geq 0$$

Or $\qquad a^2 E(X^2) + Y^2 - 2aE(XY) \geq 0$ ………(3.2.1)

(As constant can be written outside expectation sign)

Taking 'a' as a = $\frac{E(XY)}{E(X^2)}$ and putting it in (3.2.1), we get

$$\frac{\{E(XY)\}^2}{E(X^2)} + E(Y^2) - 2\frac{\{E(XY)\}^2}{E(X^2)} \geq 0$$

Or

$$E(Y^2) \geq \frac{\{E(XY)\}^2}{E(X^2)}$$

i.e., $\quad \{E(XY)\}^2 \leq E(X^2).E(Y^2)$ Proved

Problem 3.3.1: For two random variables X and Y, show that

$$\{E(X+Y)^2\}^{1/2} \leq \{E(X^2)\}^{1/2}.\{E(Y^2)\}^{1/2}$$

Solution: Here we are to prove that

$$\{E(X^2) + E(Y^2) + 2E(XY)\}^{\frac{1}{2}} \leq \{E(X^2)\}^{1/2}.\{E(Y^2)\}^{1/2}$$

Squaring both sides, we have

$$E(X^2) + E(Y^2) + 2E(XY) \leq E(X^2) + 2\sqrt{E(X^2).E(Y^2)}$$

Or

$$E(XY) \leq \sqrt{E(X^2).E(Y^2)}$$

Or

$$\{E(XY)\}^2 \leq E(X^2).E(Y^2)$$

Which is true by virtue of Cauchy-Schwarz Inequality. Hence, the given inequality is true.

Problem 3.3.2: For a random variable X whose all moments are finite, we have

$$E|X|^{\frac{m+n}{2}} \leq \{E|X|^m E|X|^n\}^{1/2};\ m \text{ and } n \text{ are positive integers}.$$

Solution: Let |X| = Y

Then we have to prove that

$$E|Y|^{\frac{m+n}{2}} \leq \{E|Y|^m.E|Y|^n\}^{1/2}$$

Similarly Y is a positive random variable.

Let a = m and a + b = $\frac{m+n}{2}$. Then, b = $\frac{h-n}{2}$

As Y is positive, we may write

$$E[Y^a(\lambda Y^2 - 1)^2] \geq 0 \qquad for\ all\ real\ values\ of\ \lambda.$$

Or

$$E\left[Y^a(\lambda^2 Y^{2b} + 1 - 2\lambda Y^b)\ \right] \geq 0$$

Or

$$\lambda^2 E[Y^{a+2b}] + E[Y^a] - 2\lambda E(Y^{a+b}) \geq 0$$

Or,

$$\lambda^2 E[Y^{a+2b}] + E[Y^a] - 2\lambda E(Y^{a+b})$$

$$Taking, \quad \lambda = \frac{E(Y^a)}{E(Y^{a+b})}, \quad we\ have$$

$$\left\{\frac{E(Y^a)}{E(Y^{a+b})}\right\}^2 E(Y^{a+2b}) \geq E(Y^a) \geq 2\left\{\frac{E(Y^a)}{E(Y^{a+b})}\right\} E(Y^{a+b})$$

Or,

$$\left\{\frac{E(Y^a)}{E(Y^{a+b})}\right\}^2 E(Y^{a+2b}) \geq E(Y^a)$$

Or

$$\{E(Y^a)\}^2\, E(Y^{a+2b}) \geq E(Y^a)\{E(Y^{a+b})\}^2$$

Or,

$$E(Y).E(Y^{a+2b}) \geq \{E(Y^{a+b})\}^2$$

Let a = m and b = $\frac{n-m}{2}$ (i.e., 2b = n-m)

So that a + b = $\frac{m+n}{2}$ and a + 2b = n

We get, $E(Y^m)E(Y^n) \geq \left\{E\left(Y^{\frac{m+n}{2}}\right)\right\}^2$

$$E(Y^m)E(Y^n) \geq \left\{E\left(Y^{\frac{m+n}{2}}\right)\right\}^2$$

Or,

$$\{E(Y^m) + E(Y^n)\}^{1/2} \leq E\left(Y^{\frac{m+n}{2}}\right)$$

Or,

$$\{E[|X|^m].E[|X|^n\}^{1./2} \leq R(|X|^{\frac{m+n}{2}})$$

$$[as\ Y = |X|] \qquad\qquad Proved.$$

5.4 Markov's Inequality

For a non negative r.v. X having finite mean, we have

$$\Pr(X \geq \lambda) \leq \frac{E(X)}{\lambda}$$

Where λ is any real positive number.

Proof: Let X be a continuous non-negative r.v. with p.d.f. f(x). Then,

$$E(X) = \int_0^\infty xf(x)dx$$

$$= \int_0^\lambda xf(x)dx + \int_0^\infty xf(x)dx \qquad (0 < \lambda < \infty)$$

$$\geq \int_\lambda^\infty xf(x)dx$$

$$\geq \lambda \int_\lambda^\infty xf(x)dx$$

$$\geq \lambda \Pr(X \geq \lambda)$$

$$\Pr(X \geq \lambda) \leq \frac{E(X)}{\lambda}$$

5.5 Chebyshev's Inequality

(Also spelled as Tchebycheff's Inequality)

This inequality is a deduction from Markov's Inequality. The theorem was discovered in 1853 by Chebyshev and was later on discussed in 1856 by Bienayme. Here Chebyshev has precisely interpreted the role of standard deviation (as a parameter) in Statistical Analysis to characterize variation in the variate value.

For a.r.v. X having finite mean and variance, we have

$$\mathrm{P}[\,|X - E(X)| \geq \lambda] \leq \frac{Var(X)}{\lambda^2}$$

Or,

$$\mathrm{P}[\,|X - E(X)| \geq \lambda] \geq 1 - \frac{Var(X)}{\lambda^2}$$

Where λ is any real positive integer.

Proof: For a positive r.v. X having finite mean, we have by Markov's inequality.

$$P[X \geq \lambda] \leq \frac{E(X)}{\lambda}$$

If we substitute {X-E$(X)^2$} in place of X and λ^2 in this inequality, we have

$$P[\{X - E(X)^2\} \geq \lambda^2] \leq \frac{E\{X - E(X)\}^2}{\lambda^2}$$

or,

$$P[\,|X - E(X)| \geq \lambda] \leq \frac{Var(X)}{\lambda^2}$$

We know that P(A) + P($\bar{A}$) =1. Hence the above expression may also be written as

$$P[\,|X - E(X)| < \lambda] \geq 1 - \frac{Var(X)}{\lambda^2}$$

Applications:

Problem 3.5.1: Let X denotes the number obtained on tossing of an unbiased die. Then prove that Chebyshev inequality gives.

$$P[\,|X - E(X)| > 2.5] < 0.47$$

While the actual probability is nearly zero.

Solution: We have

$$E(X) = \frac{1}{6}(1 + 2 + 3 + 4 + 5 + 6) = \frac{21}{6} = \frac{7}{2} = 35$$

and

$$V(X) = E(X^2) - \{E(X)\}^2$$

$$E(X^2) = \frac{1}{6}(1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2)$$

$$= \frac{1}{6}(1 + 4 + 9 + 16 + 25 + 36) = \frac{91}{6}$$

$$V(X) = \frac{91}{6} - \frac{49}{4} = \frac{182 - 147}{12} = \frac{35}{12} = 2.9167$$

By Chebyshev Inequality, we have

$$P\,[\,|X - E(X)| > \lambda] < \frac{V(X)}{\lambda^2}$$

Or,

$$P[\,|X - 35| > 2.5] < \frac{2.9167)}{6.25} = 0.47$$

We have

$$P[|-3.5| > 2.5] = 1 - P[|X - 3.5| \leq 2.5]$$

$$= 1 - \Pr[-2.5 \le (X - 3.5) \le 2.5]$$

$$= 1 - \Pr[\, 1 \le X \le 6]$$

$$= 1 - [\mathrm{P}(X = 1) + \mathrm{P}(X = 2) + \mathrm{P}(X = 3) + \mathrm{P}(X = 4) + \mathrm{P}(X = 5) + \mathrm{P}(X = 6)]$$

$$= 1 - \left[\frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6}\right] = 0$$

Problem 3.5.2 For a.r.v. X with finite variance.

$$\mathrm{P}[\,|X - E(X)| > k\sigma] \le \frac{1}{k^2} \qquad for\ k > 0$$

Solution: For a r.v. X having finite mean and variance, we have by Chebyshev inequality.

$$\mathrm{P}[\,|X - E(X)| > \lambda] \le \frac{V(X)}{\lambda^2}; \qquad \lambda > positive$$

$$Taking\ \ \lambda = k\sigma \qquad where\ \sigma^2 = var(X)$$

We have

$$\mathrm{P}[\,|X - E(X)| > k\sigma] \le \frac{\sigma^2}{k^2\sigma^2};$$

i.e., $\mathrm{P}[|X - E(X)| > k\sigma] \le \frac{1}{K^2}$

Proved.

Problem 3.5.3: A random variable X has a mean value 3 and variance 2. What is the least value of $\mathrm{P}[|X - 3| < 2]$?

Solution: We have (by Chebyshev's inequality)

$$\mathrm{P}[|X - E(X)| < \lambda] \ge 1 - \frac{Var(X)}{\lambda^2}$$

Here, E(X) = 2, λ = 2 and Var (X) = 2

Hence,

$$\mathrm{P}[|X - 3| < 2] \ge 1 - \frac{2}{4}$$
$$\ge \frac{1}{2}$$

So the lest value is 1/2.

Problem 3.5.4: Use Chebyshev's inequality to determine how many times an unbiased coin must be tossed in order that the probability will be at least 0.9 that the ratio of observed number of heads to the number of tosses will be between 0.4 and 0.6.

Solution: Let X be the observed number of heads and n be the number of tosses of an unbiased coin. Then,

$$E(X) = np = \frac{n}{2} \qquad \left(here\ p = q = \frac{1}{2}\right)$$

$$and\ V(X) = npq = \frac{n}{4}$$

$$Let\ F = \frac{X}{n}$$

Then,

$$E(T) = \frac{E(X)}{n} = \frac{1}{2} = 5$$

$$Var\ (T) = \frac{Var\ (X)}{n^2} = \frac{1}{4n} = .\frac{25}{n}$$

Applying Chebyshev's inequality to T, we have

$$\mathrm{P}[|T - E(T)| \leq \lambda] \geq 1 - \frac{Var(T)}{\lambda^2}$$

$$Taking\ \ \lambda = 0.1\ \ and\ E(T) = \ 0.5, \quad We\ have$$

$$\mathrm{P}[|T - 5| \leq 0.1] \geq 1 - \frac{0.25}{n \times 0.01}$$

Or,

$$\mathrm{P}[\,0.4 \leq T \leq 0.6] \geq 1 - \frac{25}{n}$$

According to the question

$$1 - \frac{25}{n} = 0.9 \quad or\ 0.1 = \frac{25}{n} \quad or \quad n = \frac{25}{.1} = 250$$

Hence the coin must be tossed 250 times.

Problem 3.5.5: For geometric distribution

$$p(x) = \frac{1}{2^x};\ \ x = 1, 2, 3, \dots \dots ..$$

Prove that the Chebyshev's inequality gives.

$$\mathrm{P}[|X - 2| \leq 2] > \frac{1}{2}$$

While actual probability is 15/16.

Solution: In the case E(X) = 2, E(X^2) = 6 and Var (X) = 2.

By Chebychev's inequality

$$\mathrm{P}[|X - E(X)| \leq \lambda] > 1 - \frac{Var(X)}{\lambda^2}$$

$$Taking\ \ \lambda = 2\ \ and\ \ substituting\ E(X)\ and\ Var\ (X)$$

$$We\ have \quad \mathrm{P}[|X - 2| \leq 2] > 1 - \frac{2}{4}$$

$$also, we\ may\ write \quad > \frac{1}{2}$$

$$\mathrm{P}[|X - 2| \leq 2] = \Pr[-2 \leq (X - 2) \leq 2]$$

$$= \Pr[0 \leq X \leq 4]$$

$$= p(1) + p(2) + p(3) + p(4)$$

$$= \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} = \frac{8 + 4 + 2 + 1}{16} = \frac{15}{16}$$

5.6 Self Assessment Exercises

1. A symmetric die is thrown 600 times find the lower bound for the probability of getting 80 to 120 sixes.

2. An unbiased coin is tossed 100 times. Show that the probability that the number of heads will be between 30 and 70 is greater than 0.93.

5.7 Summary

This unit gives an over view of some inequalities (namely Cauchy Schwarz, Markov and Chebyshev) with their applications and importance.

5.8 Further Readings

1. Cramer H, Mathematical Methods of statistics, Princeton University Press, 1946 and Asia Publishing House, 1962.
2. Hogg R.V. and Craig A.T., Introduction to Mathematical Statistics, Macmillan, 1978.
3. Prazen E. Modern Probability Theory and its Applications, John Wiley, 1960 and Wiley Eastern 1972.
4. Rao C. R., Linear Statistical Inference and Its Applications John Wiley, 1973 and Wiley Eastern 1974.
5. Rohtagi V. K., An Introduction to Probability Theory and Mathematical Statistics John Wiley, 1976 and Wiley Eastern 1985.
6. Vikas S.S., Mathematical Statistics, John Wiley, 1962 and Toppan.

U.P. Rajarshi Tandon Open University, Prayagraj

UGSTAT – 102

Probability, Distribution and Statistical Inference

Block

3

Concepts of Probability Distributions

Unit - 6
Univariate Distributions

Unit - 7
Discreet Distributions

Unit - 8
Normal Distributions

Unit - 9
Continuous Distributions

Unit - 10
Sampling Distributions

Course Design Committee

Dr. Ashutosh Gupta **Chairman**
Director, School of Sciences
U. P. Rajarshi Tandon Open University, Prayagraj

Prof. Anup Chaturvedi **Member**
Ex. Head, Department of Statistics
University of Allahabad, Prayagraj

Prof. S. Lalitha **Member**
Head, Department of Statistics
University of Allahabad, Prayagraj

Prof. Himanshu Pandey **Member**
Department of Statistics, D. D. U. Gorakhpur University,
Gorakhpur

Dr. Shruti **Member-Secretary**
Sr. Assistant Professor, School of Sciences
U.P. Rajarshi Tandon Open University, Prayagraj

Course Preparation Committee

Dr. Shruti **Course Coordinator**
Sr. Assistant Professor, School of Sciences
U. P. Rajarshi Tandon Open University, Prayagraj

Prof. B. P. Singh **Writer**
Department of Statistics, Banaras Hindu University, Varanasi

Dr. Padmakar Singh **Writer**
Department of Statistics, U.P.P.G. College, Varanasi

Dr. Alok Kumar **Writer**
Department of Community Medicine, Institute of
Medical Sciences, Banaras Hindu University, Varanasi

Dr. Shruti **Writer**
School of Sciences, U. P. Rajarshi Tandon Open University,
Prayagraj

Prof. V. P. Ojha **Reviewer**
Department of Statistics and Mathematics,
D. D. U., Gorakhpur University, Gorakhpur

Dr. Padmakar Singh **Reviewer**
Department of Statistics, U.P.P.G. College, Varanasi

Dr. D. C. Singh **Reviewer**
Department of Statistics, Harish Chandra P.G. College, Varanasi

Prof. R. C. Yadava **Editor**
Department of Statistics, Banaras Hindu University, Varanasi

Prof. S. K. Pandey **Editor**
Department of Statistics, Lucknow University, Lucknow

Prof. G. S. Pandey **Editor**
Department of Statistics, University of Allahabad, Prayagraj

First Edition: *March 2008* (Published with the support of the Distance Education Council, New Delhi)

Second Edition: ***July 2020***

ISBN :

Printed and Published by Dr. Arun Kumar Gupta Registrar, Uttar Pradesh Rajarshi Tandon Open University, 2020.

Printed By: Chandrakala Universal Pvt. 42/7 Jawahar Lal Neharu Road, Prayagraj.

Unit - 6: Univariate Distribution

Structure

6.1 Introduction

In this unit, we first introduce the concept of a distribution function of a random variable and its properties. In section 1.4 to 1.8, we have discussed discrete and continuous random variables, probability mass function and probability density function. The discrete distributions Bernoulli, Binomial and Poisson are discussed in section 6.9 to 6.23.

6.2 Objectives

After reading this unit you should be able to:

- Define the distribution function for a random variable, p.m.f. of discrete random variable, and p.d.f. of continuous random variable.
- Identify the properties of distribution function.
- Describe the situations underlying the Bernoulli, Binomial and Poisson distribution.
- Compute their means and variances.
- Obtain distribution of the sum of two independent binomial variates.
- Obtain Poisson distribution as limiting case of Binomial distribution.
- Work out the additive property of independent Poisson variates.

6.3 Distribution Function

The distribution function F for a random variable X is a function defined on the real line by

$$F(X) = P[X \leq x], \qquad where \ -\infty < X \infty.$$

The definition makes sense because if X is a random variable, then $[X \leq x]$ is an event in sample space Ω. This function is sometimes called the cumulative univariate distribution function. Corresponding random variable is one variable that is why it is called univariate.

Let us now try to understand the distribution function by looking at some of its properties.

6.4 Properties of a Distribution Function F(X)

We derive a number of properties common to all distribution functions.

Property 1: If *F* is a distribution function (d.f.) of the random variable (r.v.) X and if a < b, then

(i) $0 \leq F(x) \leq 1$ (ii) $F(x) \leq F(y)$ if $x < y$
$P[a < X \leq b] = F(b) - F(a)$

Proof: The event $a < X \leq b$ and $X \leq a$ are disjoint and their union is the event $X \leq b$. Hence by addition theorem of probability.

$$P[a < X \leq b] + P[X \leq a] = P[X \leq b]$$

$$\Rightarrow P[a < X \leq b] = P[X \leq b] - P[x \leq a] = F(b) - F(a)$$

Property 2: If F is the d.f. of one dimensional random variable X, then

(i) $0 \leq F(x) \leq 1$ (ii) $F(x) \leq F(y)$ if $x < y$

Proof: Using the axioms of certainty and non- negative for the probability function part (i) follows from the definition of F (x).

For part (ii), we have $x < y$,

$F(y) - F(x) = P\,[x < X \leq y] \geq 0$

$\Rightarrow F(y) \geq F(x)$

$\Rightarrow F(x) \leq F(y)$ when $x < y$

Property 3: If F is d.f. of one dimensional random variable X, then

$$F(-\infty) = \lim_{x \to \infty} F(x) = 0.$$

$$and\ F(\infty) = \lim_{x \to \infty} F(x) = 1$$

Proof: Let us express the whole sample space Ω as a countable union of disjoint events as follows:

$$\Omega = \left[\bigcup_{n=1}^{\infty}(-n < x \leq -n+1)\right] \bigcup \left[\bigcup_{n=1}^{\infty}(n < x \leq n+1)\right]$$

$$= P(\Omega) = \sum_{n=1}^{\infty} P(-n < x \leq -n+1) + \sum_{n=0}^{\infty} P(n < x \leq n+1)$$

$$(Since\ P\ is\ additive)$$

$$1 = \lim_{a \to \infty} \sum_{n=1}^{a} [F(-n-1) - F(-n) + \lim_{b \to \infty} \sum_{n=0}^{b} [F(n+1) - F(n)]$$

$$= \lim_{a \to \infty} [F(0) - F(-a) + \lim_{b \to \infty} [F(b+1) - F(0)]$$

$$= [F(0) - F(\infty)] + [F(0)$$

$$or, 1 = F(\infty) - F(-\infty) \qquad (1.1)$$

Since $-\infty < \infty, F\,(-\infty) < F(\infty)$.

$$Also, \quad F(-\infty) = 0 \qquad and \qquad f(\infty) \leq 1$$

$$\therefore \; o \leq F(-\infty) \leq F(\infty) \leq 1 \qquad (1.2)$$

Equ. (1.1) and (1.2) gives

$$F(-\infty) = 0 \; and \; F(\infty) = 1.$$

6.5 Discrete Random Variable

If a random variable takes at most a countable number of real values, it is called a discrete random variable.

In other words a real valued function defined on a discrete sample space is called a discrete random variable.

6.6 Probability Mass Function *(p.m.f.)*:

Suppose X is a one dimensional discrete random variable taking at most accountably infinity number of values X_1, X_2 …………… With each possible outcome X_i has a number p_i = $P(X = x_i) = P(x_i)$ called the probability of X_i, The number $p(x_i)$, i = 1, 2,…….. must satisfy the following conditions:

$$(i)\; p(x_i) \geq 0 \qquad \forall i \qquad and \qquad (ii) \sum_{i=1}^{\infty} p(x_i) = 1$$

This function p in called the probability mass function of the random variable X and set {p (Xi)} is called the probability distribution (p.d.) of the random variable X.

6.7 Discrete Distribution Function

In this case there are countable numbers of points X_1, X_2……..and

$$p_i \geq 0, \qquad \sum_{i=1}^{\infty} p_i = 1 \qquad such\ that \qquad F(x) = \sum_{ixi \leq x} p_i$$

Theorem: $p(x_i) = P\left(X = x_j\right) = F(x_j)$, where F in the d.f. of X.

Proof: Let $X_1 < X_2 <$ ……. we have

$$F(Xj) = P[X - Xj] = \textstyle\sum_{i=1}^{j} P(X = x_i) = \sum_{i=1}^{j} p(x_i) \qquad (1.3)$$

$$and\ F\left(x_{j-1}\right) = P\left(X \leq x_{j-1}\right) = \sum_{i=1}^{j-1} p(x_i) \qquad (1.4)$$

$$\therefore \qquad F(x_j) - F(x_{j-1}) = P(x_j).$$

Thus, if the d.f. discrete r.v. are given, we can calculate probability mass function.

6.8 Continuous Random Variable

If a random variable X takes all possible values between certain intervals then it is called continuous random variable.

In other words, if a variable can take an infinite set of values in given interval say, a~X~ b it is called a continuous random variable and its distribution is accordingly known as continuous distribution.

6.9 Probability Density Function *(p.d.f.)*

Consider the small interval $\left[x - \frac{1}{2}dx,\ x + \frac{1}{2}dx\right]$ of length dx round the point x. Let f(x) be any continuous function of x so that f(x) dx represents the probability that x falls in the infinitesimal interval , [x - (dx/2), x + (dx/2)] Symbolically

$$f_x(x)dx = P\left[\left(x - \frac{dx}{2}\right) \leq X \leq \left(x + \frac{dx}{2}\right)\right]$$

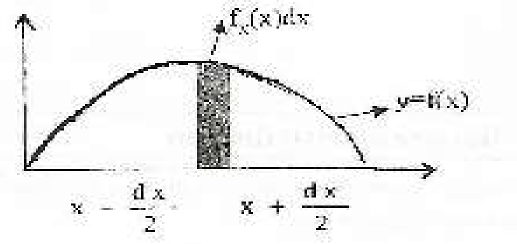


Fig.

In the figure fx (x) dx represents the area bounded by the curve y = f(x), x axis and the ordinates at points $x - \frac{dx}{2}$ and $x + \frac{dx}{2}$. The function fx(x) so defined is known as probability density function of random variable X and is usually abbreviated by p.d.f. Further since total probability is unity, we have $\int_a^b f_x(x)dx = 1$, where [a, b] is the range of the random variable X. Range may be finite or infinite.

The probability density function of random variable X usually denoted by fx (x) or simply f(x) has the following obvious properties:

(i) f(x) ~ 0 V× ER

(ii) 1= P(R) = $\int_R f(x)dx = \int_{-\infty}^{\infty} f(x)dx$

Where R is the collection of all points in the entire range of the variable X.

Example: A continuous random variable X has p.d.f.

$$F(x) \sim 5x^4, \quad 0 \le x \le 1. \quad Find\ a\ and\ b\ such\ that$$

$$P[X, S \le a] = P[X > a]$$

Solution: Since P[X, S: a] = P[X > a], each must be equal to ½ because total probability is always one.

$$\therefore P[X \le a] = \frac{1}{2}$$

$$\Rightarrow \int_0^a f(x)dx = \frac{1}{2}$$

$$\Rightarrow 5\int_0^a x^4 dx = \frac{1}{2}$$

$$\Rightarrow 5\left[\frac{x^5}{5}\right]^a = \frac{1}{2}$$

$$\Rightarrow a^5 = \frac{1}{2}$$

$$\therefore a = \left(\frac{1}{2}\right)^{\frac{1}{5}}$$

6.10 The Bernoulli Distribution

It is simplest probability distribution. It is the distribution of a r.v. X which assumes two values, 0 and 1.

Let P[X = 1] = p and P[X = 0] = 1 - p = q

Or P[X = x] = $p^x(1-p)^{1-x} = p^x q^{1-x},$ $\quad X = 0, 1$ (1.5)

Where p is the number such that 0 ≤ p ≤ 1.

The random variable X and its probability distribution specified by the p.m.f. eq. (1.5) are respectively, called Bernoulli variable and the Bernoulli distribution in honour of Jacob Bernoulli. He made a systematic study of problem connected with this distribution.

6.11 Moments of Bernoulli Distribution

The r^{th} moment about origin is

$$\mu_r = E(X^r) = o^r.q + 1^r.p = p, \qquad r = 1, 2, \dots \dots$$

$$Thus\ \mu_1^1 = E(X) = P, \qquad i.e. mean = p$$

$$and\ \mu_2^1 = E(X^2) = P$$

$$\therefore\ \mu_2 = variance = E(X^2) - [E(X)]^2$$

$$= (p - p^2) = pq$$

6.12 Moment Generating Function *(m.g.f.)*

The m.g.f. is

$$M_x(t) = E[e^{tX}] = q.e^0 + p.e' = q + pe'$$

6.13 Binomial Distribution

Definition: The binomial distribution can be defined in terms of the expression of the binomial $(q + p)^n$, where p > 0, q > 0 and n is the positive integer. The $(x + 1)^{+n}$ term in the expansion of $(q + p)^n$ is

$$\binom{n}{x} p^x q^{n-x} = \frac{n!}{x!\,(n-x)!} p^x q^{n-x}$$

The binomial distribution with parameter n, p distribution of random variable X for which,

$$P[X = x] = \binom{n}{x} p^x q^{n-x} \qquad (x = 0, 1, 2, \dots\dots, n)$$

We can interpret binomial distribution as the distribution of the total number of successes in n independent trials, each with same probability p, of success. The probability of x success and consequently (n - x) failures in n trials in a specified order (say) SSS FFS….FS (where S represents success and F failure) is given by the compound probability theorem by expression

P[SSSFFS…..FS] = P (S). P(S).P(S).P(F).P(F).P(S)………P(F).P(S)

= p.p.p.q.q.p ………………q.p

= p.p p. q.q q= p^x q^{n-x}

{x factors} {(n-x) factors}

But x successes in n trial can occur in $\binom{n}{x}$ ways and probability for each of these ways is $p^x\ q^{n-x}$. Hence the probability of x successes in n trials in any order what so ever is given by the addition theorem of probability by expression

$$\binom{n}{x} p^x\ q^{n-x}, \qquad x = 0, 1, 2, \dots\dots..n.$$

Example 1: Ten coins are throwing simultaneously. Find the probability of getting at least eight heads.

Solution:

p = probability of getting a head = 1/2

q = probability of not getting a head = 1/2

The probability of getting x heads in a throw of 10 coins is

$$p(x) = \binom{10}{x}\left(\frac{1}{2}\right)^{x}\left(\frac{1}{2}\right)^{10-x} = \binom{10}{x}\left(\frac{1}{2}\right)^{10} . x = 0, 1, 2, \dots\dots, 10$$

∴ Probability of getting x heads is given by

P[x ≥ 8] = P(8) + P(9) + P(10)

$$= \left(\frac{1}{2}\right)^{10}\left[\binom{10}{8} + \binom{10}{9} + \binom{10}{10}\right] = \frac{45 + 10 + 1}{1094} = \frac{56}{1024} = \frac{7}{128}$$

Example 2: The probability that a person recovers from a serious decrease is 0.30. Find the probability that at least one of 5 persons admitted to a hospital will survive.

Solution: If X denotes the number of persons recovered from decrease. Then we have to find P[X∼1]. X has binomial distribution with n = 5, p = 0.30.

We have

$$P[X \geq 1] = 1 - P[x = 0]$$

$$= 1 - \binom{5}{0}(0.30)^{0}(0.70)^{5} = 1 - 0.168 = 0.832$$

6.14 Moments

The first four moments about origin of binomial distribution are obtained as follows:

$$\mu_1' = E(X) = \sum_{x=0}^{n} x.\binom{n}{x} \mathrm{p}^{\mathrm{x}}\ \mathrm{q}^{\mathrm{n-x}} = \mathrm{np}\sum_{x=0}^{n}\binom{n-1}{x-1}\mathrm{p}^{\mathrm{x-1}}\ \mathrm{q}^{\mathrm{n-x}}$$

$$= np(q + p)^{n-1} = np \qquad (\because pq = 1)$$

Thus the mean of binomial distribution is np.

$$\mu_2' = E(X^2) = \sum_{x=0}^{n} x^2 \binom{n}{x} \mathrm{p^x\ \ q^{n-x}}$$

$$\sum_{x=0}^{n} [x(x-1)+x] \frac{n(n-1)}{x(x-2)} \binom{n-2}{x-2} \mathrm{p^x\ q^{n-x}}$$

$$= n(n-1)p^2 \left[\sum_{x=2}^{n} \binom{n-2}{x-2} \mathrm{p^{x-2}\ q^{n-x}}\right] + \sum_{x=0}^{n} x \binom{n}{x} \mathrm{p^x\ \ q^{n-x}}$$

$$= n(n-1)p^2(q+p)^{n-2} + np = n(n-1)p^2 + np$$

$$\mu_3' = E(X^3) = \sum_{x=0}^{n} x^3 \binom{n}{x} \mathrm{p^x\ \ q^{n-x}}$$

$$= \sum_{x=0}^{n} [x(x-1)(x-2) + 3x(x-1) + x] \binom{n}{x} \mathrm{p^x\ q^{n-x}}$$

$$= n(n-1)(n-2p^3) \sum_{x=3}^{n} \binom{n-2}{x-2} \mathrm{p^{x-3}\ q^{n-x}} +$$

$$3n(n-1)p^2 \left[\sum_{x=2}^{n} \binom{n-2}{x-2} \mathrm{p^{x-1}\ q^{n-x}}\right] + \sum_{x=0}^{n} x \binom{n}{x} \mathrm{p^x\ \ q^{n-x}}$$

$$= n(n-1)(n-2)p^3(q+p)^{n-3} + 3n(n-1)p^2(q+p)^{n-2} + np$$

$$= n(n-1)(n-2)p^3 + 3n(n-1)p^2 + np$$

$$\mu_4' = E(X^4) = \sum_{x=0}^{n} x^4 \binom{n}{x} \mathrm{p^x\ \ q^{n-x}} \qquad (1.6)$$

Now we have

$$x^4 = x(x-1)(x-2)(x-3) + 6x(x-1)(x-2) + 7x(x-1) + x \qquad (1.7)$$

Putting x^4 as from eqn. (1.7) in eqn. (1.6) we have

$$\mu_4' = n(n-1)(n-2)(n-3)p^4 + 6n(n-1)(n-2) + p^2(n-1) + p^2 + np$$

(on simplification)

Central Moments of Binomial Distribution

$$\mu_2 = \mu_2^2 - (\mu_1')^2 = n(n-1)p^2 + np - n^2p^2 \ = n^2p^2 - np^2 + np - n^2p^2$$

$$= np(1-p) = npq$$

$$\mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2(\mu_1')^3$$

$$= [n(n-1)(n-2)p^3 + 3n(n-1)p^2 + np] - 3[\{n(n-1)p^2 + np\}np] + [2(np)^3]$$

$$= np[n^2p^2 - 3np^2 + 2p^2 + 3np - 3p + 1 - 3n^2p^2 - 3np^2 - 3np + 2n^2p^2]$$

$$= [2p^2 - 3p + 1] = np(1 - p(1-2p)) = np(1-2p) = npq(p + q - 2p)$$

$$= npq(p-q)$$

$$\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'(\mu_1')^2 - 3(\mu_1')^4 = npq[1 + 3(n-2)pq]$$

Hence,

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{n^2p^2q^2(q-p)^2}{n^3p^3q^3} = \frac{(1-2p)^2}{npq}$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{[npq(1+3)(n-2)pq]}{n^2p^2q^2} = 3 + \frac{1-6pq}{npq}$$

$$\gamma_1 = \sqrt{\beta_1} = \frac{1-2p}{\sqrt{npq}}$$

$$\gamma_2 = \beta_2 - 3 = \frac{1-6pq}{npq}$$

Example 1: Comment on the following:

The mean of a binomial distribution is 4 and variance is 5.

Solution: If the given binomial distribution has parameter n and p then we have

mean = n p = 4 (1.8)

and variance = n p q = 5 (1.9)

Dividing eqn. (1.9) by (1.8), we get

q= 5/4

which is impossible, since probability cannot exceed unity. Hence given statement is wrong.

Example 2: The mean and variance of binomial distribution are 3 and 3/4 respectively. Find P[X ≥ 1].

Solution:

Mean = E(X) = nq = 3

Var (x) = npq = 3/4

Dividing we get q = 1/4 $\therefore p = \frac{3}{4}$

Now we have n p = 3

$$\Rightarrow \frac{3}{4}n = 3$$

$$\Rightarrow n = \frac{4}{3} \times 3 = 4$$

$$\therefore P[X \geq 1] = 1 - P[X = 0] = 1 - q^n$$

$$= 1 - \left(\frac{1}{4}\right)^4 = 1 - \frac{1}{256} = \frac{255}{256} = 0.996$$

Example 3: If $Y = \frac{X}{n}$ denotes the proportion of success in n independent Bernoulli trials with constant p of success then

$$E(Y) = p \qquad and \qquad V(Y) = \frac{p(1-p)}{n}$$

6.15 Moment Generating Function of Binomial Distribution

The moment generating function $M_x(t)$ of the binomial distribution with parameter n and p is

$$M_x(t) = E(e^{tx}) = \sum_{x=0}^{n} e^{tx} \binom{n}{x} \mathrm{p}^{\mathrm{x}}\ \mathrm{q}^{\mathrm{n-x}} = \sum_{x=0}^{n} e^{tx} \binom{n}{x} (pe^t)^x \mathrm{q}^{\mathrm{n-x}}$$

$$= (\mathrm{q} + pe^t)^{\mathrm{n}} = [1 + p(e^t - 1)]^n$$

6.16 Additive Property of Binomial Distribution

Let X ~ B(n_1, p_1) and Y ~ B(n_2, p_2) are independent random variables.

Then

$$M_x(t) = (q_1 + p_1 e^{'})^{n_1}, \quad M_y(t) = (q_2 + p_2 e^{'})^{n_2} \qquad (1.10)$$

We have

$$M_{x+y}(t) = Mx(t).My(t) \qquad (Since\ X\ and\ Y\ are\ indepedent)$$

$$= (q_1 + p_1 e^{'})^{n_1} (q_2 + p_2 e^{'})^{n_2} \qquad (1.11)$$

Since equ. (1.11) cannot be expresses on the form $(q + pe^t)^n$

From uniqueness theorem of m. g. f it follows that X + Y is not a binomial variate. Hence in general the sum of independent binomial variates is not a binomial variate. In other words, binomial distribution not posses the additive or reproductive property.

Corollary: However if $p_1 = p_2 = p$. (say), then from equ. (1.11) we have

$$M_{x+y}(t) = (q_1 + p_1 e^{'})^{n_1} / (q_2 + p_2 e^{'})^{n_2} = \left(q + pe^{t'}\right)^{n_1+n_2}$$

Which is m. g. f. of binomial distribution with parameter $(n_1 + n_2, p)$ i.e., when $p_1 = p_2$, binomial distribution posses the additive property.

6.17 Recurrence Relation for the Probabilities of Binomial Distribution (Fitting of Binomial Distribution)

We have

$$p(x) = \binom{n}{x} p^x\, q^{n-x}$$

$$p(x+1) = \binom{n}{n+1} p^{x+1}\, q^{n-x-1}$$

$$\therefore \qquad \frac{p(x+1)}{p(x)} = \frac{\binom{n}{n+1} p^{x+1}\, q^{n-x-1}}{\binom{n}{x} p^x\, q^{n-x}}$$

$$\therefore \qquad p(x+1) = \left[\frac{n-x}{x+1} . \frac{p}{q}\right] . p(x)$$

This is a required recurrence formula.

To fit a binomial distribution to the given data, the first step is to estimate the parameter p if not given otherwise. The parameter p is estimated $\hat{p} = \frac{\bar{x}}{n}$, $where\ \bar{x} =$ the mean of observed distribution. The expected frequency of x successes is computed by

$$N. \binom{n}{x} p^x\, q^{n-x}, \qquad x = 0, 1, 2, 3 \ldots\ldots, n$$

6.18 Poisson Distribution

Let X be a discrete random variable with probability mass function

$p_x = P[X = x] = \frac{e^{-\lambda}}{x!}$, $\quad$ x = 0, 1, 2,……, ∞

$= 0 \qquad else\ where,$

Where λ > 0 is a constant.

The probability distribution having p(x) = p(x, λ) = $\frac{e^{-\lambda}\lambda^x}{x!}$, as its probability mass function is known as Poisson distribution with parameter λ.

This distribution was discovered by Simen Denis Poisson in 1837. Poisson distribution is a limiting case of binomial distribution under the following conditions:

(i) n, the number of trails is indefinitely large, i.e., n→ ∞.

(ii) p, the constant probability of success for each trial is indefinitely small i.e., p→ 0.

(iii) np = λ (say), is finite. Thus $= \frac{\lambda}{n}\ and\ q = 1 - \frac{\lambda}{n}$, where is a positive real number.

6.19 Poisson Distribution as a limiting case of Binomial Distribution

The probability of x successes in a series of n independent trails is

$$P(X = x) = b(x; n, p) = \binom{n}{x} p^x\, q^{n-x}, \qquad x = 0, 1, 2, 3 \ldots\ldots, n$$

Can be written as

$$\frac{n!}{x!\,(n-x)!}\left(\frac{\lambda}{n}\right)^x \left(1-\frac{\lambda}{n}\right)^{n-x} = \frac{\lambda^x}{x!}\left(1-\frac{\lambda}{n}\right).\frac{n!}{(n-x)!\,n^x\left(1-\frac{\lambda}{n}\right)^x}$$

We know that n → ∞.

$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n = e^{-\lambda}, \qquad \lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^r = 1$$

Hence by making n→ ∞ , the probability of x successes is

$$= \frac{\lambda^x}{x!} e^{-\lambda} \lim_{n\to\infty} \frac{n!}{(x-x)!\,n^x}$$

Using Strling's formula for n!, viz.,

$$\lim_{n\to\infty} n! = \sqrt{2\pi n^{n+1/2}}\, e^{-n}$$

We have probability of x successes is

$$= \frac{\lambda^x e^{-\lambda}}{x!} \lim_{n\to\infty} \frac{\sqrt{2\pi n}^{\,n+1/2} e^{-n}}{n^x \sqrt{2\pi}(n-x)^{n-x+1/2} e^{-(n-x)}}$$

$$\Rightarrow \frac{\lambda^x e^{-\lambda}}{x!} \lim_{n\to\infty} \frac{n^{n+1/2}}{n^x n^{n-x+1/2}\left(1-\frac{x}{n}\right)^{n-x+1/2} e^x}$$

$$= \frac{\lambda^x e^{-\lambda}}{x!} \lim_{n\to\infty} \left[\frac{1}{\left(1-\frac{x}{n}\right)^n \left(1-\frac{x}{n}\right)^{n-x+1/2}}\right]$$

$$= \frac{\lambda^x e^{-\lambda}}{x!\, e^x} \frac{1}{e^{-x}.1} = \frac{\lambda^x e^{-\lambda}}{x!}$$

The following are some examples of Poisson Variate:

(i) The number of defective screws per box of 100 screws.

(ii) The number of printing mistakes at each page of a book.

(iii) The number of deaths in a district in one year by rare disease.

(iv) The number of air accidents in some unit of time.

(v) The number of suicides reported in a particular city.

(vi) The number of cars passing through a certain street in time.

Note 1: $p(x) = \beta_1 = 0 \; and \; \beta_2 = 3$ is probability distribution.

$$Since \sum_{x=0}^{\infty} p(x) = \sum_{x=0}^{\infty} \frac{e^{-\lambda}\lambda^x}{x!} = e^{-\lambda}\left[1 + \frac{\lambda}{1!} + \frac{\lambda^2}{2!} + \cdots.\right] = e^{-\lambda}.e^{\lambda} = 1$$

Note 2: ***Characteristics of Poisson Distribution***

(i) It is a limiting form of the Binomial Distribution.

(ii) Hence p (or it may be q as well) is very small close to zero. So it assumes a J shaped distribution.

(iii) It consists of a single parameter λ, only. Thus entire distribution can be obtained by knowing the volume of λ alone.

6.20 Moments of the Poisson Distribution

$$\mu_1' = E(X) = \sum_{x=0}^{\infty} x.p(x) = \sum_{x=0}^{\infty} x.\frac{e^{\lambda}\lambda^x}{x!} = e^{-\lambda}.\lambda \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!}$$

$$= e^{-\lambda}.\lambda\left[1+\lambda+\frac{\lambda^2}{2}+\cdots..\right] = e^{-\lambda}.\lambda e^{\lambda} = \lambda.$$

Hence the mean of Poisson Distribution is λ.

$$\mu_2' = E(X^2) = \sum_{x=0}^{\infty} x^2 = \sum_{x=0}^{\infty}\{2x(x-1)+x\}\frac{e^{-\lambda}\lambda^x}{x!} = \lambda^2 e^{-\lambda}\sum_{x=2}^{\infty}\frac{\lambda^{x-2}}{(x-2)!}+\lambda$$

$$= \lambda^2 e^{-\lambda}\left[1+\frac{\lambda}{1!}+\frac{\lambda^2}{2!}+\cdots\ldots\right]+\lambda = \lambda^2 e^{-\lambda}e^{\lambda}+\lambda = \lambda^2+\lambda$$

$$\mu_3' = E(X^3) = \sum_{x=0}^{\infty} x^3 p(x) = \sum_{x=0}^{\infty} x^3\frac{e^{-\lambda}\lambda^x}{x!} = \sum_{x=0}^{\infty}[x(x-1)(x-2)]e^{-\lambda}\ \frac{\lambda^3}{x!}+\sum_{x=0}^{\infty} x\frac{e^{-\lambda}\lambda^x}{x!}$$

$$= e^{-\lambda^3}e^{\lambda}+3e^{-\lambda}\lambda^2 e^{\lambda}+e^{-\lambda}\lambda, e^{\lambda}\ = \lambda^3+3\lambda^2+\lambda$$

$$\mu_4' = E(X^4) = \sum_{x=0}^{\infty} x^4 p(x)\ = \sum_{x=0}^{\infty} x^4\frac{e^{-\lambda}\lambda^x}{x!}$$

$$= \sum_{x=0}^{\infty}[x(x-1)(x-2)(x-3)+6x(x-1)(x-2)+7x(x-1)+x]\frac{e^{-\lambda}\lambda^x}{x!}$$

$$= e^{-\lambda}\lambda^x\left[\sum_{x=4}^{\infty}\frac{\lambda^{x-4}}{(x-4)!}\right]+6e^{-\lambda}\lambda^3\left[\sum_{x=3}^{\infty}\frac{\lambda^{x-2}}{(x-2)!}\right]+7e^{-\lambda}\lambda^2\left[\sum_{x=3}^{\infty}\frac{\lambda^{x-2}}{(x-2)!}\right]+\lambda$$

$$= e^{-\lambda}\lambda^4 e^{\lambda}\ +6e^{-\lambda}\lambda^3 e^{\lambda}+7e^{-\lambda}\lambda^2.e^2+\lambda = \lambda^4+6\lambda^4+7\lambda^2+\lambda$$

The four central moments are now obtain as follows:

$$\mu_2 = \mu_2' - (\mu_1')^2 = (\lambda^2+\lambda)-\lambda^2 = \lambda$$

Thus the mean and variance of Poisson distribution are equal to λ.

$$\mu_3 = \mu_3' - 3\mu_1'\mu_2' + 2(\mu_1')^3 = (\lambda^3+3\lambda^2+\lambda)-3\lambda(\lambda^2+\lambda)+2\lambda^3 = \lambda$$

$$\mu_4 = \mu_4' - 4\mu_1' + 6\mu_2'\mu_1'^2 - 3(\mu_1')^4$$

$$= (\lambda^4+6\lambda^3+7\lambda^2+\lambda)-4\lambda(\lambda^3+3\lambda^2+\lambda)-3\lambda^4 = 3\lambda^2+\lambda$$

Coefficients of Skewness and Kurtosis:

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{\lambda^2}{\lambda^3} \qquad and \qquad \beta_2 = \frac{\mu_4}{\mu_2} = \frac{3\lambda^2+\lambda}{\lambda^2} = 3+\frac{1}{\lambda}$$

Also,

$$\gamma_1 = \sqrt{\beta_1} = \frac{1}{\sqrt{\lambda}} \qquad and \qquad \gamma_2 = \beta_2 - 3 = \frac{1}{\lambda}$$

Note: As λ→ ∞ we have

$$\beta_1 = 0 \qquad and \qquad \beta_2 = 3$$

6.21 Moment Generating Function of Poisson Distribution *(m.g.f.)*

$$M_{\mathrm{x}}(\mathrm{t}) = E[e^{tx}] = \sum_{x=0}^{\infty} e^{tx}\left[\frac{e^{-\lambda}\lambda^t}{x!}\right] = e^{-\lambda}\sum_{x=0}^{\infty}\left[\frac{e^t\lambda}{x!}\right]$$

$$= e^{-\lambda}\left[1 + \frac{\lambda e^t}{1!} + \frac{(\lambda e^t)^2}{2!} + \cdots \ldots.\right] = e^{-\lambda}e^{\lambda et} = e^{\lambda}(e^t - 1)$$

6.22 Additive Property or Reproductive Property of Independent Poisson Variate

Statement: Sum of independent Poisson variates is also a Poisson variate. More elaborately, if X_i (i = 1, 2, 3,…., n) are independent Poisson variates with parameters A_i, i = 1, 2,…., n respectively, then $\sum_{i=1}^{n} X_i$ is a Poisson variate with parameter$\sum_{i=1}^{n} \lambda_i$

Proof: We have

$$M_{\mathrm{xt}}(\mathrm{t}) = e^{\lambda}(e^t - 1), \qquad i = 1, 2, \ldots., n$$

$$\therefore M_{\mathrm{x_1+x_2+\cdots\ldots x_n}}(\mathrm{t}) = M_{\mathrm{x_1}}(t)M_{\mathrm{x_2}}(t)\ldots\ldots M_{\mathrm{x_n}}(t)$$

$$(Since \quad X_{\mathrm{i}} \qquad i = 1, 2, \ldots.n \; are \; independent)$$

$$= e^{\lambda 1}(e^t - 1).\, e^{\lambda 2}(e^t - 1)\ldots\ldots.\, e^{\lambda n}(e^t - 1)$$

$$= e^{\lambda 1 + \lambda 2 + \cdots \lambda n}(e^t - 1) = e^{\sum_{i=1}^{n}\lambda 1}(e^t - 1)$$

Which is the m.g.f. of a Poisson variate with parameters $\sum_{i=1}^{n} \lambda_i$……Hence, by uniqueness theorem of m.g.f., $\sum_{i=1}^{n} X_i$ is also a Poisson variate with parameters$\sum_{i=1}^{n} \lambda_i$

6.23 Recurrence Relation for the Moments of the Poisson Distribution

By definition

$$\mu_r = E[X - E(X)]^r = E(X-\lambda)^r = \sum_{x=0}^{\infty} (X-\lambda)^r \frac{e^{-\lambda}\lambda^x}{x!}$$

Differentiating with respect to λ we get

$$\frac{d\mu_r}{d\lambda} = \sum_{x=0}^{\infty} \Gamma(x-\lambda)^{r-1}(-1)\frac{e^{-\lambda}\lambda^x}{x!} + \sum_{x=0}^{\infty} \frac{(x-\lambda)^r}{X!}[x\lambda^{x-1}e^{-\lambda} - \lambda^x e^{-\lambda}]$$

$$= -\Gamma\sum_{x=0}^{\infty}(x-\lambda)^{r-1}\frac{e^{-\lambda}\lambda^x}{x!} + \sum_{x=0}^{\infty}\frac{(x-\lambda)^r}{x!}[\lambda^{x-1}e^{-\lambda}(x-\lambda)]$$

$$= -\Gamma\sum_{x=0}^{\infty}(x-\lambda)^{r-1}\frac{e^{-\lambda}\lambda^x}{x!} = -\Gamma\mu_{x-1} + \frac{1}{\lambda}\mu_{x-1}$$

$$or\ \mu_{x-1} = r\lambda\mu_{x-1} + \lambda\frac{d\mu_r}{d\lambda}$$

Putting r = 1, 2 and 3 successively, we get

$$\mu_2 = \lambda\mu_0 + \lambda\frac{d\mu_1}{d\lambda} = \lambda, \quad (since\ \mu_0 = 1,\ \mu_1 = 0)$$

$$\mu_3 = 2\lambda\mu_1 + \lambda\frac{d\mu_2}{d\lambda} = \lambda$$

$$\mu_4 = 3\lambda\mu_2 + \lambda\frac{d\mu_3}{d\lambda} = 3\lambda^2 + \lambda$$

Example: If X is a Poisson variate such that; P[X = 2] = 9P[X = 4] + 90P[X = 6]

Find (i) λ, the mean of X. (ii) β_1, the coefficient of skewness

Solution: If X is a Poisson variate with parameter λ, then

$$P[X = x] = \frac{e^{-\lambda}\lambda^x}{x!}, x = 0, 1, 2, \dots\dots\infty, \qquad \lambda > 0.$$

$$\therefore P[X = 2] = \frac{e^{-\lambda}\lambda^2}{2!}$$

$$P[X = 4] = \frac{e^{-\lambda}\lambda^4}{4!}$$

$$P[X = 6] = \frac{e^{-\lambda}\lambda^6}{6!}$$

$\therefore$ *From given condition we have*

$$\frac{e^{-\lambda}\lambda^2}{2!} = 9.\frac{e^{-\lambda}\lambda^4}{4!} + 90\frac{e^{-\lambda}\lambda^6}{6!}$$

or $$\frac{e^{-\lambda}\lambda^2}{2!} = \frac{e^{-\lambda}\lambda^4}{4!}\left[9 + 90\frac{\lambda^2}{30}\right]$$

or $$1 = \frac{\lambda^2}{12}[9 + 3\lambda^2]$$

or $$1 = \frac{\lambda^2}{4}[3 + \lambda^2]$$

or $$\lambda^4 + 3\lambda^2 - 4 = 0$$

$\therefore$ $$\lambda^2 = \frac{-3 \pm \sqrt{9+16}}{2} = \frac{-3+5}{2}$$

Since λ > 0, we get

$$\lambda^2 = 1 \qquad \Rightarrow \lambda = 1$$

Hence mean = λ = 1, μ_2 = variance = λ = 1 and β_1 = coefficient of skewness = 1/λ = 1.

6.24 Recurrence Formula for the Probability of Poisson Distribution

(Fitting of Poisson Distribution)

For a Poisson distribution with parameter λ we have

$$p(x) = \frac{e^{-\lambda}\lambda^x}{x!}, x = 0,1,2,\ldots\ldots\infty,$$

and $$p(x+1) = \frac{e^{-\lambda}\lambda^{x+1}}{(x+1)!}, x = 0,1,2,\ldots\infty$$

$$\therefore \frac{p(x+1)}{p(x)} = \frac{\lambda}{x+1}$$

$$\Rightarrow p(x+1) = \frac{\lambda}{x+1}.p(x),$$

This is required recurrence formula.

Here first of all we calculate p(0) which is given by, P(0) = $e^{-\lambda}$ where 'A' is estimated from given data. The other probabilities. viz. p(1). p(2)……. can now be easily obtained as below:

$$p(1) = [p(x+1)]_{x=0} = \left[\frac{\lambda}{x+1}\right]_{x=0} p(0),$$
$$p(2) = [p(x+1)]_{x=2} = \left[\frac{\lambda}{x+1}\right]_{x=2} p(2),$$
$$p(3) = [p(x+1)]_{x=3} = \left[\frac{\lambda}{x+1}\right]_{x=3} p(3),$$

And so on

6.25 Self Assessment Exercises

1. If 1% of bolts produced by a certain machine are defective, find the probability that in a random sample of 200 bolts, all are good. [Ans: e^{-2}]
2. The probability that a certain component survives a given stock is 3/4. Find the probability that 2 of the next 4 components tested survive. [Ans.: 27/128]
3. Between 2 pm to 4 p.m., the average number of phone calls received at a telephone exchange per minute is 3. Calculate the probability that during one minute, chosen at random, there will be no incoming phone call. [Ans.: 0.0498]

6.26 Summary

This unit discusses d.f., p.m.f., Bernoulli, Binomial and Poisson distribution with their importance and applications.

6.27 Further Readings

1. Cramer H, Mathematical Methods of Statistics, Princeton University Press, 1946 and Asia Publishing House, 1962.
2. Hogg R.V. and Craig A.T., Introduction to Mathematical Statistics, Macmillan, 1978.
3. Prazen E., Modern Probability Theory and its Applications, John Wiley, 1960 and Wiley Eastern 1972.
4. Rao C.R., Linear Statistical Inference and Its Applications, John Wiley, 1960 and Wiley Eastern 1974.
5. Rohtagi V.K. (1984), An Introduction to Probability Theory and Mathematical Statistics, John Wiley, 1976 and Wiley Eastern 1985.
6. Vikas S.S., Mathematical Statistics, John Wiley, 1962 and Toppan.

Unit-7: Discrete Distribution

Structure

7.1 Introduction

Suppose we toss a coin until head turn up and denote by X and number of tosses required for this purpose. Then X = 1, 2,….., and in general we can specify an upper bound k such that $P[x \sim k] = 1$. An obvious extension for the above example is the following. Suppose we decided to toss the coin until a specified number, r (say) of heads turns up. In this situation the number x of tosses required is r, r + 1, r + 2.

Although both these examples seem mainly to be of theoretical interest, they are useful in many statistical and probabilistic problems of an advanced nature. Since they are concerned with waiting times (number of trials) required for the first or rth occurrence of specific event the associated distribution are called waiting time distribution. We shall discuss tow simple waiting time distribution in sections 7.2 to 7.8: the geometric and negative binomial distribution. In section 7.9 to 7.11 we have discussed the Hyper Geometric distribution.

7.2 Objectives

After reading this unit you should be able to:

- Define the geometric, negative binomial and hyper geometric distribution
- Calculate the mean and variance of these distributions.

- Obtain the Recurrence relations for negative binomial distribution and hyper geometric distribution.

7.3 The Geometric Distribution

In this section we discuss the geometric distribution. Let us see first how such a distribution arises.

Let p denote the probability of a success in a Bernoulli trial 0 < p < 1. Consider independent repetitions of such a trial. Let X denote the number of trials required for first success. Then r. v. X takes the values 1, 2, 3…… and by definition

P [X = 1] = p

In order to obtain P [X = x] for x ≥2, observe that the event [X = x] occurs if the first (x-1) trials result in a failure and xth trial is success. Thus the probability of first (x-1) failures followed by a success is

$$p(x) = \begin{cases} \binom{-r}{x} Q^{-r} \left(-\frac{P}{Q}\right)^{r}, x = 0, 1, 2, \dots \dots \\ 0 \qquad\qquad otherwiese \end{cases} \qquad (2.14)$$

The r. v. X here is said to have a geometric distribution.

Definition: A random variable X is said to have the geometric distribution with parameter p, 0 < p < 1, if its p.m.f. is given by

$$P[X = X] = \begin{cases} q^{x-1}p, & x = 0, 1, 2, 3 \dots \dots \\ 0 & other\ wise \end{cases} \qquad \dots\dots\dots\dots\dots\dots(2.1)$$

Remark: 1. Since the various probabilities for x = 1,2,…….are the various terms of geometric progression, hence the name geometric distribution.

2. Clearly assignment of probabilities is permissible since

$$P[X = X] = \begin{cases} q^{x-1}p, & x = 0,1,2,3 \dots \dots \\ 0 & other\ wiese \end{cases}$$

7.4.1 Mean and Variance of the Geometric Distribution

By definition,

$$E(X) = \sum_{x=1}^{\infty} x q^{x-1} p = p \sum_{x=1}^{\infty} x q^{x-1} \qquad (2.3.1)$$

$$Let\ S_1 = \sum_{x=1}^{\infty} xq^{x-1} = (1 + 2q + 3q^2 + 4q^3 + \cdots \dots)$$

$$= (1 + 2q + q^2 + q^3 + \cdots) + q(l + q + q^2 + q^3 + \cdots) + q^2\,(1 + q + q^2 + \cdots) + \cdots.$$

$$= (1 + 2q + q^2 + q^3 + \cdots)(l + q + q^2 + q^3 + \cdots)$$

$$= \frac{1}{1-q}.\frac{1}{1-q} = \frac{1}{p^2} \qquad (2.3.2)$$

Putting (2.3.2) in (2.3.1) we get

$$\therefore E(X) = p.\frac{1}{p^2} = \frac{1}{p} \qquad (2.4)$$

The variance V(X), will be obtained as

$$V(X) = E(X)^2 - [E(X)]^2$$

$$Now, E(X)^2 = E[X(X-1)] + E(X)$$

$$E[X(X-1)] = \sum_{x=1}^{\infty} x(x-1)q^{x-1} = p\sum_{x=2}^{\infty} x(x-1)q^{x-1}$$

Let

$$S_2 = \sum_{x=2}^{\infty} x(x-1)q^{x-1}$$

Put r = x-1
We have

$$S_2 = \sum_{x=2}^{\infty} r(r+1)q^r$$

Hence

$$(1-q)S_2 = (1-q)\sum_{x=2}^{\infty} r(r+1)q^r = \sum_{x=2}^{\infty} r(r+1)q^r - \sum_{x=2}^{\infty} r(r+1)q^{r+1}$$

$$= (2q + 2.3q^2 + 3.4q^3 + 4.5q^4 \dots) - (1.2q^2 + 2.3q^3 + 3.4q^4 + \cdots)$$

$$= (2q + 4q^2 + 6q^3 + 8q^4 + \cdots) = 2(q + 2q^2 + 3q^3 + 4q^4 + \cdots)$$
$$= 2q(1 + 2q + 3q^2 + 4q^3 + \cdots)$$

$$= 2q.\frac{1}{p^2}\ (since\ we\ have\ from\ (9.11),$$

$$\sum_{r=1}^{\infty} rq^{r-1} = \frac{1}{p^2} = \frac{2q}{p^2}$$

$$\therefore S_2 = \frac{2q}{p^3}$$

Thus

$$E[X(X-1)] = p.\frac{2q}{p^3} = \frac{2q}{p^2} \qquad (2.5)$$

And therefore,

$$V(X) = \frac{2q}{p^2} + \frac{1}{p} - \frac{1}{p^2}$$

$$= \frac{2q+p-1}{p^2} = \frac{2q-q}{p^2} = \frac{q}{p^2} \qquad (2.6)$$

The mean = 1/p and V(X) = q/ p^2

7.3.2 Moment Generating Function *(m.g.f.)* of the Geometric Distribution

By definition

$$M_{\text{x}}(\text{t}) = E[e^{tx}] = \sum_{x=0}^{\infty} e^{tx} q^{x-1} = pe^t \sum_{x=0}^{\infty} [e^t q]^{x-1}$$

$$= pe^t \sum_{x=0}^{\infty} [e^t q]^r, \quad where\ r = x - 1$$

$$= \frac{pe'}{1-e'q} = \frac{pe'}{1-qe'} \qquad (2.7)$$

Which is valid only if $qe^t < 1$.

7.4 Negative Binomial Distribution *(N.B.D)*

In this section we discuss the properties of negative binomial distribution which is a generalization of the geometric distribution. We know that geometric distribution gives the distribution of the number of trials required to obtain first success in independent repetitions of a Bernoulli trials. Now suppose we want to find the distribution of the number of trials required to obtain the rth success in independent repetition of a Bernoulli trial with probability p of success at every trial.

Suppose we have a success of n Bernoulli trials. We assume that (a) trial are independent (b) the probability of success p in a trial remains constant from trial to trial.

Let f (x, r, p)~ denote the probability that there are x failure preceding the rth success in (x + r) trials. Now last trial must be success whose probability is p. In the remaining (x + r - 1) trials we must have (r - 1) successes whose probability is given by

$$\binom{x+r-1}{r-1} p^{r-1} q^{(x+r-1)-(r-1)} = \binom{x+r-1}{r-1} p^{r-1} q^{x} \qquad (2.8)$$

∴ (x, r, p) is given by product of probability of (r - 1) success in (x + r - 1) trials and probability of success in last (x + r)th trial, i.e.,

$$f(x,r,p) = \binom{x+r-1}{r-1} p^{r-1} q^{x}.p = \binom{x+t-1}{r-1} p^{r} q^{x} \qquad (2.9)$$

Definition: A random variable X is said to follow a negative binomial distribution if its probability mass function is given by:

$$p(x) = P[X=x] = \begin{cases} \left\{\binom{x+r-1}{r-1} p^{r} q^{x}, & x = 0,1,2,\dots\dots \\ 0, & otherwise \end{cases} \qquad (2.10)$$

Also

$$\binom{x+r-1}{r-1} = \binom{x+r-1}{r-1} \left[Since \binom{n}{r} = \binom{n}{n=r}\right]$$

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

$$= \frac{(-1)^{x}(-r)(r-1)(-r-x+2)+\ (-r-x+1)}{x!}$$

$$= (-1)^{x}\binom{-r}{x} \qquad (2.11)$$

Therefore, an alternative form of the p.m.f of N.B.D. is:

$$\therefore p(x) = \begin{cases} \binom{-r}{x} p^{x}(-q)^{x} \\ 0 \ \ otherewise \end{cases} \qquad (2.12)$$

Which is the (x + 1)th term in the expansion of p^{r} $(1 - q)^{-r}$, binomial expansion with a negative index. Hence this distribution is known as ***Negative Binomial Distribution***. It is also known as ***Pascal's distribution***.

Note: (i) The sum of the negative binomial probabilities p(x) is one i.e.,

$$\sum_{x=0}^{\infty} p(x) = p^r \sum_{x=0}^{\infty} \binom{-r}{x} (-q)^x = p^r (1-q)^{-r} = 1 \; \{Since\ (1-q) = p\}$$

(ii) $\quad If\ p = \frac{1}{Q}\ and\ q = \frac{P}{Q}\ so\ that\ Q - P = 1\ (since\ p + q = 1) \qquad (2.13)$

Then the third form of the p.m.f of NBD is:

$$p(x) = \begin{cases} \binom{-r}{x} Q^{-r} \left(-\frac{P}{Q}\right)^r, & x = 0,1,2,\ldots. \\ 0 \;\; otherewise & \end{cases} \qquad (2.14)$$

This is the general term in negative binomial expansion $(Q-P)^{-r}$

(iii) If we take r = 1 in equ. (2.2) we have $p(x) = q^x$, $\quad x = 0, 1, 2, \ldots\ldots$

Which is the probability function of geometric distribution. Hence negative binomial distribution may be regarded as generalization of geometric distribution.

7.4.1 Moment Generating Function *(m.g.f.)* of Negative Binomial Distribution

$$M_{\text{x}}(\text{t}) = E[e^{tx}] = \sum_{x=0}^{\infty} e^{tx} p(x) = \sum_{x=0}^{\infty} \binom{-r}{x} Q^{-r} \left(-\frac{P}{Q}\right)^x = (Q - Pe^t)^{-r}$$

$$\mu_1 = \left[\frac{d}{dt} M_{\text{x}}(\text{t})\right] = [-r(-Pe^t)(Q - Pe^t)^{-r-1}]_{i=0} = [rPe^t((Q - Pe^t)^{-r-1})]_{i=0}$$

$$= rP.(Q-P)^{-r-1} = -rP \;\; (since\ Q - P = 1) \qquad (2.17)$$

$$\therefore \; \mu_2 = \mu_{2-}(\mu_1)^2 = rP + r(r+1)P^2 - r^2P^2$$

$$= rP + r^2P^2 + rP^2 - r^2P^2 = rP(1+P) = rPQ \qquad (2.18)$$

As $Q > 1$, $rP < rPQ$, i.e. Mean < Variance which is a distinguishing characteristic of this distribution.

7.4.2 Poisson Distribution as Limiting Case of the Negative Binomial Distribution

Negative binomial distribution tends to Poisson distribution as

$$P \to 0\,, r \to \infty \qquad such\ that\ \ rP = \lambda\ (finite).$$

Now we have p.m.f. of negative distribution.

$$p(x) = \binom{x+r-1}{r-1} p^r q^x$$

$$\lim p(x) = \lim_{r\to\infty} \binom{x+r-1}{r-1} Q^{-r}(P/Q)^x$$

$$= \frac{1}{x!} \lim_{r\to\infty} \frac{(x+r-1)(x+r-2)\ldots\ldots(r-1)r}{x!} (1+P)^{-r} \left(\frac{P}{1+P}\right)^x$$

$$= \frac{1}{x!} \lim_{r\to\infty} (1+P)^{-r} \left(\frac{rP}{1+P}\right)^x$$

$$= \frac{1}{x!} \lim_{r\to\infty} (1+\frac{\lambda}{r})^{-r} \lim_{r\to\infty} (1+\frac{\lambda}{r})^x \qquad \text{(Since rp = λ)}$$

$$= \frac{\lambda^r}{x!} \lim_{r\to\infty} (1+\frac{\lambda}{r})^{-r} \lim_{r\to\infty} (1+\frac{\lambda}{r})^x = \frac{\lambda^r}{x!} e^{-\lambda}$$

Which is the p.m.f. of Poisson distribution with parameter λ.

7.4.3 Recurrence formula for Negative Binomial Distribution

$$f(x,r,p) = \binom{x+r-1}{r-1} p^r q^x$$

$$f(x+1,r,p) = \binom{x+r}{r-1} p^r q^{x+1}$$

$$\therefore \frac{f(x+1,r,p)}{f(x,r,p)} = \frac{(x+1)!\,x!}{(x+1)!\,(x+r-1)!}.\; q = \frac{(x+r)}{x+1} q$$

$$\Rightarrow f(x+1,r,p) = \frac{x+r}{x+1}.qf\ (x,rp) \qquad (2.19)$$

This is the recurrence relation and is useful for fitting of the negative binomial distribution.

7.5 The Hyper geometric Distribution

Let there be N items of which M are defective and N-M are good. We select n items without replacement and want to find the probability that j of them are defective. When j of n are defective then remaining n-j must be good items. But since M defective and N-M good items. We must have j ≤ m and n-j ≤ M-N. We can choose n items of out of N in $\binom{N}{n}$ ways. The defective can be chosen out of M defectives in $\binom{M}{j}$ways and (n-j) good items can be chosen out of (N-M) good items in $\binom{N-M}{n-j}$ ways. Hence , if X denotes the number of defective items selected.

$$P[X=j]=\frac{\binom{M}{j}\binom{N-M}{n-j}}{\binom{N}{j}} \qquad (2.20)$$

for j = 0, 1, 2,………n, j ≤ N - M. Here, n, N M are positive integers, n ≤ N, M ≤ N.

Definition: A random variable X has the hyper-geometric distribution with parameters (n, N, M) if its p.m.f. is

$$P[X=j]=h(j,n,N,M)=\frac{\binom{M}{j}\binom{N-M}{n-j}}{\binom{N}{j}} \qquad \ldots\ldots\ldots\ldots\ldots.(2.21)$$

for j = 0, 1, 2,………n, j ≤ M, (n - j) ≤ N - M. Here, n, N, M are positive integers, n ≤ N, M ≤ N.

7.5.1 Mean and Variance of Hyper geometric Distribution

$$Mean=E(X)=\sum_{j=0}^{n} jh(j,n,N,M)$$

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

$$=\frac{M}{\binom{N}{n}}\sum_{j=1}^{n-1}\frac{(M-1)!}{(j-1)!\,[(M-1)-(j-1)]!}\binom{(N-M)}{(n-1)-(j-1)}$$

$$=\frac{M}{\binom{N}{n}}\sum_{j=1}^{n-1}\binom{M-1}{r}\binom{(N-1)-(M-1)}{n-1-r}, \qquad r=j-1$$

Now,

$$\sum_{j=0}^{r}\binom{m}{j}\binom{n}{r-j}=\binom{m+n}{r}$$

$$\begin{bmatrix} Expanded\ both\ sides\ of\ (1+x)^{m+n}=(1+x)^m\,(1+x)^n \\ and\ comapre\ the\ coefficient\ of\ x^r\ to\ get\ this\ results \end{bmatrix}$$

Using this result we get

$$E(X)=\frac{M\binom{N-1}{n-1}}{\binom{N}{n}}=\frac{nM}{N} \qquad \ldots\ldots\ldots\ldots\ldots \quad (2.22)$$

We have

$$V(X) = E\,(X^2) - \,[E(X)]^2$$

Now

$$E(X^2) = E[X(X-1)] + E(X)$$

Also

$$E[X(X-1)] = \sum_{j=1}^{n} j(j-1)\left[\frac{\binom{M}{j}\binom{N-M}{n-j}}{\binom{N}{n}}\right]$$

$$= \frac{M(M-1)}{\binom{M}{n}} \sum_{j=2}^{n} \left[\binom{M-2}{j-2}\binom{N-M}{n-j}\right]$$

$$= \frac{M(M-1)}{\binom{N}{n}}\binom{N-2}{n-2} = \frac{M(M-1)n(n-1)}{N(N-1)}$$

$$\therefore E(X^2) = \frac{M(M-1)n(n-1)}{N(N-1)} + \frac{nM}{N} \qquad \ldots\ldots\ldots\ldots\ldots (2.23)$$

Hence

$$V(X) = \frac{M(M-1)n(n-1)}{N(N-1)} + \frac{nM}{N} - \left(\frac{nM}{N}\right)^2$$

$$= \frac{nM(N-M)(N-n)}{N^2(N-1)} \ (on\ simplification) \quad \ldots\ldots\ldots\ldots \quad (2.24)$$

7.5.2 Recurrence Relation for the Hyper Geometric Distribution

We have

$$j(j,n,N,M) = \frac{\binom{M}{j}\binom{N-M}{n-j}}{\binom{N}{n}} \qquad \ldots\ldots\ldots\ldots\ldots \quad (2.25)$$

$$j(j+1,n,N,M) = \frac{\binom{M}{j+1}\binom{N-M}{n-j-1}}{\binom{N}{n}}$$

$$\therefore \frac{j(j+1,n,N,M)}{h(j,n,N,M)} = \frac{(n-j)(M-j)}{(j+1)(N-M-nj+1)} \ (on\ simplicication)$$

$$or h(j,n,N,M) = \frac{(n-j)(M-j)}{(j+1)(N-M-nj+1)}.h(j,n,N,M) \quad \ldots\ldots\ldots \quad (2.26)$$

7.6 Self Assessment Exercises

1. Cards are drawn at random with replacement from a well shuffled pack of 52 playing cards. Find the probability that the first ace will appear before the fifth selection. [Ans.: 0.269]
2. The probability of having a male child or female child is both 0.50. Can you find the probability that a family's fourth child is the second daughter. [Ans.: 3/16].
3. Find the probability that a person tossing an unbiased coin gets fourth head on seventh toss. [Ans. $6_{e_3}\left(\frac{1}{2}\right)^7$]
4. A quality control engineer inspects two randomly selected units from a lot of 20 units. If the both units are in working conditions, the lot is accepted; otherwise all the remaining units are inspected. Let us find the probability that a lot of 20 units containing 8 detective units are accepted without further inspection. [Ans. 0.347.]
5. For the geometric distribution with p.m.f. f(xt2- X, x= 1,2,3,… show] that chebyshev's inequality gives $P[|x-2| \leq 2] > \frac{1}{2}$ while actual probability is 15/16.

7.7 Summary

This unit provides a brief idea about Geometric, Negative, Binomial and Hyper Geometric distribution with their applications and importance.

7.8 Further Readings

1. Cramer H, Mathematical Methods of statistics, Princeton University Press, 1946 and Asia Publishing House, 1962.
2. Hogg R.V. and Craig A.T., Introduction to Mathematical Statistics, Macmillan, 1978.
3. Prazen E. Modern Probability Theory and its applications, John Wiley, 1960 and Wiley Eastern 1972.
4. Rao C.R., Linear Statistical Inference and Its Applications John Wiley, 1960 and Wiley Eastern 1974.
5. Rohtagi V.K. (1984), An Introduction to Probability theory and Mathematical Statistics John Wiley, 1976 and Wiley Eastern 1985.
6. Vikas S.S., Mathematical Statistics John Wiley, 1962 Toppan.

Unit - 8: Normal Distribution

Structure

8.1 Introduction

In order to fully understand the interpretation of the standard deviation and its relationship to the mean, it is important to learn about the shape of a distribution. A histogram, or a polygon, is a helpful tool in evaluating the shape- the pattern in which scores vary-of a distribution. Only after knowing the shape does one know the precise location of scores in relation to the central tendency and standard deviation. The distribution of scores may be shaped in many different ways. Of all these shapes, the most interesting one is called the normal curve.

8.2 Objectives

After studying this unit you will be able to understand:

- The meaning of normal distribution and its properties.
- The application of normal curve to observed distribution.
- The importance of standard scores and normal area distribution.
- The advantages and characteristics of Z scores.

8.3 Normal Probability Distribution

It has been observed that most business and economic variables result in continuous data whose behavior is often best described by a bell-shaped continuous curve. Since most populations on these variables normally yield a bell-shaped curve, such a curve has come to be universally known as a normal curve. Accordingly, the probability distribution described by normal curve is called the normal (probability) distribution.

8.3.1 The Normal Distribution and its Parameters

The normal distribution is defined by the probability density function.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}; for -\infty \leq x \leq \infty \qquad \ldots\ldots\ldots\ldots\ldots (3.1)$$

Where r, σ and > are two parameters of the distribution.

A continuous r.v. X having above pdf is called normal random variable. We write it as X~N (μ, σ^2) μ and σ^2 are the parameters of the distribution. We shall see μ that is mean ^{2}and σ^2 is variance of the distribution.

Here $\int_{-\infty}^{\infty} f(x)dx = 1.$ $\quad [\,on\ puting\ z = \frac{x-\mu}{\sigma}\ and\ \sigma\ dz = dx]$

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\frac{1}{2}z^2}\ dz = 1 \qquad \ldots\ldots\ldots\ldots\ldots.. (3.2)$$

$$\left[\int_{-\infty}^{\infty} \frac{1}{2} e^{-2^2/2}\ d_2\ = 1\right]$$

The curve

$$Y = f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\left(\frac{x-\mu}{\sigma}\right)^2}$$

is called normal probability curve.

This curve is bell-shaped and symmetric about mean μ., As the curve is symmetrical about the ordinal at x = μ, The curve is concave down wards in the centre μ, but after x = $\mu \pm \sigma$ is becomes concave upwards. The two points of inflexion are given by x = $\mu \pm \sigma$,

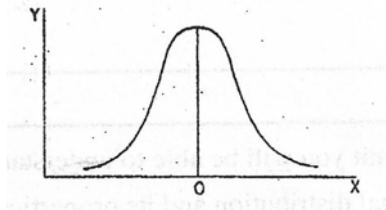


Figure: 3.1 Normal Distribution

It shows the deviations of the values of normal variables X from its mean μ. The larger these deviations, the higher are the value of standard deviations σ (or variance σ^2) which is the denominator in the exponent.

The shape of the curve shows that the observations occur most frequently in the neighborhood of the mean and their frequency decreases as they move away from the mean. From (3.1) it is seen that the distribution is symmetrical about the point $x = \mu. f(\mu + a) = f(\mu - a) = \frac{1}{\sqrt{2\pi\sigma^2}} exp(-\sigma^2 a^2)$. whatever μ may be. Hence μ is mean as well as median of the distribution. Againμ is also mode of the distribution, since $f(\mu) = 0$ and f^x μko. Also $\exp\left(-\frac{a^2}{\sigma^2}\right)$ decreases monotonically a^2 as increase from zero, i.e., as a deviates from zero in either direction.

Thus the mean, median and mode of the distribution coincide at μ, Thus the mode μ of the distribution also coincides with mean and median μ.

Approximately 68% of the area under the cure lien in the region $[\mu - \sigma,\ \mu + \sigma]$, approximately 95% in $[\mu - 2\sigma, \mu + 2\sigma]$ and almost all in the region $[\mu - 3\sigma, \mu + 3\sigma]$. In fact approximately 0.27% area lies outside region $[\mu - 3\sigma, \mu + 3\sigma]$. This fact is stated by the following relations-

Mean: By definition,

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx = \int_{-\infty}^{\infty} x \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} dx \qquad .... (3.3)$$

$$= \mu \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz + \int_{-\infty}^{\infty} Ze^{e^{-\frac{1}{2}z^2}} dz \qquad (on\, putting\ \frac{x-\mu}{\sigma} = z)$$

$$= \mu \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz + 0 \qquad \left[\because \int_{-\infty}^{\infty} Ze^{e^{-\frac{1}{2}z^2}} dz = 0\right] ......(3.3.1)$$

$$= \mu \qquad (3.4)$$

Let X be a continuous random variable with a normal distribution and with pdf given by (3.1) then a continuous random variable with

$$E(X-\mu) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x-\mu)\, e^{-(x-\mu)^2/2\sigma^2} dx$$

$$= \int_{-\infty}^{\infty} \frac{\sigma^2}{\sqrt{2\pi}} z\sigma e^{-\frac{1}{2}z^2} \sigma dz \qquad ...... (3.6)$$

$$\left(Where\ z = \frac{x-\mu}{\sigma} and\ dz = \frac{dx}{\sigma}\right)$$

$$= \frac{\sigma}{0\sqrt{2\pi}} \int_{-\infty}^{\infty} ze^{-\frac{1}{2}z^2}$$

$$= 0\ [the\ integrand\ is\ odd]$$

$$\therefore E(X) = \mu \qquad \dots\dots\dots (3.7)$$

Thus the parameter μ is the mean of the normal distribution

Variance: By definition

$$Var(X) = E(E-\mu)^2 = \frac{1}{2\sqrt{2\pi}} \int_{-\infty}^{\infty} (x-\mu)^2 e^{-(x-\mu)^2/2\sigma^2} dx \qquad \dots\dots (3.8)$$

$$= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^2 e^{-\frac{1}{2}z^2} \qquad (\text{where } z = \frac{x-\mu}{\sigma})$$

$$= \frac{2\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^2 e^{-\frac{1}{2}z^2}\ dz = \sigma^2 \qquad \dots\dots.. (3.9)$$

Thus the parameter σ^2 is the variance of the normal distribution. Normal Distribution with mean μ and variance σ^2 is also denoted by N (μ, σ^2) or N (μ, σ) Mean, Median and Mode of this distribution, coincide at theμ.

8.3.2 Standard Normal Distribution

If X~N (μ, σ^2) then its pdf is given by equ. (3.1) using transformed

$$z = \frac{x-\mu}{\sigma}$$

We obtain the pfd of ≥ as

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \qquad for\ -\infty < z < \infty \qquad (3.1a)$$

Obviously E(z) = 0 and var (Z)=1

Hence, z is standard normal variable and the pdf given by equ. (3.1a) is called standard normal distribution.

Quartiles: The first quartile Q_1 is given by

$$P[X \leq Q_L] = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{Q1} e^{-(x-\mu)^2/2\sigma^2} dx = \frac{1}{4}$$

Solving this equation with the help of Normal Probability integral Table' we found

$$Q_1 = \mu - 0.6745\sigma. \qquad \ldots\ldots\ldots\ldots\ldots \qquad (3.10)$$

Similarly, the third quartile Q_3 is

$$P[X \leq Q_3] = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{Q1} e^{-(x-\mu)^2/2\sigma^2} dx = \frac{3}{4} \qquad (3.11)$$

$$or \quad \frac{1}{\sigma\sqrt{2\pi}} \int_{Q_3}^{\infty} e^{-(x-\mu)^2/2\sigma^2} dx = \frac{1}{4} \qquad (3.12)$$

Solving the above equation with the help of 'Normal probability Integral Table' we found.

$$Q_3 = \mu + 0.6745. \ \ Thus \ \ \mu - Q_1 = Q_3 - \mu$$

$$\mu = 0 \qquad (3.13)$$

8.3 Properties of Normal Distribution

8.3.2.1 Moments of the Normal Distribution

Moments about the mean: In case of normal distribution

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}, -\infty < x < \infty, -\infty < \mu < \infty, \sigma > 0$$

It may be observed that

(i) The odd moments about mean are zero i.e., μ_{2n} + 1 = 0 and
(ii) The even moments about mean are given by the recurrence formula

$$\mu_{2n} = \sigma^2(2n-1)\mu_{2n-2} \qquad for\ n =$$

$$\mu_{2n} = (2n-1)(2n-3)\ldots 3.1\ \sigma^{2n} \quad \ldots\ldots..(3.14)$$

Thus all odd moments are zero, i.e.

$$\Rightarrow \mu_1 = \mu_3 = \mu_5 = \cdots \ldots = \mu_{2n+1} = 0 \quad \ldots\ldots(3.15)$$

The first four central moments are:

In particular: $\mu_1 = 0;\ \ \mu_2 = \sigma^2;\ \ \mu_3 = 0;\ \mu_4 = 3\sigma^4;\ \ldots\ldots\ldots.(3.16)$

Thus Beta and Gamma ratios of this distribution are

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = 0 \qquad [\,\mu_3 = 0]$$

$$\gamma_1 = \sqrt{\beta_1} = 0 \qquad \ldots\ldots\ldots\ldots\ldots\ldots(3.17)$$

⇒ The normal distribution is symmetrical about X = μ i.e., there is no skewness in normal distribution and

$$\beta_1 = \frac{\mu_4}{\mu_2^2} = \frac{3\sigma^4}{(\sigma^2)^2} = 3$$

$$\therefore\ \ \gamma_2 = \beta_2 - 3 = 0 \quad \ldots\ldots\ldots\ldots\ldots(3.18)$$

⇒ The normal distribution is mesokurtic i.e., Kurtosis of a normal distribution vanish; it is neither play kurtic nor leptokurtic.

You may remember that the Kurtosis or Flatness of the probability curve at the maximum ∅ is compared with the normal curve.

8.3.2.2 Moment Generating Function of Normal Distribution

(i) Moment generating function (about origin):

$$M_{\mathrm{x}}(\mathrm{t}) = E[e^{tx}] = \int_{x=0}^{\infty} e^{tx}.\frac{1}{\sigma\sqrt{2\pi}}e^{-(x-\mu)^2/2\sigma^2}\,dx$$

$$= \int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{t(\mu+z\sigma)}e^{-\frac{1}{2}z^2}dz, \left(putting\ \ z = \frac{x-\mu}{\sigma};\ \sigma dz = dx\right)\ldots\ldots.(3.19)$$

$$e^{\mu t}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{t(\mu+z\sigma)}e^{-\frac{1}{2}z^2}dz = e^{\mu t}+^{-\frac{1}{2}l^2\sigma^2}\ \ldots\ldots..(3.20)$$

(ii) Moment generating function (about mean)

$$M_{x-\mu}(t) = e^{\mu t+\frac{1}{2}l^2\sigma^2}\ (deducing\ from\ above)$$

Example 3.1: Show that any linear function of normal variate is also normally distributed.

Solution: Let Y∼ N (μ, σ^2). Then

$$M_Y(t) = e^{\mu t+\frac{1}{2}l^2\sigma^2} \dots\dots\dots\dots\dots\dots..(3.21)$$

Let X = a Y + b., (a≠0) be any linear function of Y.

Then $M_x(t) = E\left[e^{t(aY+b)}\right]$

$= e^{bt}E(e^{\sigma tY}) = e^{bt}\, M_Y(at) \qquad [from\ (3.21)]$

$= e^{bt}.e^{\mu at+\frac{1}{2}(\sigma t)^2\sigma^2} \qquad (3.22)$

$= e^{(\sigma\mu+b)t+\frac{1}{2}l^2a^2\sigma^2}$

On comparing (3.2) with (3.21), we find that (3.2) is the p.d.f. of normal variate with mean $a\mu + b\ and\ a^2\sigma^2$ variance. Hence

$$X{\sim}N(a\mu+b, a^2\sigma^2)$$

Example 2: If X and Y are two independent random variable with probability density function

$$f(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2 2} - \infty < x < \infty\ \ and\ f(x) = \frac{1}{\sqrt{8\pi}}e^{-(y-5)^2 8.} - \infty < y < \infty$$

Solution:

Comparing the pdf's giving by normal dist. N (μ, σ^2)given by Eqn. (3.1) we observe that

$$X \stackrel{\sim}{\sim} N(0,1)\ and\ Y{\sim}N\ (5,4) \dots\dots\dots\dots\dots\dots..(3.23)$$

Therefore

$$\mu_x = 0, \sigma_x^2 = 1$$

$$\mu_y = 0, \sigma_y^2 = 2$$

Respectively, find the variance of the random variable (U= 2X+Y.

∴ Var (U) = Var (2X+Y) = Var (2X) + Var (Y), [X and Y are independent]

= 4 Var (X)+ Var (Y) = $4\sigma_1^2 + \sigma_2^2 = 4 \times 1 + 2^2 = 8$ ….……..(3.24)

8.4 Normal Curve

As mentioned above one of the most important models in statistics is the normal curve. It is a symmetric, unimodal bell-shaped curve with a fixed proportion of area between any given points under the curve. Symmetric means half the area is on each side of the mean (which is in the centre) and unimodal bell shaped means there is only one hump (in the centre). The normal curve is a mathematical conception described by a precise mathematical formula and based on the theoretical by a precise mathematical formula and based on the theoretical distribution of the population scores. This abstract curve describes the frequency with which observations can be expected to occur. The normal curve is frequently used as a model in statistics not because of any inherent quality but because many real world variables such as IQ, height and weight, have distribution that real closely approximate it. If one is able to assume that a set of scores approximates a normal distribution, then properties of normal distribution can be used to facilitate interpretation of data.

Normal Curve: A curve by $Y = f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$ is known as normal probability curve $Y_x = Y_0 e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$ when $Y = Y_x = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}$ origin is taken at mean μ then we get.

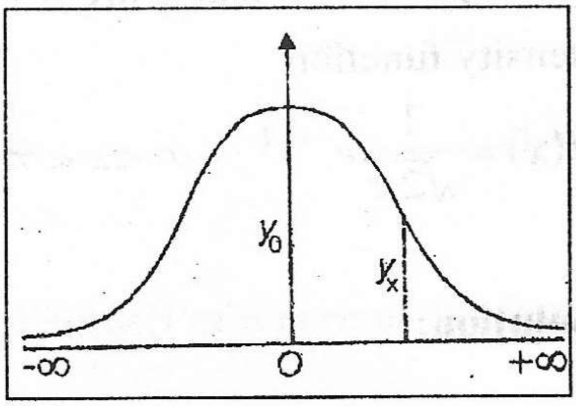


Figure 3.2: Normal Curve

Table of ordinates: The maximum ordinate or height of the normal curve is

$$Y = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2 2\sigma^2}$$

If we use transformation $Z = \frac{x-N}{\sigma}$ that is shift original the point μ and change the scale by σ then we get curve.

$$Y = \emptyset(z) = \frac{1}{\sqrt{2\pi}} e^{\frac{1}{2}z^2}$$

This is known as standard normal probability curve.

The standard deviation plays an important role in determining the shape of the curve. The mean μ lies at the centre of the normal curve and indicates the central value of the normal distribution. The standard deviation σ measures the extent of spread of X values from the central value μ. Thus, while μ fixes the position or the level of the distribution on the X- axis and σ determines the dispersion of the distribution along the X axis

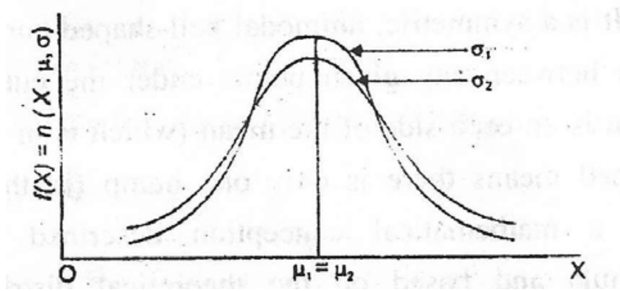


Figure 3.3: Normal Curves with different Means and Standard Deviations

In the light of the above, consider the following three situations as shown in figs. 3 and 4.

1. A change in μ standard deviation or σ remaining the same shifts the curve along the X axis without changing its spread. This is shown in fig. 3.
2. A change in σ mean μ remaining the same, changes the shape or spread of the normal curve. This is shown in fig. 4.
3. An increase in σ increases the spread of the normal curve equally on both sides of the central value it lowers is any change in μ. A decrease in σ_1 on the contrary reduces the spread of the normal curve and increase its height.

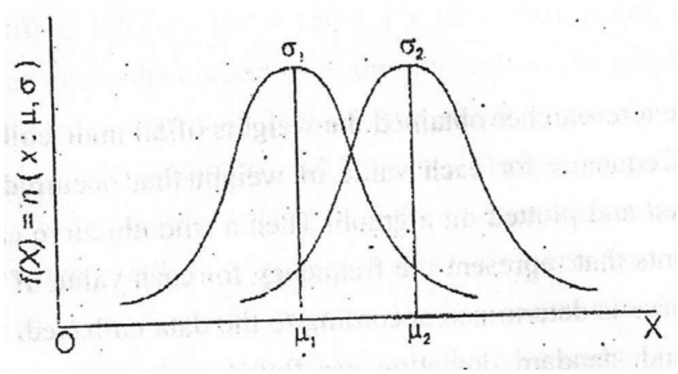


Figure 3.4: Normal Curves with different standard deviations ($\sigma_2 > \sigma_1$) and ($\mu_1 = \mu_2$)

8.4.1 Properties of a Normal Curve

(i) Normal curve is symmetric; bell shaped and unimodal.

(ii) The mean, median and mode all coincide.

(iii) Theoretically, the tails of the curve extend out to infinity in both directions without touching the horizontal axis, (x axis is asymptote to the curve) but in practice the curve can be ended at definite points. Usually at $Z = \pm 3$ infact, x- axis is asymptote to the curve.

(iv) Given the mean and standard deviation, the proportion of area (or cases between) any given points under the curve can be determined.

More specifically the normal curve is determined only by the values of μ and σ it means there are different normal curves (different in spread and positioning) for different values of these two parameters. Moreover,

(a) The normal curve is not a single curve representing only one continuous distribution. It represents a family of normal curves for different value of μ and /or σ.

(b) A change in the value of μ displays the entire curve to a different level, whereas a change in the value of σ changes the spread and determines its height.

(c) The mode of the radical distribution occurs at a point on the X-axis where the curve reaches the maximum height. Since most observation tend to cluster around the mean value, the point of mode coincides with the point of mean. That is mean and mode are equal at the point where the curve attains the maximum height.

Example 3.3: Suppose a researcher obtained the weights of all male college students in India. The frequency for each value of weight that occurred in the data could be counted and plotted on a graph. Then a smooth curve can be drawn through the points that represent the frequency for each value of weight. The shape of this curve is determined according to the data collected. Suppose the mean weight and standard deviations are found to be 150 pounds and 10 pounds respectively. Past research shows that for variable such as weight the result will be a normally shaped curve such as the one shown below.

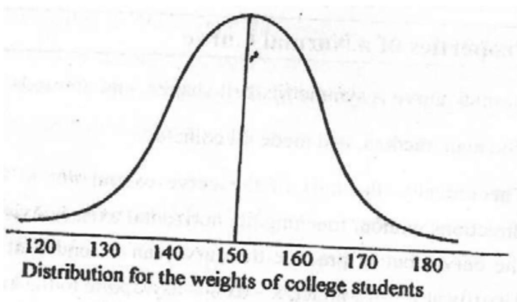


Figure 3.5

Because the above shown curve was generated by fitting it to the data, it covers all the observations. Since all the observations are given equal amounts of space on the graph, there exists a correspondence between various areas under the curve and the proportion of observations in these areas. For example, the mean weight of 150 pounds divides the area of the curve in half: 50% of the total area under the curve, whatever that might be, lies on either side of the mean. If any side of the distribution is taken the proportion of the area under the curve in that slice to the total area under the curve will be equal to the proportion of the number of observations in the slice to the total number of observations in the distribution. This means the proportion of observations that fall into any interval in the distribution can be determined by calculating the area of that interval under the curve by using some properties of the standard deviation.

It so happens that about 68.25% of the observations in a normal distribution fall within 1 standard deviation of the mean (above and below). Suppose that the distribution in the above figure was based on 1 million observations. This would mean that 682,600 college males (68.26% of 1 million) had weights fall into the interval from 140 pounds (~-10 = 150 -1×10 = 140) to 160 (~+10" = 150 + 1 × 10 = 160). It also means that 68.26% of the area under the curve lies in the interval on the graph from 140 pounds to 160 pounds. Since 68.26% of the observation lie inside this interval the remaining 31.74% (100-68.26= 31.74) must lie outside it. And because the curve is symmetric, half of 31.74% of observation lie on either side of the mean: about 15.87% of college males weight less than 140 pounds and 15.87% weigh more than 160 pounds. Similar statements hold for the corresponding area under the curve: 15.87% of the area of the area under the curve lies to the left of a line drawn at 140 pounds and 15.87% of the area under the curve lies to the right of a line drawn at 160 pounds. Similarly, the percentage of scores between any given points can be determined. For example about 95% of the observation, under a normal curve, falls within 2 standard deviations and about 99% of the observations fall within 3 standard deviations of the mean.

Example:

Suppose a population consists of 10,000 IQ scores. The distribution of IQ scores in normally shaped with a mean of 100 and SD of 16. If a person has an IQ score of 132 would he/she be happy or sad?

Solution :

Since the scores are normally distributed, 34.13% of the scores are contained, between the scores of 84 and 116 where 84 is 1 SD below the mean and 116 is 1 SD above the mean; 13.59% of the scores fall between 116 and 132; 2.15% between 132 and 148; and 0.13% are greater than 148. Similarly, 34.13% of the scores fall between 84 and 100; 13.59% between 68 and 84; 2.15% between 52 and 68; and 0.13% are lower than 52. To calculate the number of scores in each area, all that is needed is to multiply the relevant by the total number of scores. Thus there are (.3413).(10000) = 3412 scores between 100 and 116; (.1359) (10000) = 1359 scores are greater than 148. For the other half of the distribution, there are 3413 scores between 84 and 100; 1359 scores between 68 and 84; 215 scores between 52 and 68; and 13 scores are below 52. These answers would be true only if the distribution is exactly normally distributed.

In the absence of additional information it is difficult to say whether a score of 132 is good or not so good. A score is not very meaningful unless there is a reference group of compare against. Without such a group it is difficult- say whether soccer of 132 is high, average or low. Referring to the above example, assume a score out of a total of 10,000 scores. The percentage of scores less than 132 can now be determined. Recall that this is similar to determining the percentile rank of the scores of 132. It can be seen that 132 is two standard deviations above the mean. In normally shaped distribution, 34.13+13.59= 47.72% of the scores are between the mean and a score that is two standard deviations above the mean. To find the percentile rank of 132 and 47.72% to 50% because 50% of the scores fall below the mean. Thus 47.72+50.00= 97.72% of the scores an below an IQ score of 132. Thus a score of 132 is pretty good. The person should feel happy.

In order to make sense of raw data, the mean and the standard deviation of a distribution must be known. But different distributions have different means and standard deviations. It would be nice if one can find a way to control for differing means and standard deviations that is if all distribution have the same mean and SD. Luckily, this can be done by calculating the standard scores, also called the z scores.

8.4.2 Standard Scores or Standard Normal Variate

As shown above, a z scores is transformed score that designates how many standard deviation units the corresponding raw score is above or below the mean. That is

$$Z \quad = \frac{x_1 - \bar{x}}{s} \; or \; Z \quad = \frac{x_i - \mu}{\sigma} \qquad \ldots\ldots\ldots.(3.25)$$

This process of converting raw scores into scores is called scores transformation. . After converting all the X scores into z scores, the sum of all z scores is always zero. Therefore the mean will be zero. Similarly, if variance of z scores are calculated, its value turn out to be. 1.0. Thus Z scores are pure, unit free numbers ranging from -00 to 00.

i.e. if $X \sim N(\mu, \sigma^2)$ then $Z \sim N(0,1)$

The standardized Z normal variate can be understood from the following figure. It shows that when X falls between any two values x_1 and x_2. Z falls between the corresponding z_1 and z_2 values where

$$Z_1 = \frac{x_1 - \mu}{\sigma} \; and \; Z_2 = \frac{x_2 - \mu}{\sigma} \ldots\ldots\ldots.(3.26)$$

Thus, values of X falling between x_1 and x_2 will have corresponding Z values falling between z_1 and z_2. The area under the normal curve for X bound by the two ordinates at x_1 and x_2 will be the same as the area under the standardized Z normal curve bound by the corresponding two ordinates at z_1 and z_2.

$$p(x_1 < X < x_2) = p\left[\left(\frac{x_1 - \mu}{\sigma}\right)\right] < Z < \left[\left(\frac{x_2 - \mu}{\sigma}\right)\right] \ldots\ldots..(3.27)$$

$$= p(z_1 < Z < z_2)$$

Because the events [x_1 < x < x2] and [z_1 < z < z_2] are equivalent. Further the total area below the standard normal curve.

$$\int_{-\infty}^{\infty} \varphi(z)dz = 1$$

The distribution of z scores is referred to as standard distribution. For a normally distributed variable, the distribution or raw scores is called normal distribution and the corresponding distribution of z scores is called the standard normal distribution.

A standard normal curve is shown below:

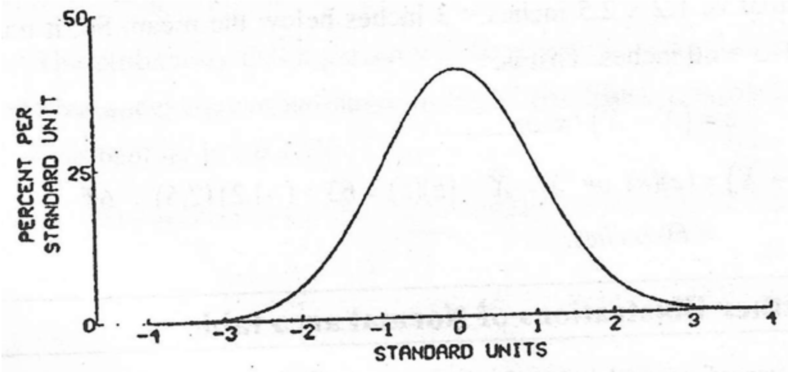


Fig. 3.6 Normal Distribution of body Weight (say) with μ =55 kg and σ_1 = 5 kg.

The standard scores express the raw scores (actual scores) in terms of standard deviation units. A z scores tell how for away (above or below) from the mean does a score lie in standard deviation units. For example a z score of 1.46 means the actual score is 1.46 standard deviations above the mean. Similarly, a z score of -0.87 means the score falls 0.87 standard deviations below the mean. A positive z scores means the scores is above the mean a negative z scores means the scores is below the mean and a zero scores means the score is exactly to the mean.

Example:

The average height of women in survey was found to be 63 inches with standard deviations of 2.5 inches. Convert the following to standard scores (or standard units): 63 inches, 65.5 inches, and 58 inches. Find the height which is -1.2 in standard units.

Solution:

To answer the first question: 63 inches is the average and therefore, it is 0 standard deviations away from the mean and is 0 in standard units:

$$i.e. \quad z = \frac{X_i - \bar{X}}{s} = \frac{63 - 63}{2.5} = \frac{0}{2.5} = 0$$

The height of 65.5 inches is 2.5 inches above the mean, that is 1 SD above the mean, In standard units (or z scores units), 65.5 inches is +1:

$$z = \frac{X_i - \bar{X}}{s} = \frac{65.5 - 63}{2.5} = \frac{2.5}{2.5} = +1$$

The height of 58 inches is 5 inches below the mean that is 3 SD below the mean. In standard units (or z scores units), 58 inches is -2.

$$z = \frac{X - \bar{X}}{s} = \frac{58 - 63}{2.5} = \frac{-5}{2.5} = -2$$

To answer the second question: This height is 1.2 standard deviations below the mean that is 1.2 × 2.5 inches = 3 inches below the mean. So it must be equal to 63 – 3 = 60 inches. That is,

$$z = \frac{X - \bar{X}}{s} \quad or$$

$$(X - \bar{X}) = (z)(s) \ \ or$$

$$X = \bar{X} + (z)(s) = 63 + (-1.2)(2.5) = 63 - 3 = 60 \ inches.$$

8.4.3 Further Illustration of Normal area table

The use of normal area tables may now be illustrated by obtaining the range probabilities under different situations referring to the normal distribution of the weight characteristic with μ = 55 kg and σ = 5 kg to understand and facilitate the use of these tables. It is always useful to draw a normal curve and show the desired area on it.

(a) Probability that a person will have weight between 60.5 and 64.5 kg and also of the one who will have weight between 45.5 and 49.5 kg.
Here the Z values are
Thus the desired probability is the shaded area under the right tail of the normal curve in Fig. 7(a) which is

$$z = \frac{60.5 - 55}{5} = 1.1 \ and \ z_2 = \frac{64.5 - 55}{5} = 1.9.$$

$$p(60.5 < X < 64.5) = p(1.1 < z < 1.9) = p(0 < z < 1.9) - p(0 < z < 1.1)$$
$$= 0.4713 - 0.3643 = 0.1070$$

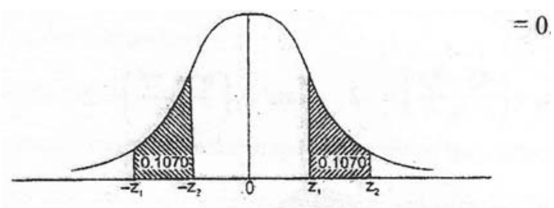


Fig. 7 (a) $p(1.1 < Z < 1.9)$ and $p(-1.9 < Z < -1.1)$

Similarly the probability that a person will have weight between 45.5 and 49.5kg is the shaded are under the left tail of the normal curve in Fig. 7(a). Here,

$$z = \frac{45.5 - 55}{5} = -1.9 \ and \ z_2 = \frac{49.5 - 55}{5} = -1.1.$$

$$the \ requirest \ probabilty \ is \ p(45.5 < X < 49.5) = p(-1.9 < X < -1.1)$$

= P(-1.9 < Z < 0) – p (1.1 < Z < 0) = 0.4713- 0.3645 = 0.1070

(b) The probability that a person will have weight more than 65kg is the shaded area under the normal curve in Fig. 7(b). Since z = (65-55)/5 =2 the required probability is = 0.1070

P(X > 65) = p (Z > 2) = 0.5000 – p (0 < Z < 2) = 0.5000-0.4772 = 0.0228

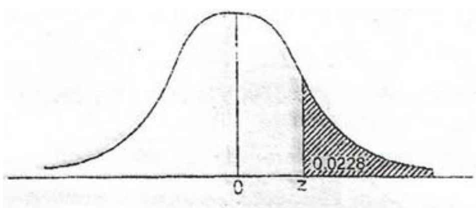


Fig. 7 (b) N (Z>2)

Similarly the probability that a person will weigh less than 50 kg is the shaded area under the normal curve in Fig. 7(c). Since z= (50-55)/5= -1, required probability. Therefore,

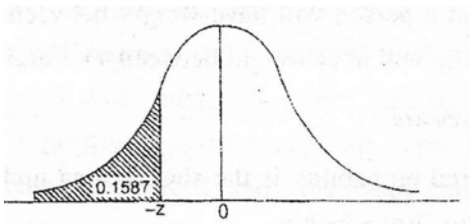


Fig. 7 (c) P(Z<-1)

P(X<50)=p(Z<-1)= 0.5000-p(-1<z<0)= 0.5000-0.3413= 0.1587.

(c) The probability that a person will weight between 45 and 65kg. the required probability is the shaded area under the normal curve in Fig. 7(d). Here Z values being.

$$z = \frac{45 - 55}{5} = -2 \ and \ z_2 = \frac{65 - 55}{5} = 2.$$

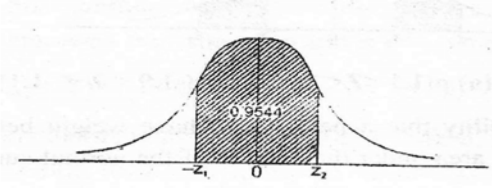


Fig. 7 (d) p(-2<Z<2)

thus the rquired probability is

$$N(45<X<65)=N(-2<Z<2)= 2N(0<Z<2) = 2(0.4772)= 0.9544.$$

(d) The probability that a person will weigh between 47 an 58kg. the corresponding Z values are

$$z = \frac{47-55}{5} = -1.6 \; and \; z_2 = \frac{58-55}{5} = 0.6$$

thus the rquired probability is the shaded are under the normal curve in fig 7(*e*)

$$N(47<X<58)=N(1.6<Z<0.6)= N(-1.6<Z<0) = N(0<Z<0.6)= 0.4452+0.2257= 0.6709$$

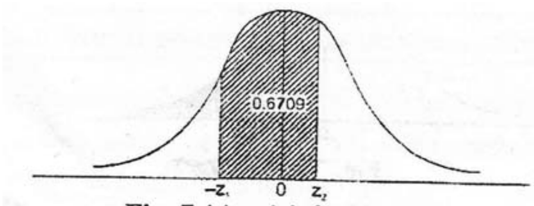


Fig. 7 (e) p (-1.6< Z < 0.6)

The area lies equally divided on two sides of the central value of the normal curve. The resultant half is then located in the body of the normal area tables to find the corresponding Z values.

8.4.3.1 Advantages of Z Scores

(i) z scores have no unit of measurement and thus they allow comparison among scores in, different distributions even when the units of measurement of the distribution are different. For example using z scores one can compare income with height because after standardizing the scores in the two distributions both have the same mean and SD.

(ii) In conjunction with a normal curve, z scores helpful in determining the number or percentage of scores that fall above or below a particular score in the distribution.

8.4.3.2 Characteristics of Z Scores

(i) The z scores have the same shape as the set of raw scoters.

(ii) z scores can be calculated for distribution of any shape using the z formula.

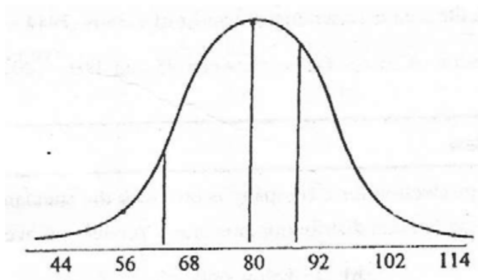
(iii) The mean of the z scores is always equal to 0; and the standard deviation of z scores is 1.

Example

The scores on a nationwide mathematics aptitude exam are normally distributed, with the mean being 80 and SD 12. What percentages of scores fall between 64 and 90?

Solution

Following diagram shows the relevant area (area of interest)



Distribution of Mathematics Aptitude Exams Scores

The shaded areas are on the either side of the mean. To solve this problem, Find the area between 64 and 80 and add up it to the area between 80 and 90. To determine area of interest, calculate the corresponding z scores for x_1 64 and $x_2 = 90$.

$$z_1 = \frac{64-80}{12} = -\frac{16}{12} = -1.33 \text{ } and \text{ } z_2 = \frac{90-80}{12} = \frac{10}{12} = 0.83$$

From standard normal tables determine the two areas: between mean and z of -1.33 on the left and the left area between mean and z of .83 on the right. The area corresponding to a z score of -1.33 is .4082 and the area corresponding to a z score of .83 is .2967. The total area equals the sum of these two areas. Thus proportion of scores falling between 64 and 90 is .4082+ .2967= .7049 or about 70%. Note that one cannot just subtract 64 from 90 and divide by 12. Also one cannot simply subtract one z value from the other because the curve is not rectangular but has differing amounts of area under various points of the curve.

Example 7

Referring to the same distribution of aptitude exam scores discussed above, find the percentage of scores falling between the scores of 95 and 110.

Solution:

First Find the areas between 110 and mean and 95 and mean. By subtracting these two areas, area between 95 and 110 can be calculated. First convert X scores into Z scores as:

$$z_1 = \frac{110-80}{12} = \frac{30}{12} = 2.50 \text{ and } z_2 = \frac{95-80}{12} = \frac{15}{12} = 1.25$$

Referring to the normal curve table, the area between mean 0 and z of 2.50 is .4938 and the area between mean 0 and z of 1.25 is .3944.

The proportion of scores falling between 95 and 110 = .4938 - .3944 = .0994 or 9.94%.

8.5 Self Assessment Exercises

1. The mean production for a company is 662 with the standard deviation 32. Assuming Normal distribution how many productions over 1000.
   a) Over 700 b) below 650
2. if Y is normal variate with mean 30 and variance 25. Find the probabilities.
   (i) $26 \le x \le 40$ (ii) $X \ge 45$ (iii) $|X-30| > 5$
3. if X is normal variate with mean 6 and variance 9. And also Y is a normal variate with mean 7 and variance 16. Then find the value of a. give a that
   $$P(2X + Y \le a) = P(4X-3Y \ge 4a)$$
4. In an admission test of B.Ed. of 100 candidates. The average marks obtained by them are 42 and standard deviation is 24. Then estimate
   (a) No. of candidates more than 60 marks.
   (b) No. of candidates who have marks between 20 and 40. (assume the normal distribution)
5. The first and third quartiles of normal distribution are 8 and 14 respectively. Calculate the mean and standard deviation of the above normal distribution.
6. In a normal distribution the 3rd deciles is 5 and the upper quartile is 63. Find mean and variance.
7. If X_1 and X_2 are two normal variate with respective mean 1 and 2 and also respective variance 4 and 3. Estimate the distribution of $X_1 + 2X_2$ with mean and variance.

8.6 Answers

1. a) 117 1.b) 352

2. a) 0.7653 b) .00135 c) 0.3174.

3) 7.51

4.a) 227 b) 289

5) mean = 11 variance = 4.4

6) mean = 30.40 variance = 47.934

7) $(x_1+2x_2) \sim N(5,16)$

8.7 Summary

Normal Distribution has been defined with its properties and applications.

8.8 Further Readings

1. Cramer H, Mathematical Methods of Statistics, Princeton University Press, 1946 and Asia Publishing House, 1962.
2. Hogg R.V. and Craig A.T., Introduction to Mathematical Statistics, Macmillan, 1978.
3. Prazen E. Modern Probability Theory and its applications, John Wiley, 1960 and Wiley Eastern 1972.
4. Rao C.R., Linear Statistical Inference and Its Applications, John Wiley, 1960 and Wiley Eastern 1974.
5. Rohtagi V.K. (1984), An Introduction to Probability Theory and Mathematical Statistics, John Wiley, 1976 and Wiley Eastern 1985.
6. Vikas S.S., Mathematical Statistics, John Wiley, 1962 Toppan.

Unit - 9: Continuous Distribution

Structure

9.1 Introduction- Uniform (Rectangular) Distributions (Definition)

It is simplest of all theoretical continuous probability distributions in which the probability density function of a continuous random variable is constant over a finite interval.

Definition:

A continuous random variable X is said to follow a uniform distribution over an interval (a, b) if its probability density function is constant over entire range (a, b) of X i.e., p.d.f. of X is given by

$$f(x) = \begin{cases} k, & if\ a \leq x < b \\ 0, & otherwise \end{cases} \text{……………………. (4.1)}$$

Here K is some positive constant. Since function given by (4.1) of p.d.f

$$\int_a^b f(x)dx = 1$$

$$\Rightarrow \int_a^b K\,dx = K[x]_a^b = K\,[b-a] = 1$$

$$i.e. \quad K = \frac{1}{b-a}$$

Therefore the p.d.f uniform distribution can be written as

$$f(x) = \begin{cases} \dfrac{1}{b-a}; & a \le x \le b \\ 0, & otherwise \end{cases} \qquad \ldots\ldots\ldots\ldots(4.2)$$

The density curve of uniform distribution is given in Fig 4.1

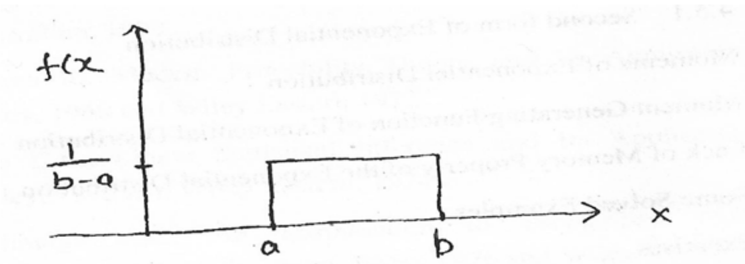


Figure 4.1 Density function of Uniform Distribution

Since density curve of a uniform distribution describes a rectangle (see Fig. 4.1), the distribution is also known as rectangular distribution. Here constants a and b are two parameters of uniform distribution and we denote by X~U (a, b) or X~R (a, b).

Some important form of uniform distribution is

(a) Uniform between 0 and 1 i.e.,

$$f(x) = \begin{cases} 1, & 0 \le x \le 1 \\ 0, & otherwise \end{cases} \qquad \ldots\ldots\ldots\ldots(4.3)$$

(b) Uniform between 0 and θ i.e.,

$$f(x,\theta) = \begin{cases} \frac{1}{\theta}, & 0 \le x \le \theta \\ 0, & otherwise \end{cases} \qquad \ldots\ldots\ldots\ldots(4.4)$$

(c) Uniform between -θ and θ i.e.,

$$f(x,\theta) = \begin{cases} \frac{1}{2\theta}, & -0 \le x \le \theta \\ 0, & otherwise \end{cases} \qquad \ldots\ldots\ldots(4.5)$$

9.2 Distribution Function of Uniform Distribution

The distribution function of a random variable is defined as

$$f(x) = P[X \le x] = \int_{-x}^{x} f(x)dx$$

The distribution function of uniform distribution between a and b is

$$f(x) = \begin{cases} 0, & if\ x < 0 \\ \dfrac{x-a}{b-a}, & if\ a \le x \le b \\ 1 & if\ x > b \end{cases}$$

The distribution function of R (a,b) is shown by Fig 4.2

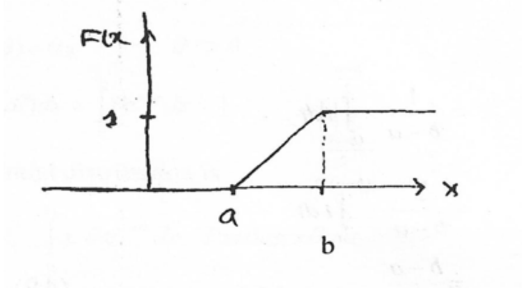


Figure 4.2 Distribution function of uniform distribution

9.3 Moments of Uniform Distribution

r^{th} raw moment of X, about origin is defined as

$$\mu_r^1 = e(x^r) = \int_a^b x^r \frac{1}{b-a} dx$$

$$= \frac{1}{b-a}\left[\frac{x^{r+1}}{r+1}\right]_a^b = \frac{1}{b-a}\left[\frac{b^{r+1}-a^{r+1}}{r+1}\right] \ldots\ldots\ldots.(4.6)$$

In particular

$$Mean = \mu_1^1 = \frac{b+a}{2} \ldots\ldots\ldots\ldots\ldots..(4.7)$$

$$\mu_2^1 = \frac{1}{3}[b^2 + ab + a^2]$$

So that

$$V(x) = \mu_2 = \mu_2 - \mu_1^2 = \frac{1}{12}(b-a)^2 \quad \ldots\ldots\ldots\ldots\ldots\ldots(4.8)$$

Therefore

$$S.D. of\ x = \frac{(b-a)}{\sqrt{12}}$$

Mean deviation about mean

$$M.D. = E\left|x - \frac{(a+b)}{2}\right| = \int_a^b \frac{1}{(b-a)}\left|x - \frac{(a+b)}{2}\right| dx;\ put\ t = x - \frac{(a+b)}{2}$$

$$= \frac{1}{b-a}\int_{-\frac{(b-a)}{2}}^{\frac{b-a}{2}} |t|dt = \frac{2}{b-a}\int_0^{\frac{b-a}{2}} tdt = \frac{b-a}{4} \qquad \ldots\ldots\ldots\ldots\ldots\ldots(4.9)$$

9.4 Moment generating function of Uniform Distribution

Moment generating function of a random variable X is defined as

$$M_x(t) = E(e^{tx})$$

Where t is any arbitrary real constant.

If X is uniform variate in (a, b)

$$M_x(t) = \int_a^b \frac{1}{b-a} e^{tx}$$

$$= \frac{1}{b-a}\left[\frac{e^{tx}}{t}\right]_a^b = \frac{e^{bt} - e^{et}}{t(b-a)} \qquad \dots\dots\dots\dots(4.10)$$

9.5 Exponential Distribution (Definition)

Exponential distribution plays important role in statistics. It has been used as potential model for lifetimes of many things. Exponential distribution is also used for continuous waiting time random variable of various events whereas geometric distribution is used for discrete waiting time random variable.

Definition: A continuous random variable X is said to follows an exponential distribution if it assumes non negative values with probability density function (p.d.f.) given by

$$f(x,0) = \begin{cases} \theta\, e^{-\theta x}, & x \geq 0, \theta > 0 \\ 0, & otherwise \end{cases} \dots\dots\dots\dots(4.11)$$

Here θ is the parameter of exponential distribution it is easy to verify that the function given in (4.11) is a p.d.f. i.e.

$$f(x,\theta) = e^{-\theta} \;\; ; \quad \theta \geq 0$$

$$and \int_{-\infty}^{\infty} f(x,\theta)dx = \int_0^{\infty} \theta e^{-\theta x} dx = 1$$

Mean of the exponential distribution is

$$\mu_1 = E(x) = \int_0^{\infty} x\theta e^{-\theta x} dx; \; putting\; x\theta = 1 \quad \text{and } \theta\, dx = dt$$

$$= \int_0^{\infty} t\;\, e^{-t} \frac{dt}{\theta} = \frac{1}{\theta} = \;\theta^l \;\, (say)$$

Another form of exponential distribution can be written in terms of meanθ'.

9.5.1 Second form of Exponential distribution:

A non negative continuous random variable X is said to follow an exponential distribution with if its p.d.f. is given by

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

$$\theta' > 0 \qquad \dots\dots..(4.12)$$

The graph of p.d.f given in (4.12) is given in Fig 4.3. The f(x, θ') is maximum at x = 0 and then decreases as x increased and tends to zero as x → ∞. Similar shape is also obtained for p.d.f. is given in (4.11).

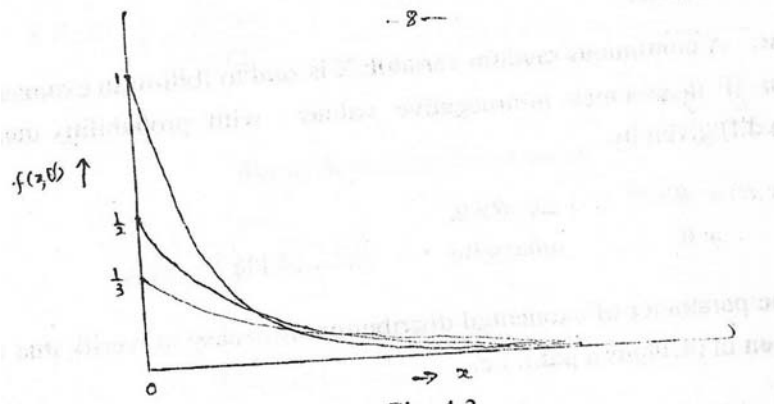


Fig. 4.3

The distribution function of exponential distribution is given by

$$F(x) = P[X \leq x] = \int_0^\infty \theta e^{-\theta} \, dt \;\; = 1 - e^{-\theta} \;\; ; x \geq 0 \qquad \dots\dots\dots..(4.13)$$

9.6 Moments of Exponential Distribution

Since the two forms of exponential distribution given in (4.11) and (4.12) are reparametrization by putting $= 1/\theta'$, it is sufficient to obtain the moments for the p.d.f. given in (4.12). The rth raw moments about origin is defined as

$$\mu_r^1 = E(x^r) = \int_0^\infty x^r \frac{1}{\theta'} e^{-\frac{e}{\theta}} \, dx;$$

$$Put \frac{x}{\theta'} = t \;\; or \; x = \; \theta' t \;\;\; and \;\; dx = \; \theta' dt$$

$$= \int_0^\infty (\theta' t)^r \, e^{-t} \, dt = \theta'^r \int_0^\infty t^r e^{-t} dt = \theta'^r \, \Gamma\,(r+1)$$

$$= \gamma!\, \theta'^r \qquad \ldots\ldots\ldots\ldots\ldots\ldots(4.14)$$

Putting r = 1, 2, We get

$$Mean = \mu_1 = E(x) = \theta'$$

$$\mu_2 = 2\theta'^2$$

$$Variance = \mu_2 = \mu_1^2 = 2\theta'^r - \theta'^r = \theta'^r$$

$$For\ p.d.f.(4.11)$$

$$\left.\begin{aligned} Mean &= \frac{1}{\theta} \\ Variance &= \frac{1}{\theta^2} \end{aligned}\right\} \qquad \ldots\ldots\ldots\ldots\ldots\ldots(4.15)$$

9.7 Moment Generating Function of Exponential Distribution

M.G.F. of X is defined as

$M_x(t) = E\,(e^{tx})$

For p.d.f. given in (4.11)

$$M_x(t) = \int_0^\infty e^{tx} \theta e^{-tx} \, dx$$

For p.d.f. given in (4.12)

$$= \theta \int_0^\infty e^{-(\theta - t)x} \, dx$$

$$= \frac{\theta}{\theta - t} = \left(1 - \frac{t}{\theta}\right)^{-1}$$

$$M_x(t) = \left(1 - \frac{t}{\theta}\right)^{-1} \qquad \ldots\ldots\ldots\ldots\ldots\ldots(4.16)$$

$$M_x(t) = (1 - \theta' t)^{-1} = \sum_{r=0}^{\infty} (\theta' t)^r$$

$$\mu_r = \theta'^r \gamma!$$

9.8 "Lack of Memory" Property of the Exponential Distribution

The exponential distribution is said to have "lack Memory" property in a certain sense. Suppose an event E can occur at any time x and occurrence (waiting) time has exponential distribution. Suppose we know that E has not occurred before K i.e., X ≥ K. Let Y = X - K, thus Y is additional time needed for E to occur. Then a distribution is said to have "lack memory" property if conditional distribution of Y given X ≥ K is same as distribution of X i.e.,

$$P\left[Y \leq \frac{x}{X} \geq K\right] = P[X \leq x] \qquad for\ all\ x \qquad \dots\dots(4.17)$$

As an illustration if Mr. A were waiting for the event E and is relieved by Mr. B after time K, the waiting time distribution of Mr. B is same as that of Mr. A. Here we will prove that exponential distribution "Lacks Memory".

Proof: The p.d.f. of an exponential distribution with parameter θ is given by

$$f(x,\theta) = e^{-\theta} \ ; \ \mathrm{x} \ \geq \ 0, \quad \theta \geq 0$$

Then

$$P[X \leq x] = \int_0^\infty \theta e^{-\theta t} dx = 1 - e^{-\theta x} \qquad \dots\dots(4.18)$$

$$P\left[Y \leq \frac{x}{X} \geq K\right] = \frac{P[Y \leq x \cap X \geq K]}{P[X \geq K]}$$

Here

$$P[Y \leq x \cap X \geq K] = P[X - K \leq x \cap X \geq K]$$

$$= P[X \leq x + K \cap X \geq K] = P[K \leq x \leq x + K]$$

$$= \theta \int_K^{K+x} e^{-\theta x}\, dx = \theta \left[\frac{e^{-\theta x}}{\theta}\right]_K^{K+x}$$

$$= e^{-\theta(x+k)} + e^{-\theta k} = e^{-\theta k}\left(1 - e^{-\theta x}\right) \qquad \dots\dots\dots(4.19)$$

From (4.18), we have

$$P[X < k] = 1 - e^{-\theta k}$$

$$\Rightarrow P[X \geq k] = e^{-\theta k}$$

$$\therefore \ P\left[Y \leq \frac{x}{X} \geq k\right] = \frac{e^{-\theta k}\left(1 - e^{-\theta k}\right)}{e^{-\theta k}} = \left(1 - e^{-\theta k}\right) = P[X \leq x]$$

$\therefore$ *Exponential distribution* Lacks Memory" *proved.*

9.9 Some Solved Examples

Example 1: Suppose that X is a uniform random variable over the interval (0,1). Find-

(i) $P\left[X > \frac{1}{3}\right]$

(ii) $P[X < 0.7]$

(iii) $P[0.3 \leq X \leq 0.8]$

Solution:

X~ U (0,1) its p.d.f. is given by

$$f(x) = 1 \quad 0 \leq x \leq 1$$

$$= 0 \quad otherwise$$

(i) $P\left[X > \frac{1}{3}\right] = \int_{1/3}^{1} 1\, dx = [x]_{1/3}^{1} = 1 - \frac{1}{3} = \frac{2}{3}$

(ii) $P[X < 0.7] = \int_{0}^{0.7} 1\, dx = [x]_{0}^{0.7} = 0.7$

(iii) $P[0.3 \leq X \leq 0.8] = \int_{10.3}^{0.8} 1\, dx = 0.8 - 0.3 = 0.5$

Example 2: Subway trains on a certain line run every half hour. What is probability that a man entering the station at a random time will have to wait at least twenty minute? Assume that waiting time is uniformly distributed between (0,30).

Solution: Let us denote X, the waiting time in minutes and it follows U (0,30) i.e., of X is given by

$$f(x) = \frac{1}{30}; \quad 0 \leq x \leq 30$$

$$= 0 \quad otherwise$$

$$\therefore p[X \geq 20] = \int_{20}^{30} \frac{1}{30} dx = \frac{1}{30}[x]_{20}^{30}$$

$$= \frac{1}{30}[30 - 20] = \frac{1}{3}\ ans.$$

Example 3: If X is uniformly distribution with mean 1and variance 1/3. Find P[X<1].

Solution:

$$Let\ X \sim U\ (a, b)$$

$$E(x) = \frac{a+b}{2} = 1 \Rightarrow a + b = 2 \ldots\ldots\ldots.(1)$$

$$and\ V(x) = \frac{(b-a)^2}{12} = \frac{1}{3} \Rightarrow (b-a)^2 = 4$$

$$or\ \ (b-a) = \pm 2.\ since\ b-a>0$$

$$b-a=2 \qquad \qquad \ldots\ldots\ldots(2)$$

From equation (1) and (2)

a = 0, b = 2.

$$\therefore f(x) = \frac{1}{2}; \qquad 0 \le x \le 2$$

$$= 0 \quad otherwise$$

$$p[X<1] = \int_0^1 \frac{1}{2}dx = \frac{1}{2}[x]_0^1 = \frac{1}{2}\ \ Ans.$$

Example 4.4: If X has exponential distribution with mean 2. Find P[X<1].

Solution: From (4.12) the p.d.f. of exponential distribution with mean $\theta' - 2$ is

$$f(x,\theta') = \frac{1}{\theta'}e^{-x/\theta} = \frac{1}{2}e^{-\frac{x}{2}} \quad x \ge 0$$

Then

$$p[X<1] = \int_0^1 \frac{1}{2}e^{-\frac{x}{2}}dx \quad Put\frac{x}{2} = t$$

$$\therefore dx = 2dt$$

$$= \int_0^{\frac{1}{2}} e^{-1}dt = [-e^{-t}]_0^{\frac{1}{2}} = 1 - e^{-\frac{1}{2}} \qquad Ans.$$

Example 4.5: If X has a uniform distribution on (0, 1). Show that Y= - log x follows exponential distribution. Obtain mean and variance of Y.

Solution: The p.d.f. of uniform distribution on (0, 1) is

$$f(x)\ = 1 \quad if\ 0 \le x \le 1$$

$$= 0\ otherwiese$$

Consider transformation Y = - log x

$\Rightarrow x = \ e^{-y}\ \ i.e.\ \ y = -\log x\ is\ one\ one\ transformation.$

$Jacobian\ of\ transformation\ is$

$$|J| = \left|\frac{dx}{dy}\right| = |-e^{-y}| = e^{-y} \; and \; as \; x \to 1 \quad y \to 0$$

Thus range of Y is (0, ∞)

The p.d.f. of Y is given by

$$G(y) = f(x)|J| = 1.e^{-y} \quad if \;\; 0 \le y \le \infty$$

$$or \; g(Y) = e^{-y}; \quad y \ge 0$$

Which is exponential distribution with parameter 1. Thus

Mean = E(Y) = 1

Variance = E(Y) = 1 Ans.

Example 4.6: If X is a continuous random variable with distribution function F, prove that Y= F(x) has a uniform distribution on [0, 1].

Solution: Since F is a distribution function, it is non decreasing function and takes values value in the range [0, 1] and inverse also exists. The distribution function of Y is given by

$$G_Y(y) = P[Y \le y] = P[F(x) \le y] = P[X \le F^{-1}(y)] = y$$

$\therefore p.d.f. of \; Y \; is \; given \; by$

$$g(y) = \frac{d}{dy} G_Y(y) = \frac{dy}{dy} = 1$$

Thus range of y is [0, 1]

$\therefore p.d.f. of \; Y \; is \; given \; by$

$$g(y) = 1 \qquad if \; 0 \le y \le 1$$

$$= 0 \qquad \text{otherwise}$$

$\therefore Y \; has \; unifrom \; distribution \; on \; [0,1]. Proved.$

9.10 Self Assessment Exercises

1. A random variable X has a uniform distribution over (-3, 3), compute

   (i) P(X=2), P(X<2), P(|X|<2)

   (ii) Find K for which P [X>k]= 1/3

   Ans. (i): 0, 5/6, 4/6 (ii): K=1

2. Suppose that X is uniformly distributed over $(-\infty, \infty)$ where$(\infty > 0.)$ Determine ∞ so that

(i) P (X > 1) = 1/3 (ii) P[X < 1/2] = 0.3 and (iii) P(|X| < 1) = P(|X| > 1)

Ans. (i) $\infty = 3$ $(ii)\infty = \frac{5}{6}$ $(iii)\infty = 2.$

3. Calculate the coefficient of coefficient of variation for rectangular distribution in [0,b]
4. Show that whatever be the distribution function F(x) of a continuous random variable X

$$P_r[a \leq F(x) \leq b] = b - a: \quad 0 \leq a, \ b \leq 1$$

5. If X has uniform distribution in (-a, a), show that odd order moments (central) are zero and even order moments $\mu_{2r} = \frac{a^{2r}}{(2r+1)}$.
6. If X has exponential distribution with parameter θ such that p[x ≤ 1] = P[X > 1]. Find mean and variance of X.

Ans. Mean = 1/log 2, and var (X) = $1/(\log 2)^2$

7. Show that Y= - (1/λ) log F(x) is exponential variate with parameter λ. F(x) is distribution function.
8. A random variable X has density function

$$(a)\frac{500}{500-t} \quad (b)\ \frac{1}{500-t} \quad (c)\ \frac{1}{500-t} \quad (d)\ \frac{1}{1-500t}$$

$$f(x) = \begin{cases} Ce^{-3x}, & x > 0 \\ 0, & x < 0 \end{cases}$$

The value of the constants C is

(a) 3 (b) 4 (c) 1/3 (d) 9.

9. If X is uniformly distributed in -2≤ $X \leq 2, then\ P[X \leq 1] is$

(a) 3/4 (b) 1/4 (c) 1/2 (d) 0.

10. The number of hours of satisfactory operation of a certain brand of T.V. set is a r.v. with p.d.f.

$$f(x) = -500e^{-500x}, \quad x > 0. The\ mgf\ of\ X\ is$$

$$(a)\frac{500}{500-t} \quad (b)\ \frac{1}{500-t} \quad (c)\ \frac{1}{500-t} \quad (d)\ \frac{1}{1-500t}$$

9.11 Summary

Uniform and Exponential distributions have been defined with their properties and applications.

9.12 Further Readings

1. Cramer H, Mathematical Methods of Statistics, Princeton University Press, 1946 and Asia Publishing House, 1962.
2. Hogg R.V. and Craig A. T., Introduction to Mathematical Statistics, Macmillan, 1978.
3. Prazen E. Modern Probability Theory and its applications, John Wiley, 1960 and Wiley Eastern 1972.
4. Rao C.R., Linear Statistical Inference and Its Applications, John Wiley, 1960 and Wiley Eastern 1974.
5. Rohtagi V. K. (1984), An Introduction to Probability Theory and Mathematical Statistics John Wiley, 1976 and Wiley Eastern 1985.
6. Vikas S. S., Mathematical Statistics John Wiley, 1962 Toppan.

Unit-10: Sampling Distribution

Structure

10.1 Introduction

if n is small the 'normal test' cannot be applied and if the sample size n is small the distribution of the various variates follows as are far from normally. In such cases sample distributions are useful. In the following sections we shall discuss: Chi square distribution, t distributions, F distribution and z distribution.

These sample distributions can; however be applied to large samples also though the converse is not true. In all the sample distributions have a basic assumptions which is that *"the population(s) from which sample(s) is (are) drawn is (are) normal, i.e. the parent population(s) is (are) normally distributed."*

A sampling distribution is also a probability distribution of any test statistic which obtained from samples drawn from a specific population. ***The study of sampling distributions is very important for inferential statistics***. In practice, if we collect sample data and, from these data, estimate parameters of the population distribution. Thus, knowledge of the sampling distribution can be very useful in making inferences about the overall population.

10.2 Objectives

After reading this unit you should be able to:

- Define the chi square, t, F and z distribution.
- Calculate the different parameters of these distributions.
- Obtain the relations between them.

10.3 The Chi Square Distribution

In this section we discuss the chi square distribution.

The Chi-square variate (χ^2 is pronounced as chi square) with 1 degree of freedom (d.f.) is the square of a standard normal variate. Mathematically

Thus if $X \sim N(\mu, \sigma^2)$, then $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$ and $Z^2 = \left(\frac{X-\mu}{\sigma}\right)^2$ is a chi-square variate with 1 d.f.

In general if Xi (i=1,2,…,n) are n independent normal variates with means and variance σ_i^2 (i= 1,2,…,n) then

$$\chi^2 = \sum_{t=1}^{n} \left(\frac{X_i - \mu_i}{\sigma_i}\right), \text{is a chi} - \text{square variate with n d.f.}$$

Through Method of Moment Generating Function the ***Derivation of the Chi-Square (χ^2) Distribution*** is

If X_i (i=1,2,…,n) are independent $N(\mu_i, \sigma_i^2)$ we want the distribution of

$$\chi^2 = \sum_{t=1}^{n} \left(\frac{X_i - \mu_i}{\sigma_i}\right)^2 = \sum_{t=1}^{n} U_t^2, where\ U_i = \frac{X_i - \mu_i}{\sigma_i} \sim N(0,1)$$

Since $X_i^{'s}$ are independent, $U_i^{'s}$ are also independent. Therefore,

$$M_{\chi^2}(t) = m_{\sum u_i^2}(t) = \prod_{i=1}^{n} M_{u_i^2}(t) = \left[M_{u_i^2}(t)\right]^n, \qquad [U_i^{'s}\ are\ i.i.d.N(0,1)]$$

$$M_{u_i^2}(t) = E[exp(tu_i^2)] = \int_{-\infty}^{\infty} \exp(tu_i^2)\, f(x_i) dx_i$$

$$= \int_{-\infty}^{\infty} \exp(tu_i^2) \frac{1}{\sigma\sqrt{2\pi}} \exp\{-(x_i - \mu)^2/2\sigma^2\}\, x$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp(tu_i^2)\ exp(-\mu_i^2/2)\, du_i\ \left[u_i = \frac{x_i - \mu}{\sigma}\right]$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left\{-\left(\frac{1-2t}{2}\right) u_t^2\right\} d_{u_i} = \frac{1}{\sqrt{2\pi}} . \frac{\sqrt{\pi}}{\left(\frac{1-2t}{2}\right)^{1/2}} = (1-2t)^{-1/2}$$

$$\left[as\ \because \int_{-\infty}^{\infty} e^{-a^2x^2} dx = \frac{\sqrt{\pi}}{a}\right]$$

$$\therefore \quad M_{\chi^2}(t) = (1-2t)^{-n/2}$$

Which is the m.g.f. of a Gamma variate with parameters ½ and $\frac{1}{2}n$. Hence by uniqueness theorem of m.g.f.'s,

$\chi^2 = \sum_i^n \left(\frac{x_i - \mu_i}{\sigma_i}\right)^2$ is a Gamma variate with parametes ½ and $\frac{1}{2}n$

$$\therefore dp(\chi^2) = \frac{\left(\frac{1}{2}\right)^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)} . \left[\exp\left(-\frac{1}{2}\chi^2\right)\right] (\chi^2)^{\left(\frac{n}{2}\right)-1} d\chi^2$$

$$= \frac{1}{2^{n/2}\Gamma\left(\frac{n}{2}\right)} [\exp(\chi^2/2)](\chi^2)^{\left(\frac{n}{2}\right)-1} d\chi^2, 0 \le x < \infty$$

Which is the p.d.f of chi-square distribution with n degrees of freedom.

Note: 1. if a r.v. X has a chi-square distribution with n.d.f we write $X \sim \chi^2{}_{(n)}$ and its p.d.f. is:

$$f(x) = \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} e^{-\frac{x}{2}} x^{\left(\frac{n}{2}\right)-1}, 0 \le x < \infty$$

Theorem: If $X \sim \chi^2{}_{(n)}$, $then\, \frac{1}{2}X \sim \gamma\left(\frac{1}{2}n\right)$.

Proof: The p.d.f. of $Y = \frac{1}{2}X$, is given by:

$$g(y) = f(x).\left|\frac{dx}{dy}\right| = \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}e^{-y}.(2y)^{\left(\frac{n}{2}\right)-1}.2 = \frac{1}{\Gamma\left(\frac{n}{2}\right)}e^{-y}\, y^{\left(\frac{n}{2}\right)-1};0 \le y < \infty$$

$$Y = \frac{1}{2}X \sim \gamma\left(\frac{1}{2}n\right)$$

10.3.1 M.G.F. of χ^2 Distribution

If X~χ^2 then

$$M_X(t) = E(\exp(tX)\} = \int_0^\infty \exp(tx)f(x)dx$$

$$= \frac{1}{2^{\frac{n}{2}}\Gamma(n/2)}\int_0^\infty exp\left\{-\left(\frac{1-2t}{2}\right)\right\}(x)^{\frac{n}{2}-1}dx = \frac{1}{2^{\frac{n}{2}}\Gamma(n/2)}\frac{\Gamma(n/2)}{[(1-2t)/2]^{n/2}}$$

$$\cong \qquad M_X(t) = (1-2t)^{-n/2}$$

Note: mean and variance of the chi-square distribution are n and 2n.

10.3.2 Skewness of χ^2 Distribution

Karl Pearson's coefficient of skewness is given by:

$$Skewness = \frac{Mean - Mode}{S.D.} = \frac{n-(n-2)}{\sqrt{2n}} = \sqrt{\frac{2}{n}}$$

Since Pearson's coefficient of skewness is greater than zero for n≥1, the$n$ χ^2 distribution is ***positively skewed*** distribution. Further since skewness is inversely proportion to the square root of d.f., it rapidly tends to symmetry as the d.f. increases.

10.3.3 Additive Property of χ^2 $-$variates

Theorem: The sum of independent chi-square variates is also a χ^2–variates. More precisely, if X_i, $(1,2,\dots,k)$are independent χ^2–variates with n_id.f. respectively, then the sum $\sum_{i=1}^{k} X_i$ is also a chi-square variate with $\sum_{i=1}^{k} n_i$ d.f.

Proof: We have

$$M_{X_1}(t) = (1-2t)^{-n/2}; i = 1,2, \dots . k.$$

The m.g.f. of the sum $\sum_{i=1}^{k} X_i$ is given by:

$$M_{\sum X_i}(t) = M_{X_1}(t) M_{X_2}(t) \dots . . M_{X_i}(t) \qquad [X_i^s \, are \; indepndent]$$

$$= (1-2t)^{-\frac{n_1}{2}} (1-2t)^{-\frac{n_2}{2}} \dots \dots (1-2t)^{-\frac{n_1}{2}} = (1-2t)^{(n_1+n_2+\dots .+n_k)/2}$$

Which is the m.g.f. of χ^2 −variate with $(n_1 + n_2 + \cdots . + n_k)$ d.f. Hence by uniqueness theorem of m.g.f.'s $\sum_{i=1}^{k} X_i$is a χ^2 variate with $\sum_{i=1}^{k} n_i$d.f.

Note: ***Converse is also true***,

Mathematically X_i: i=1,2,……, k are χ^2 variates with n_i;=1,2,…., and k d.f. respectively and if $\sum_{i=1}^{k} X_i$ is a χ^2 variate with $\sum_{i=1}^{k} n_i$d.f., then $X_i, (1,2, \dots, k)$are independent.

10.3.5 Probability Curve of Chi-Square Distribution

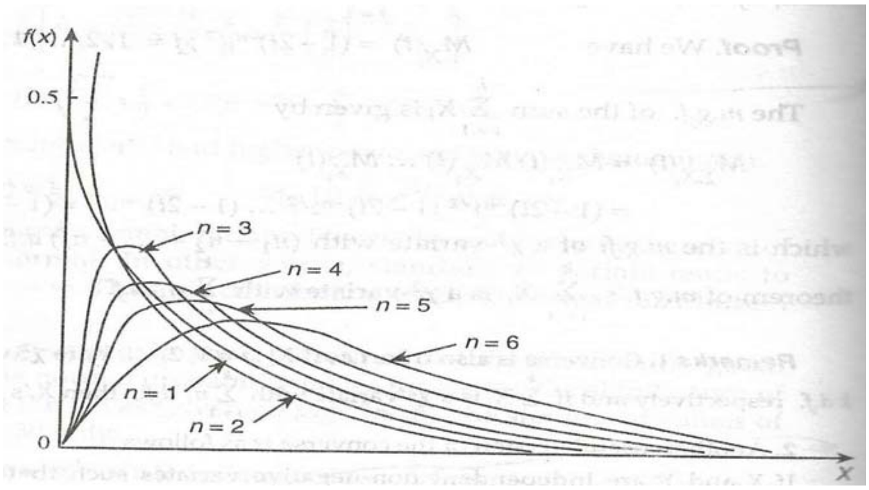


Figure: Probability Curve of Chi-Square Distribition

from pdf,

$$f'(x) = \left[\frac{n-2-x}{2x}\right] f(x)$$

Since x>0 and f(x) being p.d.f is always non-negative,

$$f'(x) < 0 \; if \; (n-2) \leq 0,$$

For all values of x. thus the χ^2 probability curve for 1 and 2 degrees of freedom is monotonically decreasing When n>2.,

$$f(x) = \begin{cases} > 0, & if\ x < (n-2) \\ = 0, & if\ x = \ (n-2) \\ < 0, & if\ x > \ (n-2) \end{cases}$$

This implies that fro n>2, f(x) is monotonically increasing for 0<x<(n-2) and monotonically decreasing for (n-2) <x<∞, while at x= n-2, it attains the maximum value.

For n ≥1 as x increase f(x) decrease rapidly and finally tends to zero as x→ ∞. Thus for n>1, the χ^2probability curve is positively skewed towards higher values of x. Moreover, x-axis is asymptote to the curve. The shape of the curve for n=1,2,3,….,6 is given in figure. For n=2 the curve will meet y = f(x) axis at x= 0, i.e., at f(x)= 0.5. For n=1, it will be an inverted J-shaped curve.

Note: In practice for ***n ≥ 30*** the ***chi-square distribution tends to normal distribution***.

10.3.5 Application of Chi-Square Distribution

In Statistics χ^2-distribution has a large number of applications, some of which most popular are given below:

(i) To test if the hypothetical value of the population variance $\sigma^2 = \sigma_0^2 (say)$ is.
(ii) To test the 'goodness of fit'.
(iii) To test the independence of attributes.
(iv) To test the homogeneity of independent estimates of population variance.
(v) To combine various probabilities obtained from independent experiments to give a single test of significance.
(vi) To test the homogeneity of independent estimates of the population correlation coefficient.

10.4 t Distribution

10.4.1 Student's t Distribution

Suppose $x_i (i = 1,2, \dots., n)$ be a random sample of size *n* from a normal population with mean μ and variance σ^2. Then the student's t test is defined by the statistic:

$$t = \frac{\bar{x} - \mu}{S/\sqrt{n}}$$

Where $\bar{x} = \frac{1}{n}\sum_{t=1}^{n} x_i$ is the sample mean and $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ is an unbiased estimate of the population varianceσ^2 . the above statistic follows students's t-distribution with $v = (n-1)$ d.f. with probability density function:

$$f(t) = \frac{1}{B\left(\frac{1}{2},\frac{1}{2}\right)}.\frac{1}{(l+t^2)} = \frac{1}{\pi}.\frac{1}{(l+t^2)}; -\infty < t < \infty \qquad \left[\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}\right]$$

Note: 1 A statistic t following student's t-distribution with n d.f. will be defined as $t \sim t_n$

2. in above pdf, if we take v=1, we get:

$$f(t) = \frac{1}{B\left(\frac{1}{2},\frac{1}{2}\right)}.\frac{1}{(l+t^2)} = \frac{1}{\pi}.\frac{1}{(l+t^2)}; -\infty < t < \infty, \quad \left[\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}\right]$$

Which is the p.d.f of standard Cauchy distribution. Hence when v=1, Student's distribution reduces to Cauchy distribution.

Derivation of Student's t-distributon: The above test statistic can be rearranged as:

$$t^2 = \frac{n(\bar{x}-\mu)^2}{s^2} = \frac{n(\bar{x}-\mu)^2}{\frac{ns^2}{n-1}}$$

$$\Rightarrow \frac{t^2}{(n-1)} = \frac{(\bar{x}-\mu)^2}{\sigma^2/n}.\frac{1}{ns^2/\sigma^2} = \frac{(\bar{x}-\mu)^2/(\sigma^2/n)}{ns^2/\sigma^2}$$

Since xi (i=1,2,……..,n) is a random sample from the normal population with mean μ and variance σ^2 $\bar{x} \sim N(\mu, \sigma^2/n) \Rightarrow \frac{(\bar{x}-\mu)}{\sigma/\sqrt{n}} \sim N(0,1)$

Hence $\frac{(\bar{x}-\mu)^2}{\sigma^2/n}$ being the square of a standard normal variate is a chi-square variate with 1 d.f. Also $\frac{ns^2}{\sigma^2}$ is a χ^2 variate with (n-1) d.f.

Further since $\bar{x}\ and\ s^2$ are independently distributed, $\frac{t^2}{n-1}$ being the ratio of two independent χ^2 variates with 1 and (n-1) d.f. respectively is a $\beta_2\left(\frac{1}{2},\frac{n-1}{2}\right)$ variate and its distribution is given by:

$$dF(t)=\frac{1}{B\left(\frac{1}{2},\frac{v}{2}\right)}\cdot\frac{\left(\frac{t^2}{v}\right)^{\frac{1}{2}-1}}{\left(1+\frac{t^2}{v}\right)^{\frac{v+1}{2}}}d\left(\frac{t^2}{v}\right),0\le t^2<\infty \qquad [where\ v=(n+1)]$$

$$=\frac{1}{\sqrt{v}B(\frac{1}{2},\frac{v}{2})}\cdot\frac{1}{\left(1+\frac{t^2}{v}\right)^{\frac{v+1}{2}}}dt;\ -\infty<t<\infty$$

Which is the required probability density function of student t- distribution with v = (n-1) d.f.

10.4.2 Fisher's t Distribution

It is the ratio of a standard normal variate to the square root of an independent chi-square variate divided by its degrees of freedom. if ξ is a N(0,1) and χ^2 is an independent chi-square variate with n.d.f. then Fisher's t is given by:

$$t=\frac{\xi}{\sqrt{\chi^2/n}}$$

and it follows student's 't' distribution with n degrees of freedom.

Derivation of Fisher's t-distributon: Since ξ and χ^2 are independent, their joint probability differential is given by:

$$dF(\xi,\chi^2)=\frac{1}{\sqrt{2\pi}}exp(-\xi^2/2)\frac{\exp\left(-\frac{\chi^2}{2}\right)(\chi^2)^{\frac{n}{2}-1}}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}d\xi\ d\chi^2$$

Let us transform to new variates t and u by the substitution:

$$t=\frac{\xi}{\sqrt{\chi^2/n}}\ and\ u=\chi^2\Rightarrow\xi=t\sqrt{u/n}\ and\ \chi^2=u$$

Jacobian of transformation J is given by:

$$J=\frac{\partial(\xi,\chi^2)}{\partial(t,u)}=\begin{vmatrix}\sqrt{\frac{u}{n}} & \frac{t}{2\sqrt{un}}\\ 0 & 1\end{vmatrix}=\sqrt{\frac{u}{n}}$$

The joint p.d.f g(t, u) of t and u becomes:

$$g(t,u)=\frac{1}{\sqrt{2\pi}\,2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)\sqrt{n}}exp\left\{-\frac{u}{2}\left(1+\frac{t^2}{n}\right)\right\}u^{\frac{n}{2}-\frac{1}{2}}du$$

$$Since\ \psi^2 \geq 0\ and - \infty < \xi < \infty, u \geq 0\ and - \infty < t < \infty.$$

Integrating w.r. to 'u' over the range 0 to , the marginal p.d.f. g_1 (.) of t becomes:

$$g_1(t) = \frac{1}{\sqrt{2\pi}\, 2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right) \sqrt{n}} \left[\int_0^\infty exp\left\{ -\frac{u}{2}\left(1 + \frac{t^2}{n}\right) \right\} u^{(n-1)/2} du \right]$$

$$= \frac{1}{\sqrt{2\pi} 2^{\frac{n}{2}} \Gamma\left(\frac{1}{2}\right)} \cdot \frac{1}{\left[\frac{1}{2}\left(1 + \frac{t^2}{n}\right)\right]^{(n+1)/2}}$$

$$= \frac{\Gamma(n+1)/2}{\sqrt{n}\Gamma\left(\frac{n}{2}\right)\Gamma(1/2)} \cdot \frac{1}{\left(1 + \frac{t^2}{n}\right)^{(n+1)/2}}, -\infty < t < \infty$$

$$= \frac{1}{\sqrt{n} B\left(\frac{1}{2}, \frac{n}{2}\right)\left(1 + \frac{t^2}{n}\right)^{(n+1)/2}}, -\infty < t < \infty$$

Which is the probability density function of student's t-distribution with n d.f.

Note 1: In Fisher's 't' the d.f. is the same as the d.f. of chi-square variate.

Theorem: Students's 't' is the particular case of Fisher's 't'.

Proof: Since $\bar{x} \sim N\,(\mu, \sigma^2/n), \xi = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$ $\quad and\ x^2 = \frac{ns^2}{\sigma^2} = \sum_{i=1}^{n} \frac{(x_i-\bar{x})^2}{\sigma^2}$ is independently distributed as chi-square variate with (n-1) d.f. Hence Fishers's is given by:

$$t = \frac{\xi}{\sqrt{x^2/(n-1)}} = \frac{\sqrt{n}(\bar{x}-\mu)}{\sigma} \cdot \frac{\sigma}{\sqrt{\sum(x_i - \bar{x})^2/(n-1)}} = \frac{\sqrt{n}(\bar{x}-\mu)}{s} = \frac{\bar{x}-\mu}{s/\sqrt{n}}$$

Which is Student's t-distribution with (n-1) d.f. Hence Students 't' is the particular case of Fisher's 't'.

10.4.3 Probability Curve of t Distribution

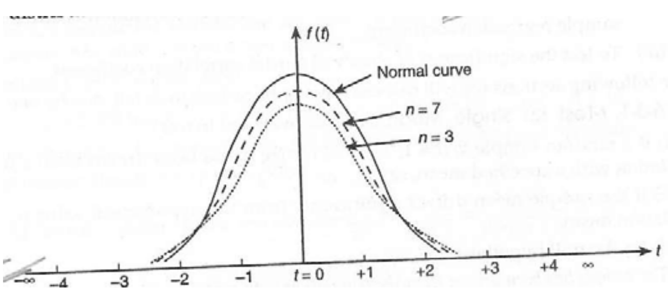


Figure: Probability curve of t distribution

The p.d.f. of t - distribution with n d.f. is:

$$f(t) = C.\left(1+\frac{t^2}{n}\right)^{-(n+1)/2} \quad ,-\infty < t < \infty$$

Since f (-t) = f(t) the probability curve is symmetrical about the line t = 0. As t increases, f(t) decreases rapidly and tends to zero at $t \to \infty$ so that t-axis is an asymptote to the curve. as

$$u_2 = \frac{n}{n-2}, n > 2;\ \beta_2 = \frac{3(n-2)}{(n-4)}, n > 4$$

Hence for n > 2, $u_2 > 1$ i.e. the variance of t- distribution is greater than that of standard normal distribution and for n > 4, $\beta_2 > 3$ and thus t- distribution is more flat on the top that the normal curve. In fact for small n, we have

$$P(|t| \geq t_0) \geq P(|Z| \geq t_0), \qquad Z \sim N(0,1)$$

i.e., the tails of the t - distribution have a greater probability (area) than the tail of standard normal distribution. ***Moreover it has been seen, that for large n ≥ 30, t - distribution tends to standard normal distribution.***

Note: **1.** Mean and Variance of t distribution is 0 and $\frac{n}{n-2}$

2. The MGF of t distribution does not exist.

10.4.4 Application of t- Distribution

The t - distribution has a large number of applications in Statistic, some of which are given below:

(i) To test the sample mean ($\bar{x}$) differs significantly from the hypothesis value μ of the population mean.
(ii) To test the significance of the difference between two samples means.
(iii) To test the significant of an observed sample correlation coefficient sample regression coefficient.
(iv) To test the significance of observed partial correlation coefficient.

10.5 F- Distribution

If X and Y are two independent chi-square variate with v_1 and v_2 d.f. respectively then F-statistic is defined as

$$F = \frac{X/v_1}{Y/v_2}$$

means, F is defined as the ratio of two independent chi-square variates divided by their respective degrees of freedom and it follows Snedector's F- distribution with (v_1, v_2) d.f. with probability function given by:

$$g_1(F) = \frac{(v_1/v_2)^{(v_1/2)}}{B\left(\frac{v_1}{2}, \frac{v_2}{2}\right)} \cdot \frac{F^{\left(\frac{v_1}{2}\right)-1}}{\left(1 + \frac{v_1}{v_2}F\right)^{(v_1+v_2)/2}}, 0 \le F < \infty$$

Note: **1.** The sampling distribution of F-statistic does not involve any population parameters and depends only on the degrees of freedom v_1 and v_2.

2. A statistic F following Snedector's F distribution with (v_1, v_2) d.f. will denoted by F ~ F (v_1, v_2).

Derivation of Snedecor's F-distribution: Since X and Y are independent chi-square variates with v_1 and v_2 d.f. respectively, their joint probability density function is given by:

$$F(,y) = \left\{\frac{1}{2^{(v_1+v_2)}\Gamma\left(\frac{v_1}{2}\right)} \exp\left(-\frac{x}{2}\right) x^{(v_1+v_2)-1}\right\} \times \left\{\frac{1}{2^{(v_1+v_2)}\Gamma\left(\frac{v_2}{2}\right)} \exp\left(-\frac{y}{2}\right) y^{(v_1+v_2)-1}\right\}$$

$$= \frac{1}{2^{(v_1+v_2)/2}\Gamma\left(\frac{v_1}{2}\right)\Gamma\left(\frac{v_2}{2}\right)} exp\{-(x+y)/2\} \times x^{(v_1+v_2)-1} y^{(v_1+v_2)-1}, 0 \le (x-y) < \infty$$

make the following transformation of variables:

$$F = \frac{\frac{X}{v_1}}{\frac{Y}{v_2}} \text{ and } u = y, \text{so that } 0 \le F < \infty, 0 < u < \infty \quad \therefore x = \frac{v_1}{v_2}Fu \text{ and } y = u$$

$$Jacobian\ of\ transformation\ l\ is\ given\ by\ l = \frac{\partial(x,y)}{\partial(F,u)} = \begin{vmatrix} \frac{v_1}{v_2}u = 0 \\ \frac{v_1}{v_2}F = 1 \end{vmatrix} = \frac{v_1 u}{v_2}$$

Thus the joint p.d.f. of the transformed variable is:

$$g(F,u) = \frac{1}{2^{(v_1+v_2)/2}\Gamma\left(\frac{v_1}{2}\right)\Gamma\left(\frac{v_2}{2}\right)} exp\left\{-\frac{u}{2}\left(1 + \frac{v_1}{v_2}F\right)\right\}\left(\frac{v_1}{v_2}Fu\right)^{(v_1/2)-1} u^{(v_2/2)-1}$$

$$= \frac{\left(\frac{v_1}{v_2}\right)^{(v_1/2)}}{2^{(v_1+v_2)/2}\Gamma\left(\frac{v_1}{2}\right)\Gamma\left(\frac{v_2}{2}\right)} exp\left\{-\frac{u}{2}\left(1 + \frac{v_1}{v_2}F\right)\right\} \times \left(\frac{v_1}{v_2}Fu\right)^{(v_1/2)-1} u^{(v_2/2)-1}|J|$$

$$= \frac{\left(\frac{v_1}{v_2}\right)^{\left(\frac{v_1}{2}\right)}}{2^{\frac{(v_1+v_2)}{2}}\Gamma\left(\frac{v_1}{2}\right)\Gamma\left(\frac{v_2}{2}\right)} exp\left\{-\frac{u}{2}\left(1+\frac{v_1}{v_2}F\right)\right\} \times u^{\left\{\frac{(v_1+v_2)}{2}\right\}-1} F^{\left(\frac{v_1}{2}\right)-1}$$

$$; 0 < u < \infty, 0 \leq F < \infty$$

Integrating w.r.t μ over the range 0 to ∞ , the p.d.f of F becomes:

$$g\ \ (F) = \frac{(v_1/v_2)^{(v_1/2)} F^{\left(\frac{v_1}{2}\right)-1}}{2^{(v_1+v_2)}\Gamma(v_2/2)} \times \left[\int_0^\infty exp\left\{-\frac{\mu}{2}\left(1+\frac{v_1}{v_2}F\right)\right\}\right] \mu^{(v_1+v_2)/2}$$

$$= \frac{(v_1/v_2)^{(v_1/2)} F^{\left(\frac{v_1}{2}\right)-1}}{2^{(v_1+v_2)}\Gamma(v_2/2)} \times \frac{\Gamma[(v_1+v_2)/2]}{\left[\frac{1}{2}\left(1+\frac{v_1}{v_2}F\right)\right]^{(v_1+v_2)/2}}$$

$$g\ \ (F) = \frac{(v_1/v_2)^{(v_1/2)}}{B\left(\frac{v_1}{2},\frac{v_2}{2}\right)} . \frac{F^{\left(\frac{v_1}{2}\right)-1}}{\left(1+\frac{v_1}{v_2}F\right)^{(v_1+v_2)/2}}, 0 \leq F < \infty$$

Which is the probability density function of F-Distribution with (v_1, v_2)d.f.

10.5.1 Mean, Mode, Variance and skewness of the F Distribution

1. Mean = v_1, which is always greater than unity.
2. Variance = $\frac{2v_2^2(v_2+v_1-2)}{v_1(v_2-2)^2(v_2-4)}$
3. Mode = $\left(\frac{v_2}{(v_2+2)}\right)\left(\frac{(v_1-2)}{v_1}\right)$, which is always less than unity.
4. Since mean is always less than unity and mode is always less than unity than, the curve of F distribution is highly positively skewed.

10.5.2 Relation between t and F- Distribution

In F-distribution with (v_1,v_2) d.f., take v_1= 1, v_2= v and t^2 = F, i.e., dF= 2t dt. Thus the probability differential of F- transforms to:

$$dG(t) = \frac{\left(\frac{1}{v}\right)^{\frac{1}{2}}}{B\left(\frac{1}{2},\frac{v}{2}\right)} . \frac{(t^2)^{\frac{1}{2}-1}}{\left(1+\frac{t^2}{v}\right)^{\frac{(v+1)}{2}}} 2t\ dt, 0 \leq t^2 < \infty$$

$$= \frac{1}{\sqrt{v}B\left(\frac{1}{2},\frac{v}{2}\right)} \cdot \frac{1}{\left(1+\frac{t^2}{v}\right)^{\frac{(v+1)}{2}}} dt, -\infty < t < \infty$$

since the total probability in the range $(-\infty, \infty)$ is unity, the probability function of Students's t-distribution with v d.f.

If a statistic t follows Student's distribution with n d.f. then t^2 follows snedecor's F-distribution with (1,n) d.f. mathematically,

$$\left.\begin{array}{r} if \quad t \sim t_{(n)} \\ then \;\; t^2 \sim F_{(1,n)} \end{array}\right\}$$

10.5.3 Relation between F and χ^2 Distribution

In F (n_1, n_2) distribution if let $n_2 \to \infty$ then $\chi^2 = n_1$ F follows χ^2 distribution with n_1 d.f.

Proof. It is given

$$f(F) = \frac{(n_1/n_2)^{n_1/2} F^{(n_1/2)-1}}{\Gamma(n_1/2)\Gamma(n_2/2)} \to \frac{\Gamma[(n_1+n_2)/2]}{\left(1+\frac{n_1}{n_2}F\right)^{(n_1+n_2)/2}}, 0 < F < \infty$$

In the limit as $n_2 \to \infty$, we have

$$\frac{\Gamma[(n_1+n_2)/2]}{{n_2}^{n_1/2}\Gamma(n_2/2)} \to \frac{(n_2/2)^{n_1/2}}{{n_2}^{n_1/2}} = \frac{1}{2^{n_1/2}}$$

$$\left[\because \frac{\Gamma(n+k)}{\Gamma(n)} \to n^k as\; n \to \infty\right]$$

$$\lim_{n_2\to\infty}\left(1+\frac{n_1 F}{n_2}\right)^{(n_1+n_2)/2} = \lim_{n_2\to\infty}\left[\left(1+\frac{n_1 F}{n_2}\right)^{n_2}\right]^{1/2} \times \lim_{n_2\to\infty}\left(1+\frac{n_1 F}{n_2}\right)^{n_1/2}$$

$$= \exp\left(\frac{n_1 F}{2}\right) = \exp\left(\frac{\chi^2}{2}\right) \qquad (\because n_1 F = \chi^2)$$

Hence in the limit, on using both equations the p.d.f of $\chi^2 = n_1 F$ becomes:

$$dP\chi^2 = \frac{\left(\frac{n_1}{2}\right)^{\frac{n_1}{2}} e^{-\frac{\chi^2}{2}}}{\Gamma\left(\frac{n_1}{2}\right)} \cdot \left(\frac{\chi^2}{n_1}\right)^{\left(\frac{n_1}{2}\right)-1} d\left(\frac{\chi^2}{n_1}\right)$$

$$= \frac{1}{2^{\frac{n_1}{2}}\Gamma\left(\frac{n_2}{2}\right)} \cdot e^{-\frac{\chi^2}{2}} (\chi^2)^{(n_1/2)-1} d\chi^2, 0 < \chi^2 < \infty$$

Which is the required p.d.f. of chi-square distribution with n_1 d.f.

10.5.4 Application of F Distribution

The F distribution has a large number of applications in Statistic, some of which are given below:

(i) To test the equality of two population variance.
(ii) To test the significance of the observed multiple correlation coefficient.
(iii) To test the significance of the observed sample correlation coefficient
(iv) To test the linearity of regression.

10.6 z Distribution

In F distribution if $F = \exp(2z) \cong Z = \frac{1}{2}\log F$

Than the distribution of Z becomes

$$g(Z) = p(F).\left|\frac{dF}{dZ}\right| = \frac{(v_1/v_2)^{(v_1/2)}}{B\left(\frac{v_1}{2},\frac{v_2}{2}\right)}.\frac{(e^{2x})^{(v_1/2)-1}2e^{2:}}{\left[1+\frac{v_1}{v_2}e^{2x}\right]^{(v_1+v_2)/2}}$$

$$= 2\frac{(v_1/v_2)^{(v_1/2)}}{B\left(\frac{v_1}{2},\frac{v_2}{2}\right)}.\frac{e^{v_1 2}}{\left[1+\frac{v_1}{v_2}e^{2x}\right]^{(v_1+v_2)/2}}; -\infty < Z < \infty$$

Which is the probability function of Fisher's z-distribution with (v_1,v_2) d.f.,

10.6.1 Moment Generating Function of z-distribution

$$M_Z(t) = E(e^{tz}) = \int_{-\infty}^{\infty} e^{tz}(gz)dz = \int_{0}^{\infty} F^{\frac{1}{2}}f(F)df \qquad [\therefore e^2 = F]$$

Since u_r' (about origin) for F-distribution is $\int_0^\infty F^r f(F)dF$, we can find m.g.f of the z-distribution by putting r= t/2 in the expression for u_r' for-distribution.

$$M_Z(t) = \left(\frac{v_1}{v_2}\right)^{t/2}.\frac{\Gamma\{(v_1+t)/2\}\Gamma\{(v_2-t)/2\}}{\Gamma(v_1/2)\Gamma(v_2/2)}$$

$$mean = \frac{1}{2}\left(\frac{1}{v_2}-\frac{1}{v_1}\right),$$

$$variance = \frac{1}{2}\left(\frac{1}{v_2} + \frac{1}{v_1} + \frac{1}{v_1^2} + \frac{1}{v_2^2}\right)$$

Note: as large v_1 *and* v_2, the Z distribution tends to be normal.

10.6.2 Applications of z distribution

The z distribution has a large number of applications in Statistic, some of which are given below:

(i) To test the significance of hypothetical population correlation coefficient.

(ii) To test the significance of the difference between two independent sample correlation coefficients.

10.8 Summary

This unit provides a brief idea about chi square, t, F and z distribution with their applications and importance. In detail the applications of all distributions will be studied in the next block.

10.9 Further Readings

1. Cramer H, Mathematical Methods of Statistics, Princeton University Press, 1946 and Asia Publishing House, 1962.
2. Hogg R.V. and Craig A.T., Introduction to Mathematical Statistics, Macmillan, 1978.
3. Prazen E. Modern Probability Theory and its Applications, John Wiley, 1960 and Wiley Eastern 1972.
4. Rao C.R., Linear Statistical Inference and its Applications John Wiley, 1960 and Wiley Eastern 1974.
5. Rohtagi V.K. (1984), An Introduction to Probability Theory and Mathematical Statistics John Wiley, 1976 and Wiley Eastern 1985.
6. Vikas S.S., Mathematical Statistics, John Wiley, 1962 Toppan.
7. Gupta S.C.and V. K. Kapoor, Fundamentals of Mathematical Statistics, Sultan Chand and Sons.

U.P. Rajarshi Tandon Open University, Prayagraj

UGSTAT – 102

Probability, Distribution and Statistical Inference

Block

4

Basic Principles of Statistical Inference

Unit - 11
Estimation

Unit - 12
Methods of Estimations

Unit - 13
Testing of Hypothesis

Course Design Committee

Dr. Ashutosh Gupta **Chairman**
Director, School of Sciences
U. P. Rajarshi Tandon Open University, Prayagraj

Prof. Anup Chaturvedi **Member**
Ex. Head, Department of Statistics
University of Allahabad, Prayagraj

Prof. S. Lalitha **Member**
Head, Department of Statistics
University of Allahabad, Prayagraj

Prof. Himanshu Pandey **Member**
Department of Statistics, D. D. U. Gorakhpur University,
Gorakhpur

Dr. Shruti **Member-Secretary**
Sr. Assistant Professor, School of Sciences
U.P. Rajarshi Tandon Open University, Prayagraj

Course Preparation Committee

Dr. Shruti **Course Coordinator**
Sr. Assistant Professor, School of Sciences
U. P. Rajarshi Tandon Open University, Prayagraj

Prof. B. N. Pandey **Writer**
Department of Statistics, Banaras Hindu University, Varanasi

Prof. Ram Ji Tiwari **Writer**
Department of Statistics, University of Jammu, Jammu

Dr. Sanjay Singh **Writer**
Department of Statistics, Banaras Hindu University, Varanasi

Prof. V. P. Ojha **Reviewer**
Department of Statistics and Mathematics,
D. D. U., Gorakhpur University, Gorakhpur

Prof. S. K. Upadhyay **Reviewer**
Department of Statistics, Banaras Hindu University, Varanasi

Dr. Shruti **Reviewer**
School of Sciences, U.P. Rajarshi Tandon Open University,
Prayagraj

Prof. V. P. Ojha **Editor**
Department of Statistics and Mathematics, D. D. U.,
Gorakhpur University, Gorakhpur

First Edition: *March 2008* (Published with the support of the Distance Education Council, New Delhi)

Second Edition: ***July 2020***

ISBN :

Printed and Published by Dr. Arun Kumar Gupta Registrar, Uttar Pradesh Rajarshi Tandon Open University, 2020.

Printed By: Chandrakala Universal Pvt. 42/7 Jawahar Lal Neharu Road, Prayagraj.

Unit-11: Estimation

Structure

11.1 Introduction

The investigator may be interested in the study of the behavior of some of the characteristics of the elements of a population. A collection of objects (some elements or individuals) under a statistical enquiry, such that each element can be characterized by one more characteristics, is called a population. Thus if each element of a population is characterized by K-characteristics, then they constitute a K- variate population. For K = 1, we get a Uni- variate population. A characteristic describing the behavior of the elements is represent by a random variable say X, Y, Z, …. A characteristic may be discrete or continuous according as the corresponding random variable (r.v.) is discrete or continuous. Similarly, a population is discrete or continuous, if its elements are associated with a discrete or continuous r.v. In most of the surveys an the experiments it is not possible to take observations on each elements of the population, that is complete enumeration is not possible, but observations can be taken on a part or subset of population called a sample. A sample consists of a small collection of elements drawn from a larger aggregate (population) of elements, about which we wish we wish to take information. Obviously, it is the sample which we observe but it is the population about which we went to study. The reason for studying a sample may be the shortage of resources, such as budget, manpower and time or on the nature of the experiment. If the population is infinite (viz, number of cement bags, produced in a previous year, number of candidates applying for admission in IIT, Management Courses and medical colleges taken together voters in General Election in India for opinion survey etc.) hypothetical (tossing of coins of die etc.) or destructive (production of bulbs, electronic goods, etc.) then, too, complete enumeration cannot be undertaken. We have to resort to sampling in such cases. The sampled data is analyzed and the conditions are drawn about the population.

Statistical inference is a process of drawing conclusions about the population characteristic on the basis of available information in the sample obtain from the population. Suppose the population characteristic X follows normal distribution N (μ , σ^2), where both parameters and are unknown. We shall write it as follows: "X ~ N (μ , σ^2), both μ *and* σ^2 unknown". We cannot proceed further unless we have some estimates $\hat{\mu}$ *and* $\hat{\sigma}^2$ of μ *and* σ^2 respectively. For this a random sample X_1, X_2,…..X_n of size n is drawn from the population. Let $X_1 = x_1$, $X_2 = x_2$,………$X_n = x_n$ be the observed values of the sample. Since in normal population mean, median and mode are at μ *and* σ^2 is population variance. Therefore, either sample estimate mean $\bar{x} = \frac{1}{n}\sum_{i=1} x_1$ or median $\tilde{x}$ may be taken as an estimate of μ where as sample variance $S^2 = \frac{1}{n-1}\sum_{i=1}^{"}(x_i - \bar{x})^2$ a an estimate of population variance. In this way the population characteristic (parameters μ *and* σ^2) are estimated by the available information/ conclusions (sample mean $\bar{X}$ and the sample variance S^2) provided by the sample drawn from the population. We have obtained a point estimates $\bar{x}$ of unknown μ *and* S^2 for σ^2 respectively. Of course, one have to decide between $\bar{x}$ and $\tilde{x}$ for choosing an estimate μ of we shall in this unit discuss some of the criteria for choosing between contending estimators of a parameters say θ.

We may choose a random interval

$$\left(\bar{x} - 1.96\sqrt{\frac{S^2}{n}},\ \bar{x} + 1.96\sqrt{\frac{S^2}{n}},\right)$$

as a possible interval estimator of which includes the actual mean μ in 95% cases is repeated sampling if sample size n is large enough. This type of estimation is known as *confidence interval estimation*. Here, too, the sample information is used to gather the information about the population (parameter). A third type of problem of inference is on "test of hypothesis", a subject matter of Unit III. An electric light bulb manufacturer may claim that the average length of life of manufactured bulb is more than 2 years. To test this claim a sample of bulb of size say n =20 is taken and their mean life is computed. This $\bar{x}$ is used to test whether the hypothesis is true or false. In this case the sample the sample observation is used to justify the claim of the company. Thus the problems of statistical inference are classified into "problems of Estimation" and "test of hypothesis." The "estimation" can be done in either of the two ways: (1) "Point estimation" and (2) Interval Estimation."

11.2 Objectives

After going through this unit you will be able to understand:

- The role of statistical inference in statistics.
- Consisting property of estimators
- Unbiasedness property of estimators
- Sufficiency of estimators

- How to compare among two estimators
- How to construct interval estimate of a parameter.

11.3 Point Estimation

The theory of estimation was first founded by Prof. R.A. Fisher through a series of papers around 1930.

Some definitions are given below to prepare a background to understand the problem of point estimation.

Let us suppose that our interest lies in a characteristic X of the population. Let X be random variable having a probability density function (p.d.f.) $(x;\ \theta l, \theta \in \theta.$. Suppose further that the functional form f (x; θ) is known except for a finite number of parameters out of $\theta = (\theta_1, \theta_{2,} \ldots\ldots \theta_k)$, we shall write

Definition 1.1

The set of all admissible values of the parameters θ of a distribution f(x,θ) is called the parametric space θ of θ.

If $\chi{\sim}N(\mu\ , \sigma^2)$ both $\mu\ and\ \sigma^2$ be unknown then parametric space is Q: $\{(\mu\ , \sigma^2)\colon -\infty < \mu < \infty, \sigma^2 > 0\}$

For the different combinations of $= (\theta_1, \theta_{2,} \ldots\ldots \theta_k)$, we get different distributions having the form of p.d.f (x; θ) and as such get a family of distribution. For example $\{N(\mu\ , \sigma^2), (\mu\ , \sigma^2)\colon -\infty < \mu < \infty, \sigma^2 > 0\}$ is a family of normal distributions.

Definition 1.2 Random Sample: A set of n random variables $X_1, X_2, \ldots X_n$ are said to be a random sample of size n from the population f (x; θ), if their joint distribution f ($X_1, X_2, \ldots X_n$; $X_1, X_2, \ldots X_n$) can be factorise as

f $X_1, X_2, \ldots X_n$ ($X_1, X_2, \ldots X_n$) = $f(X_1)\ f(X_2) \ldots\ldots, f(X_n)$

or

$x_1, x_2, \ldots\ldots\ x_n$ are independent identically distribution random variables having common p.d.f. f (x; θ),

If $X_1 = x_1$, $X_2 = x_2$, $X_3 = x_3, \ldots\ldots X_n = x_n$ then ($X_1, X_2, \ldots X_n$) is an observed sample.

Definition 1.4 ***Statistic*** Any function T= T($X_1, X_2, \ldots X_n$) of a random sample $X_1, X_2, \ldots X_n$ which is independent of the unknown parameters is called a statistic.

Clearly a statistic T is a random variable having its distribution function say g (t, θ): although statistic is independent of any unknown parameters, its distribution may dependent upon unknown parameters.

Definition 1.5 ***Sampling distribution of a statistic.*** Let T= T(X_1, X_2,…..X_n) be a statistic where X_1, X_2,…..X_n is a random variable from the popular $X \sim f(x,\theta)\theta \in \sim$. The probability distribution is known as the sampling g_T(t, θ) distribution of Statistic T.

If T is continuous r.v. then

$$E(T) = \int_{-\infty}^{\infty} g_t(t,\theta)dt.$$

and

$$Var(T) = E[TE(T)^2] = \int_{-\infty}^{\infty} [t - E(T)^2]\, g_t(t,\theta)dt.$$

An important Result (without proof):

If X_1, X_2,…..X_n be a random sample of size n from the normal population $N(\mu\,,\sigma^2)$ both μ *and* σ^2 be unknown, then

(i) Sample mean $\bar{x} = \frac{1}{n}\sum_{i=1} x_1$ and sample variance $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ are independent.

(ii) $\bar{X} \sim N(\mu\,,\sigma^2/n)$

(iii) $\frac{(n-1)S^2}{\sigma^2} = \sum_{i=1}^{n}(x_i - \bar{x})^2 \sim \chi^2(n-1)$

Where X stands for χ^2 distribution with (n-1) d.f.

This result gives sampling distribution of sample mean X and sample variance S^2. Here

$$E(\bar{X}) = \mu \text{ and } Var\,(\bar{X}) = \frac{\sigma^2}{n}$$

And

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

So that

$$E(S^2) = \sigma^2 \text{ and } var\,(S^2) = \frac{2\sigma^2}{n-1}$$

Definition 1.6 ***Standard Error of statistic*** The Standard deviation, calculated from the sampling distribution if a statistics is called its standard error (S.E.), that is,

$$SE(T) = \sqrt{Var\,(T)}$$

So that for sampling from normal population $N(\mu\,,\sigma^2)$

$$SE(\bar{X}) = \sqrt{Var\,(\bar{X})} = \frac{\sigma}{\sqrt{n}}$$

And

$$SE(S^2) = \sqrt{var\ (S^2)} = \sigma^2\ = \sqrt{\frac{2}{n-1}}$$

Definition 1.7 ***Estimator and Estimate***

Any Statistic T= T(X_1, X_2,…..X_n) of a random sample X_1, X_2,…..X_n is a random variable from the population $f(x,\theta)$, which is used to estimate unknown parameter θ of the population is called are estimator of θ whereas the value of estimator T is an estimate of θ.

The estimate T(X_1, X_2,…..X_n) of θ is a point estimate of θ.

There may be more than one point estimators of a parameter. For example in normal population $N(\theta\ ,\sigma^2)$ with known σ^2 sample mean$\bar{X} = \frac{1}{n}\sum_{i=1} x_1$

Sample mean $\tilde{X}$ mid range $T = \frac{X_{(1)}+X_{(n)}}{2}$ are the estimator of mean μ .

Where $X_{(1)}$ = min (X_1, X_2,…..X_n) and max (X_1, X_2,…..X_n)

Thus the problem of estimation may be stated as follows;

Let the characteristic X having density function f (x; θ) , $\theta = (\theta_1, \theta_{2,} \dots . . \theta_k)$, $\in \theta$ of the population be under investigation. Suppose that the functional form f (x; θ) be either known or can be determined from the conditions of the experiment but some of (1≤m≤k) or all the parameters $\theta_1, \theta_{2,} \dots . . \theta_k$ may not be known.

The density f (x; $\theta_1, \theta_{2,} \dots . . \theta_k$function is said be completely specified if all the parameters $\theta_1, \theta_{2,} \dots . . \theta_k$ are known. In this case there is no need of estimation of parameter(s). For example, normal distribution N (3,4) is completely specified.

The problem of estimation arises if some or all of the parameters are unknown

- Draw a random sample X_1, X_2,…..X_n of size n form the population.
- Choose m (one- dimensional) statistics T_1= $\hat{\theta}_1$(X_1, X_2,…..X_n), T_2= $\hat{\theta}_2$(X_1, X_2,…..X_n)…… T_m= $\hat{\theta}_1$(X_1, X_2,…..X_n),

As point estimators of unknown parameters $\theta_1, \theta_{2,} \dots . . \theta_m$ where 1≤m≤k,

Such that their distributions are concerted as closely as possible about the true values of the parameters $\theta_1, \theta_{2,} \dots . . \theta_m$

By putting x_1, x_2,…..x_n in place of X_1, X_2,…..X_n an estimate of is obtained as T_1= $\hat{\theta}_1$(x_1, x_2,…..x_n) for i= 1,2,……m.

11.4 Properties of Good Estimator

Some of the criteria for a good estimators are

(i) Unbiasedness
(ii) Consistency
(iii) Sufficiency
(iv) Efficiency

The last three are known as Fisher's criteria of a good estimator.

11.5 Consistency

Definition 1.8

Let X_1, X_2,…..X_n be a sequence of independent identically distributed (i.i.d.) random variables r.v.'s with common p.d.f. function f (x, θ), $\theta \in \theta$, A sequence of point estimators T_1, T_2,…..T_n will be called consistent for θ,if

$$T_n \xrightarrow{p} \theta \text{ as } n \to \infty \text{ for each fixed } \theta \in \theta \qquad (i)$$

Or

If for every $\in > 0$.

$$P(|T_n - \theta| > \in) \to o \text{ as } n \to \infty \qquad \ldots\ldots\ldots\ldots\ldots..(ii)$$

Or

$$\lim_{n\to\infty} P(|T_n - \theta| < \in) \to 1 \text{ for every } \in > 0 \ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(iii)$$

Or

$$\lim_{n\to\infty} P(|\theta - \in < T_n| < \theta + \in) \to 1 \text{ for every } \in > 0 \ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(iv)$$

Where estimator

$$T_n = \hat{\theta}\,(X_1, X_2, X_3, \ldots\ldots..X_n)$$

Is a function of sample observations (or a random sample of size n) X_1, X_2,…..X_n, The expressions (i), (ii), (iii) and (iv) are equivalent, Since they have same meanings.

It is pointed out that an estimate of parameter θ obtained on the basis of a random sample is often different from true value of θ. A consistent estimator provides an estimate of θ that lies close to the true value of θ with probability one as the sample size increase indefinitely.

Consistency is limiting property of an estimator, since it explains the behavior of an estimator T_n, based on n independent observations, as n increases indefinitely.

If there exists a consistent estimator T_n of θ then infinitely many consistent estimators can be constructed for example.

$$T_n = a$$

Are all consistent estimator of θ. Hence consistent estimators are not unique.

11.5.1 Sufficient conditions for consistency

Theorem- (without proof): Let $\{T_n\}$ be a sequence of consistent estimators of θ such that for $\theta \in \theta$,

$$(i) \qquad E_\theta(T_n) \to \theta$$

$$and\ (ii) \quad Var_\theta(T_n) \to 0\ as\ n \to \infty$$

Then $T_n = \hat{\theta}_n(x_1, x_2, \ldots.. x_n)$ is consistent for θ.

Example 1.1

Let be a sequence of i.i.d. normal $N(\mu, \sigma^2)$ variables. Then sample mean $\bar{x} = \frac{1}{n}\sum_{i=1} x_i$ is consistent estimator of μ.

Solution:

Here, $X_i \sim N(\mu, \sigma^2)$

Therefore,

$$E(X_i) = \mu\ and\ var.(X_i) = \sigma^2\ for\ all\ i = 1,2,3 \ldots$$

$$E(\bar{X}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n}(X_i) = \mu$$

And

$$Var(\bar{X}) = var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} var(X_i) = \frac{1}{n^2}\sum_{i=1}^{n} \sigma^2 = \frac{\sigma^2}{n}$$

Obviously,

$$E(\bar{X}) = \mu\ Var(\bar{X}) = \frac{\sigma^2}{n} \to 0 as\ n \to \infty.$$

Hence, sample mean $\bar{X}$ is a consistent estimator for μ.

Example 1.2

If $X_1, X_2, \ldots. X_n$ be a random observations of size n on Bernoulli Variable X taking the value 1 with probability p and the value 0 with the probability (1-p), show that

$$\frac{\sum X_i}{n}\left(1-\frac{\sum X_i}{n}\right) = \bar{X}(1-\bar{X})$$

Is a consistent estimator or p (1-p).

Solution:

Since $X_1, X_2, \ldots. X_n$ are i.i.d. Bernoulli variable with parameters p, therefore

$$\sum_{i=1}^{n} X_i \sim Binomial\ B\ (n, p)$$

So that

$$E\left(\sum_{i=1}^{n} X_i\right) = np$$

$$E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = p = E(\bar{X}),$$

Whereas

$$Var.\left(\sum_{i=1}^{n} X_i\right) = np(1-p)$$

Or

$$Var.\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}$$

Or

$$Var.(\bar{X}) = \frac{p(1-p)}{n} \to 0 as\ n \to \infty$$

Therefore

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

is a consistent estimator for p. Hence by the invariance property of consistent is a consistent estimator $\bar{X}(1-\bar{X})$ for p(1-p).

Example 1.3:

Show that the sample mean $\bar{X}$ is a consistent estimator of the population mean even for non-normal population.

Solution:

Here parent population is not mentioned. By chebyshev's inequality, we have for ∈ > 0.

$$P(|\bar{X}-\theta| < \epsilon) \geq 1 - \frac{var(\bar{X})}{\in^2} = 1 - \frac{\sigma^2}{n}$$

Where

$$\sigma^2 = var\ (X_i) for\ all\ i = 1,2,3, \ldots.$$

Since

$$E(\bar{X}) = \left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\sum_{i=1}^{n} \theta = \theta$$

And

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

Therefore, as n→ ∞

$$P(|\bar{X}-\theta| < \epsilon) \to 1$$

Thus, $\bar{X} \xrightarrow{P} \theta$

Hence the result.

Alternative Proof;

By central limit theorem,

$$Z = \lim_{n\to\infty} \frac{\bar{X}-\mu}{\sigma} \sim N(0,1)$$

Is true whatever to be form of the parent population where μ *and* σ^2 are population mean and variance respectively, Now for $\epsilon > \theta$,

$$\lim_{n\to\infty} P(|\bar{X} - \theta| < \epsilon) \Rightarrow \lim_{n\to\infty} P\left(|Z| < \frac{\in \sqrt{n}}{\sigma}\right)$$

$$\Rightarrow \lim_{n\to\infty} P\left(-\frac{\in\sqrt{n}}{\sigma} < Z < \frac{\in\sqrt{n}}{\sigma}\right)$$

$$\Rightarrow \lim_{n\to\infty} \int_{\frac{\in\sqrt{n}}{\sigma}}^{\frac{\in\sqrt{n}}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}^{Z^2}}\, dz$$

Hence sample mean $\bar{X}$ is a consistent of μ. Proved

Example 1.4:

Let X_1, X_2,.....X_n be a sequence of i.i.d. r.v.'s distributed as normal $N(\mu, \sigma^2)$, then show that the sample median $\tilde{X}_n$ is a consistent estimator of μ.

Or

Prove that the sample median $\tilde{X}_n$ is a consistent estimator of the of a normal population μ.

Solution:

For large samples the sample median $\tilde{X}$ is distributed normally with mean μ and variance $\frac{\pi\sigma^2}{2n}$.

Thus for large n,

$$Z = \frac{\tilde{X} - \mu}{\sqrt{\frac{\pi\sigma^2}{2n}}} \sim N(0,1)$$

Therefore, for $\in > 0$.

$$\lim_{n\to\infty} P(|\tilde{X} - \mu| < \epsilon) \Rightarrow \lim_{n\to\infty} P\left(|Z| < \frac{\in \sqrt{2n}}{\sigma\sqrt{\pi}}\right)$$

$$\Rightarrow \lim_{n\to\infty} \int_{-\frac{\epsilon\sqrt{2n}}{\sigma\sqrt{\pi}}}^{\frac{\epsilon\sqrt{2n}}{\sigma\sqrt{\pi}}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}^{z^2}} dz$$

Hence Proved

1. ***You may attempt the following problem:***

E-1.1 Show that the sample variance is consistent estimator for the population variance of a normal population.

E-1.2 Fill up the suitable word(s) or phrase (s) in the blanks-

(a) If a estimator T_n converges in probability to the parametric function $\gamma(\theta)$. Then T_n is said to be a ……………………estimator of $\gamma|\theta|$.
(b) ………………………estimator may not be true.
(c) The best estimator implies that the distribution of an estimator be …………around the true parameter.
(d) Consistency ensure that the difference between the estimator T_n and parametric function $\gamma(\theta)$ …………………………as n increases.
(e) An estimator T_n which is most concentrated about a parameter θ is the ………………………. estimator.
(f) An estimator is itself a …………..
(g) A value of the estimator is called………………………………

11.6 Unbiasedness

Consistency is a large sample property of an estimator that is it holds when n is sufficiently large.

Definition 1.9 Let $X_1, X_2, \ldots X_n$ be a random sample from a population having p.d.f. f(x, θ). Let $\tau(\theta)$ be some function of θ. An estimator

$\hat{\theta}_n = \hat{\theta}_n (x_1, x_2, \ldots x_n)$ is said to unbiased for $\tau(\theta)$, if

$$\text{E}\left(\hat{\theta}_n\right) \neq \tau(\theta)\ for\ all\ \theta\ then\ \hat{\theta}_n\ issaid\ to\ biased$$

If

$$\text{E}\left(\hat{\theta}_n\right) > \tau(\theta)\ estimator\ is\ said\ to\ positively\ biased$$

and

$$\text{E}\left(\hat{\theta}_n\right)| < \tau(\theta)\ estimator\ \hat{\theta}_n\ is\ said\ negatively\ biased$$

Further, if

$$\lim_{n\to\infty} \text{E}\left(\hat{\theta}_n\right) \neq \tau(\theta)$$

Then the estimator $\hat{\theta}_n$ is said to asymptotically unbiased for (θ) . this property holds for large samples only.

Here,

$$Bias\ (T_n) = \quad \tau(\theta) - E(T_n)$$

It is observed that there is no perfect estimator of $\tau(\theta)$ which always gives an estimate. $\hat{\theta}$ Equal to $\tau(\theta)$, but an unbiased estimator does so on the average since its E $(\hat{\theta})$ mean coincides with $\tau(\theta)$. It means that "if we go on drawing various samples of size n form a population and evaluate the estimates of $\tau(\theta)$ using the proposed estimator $\hat{\theta}$ then the average E$(\hat{\theta})$ of these values of the unbiased estimator $\hat{\theta}$ or the mean of the sampling distribution of $\hat{\theta}$, is equal to the parameter value $\tau(\theta)$.

Unbiasedness of an estimator $\hat{\theta}$ of $\tau(\theta)$dose not mean that the estimate given by $\hat{\theta}$ is close to the correct value of the parameter $\tau(\theta)$. the chance of being close to $\tau(\theta)$ depends not only upon the mean E$(\hat{\theta})$ of the sampling distribution g (t, $\tau(\theta)$) but also upon the mean distribution. We know that the standard deviation is frequently used to measure dispersion. In case of unbiasedness of the statistics $\hat{\theta}$ a small standard error indicates a large chance of the estimate to be close to $\tau(\theta)$.

Example 1.5:

Show that sample mean is unbiased estimate of population mean in binomial population.

Solution:

Let $X_1, X_2, \ldots X_n$ be a random sample from a binomial distribution with parameters n and p. then

$$\left.\begin{array}{r} X_i \sim b(x_i; n, p) \\ and \quad E(x_i) = np \end{array}\right\} for\ i = 1,2, \ldots \ldots . . m$$

So that

$$E(\bar{X}) = \left(\frac{1}{m}\sum_{i=1}^{m} X_i\right) = \frac{1}{m}\sum_{i=1}^{m} E(X_i) = \frac{1}{n}\sum_{i=1}^{m}(np)$$

np= population mean

hence proved

Example 1.6:

If $\hat{\theta}$ is an unbiased estimator of θ then show that $\hat{\theta}^2$ is a biased estimator of θ.

Solution:

If $\hat{\theta}$ is unbiased for θ, then,

$$E(\hat{\theta}) = \theta \; and \quad var\,(\hat{\theta}) > 0$$

But

$$var\,(\hat{\theta}) = E\,[\,\hat{\theta} - E(\hat{\theta})]^2$$

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

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

Since $var\,(\hat{\theta}) > 0, therefore, E(\hat{\theta}^2) > \theta^2$

Which means that $\hat{\theta}^2$ is a positively biased estimator of θ^2.

Proved

Example 1.7

Let $X_1, X_2, \ldots . X_n$ be a random samples of size n from a normal population P N(μ, σ^2), then show that-

$$(i) \qquad S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \quad is\ a\ biased\ estimator\ of\ \sigma^2$$

$$(ii) \qquad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \ is\ unbiased\ for\ \sigma^2$$

Solution:

Here

$$X_i = N\,(\mu, \sigma^2)$$

So that for i= 1,2,….n

$$E\,(X_i) = \mu\ and\ var\,(x_i) = \sigma^2$$

giving $E\,(X_i) = \mu$

So that is an unbiased estimator of mean μ.

and $var\ (\bar{X}) = \sigma^2/n$

but E($\bar{X}^2$) = Var ($\bar{X}$) + $E^2(\bar{X}) = \frac{\sigma^2}{n} + \mu^2$

Therefore,

$$E(S^2) = E\left[\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2\right] = \frac{1}{n}E\left[\sum_{i=1}^{n}\{(X_i - \mu) - (\bar{X} - \mu)\}^2\right]$$

$$= \frac{1}{n}\left[\sum_{i=1}^{n} E(X_i - \mu) - nE(\bar{X} - \mu)^2\right] = \frac{1}{n}\left[\sum_{i=1}^{n}\sigma^2 - n\frac{\sigma^2}{n}\right]$$

$$= \frac{1}{n}[n\sigma^2 - \sigma^2] = \frac{n-1}{n}\sigma^2 \quad \neq \sigma^2$$

Hence, S^2 is not unbiased for σ^2.

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

Shows that S^2 is unbiased for σ^2.

Proved

Alternative Proof: Form result (A), for sampling from normal population N(μ, σ^2)

$$\sum_{i=1}^{n}\frac{(X_i - \bar{X})^2}{\sigma^2} \sim \chi^2 n - 1$$

So that

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2 n - 1$$

There,

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

$$\Rightarrow E(S^2) = \sigma^2$$

That is S^2 is unbiased for σ^2

Since $S^2 = \frac{(n-1)}{n} S^2$

Therefore E(S^2)=$\frac{(n-1)}{n} S^2 = \frac{(n-1)}{n} \sigma^2 = (1 - \frac{1}{n})\sigma^2 \neq 0$

Thus, S^2 is a biased estimator of σ^2

Hence proved.

You may attempt the following problems.

E-1.3 Show that the sample $\bar{X}$ mean is unbiased for estimate for estimate for population means λ is passion distribution.

E-1.4 Let $X_1, X_2, \ldots X_n$ be a random sample from a normal population with unknown mean μ and standard deviation 1. Show that.

$$T = \frac{1}{n}\sum_{i=1}^{n} X_i^2 \ is\ unbiased\ estimator\ of\ \mu^2 + 1.$$

E-1.5 Let T= T($X_1, X_2, \ldots X_n$) is an unbiased estimator of θ . Show that $\sqrt{T}$ is not unbiased for $\sqrt{\theta}$.

E-1.6 (i) If θ is a consistent estimator of $\theta_1 then\ \hat{\theta}^2$

(a) Would also be (b) would not be

a consistent estimator of θ^2

E-1.7 (i) If $\hat{\theta}$ is a consistent estimator of $\theta\ then\ \hat{\theta}^2$

(a) Would also be (b) would not be

a unbiased estimator of θ^2

11.7 Efficiency

Consider the sample from a normal population N (μ, σ^2), where is known. Then sample mean $\bar{X}$ and sample median $\tilde{x}$ are two unbiased and consistent estimators of the population mean μ, since

$$E\ (\bar{x}) = \mu,\ \ var\ (\bar{x}) = \frac{\sigma^2}{n}\ and \lim_{n\to\infty} var\ (\bar{x}) \to 0$$

$$E\ (\tilde{X}) = \mu,\ \ var\ (\tilde{x}) = \frac{\pi\sigma^2}{2n}\ and \lim_{n\to\infty} var\ (\tilde{x}) \to 0$$

This leads us to conclude that a need of some further criterion to be satisfied by the contending estimators. of course, in the above example of Sampling from Normal population, we observe that

For all n, $var\ (\bar{x}) = \frac{\sigma^2}{n}$

And for a large n $var\ (\tilde{x}) = 1.57\ \left(\frac{\sigma^2}{n}\right)$

So that

$$var\ (\tilde{x}) < var\ (Md)$$

Therefore, the sample mean has less spread than the sample median, Thus sample mean is more efficient than the same median. Hence sample mean of $\bar{x}$ should be preferred against sample medina as an estimator of population mean μ in N (μ, σ^2) population.

Prof. R.A. Fisher (1921) introduced the concept of efficiency for estimator, which is based on the variances of the sampling distribution of the contending estimators.

Definition 1.10:

If $\hat{\theta}_1\ and\ \hat{\theta}_2$ be two consistent estimators of some parameter θ and

$$var\ (\hat{\theta}_1) < var\ (\hat{\theta}_2)\ for\ all\ n$$

Then for all samples sizes $\hat{\theta}_1$ is more efficient than $\hat{\theta}_2$

The relative efficiency of $\hat{\theta}_1$ w.r.t. $\hat{\theta}_2$is defined as $\hat{\theta}_2$

$$E = \frac{var\ (\hat{\theta}_2)}{var\ (\hat{\theta}_1)} \times 100$$

Definition 1.11 Mean square error (MSE).

Let $\hat{\theta} = \hat{\theta}$ (X_1, X_2,X_3…....X_n) be an estimator of $\tau\theta$, where $\tau\theta$ is some function of . Then the mean square error of $\hat{\theta}$ is defined as

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

Which is the second moment of $\hat{\theta}$ about $\tau\theta$

Now,

$$MSE\ (\hat{\theta}) = E\ [\hat{\theta} - E(\hat{\theta}) + E(\hat{\theta}) - \tau\theta]^2$$

$$= E\left[\left(\hat{\theta}\right) - E\left(\hat{\theta}\right)\right]^2 + E^2\left[\hat{\theta} - \tau\theta\right]^2$$

$$= Var\left(\hat{\theta}\right) + \left[Bias\left(\hat{\theta}\right)\right]^2$$

If $\hat{\theta}$ is unbiased for $\tau\theta$. Then

$$MSE\left(\hat{\theta}\right) = Var\left(\hat{\theta}\right)$$

In fact $MSE\left(\hat{\theta}\right)$ is also a measure of the concentration of the sampling distribution of $\hat{\theta}$ about $\tau\theta$. $Var\left(\hat{\theta}\right)$ is also used for the same purpose. Hence the relative efficiency of $\hat{\theta}_1$ w.r.t. $\hat{\theta}_2$ may in this cases of biased estimators is

$$E = \frac{MSE\left(\hat{\theta}_2\right)}{MSE\left(\hat{\theta}_1\right)} \times 100$$

The measures of efficiency is the relative number of observations required to achieve equality same variance (or mean square error) by the two estimators.

For example consider the above example of estimation of mean μ of normal population,

$$= \frac{var\ (\bar{x})}{var\ (\tilde{x})} = \frac{\frac{\sigma^2}{n}}{\frac{\pi\sigma^2}{2n}} = \frac{2}{\pi} = \frac{2}{\frac{22}{7}} = \frac{7}{11} = 0.637 < 1.$$

This mean that we may obtain same precision from a sample mean based of 637 observations as we do from a sample median based on 1000 observations.

The sample mean $\bar{x}$ is more efficiency than the sample median $\tilde{x}$.

Definition 1.12 ***Efficiency of the estimator***

If $\hat{\theta}_1$ is the most efficient estimator with variance V_1 and $\hat{\theta}_2$ is any other estimator with variance V_2, then the efficiency $\hat{\theta}_2$ of is given as

$$E = \frac{V_1}{V_2} = \frac{var\left(\hat{\theta}_2\right)}{var\left(\hat{\theta}_1\right)}$$

Obviously E cannot exceed unity (≤ 1).

Example 1.8

Show than the estimator $T = \frac{1}{n}\sum_{i=1}^{n} X_i$ computed from a random sample of size n for normal population N (μ, σ^2) with known is more efficiency for than sample mean $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

through biased.
Solution:

$$E(\bar{X}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\sum_{i=1}^{n} \mu = \mu$$

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

Now,

$$T = \frac{1}{n}\sum_{i=1}^{n} X_i \ = \frac{n\bar{X}}{n+1}$$

gives

$$E(T) = \frac{n}{n+1} \ \ E(\bar{X}) = \frac{n}{n+1}\mu \neq \mu$$

Therefore T is biased estimator μ but

$$Var(T) = var\left(\frac{n\bar{X}}{n+1}\right) = \frac{n^2}{(n+1)^2} \ \ Var(\bar{X})$$

Therefore,

$$Var(T) < \ Var(\bar{X})$$

Hence, T is more efficiency than sample mean $\bar{X}$ for μ through it is a biased estimator.

You may attempt the following problems.

E-1.7 Let X_1, X_2, and X_3 be a random sample from a population with unknown mean μ and known variance. T_1, T_2, T_3 are the estimators used to estimate where $T_1= X_1-X_2+X_3$, $T_2= 5X_1 + 3X_2- 7X_2$ and $T_3= ¼ (\lambda X_1+X_2+X_3)$

(a) Are T_1 and T_2 unbiased estimators.
(b) Find he value of λ so that T_3 is unbiased for μ.
(c) Which is the best estimator?

E-1.8 If $\hat{\theta}_1\ and\ \hat{\theta}_2$ are two is independent estimators of then is less efficient then $\hat{\theta}_3 = \ \hat{\theta}_1 + \hat{\theta}_2$ both $\hat{\theta}_1 and\ \hat{\theta}_2$.

11.8 Sufficiency

We know that in estimation of parameter θ of the population f (x; θ), $\theta \in \theta,$ a random sample X_1, X_2,X_3…....X_n of known size, say n is drawn from the population f(x,θ). A statistics (X_1, X_2,…....X_n) is chosen as an estimator θ of which satisfies some nice properties. The statistic is a function of sample observations X_1, X_2,…....X_n which condenses the random variables (X_1, X_2,…....X_n) in a single random variable $\hat{\theta} = \hat{\theta}$ (X_1, X_2,…....X_n). The domain $\hat{\theta}$ of is the set Ж which is the range of values that (X_1, X_2,…....X_n) may take while its range is real time R. Thus this condension reduce or maps the n-dimensional quantities $\underline{x}$ = (x_1, x_2,…....x_n) ∈ Ж into a one dimensional point $\hat{\theta} = \hat{\theta}$ (X_1, X_2,…....X_n) $\in R.$ through the statistic$\hat{\theta}$ there is a possibility that some information about the parameter contained in the sample may be lost in this process. A statistic will be preferred if it contains as much information about the unknown parameter θ as contained in the sample and no information is lost.

A statistic S =S (X_1, X_2,…....X_n) is said to be a *sufficient statistic* for θ if it contains (or exhausts) all the information about θ that is contained in the sample. It gives as much about θ as the sample itself and no information is lost.

Definition 1.13 ***Sufficient statistic:***

Let X_1, X_2,…....X_n be a random sample from the population having density function f (x; θ), $\theta \in \theta.$ A statistic T= T (X_1, X_2,…....X_n) is said to be a sufficient statistic for θ or for a family of distribution $\{\gamma$ (x; θ), $\theta \in \theta\}$ if and only if, the conditional distribution of X_1, X_2,…....X_n given T=t does not depend on θ .

The above definition asserts that if the value of the sufficient statistic T is known then we should concentrate on T itself. The sample values themselves are not needed thereafter. It can tell you nothing more about θ because by the conditional distribution of the sample X_1, X_2,…....X_n given the sufficient statistics does not depend on θ. The above definition is not helpful for practical application. It requires to chose a statistics and then to test whether it is sufficient or not. Secondly, how should you choose such as statistic is not suggested in the definition.

Example 1.9

Let X_1, X_2,X_3……....X_n be a random sample from Bernoulli population with parameter p, that is

$$x_i = \begin{cases} 1, & \text{with probability p} \qquad 0 < p < 1. \\ 0, & \text{with probability } 1 - \text{p} \end{cases}$$

Show that the statistic $T = \sum_{i=1}^{n} X_i$ is sufficient for p.

Solution:

Let,

$$T = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Then $X_1 + X_2 + X_3 + \ldots\ldots.. + X_n$ follows binomial distribution with parameter (n, p).

The p.m.f. of T is

$$P[T = k] = \binom{n}{k} p^k (1-p)^{n-1}, \quad k = 1,2, \ldots., n$$

$$P\left\{X_1 = x_1, X_2 = x_2, X_n = x_n \left| \sum_{i=1}^{n} X_i = k \right.\right\}$$

$$= \begin{cases} \dfrac{P\{X_1 = x_1, X_2 = x_2, X_n = x_n, T = k\}}{\binom{n}{k} p^k (1-p)^{n-1}} & if \sum_{i=1}^{n} X_i = k \\ o & otherwise \end{cases}$$

Thus for $\sum_{i=1}^{n} X_i = k$ we have

$$P\left\{X_1 = x_1, X_2 = x_2, X_n = x_n \left| \sum_{i=1}^{n} X_i = k \right.\right\}$$

$$= \frac{P^{\sum_{i=1}^{n} X_i} (1-p)^{\sum_{i=1}^{n} X_i}}{\binom{n}{k} p^k (1-p)^{n-k}}$$

$$= \frac{1}{\binom{n}{k}}$$

Which is independent of p.

Hence, $\sum_{i=1}^{n} X_i$ is sufficient for p.

Definition 1.13 Joint sufficient statistics

Let $X_1, X_2, \ldots\ldots X_n$ be a random sample of size n from the population having density function f (x; $\theta_1, \theta_2, \ldots \ldots. \theta_r$); $\theta = (\theta_1, \theta_2 \ldots \ldots \theta_r) \in \theta$. The statistic $T_1, T_2, \ldots\ldots T_r$ are jointly sufficient for $(\theta_1, \theta_2, \ldots \ldots. \theta_r)$ if and only if the conditional distribution of $X_1, X_2, \ldots\ldots X_n$ given $T_1 = t_1, T_2 = t_2, T_3 = t_3, \ldots\ldots T_r = t_r$ does not depend on θ.

It may be noted that

(i) By conditional p.d.f. being independent of θ, it means that the p.d.f. does not involve θ. it also means that the domain does not involve θ. For example the p.d.f.

$$f(x) = \frac{1}{2a}\ for\ a - \theta < x < a + \theta;\ -\infty < \theta < \infty$$

Depends on θ. Since θ appears in the range.

(ii) The most general form of the distribution that admits sufficient statistic is Koopman' form of exponential family of distributions.

Definition 1.15

A family of density functions f (x; θ), $\theta \in \theta$ is said to be a one parameter exponential family of densities if p.d.f. f (x; θ) can be expressed as

$$f(x) = a(\theta)b(x)\ e^{[c(\theta)d(x)]}\ for - \infty < x < \infty; \theta \in Q$$

When a (θ) & c (Ə) and b (x) & d(x) are functions of a and x.

Note: if $X_1, X_2, \ldots\ldots X_n$ be a random sample from an exponential family of distribution f(x; θ) then the joint p.d.f. of $X_1, X_2, \ldots\ldots X_n$ can be written as

$$\prod_{i=1}^{n} \mathrm{f}(\mathrm{x};\ \theta) = [\mathrm{a}(\theta)]^{\mathrm{n}} \left[\prod_{\mathrm{i}=1}^{\mathrm{n}} b(x_i)\right] \exp\left[\mathrm{c}(\theta) \sum_{i=1}^{n} d(X_i)\right]$$

Hence by factorization criterion

$$T = \frac{1}{n}\sum_{i=1}^{n} d(X_i)$$

is a sufficient statistic for θ.

Thus under random sampling there exist a sufficient statistic for parameter θif the density belongs to the one parameter exponential family.

The Neyman factorization criterion given below is of helps in determining sufficient statistics.

Theorem (Factorization theorem):

Let $X_1, X_2, \ldots\ldots X_n$ be a random sample of size n from the densities f (x; θ), θ may be a rector. A statistic T- T ($X_1, X_2, \ldots\ldots X_n$) is sufficient for if and only if the joint p.d.f. of X_1, $X_2, \ldots\ldots X_n$ can be factorised as

$$f_{X1,X2,.......Xn}(x_1, x_{2,} x_3, \dots \dots . x_n;\ \theta) = \prod_{i=1}^{n} \mathrm{f}\ (\mathrm{x};\ \theta) = \mathrm{g}\ [\mathrm{T}(x_1, x_2, \dots \dots . x_n); \theta]$$

$$\mathrm{h}\ (x_1, x_2, \dots \dots . x_n) = \mathrm{g}(\mathrm{T}; \theta)\mathrm{h}(x_1, x_{2,} x_3, \dots \dots . x_n)$$

where the function h $(x_1, x_{2,} x_3, \dots \dots . x_n)$ is a non negative function of $x_1, x_{2,} x_3, \dots \dots . x_n$ only and does not depend on and g the function g $(\mathrm{T}; \theta)$ is non negative and depends on θ and $\mathrm{T}(x_1, x_{2,} x_3, \dots \dots . x_n)$ only.

Remarks 1. If T is sufficient for θ any one to one function of T is also sufficient for θ.

Remarks 2. If T_1 and T_2 are two distinct sufficient statistics then T_1 is a function of T_2.

For example, If $\sum X_i$ *and* $\sum X_i^2$ are jointly sufficient for mean μ and variance σ^2 then $\bar{X}$ and $\sum X_i - \bar{X}^2 = \sum X_i^2 - n\bar{X}^2$ are also jointly sufficient for (μ, σ^2)

Example 1.10:

Let $X_1, X_2, \dots\dots X_n$ be a random sample from a population with p.d.f.

$$f(x;\theta) = \theta x^{\theta-1} \qquad o < x < 1,\ \ \theta > 0$$

Show that

$$T = \prod_{i=1}^{n} X_{i,}\ is\ sufficient\ for\ \theta.$$

Solution:

The joint p.d.f. for $X_1, X_2, \dots\dots X_n$ is

$$l(x,\theta) = \prod_{i=1}^{n} f(x,\theta) = \theta^n \left(\prod_{i=1}^{n} X_i\right)^{\theta-1}$$

$$= \ \theta^n \left[\prod_{i=1}^{n} X_i\right]^{\theta} , \frac{1}{[\prod_{i=1}^{n} X_i\ ]}$$

$$= g\ (t,\theta) h\ (x_1, x_2, \dots \dots . x_n)$$

Where

$$g\ (t,\theta) = \ \theta^n \left(\prod_{i=1}^{n} X_i\right)^{\theta}\ depends\ on\ t\ and\ \theta$$

And

$$h\ (x_1, x_2, \dots \dots . x_n) = \frac{1}{(\prod_{i=1}^{n} X_i\ )}\ is\ independent\ of\ \theta$$

Hence by factorization criterion,

$$T = \prod_{i=1}^{n} X_{i,}\ as\ is\ sufficient\ for\ \theta.$$

Example 1.11:

Let X_1, X_2,…....X_n be a random sample of size n from a normal population N (μ, σ^2) show that

(a) Sample mean $\bar{X}$ is sufficient for mean μ if σ^2 is known,
(b) $S_0^2 = \frac{1}{n}\sum(X_i - \mu)^2$ Sufficient for variance $\sigma^2\ if\ \mu = \mu_0$ is known,
(c) $(\bar{X}, S^2)$Are jointly sufficient for (μ, σ^2) if μ both μ and σ^2 are unknown.

Solution:

the joint p.d.f. of X_1, X_2,…....X_n is

$$f\ (x_1, x_2, \dots \dots . x_n;\ \mu, \sigma^2) = \left[\frac{1}{(2\pi\sigma^2)}\right]^{\frac{n}{2}} exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right] \dots . (A)$$

Expression (A) may be re-arranged as below:

Part (a)

$$f\ (x_1, x_2, \dots \dots . x_n;\ \mu, \sigma^2) = \left\{\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^2 exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right]\right\} \times$$

$$exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right]$$

= (factor independent of μ) × (factor involving $\bar{x} e \mu$) only

Thus by factorization criterion, sample mean $\bar{X}$ is sufficient for μ if σ^2 is known.

Part (b)

Let |mean| $\mu = \mu_0$ be known

$$f\ (x_1, x_2, \ldots \ldots . x_n;\ \mu, \sigma^2) = \left(\frac{1}{\sigma^2}\right)^n exp\left(-\frac{nS_0^2}{2\sigma^2}\right)\ where\ S_0^2 = \frac{1}{n}\sum_n^n (x_i - \mu_0^2)$$

$$= g\ (S_0^2, \sigma^2) h\ (x_1, x_2, \ldots \ldots . x_n)$$

Where

$$g\ (S_0^2, \sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} exp\left[-\frac{nS_0^2}{2\sigma^2}\right]\ and\ h(x_1, x_2, \ldots \ldots . x_n) = 1$$

Hence

$$S_0^2 = \frac{1}{n}\sum_n^n (x_i - \mu_0)^2\ is\ sufficient\ for\ \sigma^2$$

Part (C) when both mean μ and variance σ^2 are unknown.

$$f\ (x_1, x_2, \ldots \ldots . x_n;\ \mu, \sigma^2) = \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \left(\frac{1}{\sigma^2}\right)^n exp\left\{-\frac{1}{2\sigma^2}[(n-1)S^2 + n(\bar{x} - \mu)^2]\right\}$$

$$= g\ (\bar{x}\ , S^2\ , \mu, \sigma^2) \qquad\qquad h(x_1, x_2, \ldots \ldots . x_n)$$

Functions of $\overline{x}\ , S^2$ Independent of μ, σ^2

Where

$$g\ (\bar{x}\ , S^2\ , \mu, \sigma^2) = \left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}} exp\left\{-\frac{1}{2\sigma^2}[(n-1)S^2 + n(\bar{x} - \mu)^2]\right\}$$

Is a function of $\overline{x}\ , S^2$, $\mu\ and\ \sigma^2$ and $h(x_1, x_2, \ldots \ldots . x_n) = \left(\frac{1}{2\pi}\right)^{\frac{n}{2}}$ is independent of $\mu\ and\ \sigma^2$

$$Hence\ \bar{X}\ and\ S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2\ are\ jointly\ sufficient\ for\ \mu\ and\ \sigma^2$$

Proved

From the above example, it may be noted that

(i) $\bar{X}$ is not sufficient for $\mu\ if \sigma^2$ is unknown,

(ii) S^2 is not sufficient for σ^2if μ is unknown.

In the following example the whole sample is jointly sufficient for a single parameter θ.

Example 1.12:

Let $X_1, X_2, X_3 \ldots\ldots\ldots X_n$ be a random sample of size n from a normal population following Cauchy distribution.

$$f(x,\theta) = \frac{1}{\pi}, \frac{1}{1+(x-\theta)^2} - \infty < x < \infty, -\infty < \theta < \infty$$

Examine of there exist a sufficient statistics for parameterθ.

Solution:

The joint p.d.f. of $X_1, X_2, X_3 \ldots\ldots\ldots X_n$ is

$$f\ (x_1, x_2, \ldots \ldots . x_n;\ \theta) = \left(\frac{1}{\pi}\right)^2 \prod_{i=1}^{n} \left(\frac{1}{1+(x-\theta)^2}\right) \neq\ g_1(t_1,\theta)k\ (x_1, x_2, \ldots \ldots . x_n)$$

Hence by factorization theorem there is no single statistic which alone is sufficient estimator of parameterθ.

However,

$$f\ (x_1, x_2, \ldots \ldots . x_n;\ \theta)k\ (x_1, x_2, \ldots \ldots . x_n;\ \theta)h\ (x_1, x_2, \ldots \ldots . x_n)$$

Holds which implies that the whole set ($X_1, X_2, \ldots\ldots X_n$) is jointly sufficient forθ.

The following example is on the discrete population.

Example 1.13:

Let $X_1, X_2, \ldots\ldots X_n$ be three independent observations drawn from a Poisson distribution with parameter θ. Show that

$$T = \sum_{i=1}^{n} X_{i,} =\ X_1 + X_2 + X_3\ \ is\ sufficient\ for\ \theta.$$

Solution:

Here, X_1, X_2, X_3 are i.i.d. r.v.s's having common p.m.f.

$$P[X = x_i] = \frac{e^{-\theta}.\theta^{x_i}}{x_i!}, i = 1,2,3:\ \ x_i = 0,1,2, \ldots \ldots, \theta\ and > 0$$

Now the conditional p.m.f. X_1, X_2, X_3 of given $x_1 + x_2 + x_3 = t$ can be expressed as

$$P[X_1 = x_1, X_2 = x_2, X_3 = x_3, |X_1 + X_2 + X_3 = t]$$

$$= \begin{cases} \dfrac{P\{X_1 = x_1, X_2 = x_2, X_3 = t - x_1 - x_2\}}{P[X_1 + X_2 + X_3 = t} - & if x_1 + x_2 + x_3 = t, x_i = 0,1,2,.. \\ o & otherwise \end{cases}$$

$$= \frac{\left[\dfrac{e^{-\theta}.\theta^{x_1}}{x_1!}\right]\left[\dfrac{e^{-\theta}\theta^{x_2}}{x_2!}\right]\left[\dfrac{e^{-\theta}\theta^{t-x_1-x_2}}{x_i!}\right]}{\left[\dfrac{e^{-\theta}(3\theta)^t}{t!}\right]}$$

$$= \left[\frac{Lt}{x_1!\,x_2!\,(t - x_1 - x_2)!}\right]\left[\left(\frac{1}{3}\right)^t\right]$$

Which is independent of hence T= $X_1 + X_2 + X_3$ is sufficient for θ.

Proved

You may attempt the following exercises.

E-1.9 (a) Is sufficient estimator is always consistent/

(b) Is sufficient estimator is always unbiased?

E-1.10 Obtain a sufficient statistic for θ in the following population.

$$f(x, \theta) = \frac{1}{\theta}, \quad 0 < x < \theta$$

11.9 Confidence Interval Estimation

Let X be some characteristic of the population under study, where X is a random variable having the p.d.f $f(x, \theta), \theta \in \theta.$ suppose that the function form $f(x, \theta)$ known but the parameter θ is unknown, In the point estimation, a random sample X_1, X_2,X_3……....X_n is drawn from the population and a point estimate $\hat{\theta}$ (x_1, x_2,x_3……....x_n) of θ is the value of estimator$\hat{\theta} = \hat{\theta}$ (X_1, X_2,X_3……....X_n). The point estimate$\hat{\theta}$, based on the sample observations is a single number does not always equal to the true value of the parameter and may vary from Sample to sample. The estimator $\hat{\theta}$ is a random variable having its own sampling distribution g(t; θ) with P[$\hat{\theta}$=θ]= 0. We do not have an idea how close $\hat{\theta}$ is to θ. Of course one may like to have some measure of closeness $\hat{\theta}$of to θ. The method of confidence interval estimation provides an answer to this objection. In this technique a point estimate together with some measure of assurance that the true value of parameter θ lies within the interval. Thus, in interval estimation of parameter one estimates an interval estimation of a parameter one estimates an interval which contains the true value of the parameter and specify the confidence with which it is suppose to do so.

As an illustration, consider the problem of interval estimation of population mean μ with variance σ^2.

We know that sample mean $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ of a random sample X_1, X_2,X_3……....X_n of size n from the population is point estimator of the unknown mean μ.

If population is normal N (μ, σ^2) then

$$\bar{X} = N\left(\frac{\sigma^2}{n}\right)$$

So that $E(X) = \mu,\qquad Var\ (X) = \ \sigma^2/n$

and

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

Even if the understanding distribution is non normal we have seen that for large samples

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \approx N(0,1)$$

These facts are used to construct of confidence intervals for unknown mean μ.

Form the normal area table, Zo could be obtained such that

$$P\left(-Zo \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq Z\right) = \gamma \ldots\ldots\ldots\ldots\ldots\ldots..(A)$$

For example, if γ = 0.95 then Zo= 1.96 and γ = 0.90 then Zo= 1.645

Since σ > 0 therefore the following four inequalities

$$\left(-Zo \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq Z\right), P\left(-Zo\left(\frac{\sigma}{\sqrt{n}}\right) \leq \bar{X}[\,] - \mu \leq Zo\frac{\sigma}{\sqrt{n}}\right),$$

$$\left[\left(-\bar{X} - Zo\left(\frac{\sigma}{\sqrt{n}}\right) \leq \mu \leq \bar{X} + Zo\ \left(\frac{\sigma}{\sqrt{n}}\right)\right)\ and\ \left(-\bar{X} - Zo\left(\frac{\sigma}{\sqrt{n}}\right) \leq \mu \leq \bar{X} + Zo\ \left(\frac{\sigma}{\sqrt{n}}\right)\right)\right]$$

are equivalent. Here the last inequality is true if and only if the first inequality is true therefore probabilities of occurrence of both the inequalities must be equal to .

thus, we may write (A) as:

$$P\left(-\bar{X} - Zo\left(\frac{\sigma}{\sqrt{n}}\right) \leq \mu \leq \bar{X} + Zo\ \left(\frac{\sigma}{\sqrt{n}}\right)\right) = \gamma \ldots\ldots\ldots\ldots\ldots\ldots..(B)$$

In equation (B), the ends limits $\bar{X} - Zo\left(\frac{\sigma}{\sqrt{n}}\right)$ and $\bar{X} + Zo\ \left(\frac{\sigma}{\sqrt{n}}\right)$ are known as lower and upper confidence limits, respectively, of (parameter) mean μ and γ is called the confidence of the confidence interval estimate $\left[\bar{X} - Zo\left(\frac{\sigma}{\sqrt{n}}\right), \bar{X} + Zo\ \left(\frac{\sigma}{\sqrt{n}}\right)\right]$ of mean μ.

For non-normal population, Equation (B)

$$P\left[\bar{X} - Zo\left(\frac{\sigma}{\sqrt{n}}\right) \leq \mu \leq \bar{X} + Zo\ \left(\frac{\sigma}{\sqrt{n}}\right)\right] \cong \gamma \ldots\ldots\ldots\ldots\ldots\ldots..(c)$$

Holds for large n,

$$\left[\bar{X} - Zo\left(\frac{\sigma}{\sqrt{n}}\right), \bar{X} + Zo\ \left(\frac{\sigma}{\sqrt{n}}\right)\right]$$

As well as its estimate

$$\left[\bar{x} - Zo\left(\frac{\sigma}{\sqrt{n}}\right), \bar{x} + Zo\ \left(\frac{\sigma}{\sqrt{n}}\right)\right]$$

Obtained by sample values called 100 γ % confidence Interval Estimates of population mean μ.

It is customary to write interval estimates as

$$\bar{X} + Zo\ \left(\frac{\sigma}{\sqrt{n}}\right)$$

It may be of worth to note that γ is not the probability that the mean a constant takes values from $\left(\bar{X} + Zo\ \left(\frac{\sigma}{\sqrt{n}}\right)\right)$ to $\left(\bar{X} - Zo\ \left(\frac{\sigma}{\sqrt{n}}\right)\right)$. In fact the probability that lies in a certain interval is either 1 or 0 and it cannot be $0 < \gamma < 1$ is a measure of confidence (or our belief) allotted to the statement that the random interval $\left[\bar{X} - Zo\left(\frac{\sigma}{\sqrt{n}}\right), \bar{X} + Zo\ \left(\frac{\sigma}{\sqrt{n}}\right)\right]$ includes the unknown mean μ.

Once the sample is obtained and the sample mean is computed, the interval $\bar{x} \pm Zo\left(\frac{\sigma}{\sqrt{n}}\right)$ is known.

Example 1.14:

Suppose that the sources of certain group of candidates in a competitive examination follows normal distribution with σ = 16. A random sample of size 25 has yielded mean $\bar{x}$ = 69.6. Obtain 90% confidence interval for average score μ.

Solution:

Let X= Score of the group of candidates in the examination.

Then,

$$X \sim N\left(\mu, \frac{\sigma^2}{n}\right) \; where \; n = 25, \quad \sigma^2 = (16)^2 = 256.$$

For γ = 0.90 Zo= 1.645, therefore, 90% confidence interval for mean μ is

$$\bar{x} \pm 1.645\left(\frac{\sigma}{\sqrt{n}}\right) = 69.6 \pm 1.645\frac{16}{\sqrt{25}} = [64.336, 74, 864]$$

Definition 1.16:

Let $X_1, X_2, X_3 \ldots\ldots\ldots X_n$ be a random sample of size n from the densities f (x; θ), $\theta \in Q$. Let T_1= T ($X_1, X_2, \ldots\ldots X_n$) and T_2= T ($X_1, X_2, \ldots\ldots X_n$) be two statistic satisfying $T_1 = T_2$ for which

$$P(T_1 \leq \theta \, T_2) = \gamma$$

Where γ does not depend upon θ, then the random interval (T_1, T_2) is called a 100 γ percent confidence interval for.

γ is called the confidence coefficient, T_1 and T_2 are called the lower and upper confidence limits, respectively for θ and the difference$(T_1 - T_2)$ is known as width (or length) of the interval.

A value (t_1, t_2) of the random interval (T_1, T_2) is also called 100 γ percent confidence interval for θ.

One of the two ends point T_1 and T_2 of the interval (T_1, T_2) may be constant. In other words, there may be one-sided confidence interval.

Further the width of the confidence interval is controlled by two factors:

(i) As sample size n increase, the interval gets narrower due to term involving $\left(\frac{1}{\sqrt{n}}\right)$.

(ii) The larger the sample standard deviation, the larger is the confidence interval.

Example 1.15

Obtain 100 (1-α)% confidence intervals for the parameters (a) mean μ of the normal population N (μ, σ^2) if both μ and σ^2 are unknown,

Solution:

Let $X_1, X_2, X_3 \ldots\ldots\ldots X_n$ be a random sample of size n from the density function

$$f(x;\ \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\,\sigma^2}} e^{\frac{1}{2\,\sigma^2}(x-\mu)^2} - \infty < x\infty, \sigma^2 > 0, -\infty < \mu < \infty$$

and define

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i\,, S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 \; and \;\; S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

The statistic

$$t = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim Students\ t\ distribution\ with\ (n-1)\ d.f.$$

Hence 100 (1- α%) confidence interval for μ is given by

$$P[|t| \leq t_\alpha] \qquad\qquad = 1 - \alpha$$

$$\Rightarrow P\left[|\bar{X} - \mu| \leq \frac{S}{\sqrt{n}}.t_\alpha\right] = 1 - \alpha$$

Where gives

$$P\left[\bar{X}\, t_\alpha \frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + t_\alpha \frac{S}{\sqrt{n}}\right] = 1 - \alpha$$

Where $t_\alpha$is the tabulated valued of t for (n-1) df at significance level α**.** Hence the required confidence interval for μ is given by.

$$\left(\bar{X} - t_\alpha \frac{S}{\sqrt{n}}, \bar{X} + t_\alpha \frac{S}{\sqrt{n}}\right)$$

Example 1.16:

Obtain 100 (1- α)% confidence interval for of the normal population N (μ, σ^2) if

(a) μ is unknown
(b) $\mu = \mu_0$ is known

Solution:

Part (a) Here the p.d.f of r.v. X is

$$f(x;\ \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\,\sigma^2}} e^{\frac{1}{2\,\sigma^2}(x-\mu)^2} - \infty < n < \infty, -\infty < \mu < \infty,\ \sigma^2 > 0$$

You know that

$$\sum_{i=1}^{n}\left(\frac{X_i-\bar{X}}{\sigma^2}\right)^2 = \frac{ns^2}{\sigma^2} = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

If $\chi^2(n-1)$ be valueχ^2 of such that

$$P\left(\chi^2 - \chi^2_{\alpha_{1n-1}}\right) = \int_{\chi^2_{\alpha_{1n-1}}}^{\infty} p(\chi^2)d\chi^2 = \alpha$$

Where p(χ^2) is the p.d.f of χ^2 with n.d.f. then the required confidence interval is given by

$$P\left(\chi^2 - \chi^2_{\alpha_{1n-1}} < \frac{(n-1)S^2}{\sigma^2} < \chi^2_{\alpha_{1n-1}}\right) = 1 - \frac{\alpha}{2} - \frac{\alpha}{2} = 1 - \alpha$$

$$\Rightarrow P\left(\frac{(n-1)S^2}{\chi^2_{\alpha_{/2;n-1}}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{\alpha/2_{;n-1}}}\right) = 1 - \alpha$$

Hence the random interval for the variance of a normal population is

$$\left(\frac{(n-1)S^2}{\chi^2_{\alpha_{/2;n-1}}}, \frac{(n-1)S^2}{\chi^2_{\alpha/2_{;n-1}}}\right)$$

Part (b): If menu μ is known and equal to μ_0 then

$$\frac{n\hat{\theta}^2_{\sigma}}{\sigma^2} = \sum_{i=1}^{n}\left(\frac{X_i-\bar{X}}{\sigma^2}\right)^2 \sim \chi^2$$

Where

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu_0)^2$$

The required 100(1-α)% confidence interval is given by

$$P\left[\chi^2_{1-\alpha;n} < \chi^2 < \chi^2_{\alpha/2;n}\right] = 1 - \alpha$$

$$P\left[\chi^2_{1-\alpha/2} < \frac{n\sigma^2}{\sigma^2} < \chi^2_{\alpha/2;n}\right] = 1 - \alpha$$

$$P\left[\frac{n\hat{\sigma}^2}{\chi^2_{\alpha/2;n}} < \sigma^2 < \frac{n\hat{\sigma}^2}{\chi^2_{1-\alpha/2;n}}\right] = 1 - \alpha$$

Where $\chi^2_{\alpha/2;n}$ *and* $\chi^2_{1-\alpha/2;n}$ are obtain from the Table fro Critical values under Chi-square distribution.

You may try the following exercises.

E- 1.11 A random sample of size 40 from a normal population with known σ^2= 10 has yielded mean $\bar{x}$ = 7.164 obtain 80% confidence interval of population mean μ .

E- 1.12 Let a random sample of size 17 form the normal distribution N (μ, σ^2) yields $\bar{x}$= 4.7 and S^2= 5.76. Determine a 90% confidence interval for mean μ .

11.10 Solution/ Answers

E- 1.2) (a) (b) (c) (d) (e) (f) (g) (h)

E- 1.7) (a) E (T_1) =θ = E(T_2); T_1 and T_2 are unbiased

(b) λ =2 (c) Var (T_1) = $3\sigma^2$, Var (T_2) = $83\sigma^2$, Var (T_3)= ¼ σ^2 T_3 is best among them.

E- 1.9) (a) (b)

E- 1.10)

E-1.11) $7.164 \pm 1.284\sqrt{\frac{10}{40}}$ or (6.523, 7.805)

E- 1.12) (3.7, 5.7)

11.11 Summary

In this unit an attempt is made to cover the concepts of statistical inference and its role in deciding the unknown value of parameter. We understood how to decide the quality of an estimator. In this respect the unbiasednedd and the Fisher's Criteria of a good estimator namely consistency, efficiency are dealt with. A brief we have also studied the method of development of confidence interval estimates for a population parameter with a given confidence level, which is conceptualized as the probability that a random interval will contain the true value of the parameter.

11.12 Further Readings

1. Rahtagi V.K. (1984): An Introduction to Probability theory and Mathematical Statistics chapter VIII, IX & X Pub; John Wiley & Sons, New York.
2. Goon A.N., Gupta M.K. & Das Gupta B (1987) Fundamentals of Statistics Vol. I The World Press Pvt. Ltd., Kolkata.
3. Kapoor V.K. & S.C. Saxena: Fundamentals of Mathematical Statistics, Chapter Seventeen, Pub: S. Chand.

Unit-12: Methods of Estimation

Structure

12.1 Introduction

An investigator may be interested in the study of the behavior of some of certain characteristics say X of the elements of a population under consideration. The behavior of the random variable X is explained by its

$p.d.f.\ f_x(x,\theta) = f(x,\theta)\ (if\ x\ is\ a\ Continuous\ \ r.v.)$

Or

$p.m.f.\ P_x(x,\theta) = f(x,\theta)\ (if\ x\ is\ a\ discrete\ \ r.v.)$

For simplicity we shall use the same notation $f(x;\theta)$ for both pdf and pmf and confine our discussion on estimation problem on univariate (one dimensional) case only.

The density function f (x, θ) of a r.v. X may depend upon certain number say k, of parameters $\theta_1, \theta_2 \dots \dots \theta_k. Thus\ \theta = (\theta_1, \theta_2 \dots \dots \theta_k)$may be vector.

$\theta_1, \theta_2 \dots \dots \theta_k$ may take any value on a set θ.

The set θ is the collection of all possible values of $\theta_1, \theta_2 \dots \dots \theta_k$ associated with the population f (x, θ) it is known as parameter space . For different combinations $\theta_1, \theta_2 \dots \dots \theta_k \in Q$ we get different densities having the given form f $(x, \theta_1, \theta_2 \dots \dots \theta_k)$.

The set { f (x, θ): $\theta \in \theta$ } is called the family of distribution of r.v. X.

Ex 2.1 Let X~ b (n,p.) with known n and unknown p. Then $\Theta = \{p: < p < 1\}$ is he family of possible p.m.f. s of the Poisson variable X.

Ex 2.2 Let X~ Poisson P(x; λ) Then $\Theta = \{\lambda: 0 < \lambda < \infty\}$ is the parameter space and $\{P(x;\lambda): 0 < \lambda < \infty\}$ or $\{P(x;\lambda): \lambda > 0\}$is the family of possible p.m.f. of the Poisson variable X.

Ex 2.3 Let X ~ Normal N(μ, σ^2) and μ and σ^2 be the parameters of the distribution

(a) If both μ and σ^2 be unknown then the parameters space is Θ ={ $(\mu, \sigma^2): \infty - \theta < \mu < \theta;\ \sigma^2 > 0\}$and the family of normal distribution is N(μ, σ^2) : $-\theta < \mu < \theta;\ \sigma^2 > 0\}$.
(b) If σ^2= 1 the family of normal distribution is given by $(N\ (\mu, 1): \in \Theta\}\ where\ \Theta = \{\mu: -\infty < \mu < \infty\}.)$
(c) If $\mu = \mu_0$ (known), but $\sigma^2$is unknown then the parameter space is$\{\Theta = (\mu, \sigma^2): \sigma^2 > 0\}$ or simply $\Theta = \{\sigma > 0\} = (0, \infty)$, while family of normal distribution is { N(μ, σ^2) : $\sigma^2 > 0$ }

It may be noted that N (2.1), N (-3,1), N (1.5.1) are member of the family given in Ex. 2.3 (b) and N (2,1), N (2,4), N (2,16) are the members of the family of distributions { N (2, σ^2) : σ^2>0}.

The general family of distributions may be expressed by $\{\ \mathrm{f}(x, \theta_1, \theta_2 \ldots \ldots \theta_k) : \theta_i\ \in \Theta; i = 1{,}2 \ldots \ldots k\}$.

In the most of the practical applications, the functional from f (x, θ) of the density of the r.v. X is either known or assumed to be known from the past experience or the conditions of the experiment, but the a few or all the past experience or the conditions of the experiment but a few or all the parameters $\theta_1, \theta_2 \ldots \ldots \theta_k$ may not be known.

If all the parameters $\theta_1, \theta_2 \ldots \ldots \theta_k$ as well as the functional form of the density $f(x;\ \theta_1, \theta_2 \ldots \ldots \theta_k)$ are known then we say that the density $f(x;\ \theta_1, \theta_2 \ldots \ldots \theta_k)$ is completely specified and there is no need to make inference about it.

The problem of estimation of parameters arises if either some or all the parameters of the density of the r.v.X are unknown although the functional form of the density function { f (x; θ) : $\theta \in \Theta$ be known.

The problem of estimation may be stated as follows:

Suppose that the random variable of interest X has p.d.f. f (x; θ) : $\theta \in \Theta$ where the functional form f (x; θ) of is known but some or all the parameters $\theta_1, \theta_2 \ldots \ldots \theta_k$ are unknown. We have to get estimates of unknown parameters.

Let $X_1, X_2, \ldots\ldots X_n$ be a random sample from the population under consideration. Let $X_1=x_1$ $X_2=x_2\ldots\ldots X_n= X_n$ be the observed values of a random sample $X_1, X_2, \ldots\ldots X_n$ .There will be a infinite number of functions T $(X_1, X_2, \ldots\ldots X_n)$ of the sample $X_1, X_2, \ldots\ldots X_n$ .These functions T $(X_1, X_2, \ldots\ldots X_n)$ of sample observations are called Statistics. The statistics used to estimate a parameter θ is known as estimator of θ.

The problem of estimate is to choose a statistic $\hat{\theta}_r = \hat{\theta}_r$ $(X_1, X_2, \ldots\ldots X_n)$ for r = 1,2,…k to be used as estimator of unknown parameters $\theta_1, \theta_2 \ldots\ldots \theta_k$or of the functions $\gamma_1(\theta), \gamma_2(\theta) \ldots.\gamma_m(\theta)\ of\ \theta = (\theta_1, \theta_2 \ldots\ldots \theta_k)$ as the case may be. The numerical value of $\hat{\theta}_r = \hat{\theta}_r$ $(X_1, X_2, \ldots\ldots X_n)$ is called the estimate of $\hat{\theta}_r$, r= 1,2,…k.

There may be more than one estimator of a parameter. We prefer the best estimator among them according to certain criterion. One of the criteria is that the estimate should fall nearest to the true value of the parameters to be estimated. In other worlds the distribution of the statistic should concentrate as closely as possible near the true value of the parameter. Some of the other properties like unbiasedness, consistency, sufficiency and efficiency, MVUE have been discussed in UNIT-1.

Our aim in this section is to determine the functions of the sample observations $\hat{\theta}_1 = \hat{\theta}_1$ $(X_1, X_2, \ldots\ldots X_n)$, $\hat{\theta}_2 = \hat{\theta}_2$ $(X_1, X_2, \ldots\ldots X_n)$……. $\hat{\theta}_k = \hat{\theta}_k$ $(X_1, X_2, \ldots\ldots X_n)$ population under consideration such that their sampling distributions are concentrated as closely as possible near the true value of $\theta_1, \theta_2 \ldots\ldots \theta_k$ the parameters.

12.2 Objectives

After going through this unit you will be able to -

- Obtain estimators of the unknown parameters based on method of moments estimation.
- Obtain estimators of the unknown parameters which maximizes the likelihood function L $(\theta | x_1, x_2, \ldots\ldots x_n)$ of the observed sample $x_1, x_2, \ldots\ldots x_n$
- Know the properties of the estimate obtained by above methods of estimation.

12.3 Procedures of Estimation

Suppose that we are interested in the study of behavior of the random variable X which has the density function $f(x;\ \theta_1, \theta_2 \ldots\ldots \theta_k)$ where $=\theta = \theta_1, \theta_2 \ldots\ldots \theta_k \in \Theta$. Suppose we know the function form of $f(x;\ \theta_1, \theta_2 \ldots\ldots \theta_k)$. But some or all of the parameters are unknown.

For making any inference about the population having pdf f $(x; \theta)$ random sample $X_1, X_2, \ldots\ldots X_n$ of predetermined size n is drawn from the population. Let $X_1=x_1$ $X_2=x_2\ldots\ldots X_n= X_n$ be the observed values of the sample. Our objective is to determine estimators.

$\hat{\theta}_1 = T_1 = \hat{\theta}_1$ (X_1, X_2,…....X_n), $\hat{\theta}_2 = T_2 = \hat{\theta}_2$ (X_1, X_2,…....X_n)……. $\hat{\theta}_k = T_k = \hat{\theta}_k$ (X_1, X_2,…....X_n) and as such the point estimates $\hat{\theta}_1 = \hat{\theta}_1$ (X_1, X_2,…....X_n), $\hat{\theta}_2 = \hat{\theta}_2$ (X_1, X_2,…....X_n)……. $\hat{\theta}_k = \hat{\theta}_k$ (X_1, X_2,…....X_n) of the unknown parameters $\theta_1, \theta_2 \dots \dots \theta_k \in \Theta$ whose distribution $g_i\,(\hat{\theta}_r;\ \theta_1, \theta_2 \dots \dots \theta_k)$ for i= 1,2,……, k is concentrated as closely as possible near the true value of the parameters and satisfy other desired nice properties.

The following are the methods of estimation:

(i) Methods of Moments (MME)
(ii) Maximum Likelihood Method of Estimation (MLE)
(iii) Method of Minimum Variance (MMV)
(iv) Method of Least Squares (MLS)
(v) Method of Minimum Chi-squares (MCS)
(vi) Method of Inverse Probabilities

We shall discuss only two methods of estimation viz. Method of moments (MME) and Method of Maximum Likelihood (MLE).

12.3.1 Method of Moments (MME)

This oldest but simple method was proposed and studied, in detail by Prof. Karl Pearson.

The r-th moment about origin of the parent population $f(x;\ \theta_1, \theta_2 \dots \dots \theta_k)$ is defined as

$\mu'_r = E(X^r) = \int_{-\infty}^{\infty} x^r\, f(x;\ \theta_1, \theta_2 \dots \dots \theta_k)$ dx for r 1,2,….k (2.1)

In general, $\mu'_1,\ \mu'_2, \dots \dots .. \mu'_r$ will be the functions of the parameters $\theta_1, \theta_2 \dots \dots \theta_k$

Let m'$_1$, m'$_2$, …..m'$_r$ be the sample moments about origin of the sample x_1, x_2, …..x_n, that is $m_r = \frac{1}{n}\sum_{i=1}^{n} x_i^r \dots \dots \dots \dots \dots \dots$ (2.2)

The method of moments consists of solving the k-equations (2.1) for $\theta_1, \theta_2 \dots \dots \theta_k$ in terms of $\mu'_1,\ \mu'_2, \dots \dots .. \mu'_r$and thereafter replacing these moments μ'_r r= 1,2,3….k by corresponding sample momentsm'_r , r= 12, ….k respectively so that,

$$\hat{\theta}_r = \hat{\theta}_r(\,\mu_1, \mu_2 \dots \dots \mu_k) = \,\hat{\theta}_r(\,m_1, m_2 \dots \dots m_k)\ for\ r = 1,2, \dots \dots k \qquad (2.3)$$

are there required MME of θr , r= 1,2,……k.

The estimates of $\theta_1, \theta_2 \dots \dots \theta_k$ are the values of $\hat{\theta}_1, \hat{\theta}_2, \dots . \hat{\theta}_k$ given by replacing X_1, X_2,…....X_n by their sample values x_1, x_2,…....x_n . Some illustrations are given to explain the technique.

Example 2.4 Let the random variable X take the values

x:	0	1	2
P(x)	$\frac{\theta}{4N}+\frac{1}{2}\left(1-\frac{\theta}{N}\right)$	$\frac{\theta}{2N}+\frac{\alpha}{2}\left(1-\frac{\theta}{N}\right)$	$\frac{\theta}{4N}+\frac{1-\alpha}{2}\left(1-\frac{\theta}{N}\right)$
Observed f:	27	38	10

Estimate θ and α by the method of moments.

Solution: Here,

$$\mu_1 = E(X)$$

$$= 0\left[\frac{\theta}{4N}+\frac{1}{2}\left(1-\frac{\theta}{N}\right)\right] + 1\left[\frac{\theta}{2N}+\frac{\alpha}{2}\left(1-\frac{\theta}{N}\right)\right] + 2\left[\frac{\theta}{4N}+\frac{1-\alpha}{2}\left(1-\frac{\theta}{N}\right)\right]$$

$$= \frac{\theta}{N}+\left(1-\frac{\theta}{N}\right)\left[\frac{\alpha}{2}+(1-\alpha)\right]$$

$$= 1-\frac{\alpha}{2}\left(1-\frac{\theta}{N}\right) \qquad (2.4)$$

$$\mu_2 = E(X^2)$$

$$= 0\left[\frac{\theta}{4N}+\frac{1}{2}\left(1-\frac{\theta}{N}\right)\right] + 1^2\left[\frac{\theta}{2N}+\frac{\alpha}{2}\left(1-\frac{\theta}{N}\right)\right] + 2^2\left[\frac{\theta}{4N}+\frac{1-\alpha}{2}\left(1-\frac{\theta}{N}\right)\right]$$

$$= \frac{3\theta}{N}+\left(1-\frac{\theta}{N}\right)\left[\frac{\alpha}{2}+2(1-\alpha)\right]$$

$$= 2-\frac{\theta}{2N}-\frac{3\alpha}{2}\left(1-\frac{\theta}{N}\right) \qquad (2.5)$$

We also have,

$$m_1 = \frac{\sum fx}{\sum f} = \frac{58}{75}$$

$$m_2 = \frac{\sum fx^2}{\sum f} = \frac{78}{75} \qquad (2.6)$$

To get MME for θ and α we have to solve

$$1-\frac{\alpha}{2}\left(1-\frac{\theta}{N}\right) = \frac{58}{75}$$

and

$$2-\frac{\theta}{2N}-\frac{3\alpha}{2}\left(1-\frac{\theta}{N}\right) = \frac{78}{75}$$

We finally get the estimates as

$$\hat{\alpha} = \frac{34}{33}, \quad \hat{\theta} = \frac{42}{75} \; N \; (if \; N \; is \; a \; known \; number)$$

Example 2.5 Let the sample values from the population with pdf

$f(x) = (1+\theta)x^{\theta}. 0 < x < 1, \theta > 0 \; be \; given \; below.$

0.46, 0.38, 0.61, 0.82, 0.59, 0.53, 0.72, 0.44, 0.59, 0.60.

Find out the estimate of θ by method of moments .

Solution: Here, $\underline{\theta} = (\theta_1, \theta_2 \dots \dots \theta_k) \; which \; maximizes \; likelihood \; function \; L \; (\theta)$

$$\mu_1 = E(X) = \int_0^1 (1+\theta)x^{\theta+1}dx = \frac{1+\theta}{2+\theta} \quad and$$

$$m_1' = \frac{1}{10}\sum x = \frac{5.74}{10} = 0.574$$

Now, by solving $\frac{1+\theta}{2+\theta} = 0.574$, we get the MM estimate for $\hat{\theta}$ as 0.3474

Example 2.6: Obtain estimate of α and β for the Pearson type III distribution

$f(x;\alpha,\beta) = \frac{\beta^{\alpha}}{\Gamma\alpha} x^{\alpha-1}e^{-\beta x}, 0 < x < \infty$ by method of moments based on a random sample size n.

Solution: Here, r-th raw moments is

$$f(x;\alpha,\beta) = \frac{\beta^{\alpha}}{\Gamma\alpha} x^{\alpha-1}e^{-\beta} \; , 0 < x < \infty$$

We get

$$\frac{\mu_2'}{(\mu_1')^2} = \frac{\alpha+1}{\alpha} = 1 + \frac{1}{\alpha} \Rightarrow \alpha = \frac{\mu_1'^2}{\mu_2 - \mu_1^2} \; and \; \beta = \frac{\alpha}{\mu_1} = \frac{\mu_1}{\mu_2 - \mu_1'^2}$$

You may attempt the following problems:

E-2.1 Let X_1, X_2, ……....X_n be a sample from a population with pdf

$$f(x;a,b) = \begin{cases} \frac{1}{b-a}, & a \le x \le b \\ o, & otherwise \end{cases}$$

Obtain estimator of a and b by method of moments.

E-2.2 Let $X_1, X_2, \ldots\ldots X_n$ be i.i.d. b(x; n', p) random variable, where both n' and p are unknown. Obtain the estimates of n' and p by method of moments.

E-2.3 Let $X_1, X_2, \ldots\ldots X_n$ be a random sample from N (μ, σ^2). Obtain the method of moments estimates for (μ, σ^2).

E- 2.4 Let $X_1, X_2, \ldots\ldots X_n$ be a random sample of size 10 from N $(3, \sigma^2)$.Obtain the method of moments estimates for σ^2.

E- 2.5 Let $X_1, X_2, \ldots\ldots X_n$ be a random sample from a Poisson population P (x;λ) having pmf p(x) $\frac{e^{-\lambda}\lambda^x}{x!}$, x= 0,1,2,……… Obtain the method of moments estimates for λ.

E- 2.6 Let $X_1, X_2, \ldots\ldots X_n$ be a random sample from an exponential population

$$f(x; A, \theta) = \begin{cases} \frac{1^{\frac{x-A}{\theta}}}{\theta^e}, for\ x \geq A \\ o, \quad elsewhere \end{cases}$$

Find out estimates of A and θ by method of moments.

12.3.2 Method of Maximum Likelihood (MLE)

C.F. Gauss had initially formulated the maximum likelihood method of estimation. Prof. R.A. Fisher develop this technique as a general method of estimation by showing that it yields a sufficient estimator whenever it exists and that the estimators are asymptotically minimum variance unbiased estimator.

Definition 2.1: Likelihood Function

Let $x_1, x_2, \ldots\ldots x_n$ be the observed values of the random sample $X_1, X_2, \ldots\ldots X_n$ drawn from the population with pmf (pdf) f(x, θ).

In the discrete case, the probability of getting this sample observation is L=L(θ)= P [X_1=x_1 X_2=x_2…....X_n= X_n] = f (x_1, x_2, ……....x_n, θ)

$$= \prod_{i=1}^{n} f(x, \theta)[Since\ X_1, X_2, \ldots\ldots X_n\ are\ i.i.d.random\ variable] \quad (2.7)$$

$X_1, X_2, \ldots\ldots X_n$ at X_1=x_1 X_2=x_2…....X_n= X_n. Since the sample values $x_1, x_2, \ldots\ldots x_n$ are observed and therefore fixed number, Hence L is a function of θ.

We refer to the function (2.7) as a function of and is denoted by L(θ). In the continuous case $f(x_1, x_2, \ldots\ldots x_n; \theta)$ is the joint pdf of random variables $X_1, X_2, \ldots\ldots X_n$ therefore the likelihood function is

L= L(θ) =f (x_1, x_2, ……..x_n; θ)= $\prod_{i=1}^{n} f(x, \theta)$ (2.8)

Which is the relative likelihood that the random variables X_1, X_2, ……....X_n assumes the particular set of values x_1, x_2, ……..x_n.

The principle of maximum likelihood method of estimation (MLE) consists of choosing an estimator$\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2 \dots \dots \hat{\theta}_k)$ of the unknown parameter.

$\underline{\theta} = (\theta_1, \theta_2 \dots \dots \theta_k)$ which maximizes the likelihood function L(θ) for variations in $\underline{\theta}$.

Procedure:

- Draw a random sample X_1, X_2, ……....X_n from the population under consideration $f(x, \theta), \theta \in \Theta$ where θ is a single unknown parameter.
- The likelihood function for the observed sample values x_1, x_2, ……..x_n will be L(θ) = L($x; \theta$)= $f(x_1, \theta), f(x_2, \theta)$…….. $f(x_n, \theta)$
- The maximum likelihood principal suggests to choose that value of the estimator of in the admissible range of which maximizes L(θ).

Thus $\hat{\theta} = \hat{\theta}$ (x_1, x_2, ……..x_n) is said to be a m.l.e. of θ if

L($\hat{\theta}$)> L($\hat{\theta}$) for all $\theta \in \Theta$

⇒ L($\hat{\theta}$)= sup L(θ) for all $\theta \in \Theta$ (2.9)

- Hence if L(θ) is twice differentiable (i.e. if the first and second derivative of L(θ) and then differentiate it and solve it by equating it to zero since

$$\frac{dL(\theta)}{d\theta} = 0 \; such \; that \; \frac{d^2L(\theta)}{d\theta^2}|_{\theta=\hat{\theta}} < 0 \qquad (2.10)$$

- In practice the estimation process becomes easier if one takes the logarithm of L (θ) and then differentiate it and solve it by equating it to zero, since

$\frac{1}{L}\frac{dL(\theta)}{d\theta} = \frac{d \, log_e \, L(\theta)}{d\theta}$ Being a probability function L(θ)>0 and

$log_e \; L(\theta)$is a non decreasing function of L. Therefore $log_e \; L(\theta)$land $L(\theta)$attain extreme values (maxima or minima) at the same value of θ . Hence solving eqn. (2.10) is equivalent to solving $\frac{d \, log_e \, L(\theta)}{d\theta}$=0 such that

$$\frac{d^2 log_e \, L(\theta)}{d\theta^2}|_{\theta=\hat{\theta}} < 0 \qquad (2.11)$$

If $\Theta=(\theta_1, \theta_2 \dots \dots \theta_k)$ be a k- dimensional parametric vector. Then the estimator $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2 \dots \dots \hat{\theta}_k)$ which maximizes.

L (x; $\theta_1, \theta_2 \dots \dots \theta_k$) will be obtained by differentiating partially

{L (x; $\theta_1, \theta_2 \dots \dots \theta_k$)}w.r.t. $\theta_1, \theta_2 \dots \dots \theta_k$ respectively and equating it to zero.

The estimation $\widehat{\theta} = (\hat{\theta}_1, \hat{\theta}_2 \dots \dots \hat{\theta}_k)$ so obtained by solving the k simultaneous equations.

$$\frac{\partial}{\partial \theta_1}\{log_e \text{ L (x; } \theta_1, \theta_2 \dots \dots \theta_k)\} = 0$$

$$\frac{\partial}{\partial \theta_2}\{log_e \text{ L (x; } \theta_1, \theta_2 \dots \dots \theta_k)\} = 0$$

……………..

………………

$$\frac{\partial}{\partial \theta_k}\{log_e \text{ L (x; } \theta_1, \theta_2 \dots \dots \theta_k)\} = 0$$

In k unknowns are known as maximum likelihood estimates (m.l.e.) of $\widehat{\theta} = (\hat{\theta}_1, \hat{\theta}_2 \dots \dots \hat{\theta}_k$ provided the matrix $\left(\frac{\partial^2 log_e \text{L}}{\partial \theta_i \, \partial \theta_j}\right)_{\theta=\widehat{\theta}}$ is negative definite.

Now, we shall explain this procedure through some illustrations

Example 2.7: Obtain the MLE for parameters (μ, σ^2) is normal population N (μ, σ^2) when

(a) σ^2 is known
(b) μ is known
(c) $\mu \; and \; \sigma^2$ are unknown

Solution: $f(x;\ \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\, \sigma^2}} e^{\frac{1}{2\,\sigma^2}(x-\mu)^2} - \infty < n < \infty, -\infty < \mu < \infty,\ \sigma^2 > 0$

X_1, X_2, ……....X_n be a random sample of size n from N (μ, σ^2) population. Then the likelihood function

L= L (μ, σ^2)

$$= \prod_{i=1}^{m} f(x;\ \mu, \sigma^2) = \frac{1}{(2\mu\, \sigma^2)^{\frac{n}{2}}} e^{\frac{1}{2\sigma^2}\Sigma(x_i-\mu)^2} = (2\mu\, \sigma^2)^{\frac{n}{2}}\ exp\left[-\frac{n}{2\sigma^2}\{(\bar{x}-\mu) + S^2\}\right]$$

Where $S^2 = \frac{1}{n-1}\sum(x_i - \bar{x})^2$ and $\bar{x} = \frac{1}{n}\sum x_i$

We gives log_e

$$L = -\frac{n}{2}log_e(2\pi) - \frac{n}{2}log_e \sigma^2 - \frac{n}{2\sigma^2}[(\bar{x}-\mu) + S^2]$$

Part (a) if $\sigma^2 = \sigma_0^2$ is known

$$log_e L = const. - \frac{n}{2}log_e \sigma_0^2 - \frac{n}{\sigma_0^2}[(\bar{x}-\mu) + S^2]$$

Differentiating w.r.t μ and equating the derivative to zero we get the likelihood equation as

$$\frac{d\ log_e\ L}{d\mu} = 0 \Rightarrow -\frac{n}{2\sigma_0^2} = 0 \Rightarrow -\frac{n}{2\sigma_0^2}\{2(\bar{x}-\mu)(-1)\} = 0 \Rightarrow \frac{\bar{x}-\mu}{\frac{\sigma_0^2}{n}} = 0$$

$\Rightarrow \mu = \bar{x}$

So that maximum likelihood estimator (MLE) of μ is $\hat{\mu} = \bar{X} = \frac{1}{n}\sum X_i$ and maximum likelihood estimate (m.l.e) of μ is $\hat{\mu} = \bar{X} = \frac{1}{n}\sum X_i$

In fact

$$Var\ (\bar{X}) = \frac{\sigma_0^2}{n}. Here\ \frac{\partial^2 log_e\ L}{\partial \mu^2}|_{\mu=\bar{x}} = -\frac{n}{\sigma^2} < 0.$$

Part (b) if $\mu = \mu_0$ is known

$$log_e L = const. - \frac{n}{2}log_e \sigma^2 - \frac{n}{2\sigma^2}[(\bar{x}-\mu) + S^2]$$

Differentiating w.r.t σ^2 and equating the derivative to zero we get the likelihood equation in σ^2as $\frac{d\ log_e\ L}{d\mu} = 0$

Or

$$-\frac{n}{2\sigma^2} + \frac{n}{2\sigma^4}[(\bar{x}-\mu_0) + S^2] = 0$$

Or

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(\bar{x}-\mu_0)^2 = \frac{1}{n}\left[\sum_{i=1}^{n}(x_i + \bar{x} + \bar{x} - \mu_0)^2\right] = \frac{1}{n}[(n-1)S^2 + n(\bar{x}-\mu_0)^2]$$

As the maximum likelihood estimate of σ^2, since

$$\frac{d^2 log_e\ L}{d(\sigma^2)^2} = \frac{n}{2\sigma^4} - \frac{n}{2\sigma^6}\{(\bar{x}-\mu_0)^2 + S^2\}$$

$$\frac{d^2 log_e\ L}{d(\sigma^2)^2}|_{\sigma^2=\sigma^2} = \frac{n}{2\sigma^2} - \frac{n}{\sigma^4} = \frac{n}{2\sigma^4} < 0$$

Further Var

$$(\hat{\sigma}^2) = \frac{2\sigma^4}{n}$$

Part (c) Both μ *and* σ^2 are unknown

Corresponding simultaneous equations are

$$\frac{\partial^2 log_e\ L}{\partial\mu} = 0 \Rightarrow \frac{n}{\sigma^2}\{2(\bar{x}-\mu)(-1)\} = 0$$

And

$$\frac{\partial^2 log_e\ L}{\partial\sigma^2} = 0 \Rightarrow \frac{n}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2 - \frac{n}{2\sigma^2} = 0$$

Which gives the m.l.e. of μ *and* σ^2as

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n}X_i = \bar{X}\ and\ \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2 = m_2 = S^2$$

$$Since\left(\frac{\partial^2 log_e \mathrm{L}}{\partial\sigma^2\ \partial\mu}\right)_{\sigma^2=S^2,\ \ \mu=\bar{x}} = \frac{1}{\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2|_{\sigma^2=S^2,\ \ \mu=\bar{x}} = -\frac{1}{S^2}$$

$$\left.\frac{\partial^2 log_e \mathrm{L}}{\partial\mu^2}\right]_{\hat{\mu}=\bar{x},\sigma^2=S^2} = \frac{1}{S^2}\sum_{i=1}^{n}2(x_i-\bar{x}) = 0$$

$$\left.\frac{\partial^2 log_e \mathrm{L}}{\partial(\sigma^2)^2}\right]_{\mu=\bar{x},\sigma^2=S^2} = -\frac{N}{(S^2)^2} + \frac{N}{2(S^2)^2} = \frac{N}{2(S^2)^2}$$

Giving,

$$\begin{vmatrix} \frac{\partial^2 log_e \mathrm{L}}{\partial \mu^2} & \frac{\partial^2 log_e \mathrm{L}}{\partial \sigma^2\, \partial \mu} \\ \frac{\partial^2 log_e \mathrm{L}}{\partial \sigma^2\, \partial \mu} & \frac{\partial^2 log_e \mathrm{L}}{\partial (\sigma^2)^2} \end{vmatrix} = \begin{vmatrix} 0 & -\frac{1}{S^2} \\ -\frac{1}{S^2} & -\frac{N}{2(S^2)^2} \end{vmatrix} = -\frac{1}{(S^2)^2} < 0$$

It may be noted that the ML estimate of is the sample mean $\bar{x}$ which is unbiased consistent and Best Asymptotically Normal (BAN) estimate, but the estimate of σ^2 is $S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$ which is not unbiased but consistent. Further $\bar{x}$ is minimum variance bound estimate of μ with variance $\sigma^2/n\bar{x}$ Again, the estimators and $S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$ are jointly sufficient for and σ^2.

Example 2.8: Consider a binomial population: X ~ b (x; n,p) where n is known and p is unknown. Let X_1, X_2, ……....X_n be a random sample of size N drawn from this population. Obtain the maximum likelihood estimator of p.

Solution: The likelihood function is

$L(p) = {}^{n}C_{x1}, {}^{n}C_{x2}, \ldots\ldots {}^{n}C_{kn}, p^{(\sum_{i=1}^{n} x_i)}\ (1-p)^{Nn-\sum_{i=1}^{n} x_i}$

So that,

$$log_e \mathrm{L(p)} = \log[\mathrm{n_{C_{x1}}, n_{C_{x2}}, \ldots \ldots n_{C_{kn}}}] + \left(\sum_{i=1}^{n} x_i\right)\log \mathrm{p} + \left(Nn - \sum_{i=1}^{n} x_i\right)\log(1-\mathrm{p})$$

$$The\ likelihood\ equation\ is\ \frac{\partial^2 log_e \mathrm{L(p)}}{\partial \mu^2} = 0$$

Or

$$\frac{1}{p}\left(\sum_{i=1}^{n} x_i\right) - \left(Nn - \sum_{i=1}^{N} x_i\right)\frac{1}{(1-\mathrm{p})} = 0$$

Or

$$p\left(\sum_{i=1}^{n} x_i\right) + 1(1-\mathrm{p})\left(\sum_{i=1}^{N} x_i\right) = \mathrm{pNn}$$

Or

$$\hat{p} = \frac{1}{Nn}\sum_{i=1}^{N} x_i \quad or\ \hat{p} = \frac{\bar{x}}{n}, \quad where\ \bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

Since the sample mean $\bar{x}$is an unbiased consistent and sufficient estimate of the population mean μ therefore it implies that $\hat{p}$ is an unbiased and consistent estimate of p.

Example 2.9: Suppose that the random variable X has pdf

$$f(x) = \begin{cases} cx^{\alpha}, & for < 0 < x < 1 \qquad c\ is\ content \\ 0, & elsewhere \end{cases}$$

Obtain the maximum likelihood estimate of α if $X_1, X_2, \dots\dots\dots X_n$ be a random sample of size N from the population.

Solution: If f(x) is a pdf, then

$$\int_{-\infty}^{\infty} f(x)dx = \int_{0}^{1} f(x)dx = \int_{0}^{1} cx^{\alpha}dx - 1 \Rightarrow c = \alpha + 1$$

Likelihood function is:

$$L\,(\alpha) = \prod_{i=1}^{n} f(x_i) = (\alpha + 1)^2\ x_1^{\alpha} x_2^{\alpha} \dots \dots \dots x_n^{\alpha}$$

So that,

$$log_e\, L(\alpha) = n\ log_e\,(\alpha + 1) + \alpha \sum_{i=1}^{n} log_e\, x_i$$

Likelihood equation: $\frac{\partial log_e\, L(\alpha)}{\partial \alpha} = 0$

Or

$$\frac{n}{(\alpha + 1)} + \sum_{i=1}^{n} log_e\, x_i = 0$$

Giving m.l.e. of α as

$$\hat{\alpha} = -1 - \frac{n}{\sum_{i=1}^{n} log_e\, x_i}$$

You may try the following problems

E 2.7 If X is a Poisson random variable with parameter λ, obtain the MLE of λ based on a random sample $X_1, X_2, \dots\dots\dots X_n$ from the population.

E-2.8 Assuming known obtain the maximum likelihood estimate of for Gamma G (x; α,λ) whose pdf is given by

E 2.9 A random Variable X has a distribution with the density function f(x)= (α+1) x^{α} for 0<x<1

= 0 otherwise

A random sample of size 8 produces the data

0.2, 0.4, 0.8, 0.5, 0.7, 0.9, 0.8, 0.9

Obtain the m.l.e. of the unknown parameter α. It is given that

$$log_e\ 0.0145152 = -4.2326.$$

E 2.10 Determine the m.l.e of the parameter λ of the Weibull distribution $f\ (x) = \lambda\alpha x^{\alpha-1}e^{-\lambda x^{\alpha}}$ for x > 0 using a sample of size n, assuming that α is known.

So far we have considered the determination of the MLE of the parameters where the density $f(x;\theta)$ has range independent of the unknown parameter θ. If the range of the distribution is not independent of unknown parameter θ, the differentiation method fails to give a maximum of L (θ). We apply other method to obtain mle of θ, which maximizes L (θ).

Example 2.10 A random sample $x_1, x_2, \ldots\ldots x_n$ of n independent observations is drawn from the rectangular population

$$f(x;\alpha,\beta) = \frac{1}{\beta-\alpha}, \alpha < x < \beta < \infty$$

$$= 0\ ,\ \ otherwise$$

Obtain the maximum likelihood estimates for α and β.

Solution: The likelihood function of observed values $x_1, x_2, \ldots\ldots x_n$ is

$$L = L(\alpha,\beta) = \prod_{i=1}^{n} f(x;\alpha,\beta) = \frac{1}{(\beta-\alpha)^n}$$

And $log_e\ L(\alpha,\beta) = n\ log_e\ L(\beta-\alpha)$

The likelihood equations for α and β are $\frac{\partial log_e L}{\partial\alpha} = 0\ and\ \frac{\partial log_e L}{\partial\beta} = 0$ which leads to $\frac{n}{(\beta-\alpha)} = 0\ and\ \frac{n}{(\beta-\alpha)} = 0$

This implies the $(\beta-\alpha) \rightarrow \infty$, which is impossible

Here, the differentiation method to obtain maxima of $L(\alpha, \beta)$ fails. We have to adopt the logical approach to get MLE for α, β .

The likelihood function $L(\alpha, \beta) = \frac{1}{(\beta-\alpha)^n}$ attains maxima if $(\beta - \alpha)$ is minimum; that is β takes the maximum possible value and α takes the maximum possible value in the sample.

Let $\alpha \leq x_{(1)} \leq x_{(2)} \leq \cdots .. x_{(n)} \leq \beta$ be the ordered sample corresponding to the observed sample $x_1, x_2, \ldots\ldots\ldots x_n$.

Then $x_{(1)} = min$ ($x_1, x_2, \ldots\ldots\ldots x_n$) and $x_{(n1)} = max$ ($x_1, x_2, \ldots\ldots\ldots x_n$) will be the maximum possible value of α and possible value of β respectively.

Hence $L(\alpha, \beta) = \frac{1}{(\beta-\alpha)^n}$ is maximum if $\beta = x_{(n)}$ and $\alpha = x_{(1)}$.

Therefore, MLE for α and β are given by

$\hat{\alpha} = x_{(1)} = min$ ($x_1, x_2, \ldots\ldots\ldots x_n$) = smallest sample observations

$\hat{\beta} = x_{(n1)} = max$ ($x_1, x_2, \ldots\ldots\ldots x_n$) = largest sample observation, respectively.

Now you may attempt the following problems:

E-2.11 Obtain the maximum likelihood estimate of θ based on a random sample of size n form the population

$$f(x; \alpha, \beta) = \begin{cases} if\ \theta - \frac{1}{2} \leq x \leq \theta + \frac{1}{2} \\ 0, \qquad elsewhere \end{cases}$$

E- 2.12 The life-time of an electronic device has a pdf $(x; \theta) = 3\theta^3 y^{-4}, y \geq \theta$. For a random sample $X_1, X_2, \ldots\ldots\ldots X_n$ of size n from this population, obtain the MLE of θ.

E-2.13: A random sample $X_1, X_2, \ldots\ldots\ldots X_n$ of size n is drawn from the population having pdf

$$f(x; a, b) = \frac{1}{b} e^{-\frac{1}{b}(x-a)^2} \qquad for - \infty < x < \infty, -\infty < a, b > 0$$

$$= 0, \qquad elsewhere$$

Obtain the MLE for a and b.

E-2.14 Let X~b (1p), $p \in \left[\frac{1}{4}, \frac{1}{3}\right]$. Obtain the maximum likelihood estimate of p.

E-2.15 Let $X_1, X_2, \ldots\ldots\ldots X_n$ be i.i.d. Bernoulli variables having pmf p(x) = $p^x (1-p)^{1-x}, x = 0,1;\ \ 0 < 1p < 1$

$= 0, \quad otherwise$

Obtain the MLE for p.

What happens if either (a) (0,0,….,0) (b) (1,1,….1) is our observed sample?

12.3.3 Method of Scoring

It may sometimes happen that the maximum likelihood leads to complicated equations. In such case we may employ the iterative or successive approximation (Newton Raphson) method to get the solution. The method is known as 'Method of Scoring". The estimator so obtained is consistent but less efficient. Some correction is to be applied to bring the solution nearer to the desired form.

If t= t $(x_1, x_2, \dots . x_n)$ be MLE for θ and $t' = t'(x_1, x_2, \dots . x_n)$be another estimate of θ which is less efficient. Then$\left.\frac{\partial log_e \mathrm{L}}{\partial \theta}\right]_{\theta=t}$ = 0 and E $\left.\frac{\partial^2 log_e \mathrm{L}}{\partial \mu^2}\right]_{\theta=t} = -\frac{1}{Var(t)'}$ at least for large n.

By Taylor's expansion

$$\left.\frac{\partial log_e \mathrm{L}}{\partial \theta}\right]_{\theta=t'} = \left.\frac{\partial log_e \mathrm{L}}{\partial \theta}\right]_{\theta=t} + (t' = t')\left.\frac{\partial^2 log_e \mathrm{L}}{\partial \theta^2}\right]_{\theta=t} + tems\ of\ higher$$

Since $(t' = t')$ is small and $\left.\frac{\partial^2 log_e \mathrm{L}}{\partial \theta^2}\right]_{\theta=t} \cong E\left.\frac{\partial^2 log_e \mathrm{L}}{\partial \theta^2}\right]_{\theta=t} = -\frac{1}{Var(t)'}$

Therefore, we have$t' \cong t' + Var\ (t)\left.\frac{\partial log_e \mathrm{L}}{\partial \theta}\right]_{\theta=t'}$

Example 2.11 Obtain an approximate estimator of the parameter θ for the Cauchy distribution having pdf

$$f(x|\theta) = \frac{1}{\pi} + \frac{1}{1 + (x - \theta)^2}\ for - \infty < x < \infty$$

Solution: The likelihood function for observed random sample $(x_1, x_2, \dots . x_n)$ is given by L=L($\theta|x$)=$\prod_{i=1}^{n}\left[\frac{1}{\pi} + \frac{1}{1+(x-\theta)^2}\right]$

So that,

$$log_e \mathrm{L}(\theta) = -\mathrm{n}log_e\, \pi - \sum_{i=1}^{n}\{1 + (x - \theta)^2\}$$

Therefore,

$$\frac{\partial log_e \mathrm{L}(\theta)}{\partial\theta} = \sum_{i=1}^{n}\left[\frac{(x_i-\theta)^2}{1+(x-\theta)^2}\right].$$

We observe that $\frac{\partial log_e \mathrm{L}(\theta)}{\partial\theta}$ is a polynomial in degree in θ of degree (2n-1) but is difficult to be solved. We know that

$$\frac{1}{Var(t)} = -n\int_{-\infty}^{\infty}\left(\frac{\partial^2 log_e \mathrm{f}}{\partial\theta^2}\right)fdx = \frac{n}{2}\,\frac{n}{n-1}S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

So that Var (t) = 2/n.

Again the sample median $\hat{\theta}$ is an estimator of θ. Therefore efficiency of sample median w.r.t. t is

$$= \frac{Var(t)}{Var(\hat{\theta})} = \frac{\left(\frac{2}{n}\right)}{\left(\frac{\pi^2}{4n}\right)} = \frac{8}{\pi^2} = 0.8\,approximately\,.$$

Here, the appropriate estimator $\tilde{\theta}$ of is θ given by

$$\tilde{\theta} = \tilde{\theta}\,(x_1, x_2, \ldots .x_n) = \mathrm{t}' + \frac{4}{\mathrm{n}}\sum_{\mathrm{i}=1}^{\mathrm{n}}\frac{\mathrm{x_i}-\mathrm{t}}{1+(\mathrm{x_i}-\theta)^2}.$$

12.4 Properties of Estimators obtained by MME & MLE

Some properties of the MME and MLE are outlined below. The proof of a few of few of these properties are also given.

12.4 .1 Properties of MME

Prop. 2.1 MME's are consistent estimators.

Proof: Let (X_1, X_2,….X_n) be a random sample of size n form the population with density function f(x; θ). Therefore X_1, X_2,….X_n are i.i.d. r.v.'s with common pdf f(x;θ). So that $X_1^j, X_2^j \ldots\ldots X_r^j$ are i.i.d.r.v.'s. Thus if u_j = E(X_r^j) exists, then by WLLN we have

$$m_j = \frac{1}{n}\sum_{i=1}^{n} X_r^j \xrightarrow{P} \mathrm{E}(X_r^j) = \mu_j \quad \text{for j} = 1{,}2\,\ldots$$

i.e. the sample moments are consistent estimators of the corresponding population moments.

The MME of k unknown parameters $\theta_1, \theta_2 \dots \dots \theta_k$are obtained by solving first K simultaneous equations in k parameters $\theta_1, \theta_2 \dots \dots \theta_k$ in terms of $\mu_1, \mu_2 \dots \dots \mu_k$and then replacing these moment μ_r by sample moments $m_1, m_2 \dots \dots m_k$.

Hence $\hat{\theta}_j \xrightarrow{P} \theta_j \;\; for\ j = 1,2, \dots k$, hence the property.

Prop 2.2 Under quite general conditions the estimates obtained by method of moments are asymptotically normal.

Prop. 2.3 In general, the estimators obtained by method of moments are less efficient.

Prop 2.4 the estimates obtained by method of moments (MME) are less efficient than the those obtained by method of maximum likelihood (MLE)

Example 2.12 Let us consider U(0,θ) population having pdf

$$f(x) = \frac{1}{\theta}, 0 \le x \le \theta \quad = 0, otherwise$$

$X_1, X_2, \dots X_n$ be a random sample form the population.

Now we have E(X) = θ /2 and Var (X) = $\frac{\theta^2}{12}$

To get MME we consider $\bar{X} = \frac{\theta}{2} \Rightarrow \hat{\theta}_{MME} = 2\bar{X}$

Now Var (X) = $\frac{\theta^2}{12} \Rightarrow var\left(\hat{\theta}_{MME}\right) = var\ (2\bar{X}) = 4Var\ (\bar{X}) = \frac{\theta^2}{3n}$

Here we know that $\hat{\theta}_{MLE} = X_n = Max\ (\ X_1, X_2 \dots \dots X_k)$ The distribution of X_n is given by

$$fX_n(y) = \frac{ny^{n-1}}{\theta^n}, 0 \le y \le \theta$$

So that,

$$E(X_n) = \frac{n\theta}{n+1}\ and\ Var\ (X_n) = \frac{n\theta^2}{n+2} - \frac{(n\theta)^2}{(n+1)^2}$$

Giving

$$MSE\ X_{(n)} = \frac{2\theta^2}{(n+1)(n+2)}$$

Here $MSE\ X_{(n)} < Var\ (\hat{\theta}_{MME})$ i.e. Mme is less efficient than MLE.

Prop. 2.5 The two estimator MME and MLE are identical if the help p.d.f. (or p.m.f.) of the parent population f(x; θ) is of the form

$f(x;\theta) = \exp[b_0 + b_1 x + b_2 x^2 + \cdots.]$ where$b_j{}'$ s are independent of x but may dependent on $\underline{\theta} = (\theta_1, \theta_2 \dots \dots \theta_k)$

Note: In this case, the M.L.E.'s may be obtained as linear function of sample moments.

Remarks: $f(x;\theta) = \exp[b_0 + b_1 x + b_2 x^2 + \cdots.]$

$\Rightarrow L(x_1, x_2 \dots \dots x_k; \theta) = \exp[nb_0 + b_1 \sum x_i + b_2 \sum x_i^2 + \cdots.]$

$$\Rightarrow \frac{\partial log_e \mathrm{L}}{\partial \theta_j} = a_0 + a_1 \sum x_i + a_2 \sum x_i^2 + \cdots..$$

Thus both the methods yields identical estimators if MLE's are obtained as linera functions of the moments.

12.4.2 Properties of MLE

Regularity Conditions of MLE:

The following conditions are known as the regularity conditions of the maximum likelihood estimators. The students are advised to only note them. In the discussion rigorous proofs of the theorems are avoided.

(i) $X_1, X_2 \dots \dots X_n$ are mutually independent observations.

(ii) The distribution function f(x| θ) admits a pdf f(x| θ).

(iii) The first and second derivatives, $\frac{\partial log_e \mathrm{L}}{\partial \theta}$ and $\frac{\partial^2 log_e \mathrm{L}}{\partial \theta^2}$ viz exist and are continuous functions of θ in the range R (including true value θ_0 of the parameter θ) for almost all x.

(iv) Further for every $\theta \in$ R $\left|\frac{\partial log_e \mathrm{L}}{\partial \theta}\right| < F_1(x) and$ $\left|\frac{\partial^2 log_e \mathrm{L}}{\partial \theta^2}\right| < F_2(x)$are integrable function over $(-\infty, +\infty)$.

(v) The third order derivates $\frac{\partial^3 log_e \mathrm{L}}{\partial \theta^3}$ exists, such that $\left|\frac{\partial^3 log_e \mathrm{L}}{\partial \theta^3}\right| < M(x)$where E [M(X)] < K, a positive quality

(vi) For every $\theta \in$ R, E $\left(-\frac{\partial^2 log_e \mathrm{L}}{\partial \theta^2}\right) = \int_{-\infty}^{\infty} \left(\frac{\partial^2 log_e \mathrm{L}}{\partial \theta^2}\right) Ld\underline{x} = i(\theta)$is finite and non-zero.

Here I(θ) is called by R.A. fisher as the amount of information on supplied by the sample $(x_1, x_2 \dots \dots x_n)$ and its reciprocal 1/ I(θ) is known as the *information limit*. In fact 1/ I(θ) is the Cramer Rao lower bound for the unbiased estimator of θ.

(iii) The range of integration is independent of θ. But, if the range of integration depends on θ, then f(x; θ) vanishes at the extremes depending on θ. This assumption justifies the differentiation under the integral sign.
Under the above assumptions MLE's possess a number of important properties. Some of them are mentioned below.

Prop. 2.6 (*Cramer, 1946*) with probability approaching unity as $n \to \infty$. the likelihood equation $\frac{\partial log_e \mathrm{L}}{\partial \theta} = 0$ has a solution which converges in probability to the true value θ_0 of the parameters θ. In otherwords the MLE,s are consistent estimators.

Prop 2.7: The MLE's are always consistent (Prp.1) but need not be unbiased.

Example 2.13: In sampling from N(μ, σ^2) population with both μ *and* σ^2unknown we know that $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$ *and* $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = S^2$ are MLE of μ *and* σ^2 respectively.

Here

$$E(\bar{X}) = \mu \; and E(\hat{\sigma}^2) = E(S^2) = \frac{(n-1)\sigma^2}{n} = (1 - \frac{1}{n})\sigma^2$$

Therefore, MLE $\hat{\mu} = \bar{X}$ is a consistent and unbiased estimators of μ while MLE $\hat{\sigma}^2 = S^2$ is a consistent and biases estimator of σ^2.

Prop. 2.8 (*Huzurbazar's Theorem*) Any consistent solution of the likelihood equations provides a maximum of the likelihood with probability tending to unity as the sample size n tends to infinity.

Prop. 2.9 The maximum likelihood estimator is asymptotically normally distributed about the true value θ_0. Thus, $\hat{\theta} \sim N\left(\theta_0 , \frac{1}{\mathrm{I}(\theta_0)}\right)$ *as* $n \to \infty$. It may be noted that

$$Var\left(\hat{\theta}\right) = \frac{1}{I(\theta)} = \frac{1}{\left\{E\left(\frac{\partial^2 log_e \mathrm{L}}{\partial \theta^2}\right)\right\}}.$$

Prop. 2.10: When maximum likelihood estimator exists, it is most efficient.

Prop. 2.11: If a sufficient estimator exists, then maximum likelihood estimator is a function of the sufficient estimator.

Proof: Let t ($x_1, x_2 \dots \dots x_n$) be a sufficient estimator of θ. Then the likelihood function can be expressed as

$$L = \prod_{i=1}^{n} f(x, \theta) = g(t, \theta) h(\, x_1, x_2 \dots \dots x_n) which \; is \; independent \; of \; \theta.$$

Therefore,

(i) $log_e \mathrm{L} = log_e \mathrm{g(t, \theta)} + log_e \mathrm{h}(x_1, x_2 \dots \dots x_n|\mathrm{t})$

(ii) $\frac{\partial log_e}{\partial \theta} = \frac{\partial log_e \mathrm{g(t,\theta)}}{\partial \theta} = \Psi(t, \theta)$ which is a function of t and θ only.

(iii) The MLE of θ is the solution of the likelihood equation

(iv) $\frac{\partial lo \quad _{e\ L}}{\partial \theta} = 0$

(v) $\Rightarrow \Psi(\mathrm{t}, \theta) = 0$

(vi) Therefore $\hat{\theta} = n(t)$ = some function of sufficient estimator t of θ

Hence the property.

Remark From equn. (v) and we obtain t=k($\hat{\theta}$)

⇒The sufficient statistic t of θ is some function of MLE $\hat{\theta}$.

This property is helpful in obtaining the sufficient statistic of θ. It is further used to determine whether a sufficient statstics exists or not. In fact, if $\frac{\partial log_{e\ L}}{\partial \theta}$can be expressed in the for (ii) that is$\frac{\partial log_{e\ L}}{\partial \theta} = \Psi(t, \theta)$ a function of a statistic t and parameter θ alone, then the statistics t can be regarded as a sufficient statistic for θ.

No sufficient statistic of θ exist if $\frac{\partial log_{e\ L}}{\partial \theta}$ cannot be expressed in the form (ii).

This property does not say that an MLE is itself a sufficient statistic although this will always be the case.

Example 2.14: Let $X_1, X_2 \dots \dots X_n$ be a random sample of size n from the uniform population U$[\theta, \theta + 1]$, $\theta \in R$. Show that MLE is not sufficient.

Solution: Here,

$f(x; \theta) = 1, \theta \leq \min x_i \leq \theta + 1$

O= otherwise

It can be shown that (min X_i, max X_i) is jointly sufficient for θ. Any value of θ satisfying max X_i-1< θ<min X_i is an MLE of θ.

In particular, $\min_{1 \leq i \leq n} x_i$ is not an MLE of θ but it is not sufficient.

Theorem If for a given population with pdf f(x; θ), an MVB estimator T exists for θ, then the likelihood equation will have a solution $\hat{\theta}$ =T=T ($X_1, X_2 \dots \dots X_n$)

Proof Since T an MVB estimator of θ therefore,

$$\frac{\partial log_{e\ L}}{\partial\theta} = \frac{T-\theta}{\lambda\theta} = (T-\theta)^{\lambda-1}(\theta)$$

Hence the theorem.

Prop. 2.12 (Invariance property of MLE) Let be $\hat{\theta}$ the MLE of θ and be a one-to-one function of θ . Then MLE of $\Psi(\theta)$ is Ψ $(\hat{\theta})$.

Example 2.15 Let $x_1, x_2 \dots \dots x_n$ be normal distribution with mean θ and known variance σ^2. Obtain the MLE of (i) $\frac{\theta}{4} + 5$ (ii) θ^2

Solution: The MLE is . By the property of invariance, we have

(h) $\hat{\theta}_{MLE} = \frac{\bar{X}}{4} + 5\ and\ (ii)\hat{\theta}^2 MLE = \bar{X}^2$

The maximum likelihood solution attempts to determine the mode of the likelihood L($\theta|\ x_1, x_2 \dots \dots x_n$)over the variation in the value of θ in Θ. In estimation theory, mode is in general inferior to either mean or median for small samples. Therefore the performance of MLE is poor for small samples. However for large samples the mode tends to mean and median provide they exist. Therefore the MLE has many optimum large sample properties such as, Prop. 2.6, 2.7, 2.9 and 2.10 mentioned above. In fact the MLE is a consistent asymptotically normal and asymptotically efficient estimator of $\theta$for large samples.

Bias of MLE

For a finite population MLE $\hat{\theta}$ is biased estimate of parameter θ for f(x; θ), $\theta \in \Theta$. However a slight modification in may eliminate the bias and we may get an unbiased estimate $\hat{\theta}_1$ of θ.

Example 2.16 Consider a random sample $x_1, x_2 \dots \dots x_n$of size n form the normal population N(μ, σ^2) with both $\mu\ and\ \sigma^2$unknown.

Then MLE of $\mu\ and\ \sigma^2\ are\ \hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$ and

$$\hat{\mu}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i = \bar{X})^2 = S^2, respectively.$$

$$E(\hat{\mu}) = \mu\ and\ E(\hat{\sigma}^2) = \left(\frac{n-1}{n}\right)\sigma^2, i.e. s^2\ is\ biased\ for\ \sigma^2$$

$$But,\quad E\left(\frac{n}{n-1}\hat{\sigma}^2\right) = \sigma^2$$

$$\frac{n-1}{n}s^2 = S^2 = \frac{1}{n-1}\sum_{i-1}^{n}(x_i = \bar{x})^2 \; is\ an\ unbiased\ estimator\ of\ \sigma^2.$$

Hence$\hat{\mu} = \bar{X}$and $\frac{1}{n-1}\sum_{i-1}^{n}(x_i = \bar{x})^2$ .are the unbiased estimates of μ and σ^2respectively.

Example 2.17: Consider a set of n Bernoulli trials with probability of success p. The pmf is $(x_i|p) = p^{x_i}(1-p)^{1-x}, x_i = 0, 1; 0 < p < 1$. obtain the MLE of p.

Solution: The likelihood function of sample observations $(x_1, x_2 \dots\dots x_n)$ is $L(p) = L(p|\, x_1, x_2 \dots\dots x_n) = p^{\sum x_i}(1-p)^{n-\sum x_i}$

So that $\frac{\partial log_e L\,(p)}{\partial p} = \frac{\sum x_i}{p} - \frac{n-\sum x_i}{p}$

The likelihood equation is $\frac{\partial lo\quad _e L\,(p)}{\partial p} = 0 \Rightarrow \frac{\sum x_i}{p} - \frac{n-\sum x_i}{p} = 0$ giving solution as $\hat{p} = \frac{1}{n}\sum x_i = \bar{x}\ provided\ 0 < \sum x_i < n$

The MLE of p is $\hat{p} = \bar{x}$.

Example 2.18: Let $x_1, x_2 \dots\dots x_n$be a random sample drawn from the population having pdf $f(x|\theta) = \frac{1}{2}e^{-[x-\theta]}, -\infty < x < \infty$. Obtain the MLE for θ.

Solution: The likelihood function is $L = \left(\frac{1}{2}\right)^n e^{-\sum|x_i-\theta|}$maximum, if $\sum_{i=1}^{n}|x_i - \theta|$ is minimum.

Hence, the MLE of θ is $\hat{\theta}$= Median of $x_1, x_2 \dots\dots x_n$

You may attempt the following problems:

E-2.16 Let $x_1, x_2 \dots\dots x_n$ be a random sample from B(α, β). Obtain the method of moment estimators of (α, β).

E-2.17 Let $X_1, X_2 \dots\dots X_n$ be a random sample from Poisson population P(x; λ) with parameterλ. obtain the method of moment (MME) and maximum likelihood (MLE) Estimates of λ.

E-2.18 Let $X_1, X_2 \dots\dots X_n$ be an Bernoulli trials with parameter $p(0 \le p \le 1)\ and\ \Psi(p) = p(1-p)$ be a function of p. Obtain the MLE of Ψ (p).

E 2.19 Obtain the MLE of θ based on a random sample of size n from the population $f(x;\theta) = e^{-[x-\theta]}, 0 \le x < \infty, -\infty < \theta < \infty$

E 2.20 Let $X_1, X_2 \ldots \ldots X_n$ be i.i.d binomial variates b(k,p). Obtain estimate of k & p using method of moments.

12.5 Solutions/ Answers

Solution/ answers of some of the Exercises are given below.

E 2.1 $T_1 = \bar{X} - \sqrt{\frac{3}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}$ *and* $T_2 = \bar{X} + \sqrt{\frac{3}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}$ are MME of a and b, respectively

E 2.2 T_1 and T_2 are the MME for p and n', respectively where

$$T_1 = T_1(X_1, X_2 \ldots \ldots X_n) = \frac{\bar{X}}{T_2(X_1, X_2 \ldots \ldots X_n)}$$

$$T_2 = T_2(X_1, X_2 \ldots \ldots X_n) = \frac{(\bar{X})^2}{\bar{X} + (\bar{X})^2 - \frac{1}{n}\sum_{i=1}^{n} X_i^2} \; where \; \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

E 2.3 $\hat{\mu} = \bar{X}$ and $\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i-1}^{n}(X_i - \bar{X})^2$

E 2.4 $\hat{\sigma}^2 = \frac{1}{10}\sum_{i-1}^{10}(x_i - 3)^2$

E 2.5 $\hat{\lambda} - \bar{X}$

E 2.6 $\hat{\theta} = \bar{X} - \hat{A}$ *and* $\hat{A} = X_{(1)} = \min(X_1, X_2 \ldots \ldots X_n)$

E 2.7 $\hat{\lambda} - \bar{X}$

E 2.8 $\hat{\alpha} = \frac{n\lambda}{\sum x_i}$

E 2.9 $log_e(x_1, x_2 \ldots \ldots x_n) = -\frac{n}{\alpha+1}$ *gives MLE of* α *and* $\hat{\alpha} = \frac{3.7674}{4.2336}$

E 2.10 $\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} x_i^{\alpha}}$

E 2.11 Every statistics t = t ($x_1, x_2 \ldots \ldots x_n$) such that $x_{(n)} - \frac{1}{2} \leq t \leq x_{(1)} + \frac{1}{2}$ provides an MLE for θ.

E 2.12 $L(\theta) = 3^n\theta^{3n}[\prod_{i=1}^{n} y_i^{-4}]$ and likelihood equation

$$\frac{\partial log_e L(\theta)}{\partial p} = 0 \Rightarrow \frac{3n}{\theta} = 0 \; which \; does \; not \; yield \; a \; solution$$

$\hat{\theta} = Min(Y_1, Y_2 \ldots \ldots Y_n) = Y_{(1)}$

E 2.13 $\hat{\alpha} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \; and \; m_2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$ are the estimators of $\alpha \; and \; \beta$ respectively.

E 2.14 $L(p) = p^x(1-p)^{1-x}, for \; x = 0{,}1$. We can not differentiate L(p) to get MLE of p, since it gives $\hat{p} = x \; where \; x \nexists \left[\frac{1}{4}, \frac{1}{3}\right]$.

$$\hat{p}(x) = \frac{1}{4} if \; x = 0$$

L(p) is maximized if = 3/4 if x=1

Thus MLE of p is given by $\hat{p}(xX) = \frac{2X+1}{4}$

E 2.15 For sample (0,0,…….0) $\hat{p}$ = 0 and sample (1,1,…….1) $\hat{p}$ =1.

12.6 Summary

After studying this unit we have learnt about the different procedures of the estimation namely method of moments, method of maximum likelihood fo method of scoring. We also learnt about the properties of Estimators.

12.7 Further Readings

1. Rahtagi V.K. (1984): An Introduction to Probability theory and Mathematical Statistics chapter VIII, IX & X Pub; John Wiley & Sons, New York.
2. Goon A.N., Gupta M.K. & Das Gupta B (1987) Fundamentals of Statistics Vol. I The World Press Pvt. Ltd., Kolkata.
3. Kapoor V.K. & S.C. Saxena: Fundamentals of Mathematical Statistics, Chapter Seventeen, Pub: S. Chand.

Unit-13: Testing of Hypotheses

Structure

13.1 Introduction

There are two major areas of statistical inference namely the estimation of parameter and the testing of hypotheses. Our present aim is to introduce to concepts involved in the development of general methods for testing of hypotheses. Some illustrations are taken from population having the common known distributions. In all the problems of statistical inference there is generalization of the results or conclusions of a sample(s) from the population to the population itself. The error is possible. We shall explain the two kinds of error in the context. In testing to hypotheses a decision is taken on the basis of a samples (s) whether to accept or to reject a specified value Ho: $\theta = \theta_0$ or a set of specified values Ho: $\theta \in \theta_0$ where $Q_0 C\ \theta$ On the basis of the results of sample (s) from the population f(x; θ), $\theta \in\ \theta$ where θ may be vector and θ is the parameter space of θ, for example in exponential distribution f(x; θ)= $\theta\ e^{-\theta x}, x > 0, \theta > 0;$ θ is a single parameter and but in normal density f(x; θ)= $\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$

Where ,

$$-\infty < x < \infty, \sigma^2 > 0, -\infty < \mu < \infty$$

Unknown parameter μ and σ^2 are unknown the $\theta = (\mu,\sigma^2)$ and its parameter space is θ: $(-\infty < \mu < \infty, \sigma^2 > 0\ )$.

In the present discussion the sample size n is considered fixed in the advance, The case of sequential analysis, discussed by A. Wald, where are is used are not considered.

Some common situations where of the testing of hypotheses are is used are cited below.

Suppose a new hybrid variety of a grain is introduced. One may wish to estimate the average yield per ace of this variety, (problem of estimation) and if agriculturist decided to test whether the average yield of new variety is higher than the usual sown variety $\mu_1 < \mu_2$ then it is problem of testing of hypothesis. A Farmer may like to decide on the basis of samples whether Nitrogen fertilizer T_1 poorer than (Nitrogen & Potassium) the mixed fertilizer T_2. For the production of wheat. An engineer may want to know whether the average life time of a certain kind of type is at least 50,000 km. or the production manager of a steel factory may like to compare the breaking strength of a steel bar manufactured by process I with that produced by new process II.

A pharmaceutical concern may be interested to find if a new drug is really effective for treatment of an element say cancer or in reducing blood pressure or inducing sleep; One may want whether a new foodstuff is really effective in increasing weight; or which of the two brands of particular product say, food stuff, fertilizers, etc. is more effective; such practical problems may be quoted, where the modern probability theory plays a vital role in decision making and the branch of statistic which helps us in arriving at a criterion for such decision known as testing of hypothesis. The theory of testing of hypothesis was first given a mathematical sound footings by J. Neyman and E.S. Pearson through series of papers. They dealt with the statistical techniques to arrive at decision in certain situations where there is an element of doubt, on the basis of a sample whose size is fixed in advance. There is another technique known as sequential testing pro-founded by Abraham Wald where the sample size is not fixed I advance but is regarded as a random variables.

We shall deal here with Neyman and Pearson Approach only.

13.2 Objectives

After studying this unit you shall be able to:

- Understand testing of hypothesis.
- Decide a null hypothesis and alternative hypothesis.
- Understand the meaning of level of significance, size of a test and power of a test.
- Understand the most powerful (MP) test and uniformly most powerful (UMP) test
- We shall define some terms associated with testing of hypothesis.

13.3 Statistical Hypotheses

Definition 3.1 A statistical hypothesis is some assumption or statement about a population, or probability distribution characterizing the given population. It is frequently denoted by H.

A hypothesis is not accepted without being supported by evidence from the population. A hypothesis needs to be verified and is therefore, put to test; and based on the evidences provided by a random sample (which are set of independent observations) from the population a decisions is taken to accept or reject it. In fact the evidence provided by the sample is the value of a test statistic which will decided the course of the action – to accept or to reject H.

For example if r.v. X ~N (μ, 25) then the statement that the mean of the population is greater than 20, is a statement about the population mean with known variance σ^2 = 25 and therefore is a hypothesis. We write H: μ >20.

13.4 Simple and Composite Hypotheses

Definition 3.2: A hypothesis is known as *simple hypothesis* if it completely specifies the population; otherwise it is known as a *composite hypothesis.*

In sampling from a normal population N (μ, σ^2), the hypothesis.

(i) $Ho: \mu = \mu_0, \quad \sigma^2 = \sigma o^2$

is a simple hypothesis. It specifies values to both parameter μ and σ^2;and therefore, it completely specifies the distribution, On the other hand each of the following hypotheses is composite hypothesis:

(ii) $Ho: \mu = \mu_0$ (No statement about σ^2)

(iii) H: $\sigma^2 = \sigma o^2 \quad (\mu \; is \; not \; specified)$

(iv) H: $\mu < \mu_0, \quad \sigma^2 = \sigma o^2$
(v) $H: \mu > \mu_0, \quad \sigma^2 = \sigma o^2$
(vi) $H: \mu = \mu_0, \quad \sigma^2 > \sigma o^2$
$and \; so \; on$

13.5 Null hypothesis and Alternative Hypotheses

Definition 3.3: A *null hypothesis* is a statistical hypothesis which is put to test for possible rejection under the assumption that it is true; it is denoted by Ho.

For example in sampling from normal population N (μ, σ^2) the hypothesis $Ho: \mu = \mu_0$ is a null hypothesis if it is to be tested. It is said to be a null hypothesis since it states that there is no difference between $Ho: \mu \; and \; \mu o$.

It is very important to state the alternative hypothesis H_1 explicitly in respect to any null hypothesis Ho because the acceptance or rejection of Ho is meaningful only if it is being tested against the rival hypothesis H_1.

The concept of simple and composite hypothesis applies also to alternative hypothesis. For example in comparing the mean effect on the yield of soyabean of two fertilizer say A and B, we may formulate the null and alternative hypothesis as Ho: $\mu A = \mu B$ against H_1: $\mu A < \mu B$. Ho is a simple hypothesis and H_1 is a composite hypothesis .

If we want to test the null hypothesis Ho that the population N (μ, σ^2) has specified mean μ_0 let be σ^2 known then

H_1: $\mu = \mu_0$ is a simple hypothesis and all the possible alternative:

(i) $H_{11}: \mu \neq \mu_0$ $(i.e. \mu > \mu_0$ or $\mu < \mu_0)$
(ii) $H_{12}: \mu > \mu_0$
(iii) $H_{13}: \mu < \mu_0$

are composite hypothesis: while

(iv) $H_1: \mu = \mu_1$

is a simple hypothesis.

The normal population : X~ N (μ, σ^2) both μ and σ^2 are unknown has parameter space.

Q: $\{(\mu, \sigma^2): -\infty < \mu < \infty, \sigma^2 > 0\}$. Here

The null hypothesis

$Ho: \mu = \mu_0, \sigma^2 > 0$

and alternative hypothesis

$Ho: \mu > \mu_0, \sigma^2 < 0$

and both composite hypothesis.

If $\sigma^2 = \sigma o^2$ be known, then the null hypothesis

$Ho: \mu = \mu_0$

is a simple hypothesis.

The above definition can be symbolically written as follows:

For population: X~f (x;θ), $\theta \in \theta$

Ho: $\theta \in \theta_0 \subset \theta$

is simple if θ_0 is singleton set otherwise it is composite hypothesis. Similarly the alternative

H1: $\theta \in \theta \; where \; \theta_1 = \theta\,' - '\theta_0$

is simple if $\theta_1 = \theta\,' - '\theta_0$is a singleton set otherwise it is composite.

It is worth to note that sometimes one has to formulate the hypothesis exactly opposite of what is to be tested in the problem. For instance if it is required to show that the students of one school has a higher IQ than those of another school, i.e. $\mu_1 > \mu_2$. In this case also the null hypothesis must be Ho: $\mu_1 = \mu_2$ instead of H: $\mu_1 > \mu_2$we formulate the null hypothesis that there is "no difference" in the IQ's of the two schools. Now a days the null hypothesis is being used to "any hypothesis we may want to test."

Definition 3.5: A test of a statistical hypothesis H is a rule or procedure, based on the observed values of random sample from the population to accept or reject the hypothesis H.

Thus test is a two actions decisions – rule where the actions are either to accept or to reject the hypothesis Ho. The truth or falsity of the statistical hypothesis H depends upon whether the information contained in the sample is consistent with or hypothesis. If the sample information is inconsistent with the hypothesis then the hypothesis is rejected; otherwise it is accepted.

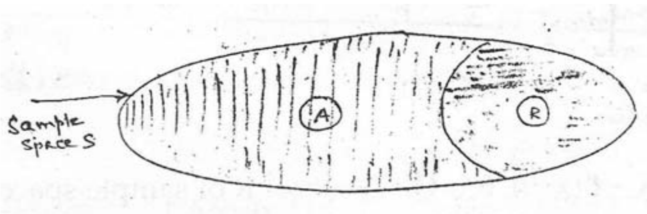
13 . 6 Critical (Rejection) Region

Let the population be X~ f (x; θ) $\theta \in Q$ where Q is the parameter space of the parameter θ .

Let $\underline{x}$: ($x_1, x_2 \dots \dots x_n$)be n independent sample observations corresponding to a random sample $\underline{X}$: ($X_1, X_2 \dots \dots X_n$)of size n from the population.

The n-dimensional space S which is the aggregate of all sample points $\underline{x}$: ($X_1, X_2 \dots \dots X_n$) is called a *sample space and* is denoted by S.

The test for a hypothesis divides the whole sample S into two disjoint (mutually exclusive) regions; one region A for acceptance of hypothesis H and another region R (or C) for rejection of hypothesis H.



A- *Figure 3.1 Sample Space, Acceptance (A) and Rejection (R) region*

B- Acceptance region (ACS)

R- Rejection (or critical) region (RCS)

With A∪R = S and A∩R = ∅

Thus the test for hypothesis H is;

Rejected Ho if $(x_1, x_2 \dots \dots x_n) \in$ R

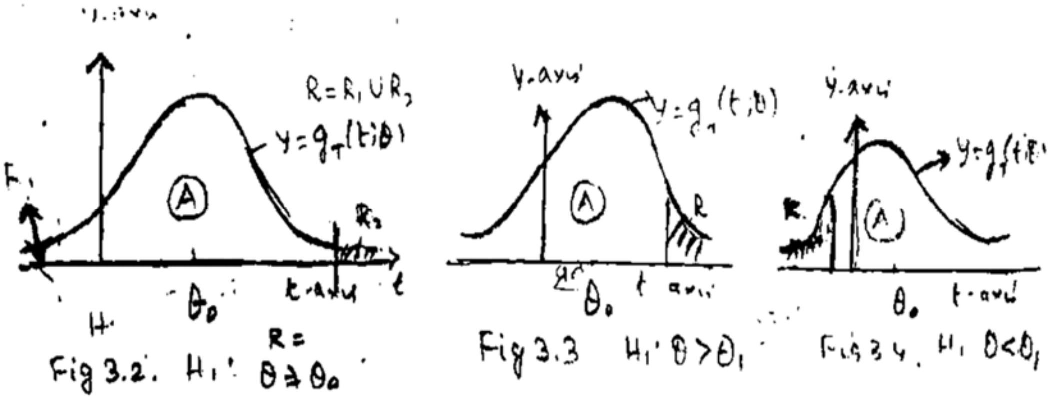
Accepted Ho if$(x_1, x_2 \dots \dots x_n) \in$ A

Definition 3.6: If a statistics T = T$(X_1, X_2 \dots \dots X_n)$ is used as an estimator for a parameter (-), then T=T ($\underline{X}$) is known as estimator of θ. if a statistic T is used to define a test of a hypothesis H, then it is known as a test statistic for H.

Thus a statistic associated with the test is called a test statistic.

A statistic R =T $(X_1, X_2 \dots \dots X_n)$ condenses the experimental data $\underline{x}$: $(x_1, x_2 \dots \dots x_n)$to a point t-T($\underline{x}$) = T $(x_1, x_2 \dots \dots x_n)$. In other words it maps the n-dimensional sample space S into a real line (one-dimensional) R_1: $(-\infty, \infty)$. There will be a region R and a region A on he real line R_1 corresponding to region R and region A, respectively in sample space S. Thus, a test γ partitions the real R_1 or the range of the test statistic T ($\underline{x}$) into two disjoint sets: The acceptance region A and rejection Region R.

If g (t, θ) be the sampling distribution of the test statistic T($\underline{X}$) or test of hypothesis H, then we may get following types of rejection regions for H: $\theta = \theta_0$



Figure

Definition 3.7: Let X~ f (x; θ) $\theta \in Q$. A subset R of sample space S, such that if R then Ho is rejected (with probability 1) is called the *critical region (or rejection region)* C of the test, where

$$C = \{\underline{x} \in S: H_0 \; is \; rejected \; if \; \underline{x} \in R\}$$

The complementary set A or R is said to *acceptance region* of the test.

13.7 Two kinds of Error

The decision of the test for hypothesis H is taken on the basis of the information of a sample from the population: X$\sim$ f (x; θ) $\theta \in Q$. As such there is an element of risk – the risk of taking wrong decisions. In any test procedure, there are four possible mutually exclusive and exhaustive decisions:

(i) Reject Ho when actually Ho is not true (false)

(ii) Accept Ho when it is true

(iii) Reject Ho when it is true

(iv) Accept Ho when it is false

The decisions in (i) and (ii) are correct while the decisions (iii) and (iv) are wrong decisions. These decisions may be expressed in the following dichotomons table.

True state in The nature	Decision		
	Ho True	Wrong (Type I error)	Correct
	Ho False (H_1 True)	Correct	Wrong (Type II error)

Thus in testing hypothesis may lead to following two kinds of errors.

Definition 3.8: An error of type I is made if the null hypothesis Ho is rejected when Ho is true; and the error of Type II is made if the null hypothesis Ho is accepted when Ho is false.

Type I Error = [Reject Ho | Ho is true]

Type II Error = [Accept Ho | Ho is false]

= [Accept Ho | Ho is true]

= [Reject H_1 | Ho is true]

The *probabilities of type I and Type II errors* are denoted by α *and* β respectively.

Definition 3.9: The *size of a Type I error is the probability of type I error* α *similarly, the size of a type II* error is the *probability of type II error* β

Thus,

α = Probability of Type I error

$= \text{Prob. [Reject Ho |Ho]}$

$= \text{Prob. } [\underline{x} \in R|\ Ho]$ where $\underline{x} = (x_1, x_2 \ldots\ldots x_n)$

$= \int_R L_0 dx$

$= \text{Prob. } [\text{T } (\underline{x}) \in R|\ Ho]$

$= \int_R T(x) g_T(t, \theta | Ho) dt$

Where Lo is the likelihood of the sample observations under Ho and $\int d\underline{x}$ stands for n-fold integral $\int \int \ldots\ldots\ldots\ldots \int dx_1 dx_2 dx_3 \ldots\ldots.. dx_n$.

g (t, $\theta| Ho$) is sampling distribution of test statistic T= T ($\underline{X}$) under Ho:

Similarly,

β = Probability of Type II error

= Prob. [Accept Ho |H_1]

$= \text{Prob. } [\underline{x} \in A|\ H_1]$

$= \int_R L_1 d\underline{x}$

$= 1 - \int_R L_1 d\underline{x}$ [Since $\int_A d\underline{x} + \int_R L_1 d\underline{x} = 1$]

$= 1 - \int_R T(x) g_T(t, \theta | H_1) dt$

When L1 is the likelihood function of the sample $Z_0 SE(\bar{X})$ observations under H_1.

Hence we define a power function of the test as

P(θ)= [Prob. (Reject Ho |H_1) = Prob. ($\underline{x} \in R|H_1$)=1-$\beta(\theta)$] and the power of the test for H0: $\theta = \theta_0$ against H_1: $\theta = \theta_1$ as P(θ_1)= [Prob. (Reject Ho |H_1) = Prob. ($\underline{x} \in R|H_1$)=1-β.

Definition 3.10: ***Level of significance:***

The *level of significance* is the maximum of probability of the type I error with which one is prepared to reject Ho when Ho is true. It is also called the size of critical region.

Definition 3.10: ***Power of the test:***

It is the probability with which the test reject Ho when H_1 is true. It is denoted by 1- β where β is the probability of type II error.

Power of the test = Prob. [Reject Ho |Ho]

= Prob. [Accept H_1 |H_1]

= 1-β

= Prob. ($\underline{x} \in R|H_1$)

The power of the test provides a basis for the comparison of two or more tests for simple hypothesis Ho against the sample alternative H_1. The power function P() is a function P(θ) is a function of θ ; therefore the power curve will be the basis for comparison betweens tests for Ho Vs. H_1.

An idea test would be one for which both of the probabilities α *and* β are zero, but there exists no test with fixed sample size n for which both α *and* β are zero. Consequently for fixed sample size n it is not possible to minimize both the error simultaneously. In general type I error is supposed to be more serious than type II error.

Hence for a fixe sample size n the usual practice in testing of a hypothesis Ho against alternative H1 is to keep α at a pre-determined low level say 0.01 or 0.05 and the test which has a more power or lesser β is said to be better than the other one.

A level of significance $\alpha = .05$, implies that if a very large number of samples, each of size n, be taken from the population the event [T ($\underline{x}$) $\in R$] is observed then in about 5 out of 100 cases, the hypothesis Ho is rejected when Ho is true.

13.8 One-sided and two sided or One-tailed and Two tailed Test.

Consider a situation in which we want to test the null hypothesis Ho: $\theta = \theta_0$ against $H_1 = \theta_1 \neq \theta_0$. it appears logical to accept the null hypothesis if the point estimate $\hat{\theta}$ of θ is closed to and to reject Ho if $\hat{\theta}$ is much larger or much smaller than θ_0, it would be reasonable to let the critical region R consists of both tails of the sampling distribution $g_{\hat{\theta}}(\hat{\theta}, \theta)$ of the chosen test statistic $\hat{\theta}$. The alternative hypothesis $H_1 = \theta \neq \theta_0$ is two sided, since the alternative consists of values of θ both below and above θ_0. Further, the critical region lies on both tails or ends of the sampling distribution of test statistic $\hat{\theta}$ Such a test is said to be *two sided test* or *two tailed test.*

On the other hand, the test for the null hypothesis Ho: $\theta = \theta_0$against the one sided alternative $H_1 = \theta < \theta_0$is said to be one sided test. In this situation it appears reasonable to reject Ho only when estimate $\hat{\theta}$ is much smaller than θ_0. Therefore, in this situation, the critical region. R consists of only left-hand tail of the sampling distribution of test statistic $\hat{\theta}$. The test for null hypothesis Ho with one sided alternative H_1: $\theta < \theta_0$ is said to be one sided test or left hand (one tailed) test.

Similarly in testing Ho: $\theta = \theta_0$ against the one sided alternative H1: $\theta > \theta_0$, the null hypothesis Ho could be rejected only for large values $\hat{\theta}$ of. So that the critical region consists only of the right tail of the sampling distribution of $\hat{\theta}$. The test for null hypothesis Ho with one sided alternative H1:$\theta > \theta_0$is said to be one sided test or right tailed test.

Definition 3.12: Any test where the critical region R consists of only of one tail of the sampling distribution of the test statistic is said to me one sided test while, it is said to be sided test if the critical region consists of the both tails of the sampling distribution of the test statistic.

13.9 Test of Significance

Suppose that the problem is to test the hypothesis that the mean μ of the normal population $N(\mu, \sigma^2)$with known variance σ^2 is different from μ_0.

As explained above the null hypothesis Ho and the alternative hypothesis H1 will be set up as follows:

$$H_o: \mu = \mu_o \ against \ alternative \ H_1 = \mu \neq \mu_o$$

Let us chosen level of the significant at α which is a smaller number.

Let X1, X2, X3, ……….Xn be a random sample of size n drawn from the population. Now the problem is to obtain a test for

$$H_o: \mu = \mu_o \qquad Vs. \quad H_1 = \mu \neq \mu_o$$

at level of significance α based on a sample random of size n. It is reasonable to accept Ho if the estimate $\hat{\mu}$ of μ is close enough to μ_o.

We know that the sample mean $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is an estimator of population mean and has sampling distribution.

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

So that

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

Further, 100 (1- α)% confidence internal for mean μ is

$$\bar{x} + z_0\left(\frac{\sigma}{\sqrt{n}}\right)$$

Where,

$$P[|Z| < z_0] = 1 - \alpha$$

Or

$$P(-z_0 < Z < z_0) = 1 - \alpha$$

The value of z_0 may be obtained from the Normal area Table for given α.

It is reasonable to take $\mu = \mu_o$if the estimate $\bar{x}\ of\ \mu$is close enough to μ_o . Obviously, it is reasonable accept Ho if lies in the interval $\bar{x} \pm z_0\left(\frac{\sigma}{\sqrt{n}}\right)$

That is,

$$\bar{x} - z_0\frac{\sigma}{\sqrt{n}} < \mu_o \leq \bar{x} + z_0\frac{\sigma}{\sqrt{n}}$$

Or equivalently,

$$\left|\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}\right| \leq z_0$$

In this case we do not reject the hypothesis Ho: if, on the other hand, is not in the interval or equivalently.

$$\left|\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}\right| \geq z_0$$

Then reject Ho.

Our test for $H_o: \mu = \mu_o\ \ Vs.\ \ H_1 = \mu \neq \mu_o$ at the level of significance α is,

$$Reject\ Ho \qquad\qquad if\ |\bar{x} - \mu_0| > z_0\left(\frac{\sigma}{\sqrt{n}}\right)$$

$$Accept\ \ Ho \qquad\qquad if\ |\bar{x} - \mu_0| \leq z_0\left(\frac{\sigma}{\sqrt{n}}\right)$$

Where, $\frac{\sigma}{\sqrt{n}}$ is the standard error of $\bar{X}$.

Prob [Type I error] = P [Reject Ho |Ho] =

$$P\left[\left|\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right| \geq z_0 | Ho\right] = P\left[\left|\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right| \leq z_0\right] = \alpha$$

Hence critical region R and Acceptance A are

$$A = \left\{(x_1, x_2 \dots \dots x_n): \mu_0 - Z_0\frac{\sigma}{\sqrt{n}} \leq \bar{x} \leq \mu_0 + Z\left(\frac{\sigma}{\sqrt{n}}\right)\right\}$$

and

$$R = \left\{(x_1, x_2 \dots \dots x_n): \left[\bar{x} < \mu_0 - \frac{Z}{\sqrt{n}}\right], U \left[\bar{x} + \mu\left(\frac{\sigma}{\sqrt{n}}\right)\right]\right\}$$

Such a test is known as test of significance.

Here the sample value of the statistic differ from the given value μ_0 of the parameter by more than certain amount in our case, Z_0SE ($\bar{X}$) is held important or significant to reject Ho at level of Significance α.

The value of |Z| = $\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$ under the assumption that $H_o: \mu = \mu_o$ holds is

|Z| $\alpha - 1\left|\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right|.if\ this\ value\ \left|\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right|\ exceeds\ tabulated\ value\ of Z_0\ of\ Z$

at level of significance α then we say that computed value of Z is significant at α and we reject the null hypothesis Ho.

It may worth to note that the rejection of a hypothesis Ho at α- level of significance does not implies the disapproval of the hypothesis. It only implies that the data or the sample information does not support the hypothesis at level of significance α. Similarly, the acceptance to Ho should be understood. The acceptance of the implies that x does not deviate from μ_o by so much amount that we reject Ho. It does not imply that is actually equal to μ_o but that it is close to μ_o .

Further the difference $|\bar{X} - \mu_0|$ between $\bar{x}$ and μ_o is inevitable produce of sampling fluctuations. The acceptance of Ho implies that this difference $|\bar{X} - \mu_0|$ is due to sampling fluctuations alone.

Definition 3.13: A test of significance for hypothesis Ho: is a procedure to assess the difference between the sample statistic and the value of parameter given by Ho or differences between two independent statistic to be significant or to reject or accept Ho at the given level of significance α.

We say that

(i) the difference between a statistic and the corresponding population parameters.

Or

(ii) the difference between two independent statistics

is not significant at the given level of significance, say α if it can be attributed only to the sampling fluctuations; otherwise it is said to be significant.

The procedure to be adopted for test of significance is outlined below-

(1) Propose the null hypothesis Ho and alternative hypothesis H_1:
(2) Fix a level of significance α for the test and a sample size n.
(3) Then choose a statistic T(x) whose sampling distribution is known under H_o.
(4) Keeping the value of α in mind decide upon those values of the test statistic (i.e. rejection region) that lead to its acceptance. In other words, define the test for H_o Vs. H_1 at level α.
(5) Now draw a random sample of size in from the population and compute the value of the test statistic.
(6) Finally on the basis of the value of the test statistic take the decision to accept or reject Ho.

Example 3.1: The mean of sample of size 25 from a normal population with mean μ and s.d. 4 is found to be 15. Do you accept or reject Ho: μ = 20 at the 10% level of significance?

Solution: Here $\bar{x}$ = 15, n = 25, σ =4, α = 0.1

Since

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) = N\left(\mu, \frac{4^2}{25}\right)$$

So that

$$|Z| = \left[\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}\right] \sim N\ (0,1)$$

Under Ho: μ = 20

$$|Z| = \left[\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}\right] = \frac{15-20}{4/\sqrt{25}} = -\frac{25}{4} = -6.25$$

Case I : Let Ho: μ = 20 against H_1: $\mu \neq 20$. We have to use a two tailed test. The test is as follows:

$$Reject\ Ho,\ if\ \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} < Z_{-\alpha/2} \quad or \quad \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} > Z_{\alpha/2}$$

$$Accept\ Ho, \qquad if - Z_{\alpha/2} < \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} > Z_{\alpha/2}$$

For α = 0.01, $\alpha/2$= 0.05, we get from normal areas table,

$$P\left[Z > Z_{\frac{\alpha}{2}}\right] = 1 - \frac{\alpha}{2} = 0.95$$

gives

$$Z_{-\alpha/2} = -1.645, \qquad Z_{\frac{\alpha}{2}} = +1.645$$

Test at α = 0.10 level of significance is

$$Reject\ Ho \quad if\ Z > -1.645 \quad or \quad Z < 1.645$$

$$Accept\ Ho \quad if \quad -1.645 < Z < 1.645$$

Here, computed Z = -6.25

Hence we reject Ho at 10% level of significance.

Case II Let Ho = μ = 20 be tested against H_1: Here right tailed test is to be used.

The test is

$$Reject\ Ho \quad if \qquad Z > \quad Z_{\alpha}$$

$$Accept\ Ho \quad if \qquad Z \leq \quad Z_{\alpha/2}$$

Where Z_{α} is obtained by,

0.1= Prob. [Reject Ho|Ho|] = P [Z > Z_{α}|Ho]

$$= Prob.\left[\frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} > Z_{\alpha}\right]$$

Which gives

Z_{α} = 1.282

Observed Z= - 6.25

We conclude that the does not support Ho at 10% level of significance. You may try the following problems-

E- 3.1 Let population be $\bar{X} \sim N(\mu, \sigma^2)$. To test Ho: μ = .5 we take a random sample of size n= 17 and observe that $\bar{x}$ = 78.8 and S = 12.8. Do you accept or reject Ho at 5% level of significance.

[Ans. 1.19<2.1 accept/ Ho 94 and times]

Hint: Here σ^2 is unknown $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Z_i - \bar{x})^2$is an unbiased estimator of σ^2therefore

$$T = \frac{\bar{X} - \mu}{\sqrt{S^2/(n/2)}} \sim student - test.$$

E- 3.2 Assume that IQ scores for a certain population are approximately $N(\mu, 100)$. To test Ho: $\mu = \mu_0$ against the one sided alternative hypothesis H_1: $\mu > 110$, we take a random sample of size n = 16 from this population and observer $\bar{x} = 113.5$ Do you accept or reject Ho at-

(a) 5% significance level
(b) 10% significance level,

Ans.: (a) 1.4< 1.1645 accept

(b) 1.4 > 1.282 reject.

13.10 Examples

Example 3.2 Given the frequency function

$$f(x;\theta) = \frac{1}{\theta} \quad if \quad 0 \leq x \leq \theta$$

$$= 0 \qquad elsewhere$$

It is required to test the null hypothesis Ho $\theta = 1$ against H_1: $\theta = 2$, b means of a single observed value of x. What would be the sizes of the type I and Type II error if you choose the intervals.

(i) $0.5 \leq x$

(ii) $1 \leq x \leq 1.5$

as the critical regions W? Also obtain the power of the test.

Solution:

Here we want to test Ho $\theta = 1$ against H_1: $\theta = 2$.

Part (i) For critical region

W = {x: 0.5≤ x}= {x:x≥ 0.5}

$$and\ acceptance\ region\ \ \bar{W} = \{x: x < 0.5\}$$

The test is as follows:

Reject Ho if $x \geq 0.5$

Accept Ho if $x < 0.5$

Now,

$\alpha \quad = Prob\ [Type\ I\ error]$

$= Prob\ [X \in W|\ Ho]$

$= Prob\ [X \geq 0.5|\ \theta = 1]$

$= P\ [0.5 \leq x \leq \theta|\ \theta = 1]$

$= P\ [0.5 \leq x \leq 1]$

$$= \int_{0.5}^{1} [f(x;\theta]_{\theta=1}^{dx}$$

$$= \int_{0.5}^{1} 1dx$$

$= 0.5$

Similarly,

$\beta \quad = Prob\ [Type\ II\ error]$

$= Prob\ [x \in W|\ H1]$

$= Prob\ [x \leq 0.5|\ \theta = 2]$

$$= \int_{0.5}^{1} [f(x;\theta]_{\theta=1}^{dx}$$

$$= \int_{0.5}^{1} 1dx$$

$= 0.25$

The power of the test

= 1-β

= 0.75

Thus the size of type I and type II errors and powers of the test are 0.5, 0.25, 0.75 respectively.

Part (ii) for critical region

W= {x: 1≤ x ≤ 1.5}and Acceptance region W= {x:x< 1, or x> 1.5}

The test is for Ho : θ = 2 is:

Reject Ho if 1≤ x ≤ 1.5

Accept Ho otherwise

Since under Ho,

F(x; θ) = 0 for 1≤ x ≤ 1.5

Therefore,

$$\alpha \quad = Prob\ [x \in W|\ \theta = 1]$$

$$= \int_{1}^{1.5} [f(x;\theta]_{\theta=1}\, dx$$

= 0.

Whereas,

$$\beta \quad = Prob\ [x \in W|\ \theta = 2]$$

$$= 1 - P\ [x \in W|\ \theta = 2]$$

$$= 1 - \int_{1}^{1.5} [f(x;\theta]_{\theta=1}\, dx$$

$$= 1 - \left|\frac{x}{2}\right|_{1}^{1.5}$$

= 0.75

Giving power of the test

= 1- β = 0.25

Thus the sizes of type I and type II errors and power of the test are 0., 0.75 and .25 respectively.

Example 3.3 If x ≤ 1, is the critical region W for testing Ho: θ = 2 against the alternative H1: θ = 1, on the basis of the single observation from the population,

$$f(x;\theta) = \theta e^{-\theta x} \qquad 0 \le x \le 1$$

$$= 0 \quad otherwise$$

Obtain the values of Type I and Type II errors.

Solution: Here, the critical region

$$\overline{W} = \{x: x \ge 1\}$$

and acceptance region

$$\overline{W} = \{x: x < 1\}$$

The test for Ho: $\theta = 2$ against H$_1$: $\theta = 1$ is:

Reject Ho if x $\geq$1

Accept Ho if x < 1.

Now,

$$\alpha \quad = Size\ of\ [Type\ I\ error]$$

$$= P\ [x \in W |\ Ho]$$

$$= P\ [x \ge 1 |\ \theta = 2]$$

$$= 2 \int_1^\infty [f(x;\theta]_{\theta=1}\, dx$$

$$= 2 \int_1^\infty e^{-2x}\ dx$$

$$= 2 \left| \frac{e^{-2x}}{(-2)} \right|_1^\infty$$

$$\alpha = e^{-2} = \frac{1}{e^2}$$

Similarly,

$$\beta \quad = size\ of\ [Type\ II\ error]$$

$$= P\ [x \in \overline{W} |\ H_1]$$

$$= P\ [x < 1 |\ \theta = 1]$$

$$= \int_{0}^{1} e^{-x} d_2$$

$$= [e^{-x}]_0^1$$

$$= 1 - e^{-1} = \frac{e-1}{e}$$

Thus,

$\alpha = e^{-2}\ and\ \beta = 1 - e^{-1}$

13.11 Most Powerful Test (MP- Test)

In the test of significance for Ho against alternative H_1, we have used an intuitive approach, where a test statistic is chosen whose sampling distribution is known, at least under null hypothesis Ho and a critical region is defined using the level of significance α. Thereafter a random sampling of given size n is drawn for the population and the value of test statistic is computed. If this value lies in the critical region then the hypothesis is rejects. Neyman and Pearson has profounded a more rational treatment to the tests of hypothesis by considering the probabilities of to two types of errors, that is $\alpha\ and\ \beta$ respectively, that one may commit in accepting or rejecting a hypothesis on the basis of sample observations. For all practical purposes, then the sample size n and the probabilities of a type I error are fixed and look for a test statistic which minimize the probability of a type II error, or, equivalently, which maximizes the power of the test 1-β.

Consider the problem of testing a simple null hypothesis Ho; $\theta = \theta_1$; For this purpose one has to draw a random sample of size n from the population: $\sim f(x;\theta)$. The present problem is to decide of whether the sample has come from the completely specified populations, $f(x;\theta_0)\ or\ f(x;\theta_0)\ or\ f(x;\theta_1)$ In other words, if X_1, X_2, …………X_n be a sample then our object is to test Ho: X_i is distribution as $f(x;\theta_0)$ Versus H_1: X_i is distributed as $f(x;\theta_1)$.

Obviously, when testing the simple null hypothesis is Ho: $\theta = \theta_0$against sample alternative hypothesis, H_1: $\theta = \theta_1$ power of the test $\theta = \theta_1$at is 1-β.

Definition 3.14: The critical region W is the best most powerful critical region of size α (and the corresponding test a most powerful test of level α) for testing simple null hypothesis Ho : $\theta = \theta_0$ against simple alternative hypothesis H_1: $\theta = \theta_1$ if the power of the test at $\theta = \theta_1$ is maximum. That is if W satisfies

$$P\left(\underline{X} \in W |\ Ho\right) - f, L_0 dx \qquad (a)$$

$$P\left(\underline{X} \in W \mid H_1\right) \geq P\left(\underline{X} \in W_1 \mid H_1\right) \qquad (b)$$

For every critical region W_1 satisfying (A).

To construct a more powerful critical region we refer to the likelihood of L_o a random sample of size n from the population under consideration when $\theta = \theta_0$ *and* $\theta = \theta_1$ say L_o and L_1 respectively that is

$$L_o = \prod_{i=1}^{n} f(x_i \theta_0)$$

$$L_1 = \prod_{i=1}^{n} f(x_i \theta_1)$$

Intuitively, speaking, it seems reasonable that L_o/L_1 should be small for sample points inside the critical region which leads to Type I errors when $\theta = \theta_0$ and to correct decision when $\theta = \theta_1$.

13.12 Uniformly Most Powerful (UMP) Test

Let us now take up the case of testing a simple null hypothesis Ho: $\theta = \theta_0$ against a composite hypothesis H_1: $\theta \neq \theta_1$ is such a case for a predetermined α the best test for Ho is called uniformly most powerful test level α,

The critical region W is called uniformly most powerful (UMP) Critical region of size α {and the corresponding test as uniformly most powerful (UMP) test of level α } for testing Ho = $\theta = \theta_0$ against H_1: $\theta \neq \theta_1$ if

$$P\left(\underline{X} \in W \mid Ho\right) - \int_{W}^{.} L_0 dx = \alpha \qquad (c)$$

$$P\left(\underline{X} \in W \mid H_1\right) \geq P\left(\underline{X} \in W_1 \mid H_1\right) \qquad (d)$$

$$for\ all\ \theta = \theta_0$$

Whatever the region W1 satisfying (c) may be.

13.13 Solution / Answers

P- 3.1 1.19<2.12; Accept Ho

P- 3.2 (a) 1.4≤ 1.645; Accept Ho

(b) 1.4> 1.28; Reject Ho

13.14 Summary

In this unit an attempt is made to explain the basis concepts related to the testing of hypotheses.

13.15 Further Readings

1. A.M. Mood, F. Ar. Graybill & D.C. Boes. Introduction to the theory of Statistics, III. Editions Pub: Mac.Graw Hill.

2. Rahtagi V.K. (1984): An Introduction to Probability theory and Mathematical Statistics chapter VIII, IX & X Pub; John Wiley & Sons, New York.

3. Goon A.N., Gupta M.K. & Das Gupta B (1987) Fundamentals of Statistics Vol. I The World Press Pvt. Ltd., Kolkata.

4. Kapoor V.K. & S.C. Saxena: Fundamentals of Mathematical Statistics, Chapter Seventeen, Pub: S. Chand.

U.P. Rajarshi Tandon Open University, Prayagraj

UGSTAT – 102

Probability, Distribution and Statistical Inference

Block

5

Test of Significance

Unit - 14
Exact tests and Fisher's Z transformer

Unit - 15
Large Sample Tests

Unit – 16
Non-Parametric Tests

Course Design Committee

Dr. Ashutosh Gupta **Chairman**
Director, School of Sciences
U. P. Rajarshi Tandon Open University, Prayagraj

Prof. Anup Chaturvedi **Member**
Ex. Head, Department of Statistics
University of Allahabad, Prayagraj

Prof. S. Lalitha **Member**
Head, Department of Statistics
University of Allahabad, Prayagraj

Prof. Himanshu Pandey **Member**
Department of Statistics, D. D. U. Gorakhpur University,
Gorakhpur

Dr. Shruti **Member-Secretary**
Sr. Assistant Professor, School of Sciences
U.P. Rajarshi Tandon Open University, Prayagraj

Course Preparation Committee

Dr. Shruti **Course Coordinator**
Sr. Assistant Professor, School of Sciences
U. P. Rajarshi Tandon Open University, Prayagraj

Prof. Ram Lal **Writer**
Department of Mathematics and Statistics, Allahabad
Agricultural Institute (Deemed University), Naini, Prayagraj

Prof. Umesh Singh **Writer**
Department of Statistics, Banaras Hindu University, Varanasi

Prof. S. K. Pandey **Writer**
Department of Statistics, Lucknow University, Lucknow

Prof. Anoop Chaturvedi **Reviewer**
Department of Statistics, University of Allahabad
Prayagraj

Prof. B. P. Singh **Reviewer**
Department of Statistics, Banaras Hindu University, Varanasi

Prof. A. K. Saxena **Reviewer**

Department of Statistics, Lucknow University, Lucknow

Prof. S. K. Pandey **Editor**
Department of Statistics, Lucknow University, Lucknow

Prof. S. K. Upadhyay **Editor**
Department of Statistics, Banaras Hindu University, Varanasi

First Edition: *March 2008* (Published with the support of the Distance Education Council, New Delhi)

Second Edition: ***July 2020***

ISBN :

Printed and Published by Dr. Arun Kumar Gupta Registrar, Uttar Pradesh Rajarshi Tandon Open University, 2020.

Printed By: Chandrakala Universal Pvt. 42/7 Jawahar Lal Neharu Road, Prayagraj.

Unit-14: Exact Tests and Fisher's z- Transformation

Structure

14.1 Introduction

The statistics on which our testing procedure is based in particular situation, is known as test statistic. If the probability distribution of this statistic is Chi-square, it is known as Chi-square test. In different situation we have different test-statistics and they follow different probability distributions and accordingly the name of the tests. If the probability distribution of the test statistic is t (or F) in a particular situation, it is known as t (or F) test. When probability distribution of test statistic is not approximated and it is exactly a specific probability distribution, these are called exact tests.

14.2 Objectives

After going through this unit you should be able to apply exact tests based on Chi-square, t and F distribution along with Fisher's z transformation.

14.3 Test of significance based on Chi-square Distribution.

(a) To test the signification of variance from a normal population.

Let X_1, X_2,…….X_n be random sample of size n taken from N (μ, σ^2) and w wish to test Ho: $\sigma^2 = \sigma_0^2$.

If μ is known then H_0

$$\frac{\sum(X_i - \mu)^2}{\sigma_0^2}$$

Follows Chi-square distribution with n degree of freedom.

If is not known then under H_o

$$\frac{\sum(X_i - \bar{X})^2}{\sigma_0^2}$$

Follows Chi-square distribution with (n-1) degree of freedom;

$\bar{X} = \frac{1}{n}\sum X$ Is the sample mean.

(b) To test goodness of fit.

Here we test the null hypothesis that data follow a particular probability distribution (like Binomial, Poisson, Normal etc.).

Let O_1, O_2, be the observe frequencies and e_1, e_2,… are the expected frequencies corresponding to these observed frequencies then

$$\sum \frac{(0_i - e_i)^2}{e_i} = \sum\left(\frac{o_j^2}{e_i}\right) - N \qquad where\ N = \sum O_i = \sum e_i$$

follows a chi-square distribution with certain say v degrees of freedom. If expected frequencies are less than 5, then pooling is done.

Example 1.1: Five dice were thrown 192 times and the number of times 4, 5 or 6 were as follows:

No. of dice throwing 4,5,6	5	4	3	2	1	0	Total
Observed frequency: O_i	6	46	70	48	20	2	192

Calculation the value of Chi-Square on the hypothesis that dice were unbiased and hence test whether the data are consistent with the hypothesis.

Solution: Probability of throwing 4,5,6 is 3/6= ½= p

Therefore from binomial distribution the theoretical frequencies of getting 5,4,3,2,1,0 successes with 5 dice re respectively the successive terms of :

N $(p+q)^5$= $192(1/2+1/2)^5$

Which are as follows: 6,30,60,30, 6 respectively

Table showing observed and expected frequency after clubbing (Pooling) the frequency which are less than 5.

No. of dice throwing 4,5,6	5	4	3	2	1or 0	Total
Observed frequency: O_i	6	46	70	48	22	192

Expected frequency: ei	6	30	60	60	36	192

$$So\ \chi^2 = \sum_i \left\{\frac{(0_i - e_i)^2}{e_i}\right\}$$

$$= \frac{(6-6)^2}{6} + \frac{(46+30)^2}{30} + \frac{(70+60)^2}{60} + \frac{(48+40)^2}{60} + \frac{(22+36)^2}{36}$$

$$= 0 + 8.53 + 1.66 + 2.4 + 5.44 = 18.03$$

There are n= 6 cells and k =1 cell, i.e., “0” is polled with cell “1”, therefore the degree of freedom of is

V=n-1-k, = 6-1-1=4.

The tabulated value of Chi-square (χ^2) on 4 degrees of freedom and at 5 % level of significance is 9.488. Since the calculated value of χ^2 is greater than the tabulated value hence null hypothesis may will be rejected. Therefore the observed frequency distribution is not consistent with the hypothesis.

(c) Testing of independence or Association between two (attributes) (characters):

If a character (factor, attribute) A is classified into A_1, A_2, A_i, …..A_r classes and second character (factor/ attribute) B into B_1, B_2,….B_i….B_c classes and if O_{ij} is the observed frequency due to Ai class of A and Bi class B which are shown in the following table:

Character B

Character A		B_1	B_2	B_j	B_c	R_1
	A_1	O_{11}	O_{12}	O_{2j}	O_{1c}	R_1
	A_2	O_{21}	O_{22}	O_{2j}	O_{2c}	R_2
	A_i	O_{i1}	O_{i2}	O_{ij}	O_{ic}	R_i
	A_r	O_{r1}	O_{r2}	O_{tkj}	O_{rc}	R_r
	Total	C_1	C_2	C_j	C_{rc}	N

Where,

$$\sum_{i=1}^{r} R_i = \sum_{j=1}^{c} C_j = N\ grand\ total$$

Then this table is called r x c contingency table.

We are interested in testing the null hypothesis.

H_0: Character (attributes) A and B are independence, i.e., there is no associated between two character A and B.

Let eij denotes the expected frequency due to i^{th} class of A and j-th class of character B, i= 1, 2,……r, j= 1,2,……,c, Then

$$e_{ij} = \frac{R_i \times C_j}{N} = N\left(\frac{R_i}{N}\right)\left(\frac{C_j}{N}\right)$$

Thus, the expected frequency of any cell is equal to the product of the class totals of the two classes to which the cell belongs divided by the total number of observations.

Hence to test the above null hypothesis, we use Chi-square as given by:

$$\sum_{i=1}^{r} \quad \sum_{j=1}^{c} \frac{\left(0_{ij} - e_{ij}\right)^2}{e_{ij}} = \sum_i \sum_j \frac{O_i^2}{e_{ij}} - N$$

It is a Chi statistic with (r-1) (c-1) degrees of freedom.

The calculated value of is compared against the table value of χ^2 on (r-1) (c-1) degrees of freedom and 5% probability level. If calculated value of χ^2 is greater than its table value then the null hypothesis H_0 is rejected, otherwise H_0 will be accepted.

Rejecting the null hypothesis means that there is association between two factors (attributes/ character).

Example 1.2: From a village 200 persons were randomly selected and data about their income and education achievement were recorded, which are given in the following table.

Education

		High	Medium	Low	Total
Income	High	60	20	20	R_1=100
	Low	20	20	60	R_2=100
Total		e_1=80	e_2=40	e_3=80	N= 200

Test whether education depends upon income.

Null Hypothesis

Against the alternative

H_1; There is association between education and income.

Expected frequencies

$$e_{11} = \frac{100 \times 200}{200} = 40$$

$$e_{12} = \frac{100 \times 40}{200} = 20$$

$$e_{13} = \frac{100 \times 80}{200} = 40$$

$$e_{21} = \frac{100 \times 80}{200} = 40$$

$$e_{22} = \frac{100 \times 40}{200} = 20$$

$$e_{23} = \frac{100 \times 80}{200} = 40$$

Table of expected frequencies:

Education

		High	Medium	Low	Total
Income	High	40	20	40	100
	Low	40	20	40	100
Total		80	40	80	200

$$\chi^2 = \sum_i \sum_j \frac{\left(0_{ij} - e_{ij}\right)^2}{e_{ij}}$$

$$= \frac{(60-40)^2}{40} + \frac{(20-20)^2}{20} + \frac{(20-40)^2}{40} + \frac{(20-40)^2}{40} + \frac{(20-20)^2}{20} + \frac{(60-40)^2}{40}$$

$$= \frac{20 \times 20}{40} + 200 + \frac{20 \times 20}{40} + \frac{20 \times 20}{40} + 0 + \frac{20 \times 20}{40}$$

$$= 10 + 0 + 10 + 10 + 0 + 10 = 40$$

d.f.= (r-1) (c-1)= (2-1) (3-1)= 1×2 = 2

The tabulated value of χ^2 and 2 d.f. and at 5% level of significance is 5.991.

Since the calculated value of χ^2 is greater than the table value of χ^2 on 2 degrees of freedom and at 5% probability level so our null hypothesis will be rejected. Therefore, it can be concluded from the above data that there between education and income.

(d) χ^2 in 2×2 contingency table:

A table in which each of the two characters or attributes are divided into two subgroups giving rise to four total number of cells is called a 2×2 contingency table. Here, r=2, c=2. Suppose there are two characters A and B each being divided into two groups A_1, A_2 and B_1, B_2 respectively and the observed frequencies of four cells are a,b,c,d which are arranged in the following table:

B				
		B_1	B_2	Total
	A_1	a	b	R_1=a+b
A:	A_2	c	d	R_2=c+d
	Total	C_1=a+c	C_2=b+d	N= a+b+c+d

We wish to test the null hypothesis.

H_0: There is no association between two character A and B.

Under H_o,

$$\chi^2 = \frac{(ad-bc)^2 \times N}{(a+b)(c+d)(b+d)} = \frac{(ad-bc)^2 N}{R_1 \times R_2 \times C_1 \times C_2}$$

Follows Chi-square distribution with one degree of freedom, i.e., r=1.

Where R_1, R_2 are row table and C_1 and C_2 are the column totals.

Derivations of

$$\chi^2 = \frac{(ad-bc)^2 N}{R_1 \times R_2 \times C_1 \times C_2}$$

We know that

$$\chi^2 = \sum_i \sum_j \frac{\left(0_{ij} - e_{ij}\right)^2}{e_{ij}}$$

So we need of find the expected frequencies of a, b, c and d which are given as

$$E(a) = N.\frac{(a+b)}{N} \times \frac{(a+c)}{N} = \frac{(a+b)(a+c)}{N}$$

Similarly,

$$E(b) = \frac{(a+b)(b+d)}{N}$$

$$E(e) = \frac{(c+d)(a+c)}{N} and \ E(d) = \frac{(c+d)(b+d)}{N}$$

Therefore,

$$\chi^2 = \frac{[a-E(a)]^2}{E(a)} + \frac{[b-E(b)]^2}{E(b)} + \frac{[c-E(c)]^2}{E(c)} + \frac{[d-E(d)]^2}{E(d)}$$

$$= \frac{\left(\frac{a-(a+b)(a+c)}{a+b+c+d}\right)^2}{\frac{(a+b)(a+c)}{N}} + \frac{\left(\frac{b-(a+b)(b+d)}{a+b+c+d}\right)^2}{\frac{(a+b)(b+d)}{N}} + \frac{\left(\frac{c-(c+d)(a+c)}{a+b+c+d}\right)^2}{\frac{(c+d)(a+c)}{N}} + \frac{\left(\frac{d-(c+d)(b+d)}{a+b+c+d}\right)^2}{\frac{(c+d)(b+d)}{N}}$$

$$= \frac{(ad-bc)^2}{N}\left\{\left(\frac{1}{(a+b)(a+c)}\right)\right\} + \left\{\left(\frac{1}{(a+b)(b+d)}\right)\right\} + \left\{\left(\frac{1}{(c+d)(a+c)}\right)\right\} + \left\{\left(\frac{1}{(c+d)(b+d)}\right)\right\}$$

$$= \frac{(ad-bc)^2.N}{N}\left(\frac{1}{(a+b)(a+c)(b+d)}\right) + \left(\frac{1}{(a+c)(b+d)(c+d)}\right)$$

$$= (ad-bc)^2\left(\frac{c+d+a+b}{(a+b)(a+c)(b+d)(c+d)}\right)$$

$$= \frac{(ad-bc)^2.N}{(a+b)(a+c)(b+d)(c+d)} = \frac{(ad-bc)^2.N}{R_1 \times R_2 \times C_1 \times C_2}$$

Example on 2×2 Contingency table

Example 1.3 From the following table, test whether flower colour is independent of flatness of leaves.

	Flat leaves	Curled leaves	Total
White flowers	80=a	20=b	R_1=100
Red flowers	40=c	60=d	R_2=100
Total	C_1=40	C_2=80	N= 200

H_1: flower colour is independent of flatness of leaves.

Against

H_1: flower color dependent on the flatness of leaves.

To test the above null hypothesis, we shall calculated the statistics as:

$$\chi^2 = \frac{(ad-bc)^2.N}{R_1 \times R_2 \times C_1 \times C_2}$$

$$= \frac{[(80\times 60)-(20\times 40)]^2}{100\times 100\times 120\times 80}\times 200 = \frac{[(4800\times 800)]^2\times 200}{100\times 100\times 120\times 80}$$

$$= \frac{4000\times 4000\times 200}{100\times 100\times 120\times 80} = \frac{100}{3} = 33.33$$

This is calculated value of Chi-square (χ^2).

The tabulated value of Chi-square at 5% level of significance for one d.f. is 3.841.

The calculated value of χ^2 is greater than the table value of χ^2 on 1 degree of freedom and at 5% probability level, so our null hypothesis is rejected and we can conclude that flower colour depends upon flatness of leaves.

Yates Correction continuity in 2×2 contingency table.

The distribution of χ^2 is continuous while the distribution is frequencies is by its very nature discontinuous. The continuous χ^2 may be regarded as the limit to which the true discontinuous distribution tends as the sample size increases.

So in case of any cell frequency is less than 5, Prof. F. Yates (1934) suggested the following adjustment in a 2×2 contingency table.

		B		
A:		B_1	B_2	Total
	A_1	a	b	R_1=a+b
	A_2	c	d	R_2=c+d
	Total	C_1=a+c	C_2=b+d	N= a+b+c+d

1) Calculate the ad and bc
2) If ad > bc, then add ½ to both b and c
   And subtrach ½ from both a and d
   So that the marginal totals are not affected.

		B		
A:		B_1	B_2	Total
	A_1	a-1/2=a^1	B+1/2=b^1	R_1=a+b
	A_2	c +1/2= c^1	d-1/2=d^1	R_2=c+d
	Total	C_1=a+c	C_2=b+d	N= a+b+c+d

3) Then the following formula of χ^2 is used.

$$\chi^2 = \frac{(ad^1 - bc^1)^2 . N}{R_1 \times R_2 \times C_1 \times C_2} \sim \chi^2 \ with\ 1\ degree\ of\ freedom.$$

Note: The alternative formula for χ^2 with Yate's correction is:

$$\chi^2 = \frac{\left[|ad - bc| - \frac{N}{2}\right]^2 \times N}{R_1 \times R_2 \times C_1 \times C_2}$$

Example 1.4: From the following data test the association between colour of flowers and character of fruit.

	Flower violet	Flower white	Total
Fruits prickly	40=a	20=b	R_1=60
Fruits smooth	16=c	04=d	R_2=20
Total	C_1=56	C_2=24	N= 80

H_0: colour of flowers and character of fruits are not associated.

ad= 40 × 4= 160

bc = 20 × 16= 320

so ad < bc

using

$$\chi^2 = \frac{\left[|ad - bc| - \frac{N}{2}\right]^2 \times N}{R_1 \times R_2 \times C_1 \times C_2}$$

$$\chi^2 = \frac{\left[|320 - 160| - \frac{80}{2}\right]^2 \times 80}{60 \times 20 \times 56 \times 14}$$

$$= \frac{120 \times 120 \times 80}{60 \times 20 \times 56 \times 24} = \frac{10}{14} = \frac{5}{7} = 0.71$$

$$the\ tablulated\ value\ of\ \chi^2_{.05,1} = 3.841$$

After comparing the calculated value of χ^2 against its tabulated value, it is found that the calculated value of χ^2 is smaller than the tabulated value at 5% probability level.

Hence it can be concluded from the above data that colour of flower and character of fruits are not associated.

You may try the following exercises.

Exercises on Chi- square

Inoculated Not-inoculated	Not attacked	Attacked	Total
	430	10	440
	60	100	160
	490	110	N=600

Is there significant association between inoculation and attack.

E 1.2) In an experiment on immunization of goats from antrax, the following results were obtained.

Inoculated with vaccine Not-inoculated	Not attacked	Attacked	Total
	4	16	20
	12	8	20

Derive your inference on the efficiency of vaccine.

E 1.3) From a village 100 persons were randomly selected and their education achievements were recorded. The data recorded are given in the following table:-

Education					
Gender		Middle	High school	Graduation	Total
	Male	10	15	25	50
	Female	25	10	15	50
					N=100

Hint: Based on the above data can you say that education depends on sex.

Ans. : χ^2 = (col)= 9.93

14.4 Tests of Significance based on t-distribution

t- distribution is used

i) To test the significance of mean from a normal population
ii) To test the significant different between two population means of normal population.
iii) To test the significance of correlation coefficient from a bivariate normal population.
iv) To test the significance of regression coefficient.

(i) Testing the significance of mean from a normal population.

Suppose x_1, x_2 , x_3……..x_n is a random sample of size n from a normal population with mean μ and variance σ^2 (σ^2 hot known) and one wishes to test. Ho:

Under Ho: $\mu = \mu_0$

$$t = \frac{(\bar{x} - \mu_0)\sqrt{n}}{S} \; follows$$

t distribution with (n-1) degrees of freedom. Here, $\bar{x}$ is the sample mean and

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_1 - \bar{x})^2$$

Where S^2 is an unbiased estimate of.

The calculated of |t| is compared against the tabulated value of t. If the calculated value of |t| is greater than the table value of t on (n-1) degrees of freedom and at ∝ % probability level, the above null hypothesis will be rejected at ∝–level of significance otherwise, H0 will be accepted.

Example 1.5

Ten rice plants were randomly selected from a small research plot having 100 plants. The height of these plants was recorded to study the effect of a bio-fertilizer on the growth behavior of plants which are given below-

Height in cm: 80,76,78,84,82,83,77,80,81,79

In the light of the above data, test whether the average height of plants in the population in 82.5 cm.

X= height in cm	$X_1 - \bar{X}$	$(X_1 - \bar{X})^2$	Ho: $\mu = 82.5$
80	0	0	
76	-4	16	$S^2 = \frac{\sum_{i=1}^{n}(x_1 - \bar{x})^2}{n-1}$
78	-2	4	
84	+4	16	= 60/9 = 6.666
82	+2	4	S= $\sqrt{6.666}$
83	+3	9	S= 2.58= 2.58
77	-3	9	
80	0	0	
81	+1	1	
79	-1	1	
Total 800= $\sum X_i$	-	60=$\sum(X_1 - \bar{X})^2$	

$$\bar{X} = \frac{\sum X_i}{n} = \frac{800}{10} = 80$$

$$t = \frac{\bar{x} - \mu_0}{\frac{S}{\sqrt{n}}} = \frac{(80 - 82)\sqrt{10}}{2.58}$$

$$= \frac{2.5 \times 3.16}{2.58} = -3.07$$

$$or\ |t| = 3.07$$

Table value of t on 9 d.f. and at 5% level of significance is 2.262.

Since the calculated value of |t| is greater than the table value of t on 9 degrees of freedom and at 5% probability level, so our null hypothesis Ho will be rejected. Therefore it can be concluded from the given data that population mean is significantly different from = 82.5 cm., in other words the average height of plants in the population cannot be regarded as 82.5 cm from which a random sample of 10 plants have been selected with sample mean 80 cm.

Exercise on t-tests

E-1.4 Ten boxes are selected at random from a godown and their weights are found to be in kgs as xi= 15.75, 16.0, 15.75, 16.25, 16.50, 17.25, 17.50, 17.50, 17.76

E-1.5 Discuss the suggestions that the mean weight in the population is 16.25kgs.

Give that t_9 (5%)= 2.262

(Ans. t calculated = 1.93)

(ii) Testing the significant different between two population means

Suppose x_1, x_2, x_3,….x_{n1}., is a random sample from 1st Normal population with mean μ_1 and variance $\sigma 1^2$ and another independent random sample y_1, y_2, ……y_{n2} from 2nd normal population with mean and variance. It is further assumed that $\sigma 1^2 = \sigma 2^2 = \sigma,^2$ (say) unknown

Which is unknown.

We wish to test the null hypothesis H_0= : $\mu_1 = \mu_2$

Under Ho

$$\frac{\bar{x} - \bar{y}}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Follows t-distribution with (n_1-n_2) degrees of freedom. Where

$$S^2 = \frac{\sum_{i=1}(x_1-\bar{x})^2 + \sum_{i=1}(y_1-\bar{y})^2}{n_1+n_2-2} = \frac{(n_2-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$$

Is an unbiased estimated of σ^2

The calculated value of |t| is compared against tabulated value of t on (n_1+n_2-2) d.f. and at $\alpha\%$ level of significance.

If cal. |t|> t_α, n_1+n_2-2then the null hypothesis Ho is rejected otherwise Ho is accepted.

The above procedure is called as two sample t-test.

Example on Two sample t-test

Example 1.6

Ten red plants were randomly selected from 1st plot and 8 yellow plants were randomly selected from second rose plot. The height of these selected were separately recorded and are given below in cm.

Heights of red rose plants in cm=xi	$X_1-\bar{X}$	$(X_1-\bar{X})^2$	Heights of yellow rose plants in cm=yi	$y_1-\bar{y}$	$(y_1-\bar{y})^2$
60	0	0	62	0	0
64	4	16	60	-2	4
61	1	1	63	+1	-
56	-4	10	64	+2	4
59	-1	1	61	-1	1
62	2	4	63	+1	1
58	-2	4	63	+1	1
60	0	0	61	-1	1
63	3	9			
57	-3	9			
$\sum X_i$=600		$\sum(X_1-\bar{X})^2$ =60	$\sum y_i$=496		$\sum(X_1-\bar{X})^2$ =12

Discuss the suggestion whether there is significant different between the mean height of red and yellow row plants.

Ho: There is no different between the mean heights of red and yellow rose plants.

$$\bar{x} = \frac{\sum x_i}{n_1} = \frac{600}{10} = 60 \qquad \bar{y} = \frac{\sum y_i}{n_1} = \frac{496}{8} = 62$$

$$S^2 = \frac{\sum_{i=1}(x_1 - \bar{x})^2 + \sum_{i=1}(y_1 - \bar{y})^2}{n_1 + n_2 - 2}$$

$$= \frac{60 + 12}{10 + 8 - 2} = \frac{72}{16} = 4.5$$

$$S^2 = 4.5$$

$$S = \sqrt{4.5} = 2.12$$

Therefore,

$$t = \frac{\bar{x} - \bar{y}}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{60 - 62}{2.12\sqrt{\frac{1}{8} + \frac{1}{10}}}$$

$$t = \frac{-2}{2.12 \times 3.5} = \frac{-2}{0.742} = -2.7 \qquad |t| = 2.7$$

The calculated value of |t| is greater than the table value of t on 16 d.f. and at 5% probability level, so our null hypothesis is rejected. Therefore it can be concluded that our null hypothesis is rejected, hence there is significant different between the mean heights or red and yellow rose plants.

You may try the following Exercise.

E-1.6 In a rat feeding experiment the following results were obtained gain in weight in gm.

High Protein	13	14	10	11	12	14	10	8	11	12	9	12
Low Protein	7	11	10	8	10	12	9					

Find if there is any evidence of superiority of one diet over the other.

Given t (5%) on 17 d.f. is 2.11

Ans. t= 1.77

(iii) Paired t-test

Let (x_1,y_1), (x_2,y_2), …… (x_n, y_n) be a random sample of size n drawn from a bivariate normal population $\left(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho\right)$ we wish to test

H_o: $\mu_1 = \mu_2$

In this case, under H_o:

$$\frac{\bar{d}\sqrt{n}}{S_d}$$

Follows t-distribution with (n-1) degrees of freedom.

Where

$$d_1 = x_1 - y_1$$

$$\bar{d} = \frac{1}{n}\sum d_1$$

$$S_d^2 = \frac{1}{n}\sum (d_1 - \bar{d})^2$$

Example of paired observations

If we want to compare the effects of two drugs D_1 and D_2 for the same disease, then first of all drug D_1 should be administered to a set of certain patients (i= 1,2,3, n) and its effect should be recorded. After a reasonable interval of time, the drug D_2 should be administered to the same set of patients and its effect should be recorded. The observations so obtained are said to be paired.

Example 1.*7:-* The scores of 10 cadets before and after training given below:

Cadet No.	1	2	3	4	5	6	7	8	9	10
Score before training x_i	8	3	2	4	6	8	5	8	7	9
Score after training y_i	7	5	7	5	3	9	5	10	6	10

Based on the above data can you say whether training is effective in improving the performance of cadets.

Solution:

Since observation under x and y are paired, so to test the null hypothesis.

H_o: Training is not effective in improving the performance of cadets i.e.,

H_o: $\mu_1 = \mu_2$ against the alternative H_1: $\mu_1 \neq \mu_2$ we use the paired t-test and test statistic.

$$t = \frac{\bar{d}}{S_d/\sqrt{n}}$$

$d_i = x_i - y_i$	-1	2	5	1	-2	1	0	2	-1	1	$\sum d_i$=7
di^2	1	4	25	1	4	1	0	4	1	1	$\sum di^2$= 47

$$\bar{d} = \frac{\sum d_i}{n} = \frac{7}{10} = 0.7$$

$$S_d^2 = \frac{\sum (d_i - \bar{d})^2}{n-1} = \frac{\sum d_1^2 - n\bar{d}^2}{n-1}$$

$$= \frac{47 - 10 \times 4.9}{9} = \frac{47 - 4.9}{9}$$

$$= \frac{4.21}{9} = 2.15$$

Here,

$$S_d = \sqrt{4.65} = 2.15$$

We have

$$t = \frac{\bar{d}\sqrt{n}}{S_d} = \frac{0.7 \times \sqrt{10}}{2.15} = \frac{0.7 \times 3.16}{2.15} = 1.28$$

Table value of t (5%) on 9 d.f. is 2.262

The calculated value of t is less than the table value of t on 9 degrees of freedom and at 5% probability level. Therefore, null hypothesis may be accepted and it can be conducted that training is not very effective in improving the performance of cadets.

You may try the following exercise.

Exercise on paired to t-test:

Ex- 1.7 Certain stimulus administered to each of 12 patients resulted in the following increases of blood pressures:

5,2,8,-1,3,0,6,-2,1,5,0,4.

Test whether stimulus can increases the blood pressure given t(5%) on 11 degrees of freedom is 2.201.

Ans. Sd= 3.08

Computed t= 2.94

iii) Test of significance of correlation coefficient

Suppose a pair of random sample (x_1,y_1); (x_2,y_2)….(x_n,y_n) is drawn from a bivariate normal population with population correlation p.

Let the sample correlation between x and y be r

Then we wish to test the null hypothesis:

H_o: p=0 (i.e. population correlation effective is zero) against the alternative H_1: $p \neq 0$.

Under H_o; $\frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$ follows a t-distribution with (n-2) degrees of freedom.

Example 1.8:

A random sample of size 18 form a bivariate normal population gave a correlation coefficient 0.6. Does it indicate the existence of correlation in the population?

H_o: p=0 Here, we with to test H_o: p=0 Vs H_1: p≠0.

The value of t=$\frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$ is

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0.6\sqrt{18-2}}{\sqrt{1-.36}} = \frac{0.6 \times 4}{.80} = \frac{2.4}{.8} = 3.0$$

The tabulated value of t on 16 degrees of freedom at 5% level of significance is 2.12.

Since the calculated value of t is greater than the table value of t on 16 d.f. and at 5% level of significance so our null hypothesis Ho is rejected. Therefore, it can be concluded that the sample correlation coefficient r=0.6 is significant and it indicates the existence of correlation in the population.

(iv) To test the significance of Regression Coefficient:

Suppose *(x_1,y_1); (x_2,y_2)....(x_n,y_n)* is a random sample from a bivariate normal population with regression coefficient β, of y and x

We know that the regression equation of y on x from the sample is: $y - \bar{y} = b_{yx}(x - \bar{x})$.

$\bar{x}, \bar{y}$ are sample means and b_{yx} is the sample regression coefficient of y or x.

So that the estimated value of y corresponding to given xi is

$$Y_i = \bar{y} + b_{yx}(X_i - \bar{X})^2$$

We wish to test the null hypothesis:

H_o: $\beta = \beta_0$ where β_0 is known

Under Ho:

$$\frac{(b_{yx} - \beta_0)\sqrt{n-2}\sum(x_i - \bar{x})^2}{\sum(y_i - \bar{y})^2}$$

Follows t- distribution with (n-2) defines of freedom.

Example 1.9:

A random sample of size 27 gave $\sum(x_i - \bar{x})(y_i - \bar{y}) = 599.62$

$\sum(x_i - \bar{x})^2 = 2247.5$ *and* $\sum(y_i - \bar{y})^2 = 1020.6$ test the significance of regression coefficient of y on x.

H_o: $\beta = 0$. Vs H_1: $\beta \neq 0$ is known

$$b_{yx} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} = \frac{599.62}{1020.6} = 0.27$$

And

$$\sum(y_i - \bar{y})^2 = \sum(y_i - \bar{y})^2 - b_{yx}\left[\sum(x_i - \bar{x})^2\right]$$

$$= 1020.6 - (.27)^2 \times 2247.5 = 1020.6 - 159.97 = 860.59$$

So

$$t = b_{yx}\frac{(n-2)\sum(x_i - \bar{x})^2}{\sqrt{\sum(y_i - \bar{y})^2}} = \frac{5 \times 2247.5}{\sqrt{860.59}} = 2.261$$

The table value of t (5%) on 25 degrees of freedom is 2.06

The calculated value of t is greater than the table value of t on 25 degrees of freedom and at 5% probability level so our null hypothesis is rejected and it is concluded that the observed value of regression coefficient b_{yx} is significant.

14.5 Tests of Significance Based on F-Distribution

(a) Testing of equality between two population standard deviations:

If $x_1, x_2, x_3, \ldots\ldots x_{n_1}$ is a random sample from a normal population with mean μ and variance $\sigma1^2$ and another random sample $y_1, y_2, y_3 \ldots\ldots y_{n_1}$ From another normal population with mean $\mu2$ and variance $\sigma2^2$ then to test the null hypothesis.

Ho: $\sigma1^2 = \sigma2^2$(on the assumption that $\mu_1 = \mu_2$)

$F = \frac{S_1^2}{S_2^2}$ $(where S_1^2 > S_2^2)$ is a F statistic on (n_1-1) and (n_2-1)

Degrees of freedom

Where

$$S_1^2 = \frac{\sum(x_i - \bar{x})^2}{n_1 - 1}$$

$$S_2^2 = \frac{\sum(y_i - \bar{y})^2}{n_2 - 1}$$

If $S_2^2 > S_1^2$ then

$F = \frac{S_1^2}{S_2^2} \sim F$ With (n_2-1) and (n_1-1) degrees of freedom.

If F calculated is greater than the table value of F at 5% probability level and on (n_1-1) and (n_2-1) degrees of freedom, the null hypothesis Ho will be rejected otherwise H0 will be accepted.

Example 1.10:

Two random samples of sizes 11 and 10 give the sum of square of deviations from their respective means equal to 180 and 144 respectively. Can they regarded as draw from two normal populations with same standard deviation.

Ho: $\sigma 1^2 = \sigma 2^2$

$$S_1^2 = \frac{\sum(x_i - \bar{x})^2}{n_1 - 1} = \frac{180}{10} = 18.0$$

$$S_2^2 = \frac{\sum(y_i - \bar{y})^2}{n_2 - 1} = \frac{144}{10} = 14.4$$

Since $S_2^2 > S_1^2$

$So \qquad F = \frac{S_1^2}{S_2^2} = \frac{18.0}{16.0} = 1.25$

Table value of F on 10 and 9 degrees of freedom and at 5% prob. Level is = 3.14

Since the calculated value of F is less than the table value of F at 5% prob. Level and on 10 and 9 degrees of freedom null hypothesis Ho is accepted and it may be concluded that there is no significant difference between standard deviations of two normal populations. Hence two sample can be regarded as drawn from two normal populations with same variances.

14.6 Tests of Significance Based on Fisher's z- Transformations

(a) To test the significance of correlation coefficient:

Let (x_1,y_1); (x_2,y_2)….(x_n,y_n) be random sample of size n taken from bivariate normal population and r be the sample correlation coefficient. We wish to test the null hypothesis.

$H_o: \rho = \rho_0$ Vs $H_1: \rho \neq \rho_0$

Where ρ is the population correlation coefficient and ρ_0 be again specified value for ρ which is different from zero.

To test H_o a testing procedure is suggested by Prof. R.A. Fisher. He suggested to use following Z: transformation.

$$z = \frac{1}{2} log_e \frac{1+r}{1-r} \quad and\ Z_0 = \frac{1}{2} log_e \frac{(1+\rho_0)}{(1-\rho_0)}$$

Or

$$z = 1.1513\ log_{10} \left(\frac{1+r}{1-r}\right) \quad and\ Z_0 = 1.15131 log_{10} \left(\frac{1+\rho_0}{1-\rho_0}\right)$$

For testing Ho

The variable z found to be normally distribution with mean z_0 and variance 1/ (n-3) under the null hypothesis for large n.

In other words, under Ho we have

$$Z = \frac{z - z_0}{\sqrt{1/(n-3)}}$$

Which is S.N.V., i.e.,

Z~ N (0,1)

So if calculated value of Z is greater than the table of Z= 1.96 at 5% significance level, then our null hypothesis will be rejected otherwise, Ho will be accepted.

B. Testing significance difference between two population correlation coefficients:

Suppose n_1, r_1, p_1 are sample size sample correlation coefficient and population correlation coefficient for one set of a random sample taken from a bivariate normal population while n_2, r_2, p_2 are the same values for another random sample taken from other bivariate normal population.

If we wish to test ho: $p_1=p_2$ then we calculate,

$$z_1 = \frac{1}{2} log_e \frac{1+r_1}{1-r_1} = 1.15131 log_{10} \frac{(1+r_1)}{(1-r_1)}$$

and

$$z_2 = \frac{1}{2} log_e \frac{1 + r_2}{1 - r_2} = 1.15131 log_{10} \frac{(1 + r_2)}{(1 - r_2)}$$

Under H_0 we have

$$Z = \frac{z_1 - z_2}{\sqrt{\frac{1}{(n_1 - 3)} + \frac{1}{(n_2 - 3)}}}$$

Following N (0,1)

Thus if calculated value of |Z| is greater than the table value of Z= 1.96 at 5% probability level, then we reject our null hypothesis H0, otherwise Ho will be accepted.

The above Z is also a Fisher's Z-transformation.

Example 1.11:

The value of sample correlation coefficient obtained from a sample of size 19 drawn a bivariate normal population is 0.8. Is this value consistent with the hypothesis that the correlation in the population is 6 at 5% level of significance?

Solution

There

$$z = \frac{1}{2} log_e \frac{1 + r}{1 - r} = 1.15131 log_{10} \frac{(1 + r)}{(1 - r)}$$

$$= 1.15131 log_{10} \frac{1 + .8}{(1 - .8)}$$

$$= 1.15131 log_{10} 9$$

$$= 1.15131(.9542) = 1.0985$$

and

$$z_0 = \frac{1}{2} log_e \frac{1 + \rho_0}{1 - \rho_0} = 1.15131 log_{10} \frac{(1 + \rho_0)}{(1 - \rho_0)}$$

$$= 1.1513\ log_{10} \frac{1 + .6}{(1 - .6)}$$

$$= 1.1513\ log_{10} 4$$

$$= 1.1513(.6021) = 1.0985$$

Under H_o

$$Z = \frac{z - z_0}{\sqrt{\frac{1}{(n-3)}}} = \frac{1{,}0985 - .6932}{\sqrt{\frac{1}{19-3}}}$$

$$= 4(.4053) = 1.6212$$

For two side alternative $H_1: \rho \neq .8$ the tabulated value of z at 5% level of significance is 1.96.

As calculated value falls in the region of acceptance, Ho may be accepted at 5% of the significance and we may say that at 5% level of significance, the value of sample correlation coefficient is consistent with the hypothesis that the population correlation is 0.6.

Example 1.12:

A sample of size 67 gave a correlation of 0.6 whereas another sample of 39 gave a correlation of 0.8 can these two samples be considered as coming from population having equal correlation coefficients?

Solution:

We have to test the null hypothesis

H_o: $\rho_1 = \rho_2$

Now, Z1= 1.1513 log

$$Z = \frac{z - z_0}{\sqrt{\frac{1}{(n-3)}}} = \frac{1{,}0985 - .6932}{\sqrt{\frac{1}{19-3}}}$$

$$= 1.1513 \times 0.6021 = 0.6932$$

$$Z_2 = 1.1513\, log_{10} \frac{1.8}{100.2} = 1.1513 \times log_{10}(9.0)$$

$$= 1.1513 \times 0.9542 = 1.099$$

$$Z = \frac{z_1 - z_2}{\sqrt{\frac{1}{(n_1-3)} + \frac{1}{(n_2-3)}}} = \frac{.6932 - 1.099}{\sqrt{\frac{1}{64} + \frac{1}{36}}}$$

$$= (-.4058) \times 4.8 = -1.948$$

|Z|= 1.948

Since the calculated value of |Z| is smaller than the table value of Z = 1.96 at 5% probability level so null hypothesis Ho will be accepted. Therefore it can be concluded that there is no significant difference between correlations of two population i.e. two samples come from population having equal correlations.

(C) To test the homogeneity of correlation coefficients

Let ri be the sample correlation coefficient based on a random sample of size ni taken from the bivariate normal population with correlation coefficient p_1, i= 1,2,3…., k (k>2). We wish to test Ho: p_1=p_2= ……p_k. We define,

$$z_i = \frac{1}{2} log_e \frac{1 + r_i}{1 - r_i} = 1.15131 log_{10} \frac{(1 + r_i)}{(1 - r_i)}$$

and

$$\bar{Z} = \frac{\sum_{i=1}^{k} (n_i - 3) z_i}{\sum_{i=1}^{k} (n_i - 3)}$$

Then under Ho:

$$\sum_{i=1}^{k} (n_i - 3)(z_i - \bar{z})^2$$

Follows chi-square distribution with (k-1) degrees of freedom.

If Ho is accepted at α% level significance, then an strate of population correlation coefficient is given by

$$\bar{z} = \frac{1}{2} log_e \frac{1 + \hat{\rho}}{1 - \hat{\rho}}$$

Or

$$e^{25} = \frac{1 + \hat{\rho}}{1 - \hat{\rho}}$$

Or

$$(1 - \hat{\rho}) e^{25} = 1 + \hat{\rho}$$

Or

$$e^{25} - \hat{\rho} e^{25} = 1 + \hat{\rho}$$

Or

$$e^{25} - 1 = \hat{\rho}(1 + e^{25})$$

Or

$$\hat{\rho} = \frac{e^{25} + 1}{e^{25} - 1}$$

= tan h ($\bar{z}$)

(D) Testing the significance of the ratio to two independent estimates of the population variance (The variance- ratio test)

Let there be two independent samples of sizes n_1 and n_2 taken from normal population with variance σ_1^2 *and* σ_2^2 respectively and s_1^2, s_2^2 are the unbiased standards of σ_1^2 *and* σ_2^2 respectively based on these samples. Under H0: the statistic $s_1^2/\ s_2^2$ follows Snedecor's F-distribution.

In fact, R.A. Fisher had originally defined the statistics z as

$$z = \frac{1}{2} log_e \frac{s_1^2}{s_2^2}$$

For testing Ho: $\sigma_1^2 = \sigma_2^2$

The distribution of this $z = \frac{1}{2} log_e \frac{s_1^2}{s_2^2}$ follows a less skewed distribution than F.

So this test may be used for testing Ho: $\sigma_1^2 = \sigma_2^2$

Remark:

Fisher's z transformation should not be confused with SNV. You may try the following Exercises.

14.7 Self Assessment Exercises

1. Talcom Powder is packed into tins by a machine. The weights of a random sample of 10 tins were taken in lbs as .46, .45, .49, .50., .49, .51, .50, .43, .44, .48. Examine at 5% level of significance if the variability in weights can be expressed by a standard deviation of 0.3 lbs.?
2. A random sample 9 out of a very large number of mass produced components gives a mean dimensions of 68 inches and an unbiased standard of population variance as (4.5) $(\text{inches})^2$. Are these data consistent with the assumption that the mean dimension in the population (assuming normal) is 68.5 inches at 5% level of significance?
3. Eight plots growing three wheat plants each were exposed to a high tension discharge while

nine similar plots were enclosed an earthen ware cage. The number of tillers in each plot were as follows:

Caged	17	26	18	25	28	26	23	17
Electrified	16	16	22	16	21	18	15	20

Discuss whether electrification makes a significant difference in the average tillers at 1% level of significance.

4. A correlation coefficient of 0.3 is obtained from a random sample of 27 pairs of observations from a bivariate normal population. Is this value is significantly different from zero correlation at 5% level of significance/
5. Random samples of sized 10 and 12 taken from two normal population with variances $\sigma_1^2 \ and \ \sigma_2^2$ respectively. These gave the values of their unbiased estimators as 144 and 324 respectively. Test the null hypothesis. Ho: $\sigma_1^2 = \sigma_2^2$ at 5% level of significance.

14.8 Solutions and Answers

9. Cal $\chi^2 = 7.83'$

Ho: may be accepted at 5% level of significance and we may concluded that s.d. of the wrights of ten's can be taken as .03 lbs.

10. cal |t| = .7

Ho: may be accepted at 5% level of significance and it ma be concluded that sample has been from the population having mean of 68.5 inches.

***11*:** Cal |t| = 2.75

Degrees of freedom = 15

Ho: may be accepted at 1% level of significance and it may be concluded that there is zero correlation in the population

12. Here , $F = \frac{s_1^2}{s_2^2} = \frac{324}{144} = 2.25$

Degrees of freedom (11,9)

F.05 (11,9) = 3.15

Ho: is accepted at 5% level of significance.

14.9 Summary

The statistic on which our testing procedure is based is known as test statistic. The probability distribution of this statistic determines the nature of the test. If the probability of this statistic is Chi-square or t or F then the tests are reffered ad Chi-square, t or F tests. The statistic differs from one situation to another. The various tests in detail have been discussed alongwith situations where they are applicable.

14.10 Further Readings

1. Fundamentals of Statistics volume I by A.N. Goon, B.D. Gupta and Das gupta. Pub; Calcutta Publishing House, Kolkata
2. Introduction to Mathematical statistics by Hogg and Craig.
3. Introduction to Mathematical Statistics by Mood, Graybill and Boes. Pub: Mac. Graw Hill

Unit-15: Large Sample Tests

Structure

15.1 Introduction

An important aspect of the sampling theory is to study the test of significance which enable us to decide (on the basis of a random sample of size n taken from the parent population) whether-

(i) The different between the observed sample statistic and the hypothesis parameter value, or

(ii) The difference between two sample statistics.

Is significant or might be attributed due to chance or the fluctuations of the sampling.

For applying tests of significance, one first sets up hull and alternative hypothesis and takes decision regarding critical region, level of significance critical (or significant) value of the test statistic as per given conditions/ situation.

For large samples, corresponding to test statistic T, the variable

$$Z = \frac{T - E(T)}{SE(T)}$$

Is assumed to be normally distributed with mean zero and variance unity.

Generally, a random sample of size more than 30 is regarded as a large sample.

15.2 Objectives

After going therefore this unit, you will be able to apply large sample tests for

- Testing significance of mean
- Testing equality of means

- Testing significance of proportion
- Testing equality of proportions
- Testing significance of s.d.
- Testing equality of standard deviations.

15.3 Testing significance of mean

Let x_1, x_2, x_3,..........x_n be random sample of size n taken from N(μ, σ^2). The sample mean $\bar{x} = \frac{1}{n}\sum x_i$ follows normal distribution with mean μ and variance σ^2/n. For large samples, (i.e. where sample size is more than 30), it is true even if population is not normal. Thus,

$$Z = \frac{(\bar{x} - \mu)\sqrt{n}}{\sigma}$$

be have like a S.N.V. incase of large samples. However, if the parent population is normal then the result is true even for small samples.

With an increased sample size, the sample variance $S^2 = \frac{1}{n}\sum(x - \bar{x})^2$ can safely be taken as an approximation to population variance $\sigma^2 =$ (if not known). Thus without any significant error approximation in large samples we have,

$$Z = \frac{(\bar{x} - \mu)\sqrt{n}}{s}$$

As S.N.V.

So if one wishes to test H_0: $\mu = \mu_0$ in case of large samples with unknown population variance one may use the test statistic

$$Z = \frac{(\bar{x} - \mu_0)\sqrt{n}}{s}$$

For known variance one uses

$$Z = \frac{(\bar{x} - \mu_0)\sqrt{n}}{s}$$

To reject or accept the above null hypothesis, the calculated value of Z is compared with 1.96 which is table value of a standard normal variate at 5% probability level for two tail test and 1.645 for one sided test.

If the calculated value of | Z | is greater than 1.96, then we reject our null hypothesis Ho and otherwise it is accepted.

Example: (2.1) A machine is expected to produce nails of length 5.00cm. A random sample of 64 nails gave an average length of 5.8 cm. with standard deviation of 0.80 cm., can it be said that the machine is producing nails as per specification.?

Here $\bar{x}$ = 5.8cm, μ_0= 5.00cm σ = 0.80cm.

The null hypothesis is

H0: $\mu = \mu_0 = 5.00cm.$

To test the above null hypothesis we use:

$$Z = \frac{(\bar{x} - \mu_0)\sqrt{n}}{\sigma} = \frac{(5.8 - 5.0)}{0.8} \times \sqrt{64} = \frac{0.8 \times 8}{0.8} = 8.0$$

The calculated value of Z is greater than the table value of Z = 1.96 at 5% probability level, so our null hypothesis Ho is rejected. The calculated value of Z = 8.0 is also greater than the table value of Z = 2.58 at 1% probability level. Therefore the value is also significant at 1% level of significance. Therefore the value is also significant at 1% level of significance. Therefore from the above available data it can be concluded that machine is not performing up to specification.

Example 2.2 A random sample of 400 male students is found to have a mean height of 168 cm. Can it be reasonable regarded as a sample from a population with mean height = 167.8cm. and standard deviation 3.25cm?

Here $\bar{x}$ = 168, μ_0= 167.8cm, σ = 3.25, n = 400.

So

$$Z = \frac{(\bar{x} - \mu_0)}{\frac{\sigma}{\sqrt{n}}} = \frac{168 - 167.8}{\frac{3.25}{\sqrt{400}}} = \frac{0.2 \times 20}{3.25} = \frac{4.00}{3.25} = 123$$

The calculated value of is smaller than the table value of Z = 1.96 at 5% probability level, so our null hypothesis will be accepted. Hence there is no significant difference between sample mean and population mean. Therefore it can be reasonably regarded as a random sample from a population with mean 167.8cm and σ = 3.25

15.4 Testing equality of means

Suppose x_1, x_2,…………x_n is an random sample from a normal population with μ_1 as mean and variance σ_1^2 and another random sample y_1, y_2,….y_{n2} from second normal population with mean μ_2 and σ_1^2 . It is assumed that and are large two random samples are independent. Then we wish to test the null hypothesis.

n_1 and n_2 are large and two random samples are independent. Then we wish to test the null hypothesis.

Ho: $\mu_1 = \mu_2$

Now we know that

$$\bar{x} \sim N\left(\mu_1, \frac{\sigma_1^2}{n_1}\right)$$

$$and\ \bar{y} \sim N\left(\mu_2, \frac{\sigma_2^2}{n_2}\right)$$

Hence,

$$Z = \frac{(\bar{x} - \bar{y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} is\ a\ standard\ normal\ variate\ (S.N.V.)$$

So, if Ho: $\mu_1 = \mu_2$ is true then

$$Z = \frac{(\bar{x} - \bar{y})}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} is\ a\ standard\ normal\ variate\ (S.N.V.)$$

Hence if calculated value of | Ƶ | is greater than 1.96 which is table value of Ƶ at 5% probability level, then our null hypothesis Ho will be rejected, otherwise accepted. The table value of Ƶ at 1% level of significance is 2.58.

Example 2.3: A random sample of 150 village was taken from district A and the average population per village of was found to be 440 and standard devotion 32. Another random population of 250 villages from the same district gave an averaged population 480 per village with standard

Here $n_1 = 150$ $\bar{x} = 440$ $s_1 = 32$

$n_1 = 250$ $\bar{y} = 480$ $s_2 = 56$

We want to test the null hypothesis

Ho: $\mu_1 = \mu_2$

Under Ho,

$$Z = \frac{(\bar{x} - \bar{y})}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{(440 - 480)}{\sqrt{\frac{(32)^2}{150} + \frac{(56)^2}{250}}}$$

$$= \frac{-40}{\sqrt{6.83 + 12.542}} = \frac{-40}{\sqrt{19.372}} = -\frac{40}{4.42} = -9.09$$

Hence | Ƶ |= 9.09

Since the calculated value of | Ƶ | is greater than the table value of Ƶ = 2.58 at 1% level of significance. Therefore the different between the means is highly significant.

You may try the following exercise.

Example P- 2.1: The potential buyer of light bulbs purchased 50 bulbs of each of two brands. After testing these bulbs he found that brand A had a mean life of 1282 hours with a standard deviation of 80 hours, where as brand B had a mean life of 1208 hours, with a standard deviation of 94 hours. Can the buyer be quite certain that the two brands differ in quality?

14.5 Testing Significance of Proportion

If a random sample of size n is drawn from a population with population proportion P. then we wish to test the null hypothesis

Ho: P= P_o where P_o is a particular specified value of P

Standard error of sample proportion is

$$SE\ (p) = \sqrt{\frac{PQ}{n}}\ where\ Q = 1 - P$$

and n is large

Then to test the above null hypothesis, under Ho,

$$Z = \frac{p - p_o}{\sqrt{\frac{P_o Q_0}{n}}}\ is\ a\ Standard\ Normal\ Variate\ (SNV);\ Q_o = 1 - P_o$$

if calculated value of | Ƶ | is greater than the table value of Ƶ at 5% level of significance, our null hypothesis will be rejected, otherwise accepted. (the table value of Ƶ = 1.96 at 5% level for two sided test and the table value of Ƶ = 1.645 for one sided test)

Example 2.4 In a sample of 400 burners, there were 12 burners whose internal diameters were not within tolerance. Is this sufficient evidence for concluding that the manufacturing process is turning out more than 2% defective burners?

Here P= .02, Q= 1-P= 0.98 and p= 12/400= .03

The null hypothesis is:

Ho: P= 0.02

To test the above null hypothesis we use:

$$Z = \frac{p - p_o}{\sqrt{\frac{P_o Q_0}{n}}} = \frac{(.03 - .02)}{\sqrt{\frac{.02 \times .98}{400}}}$$

$$= \frac{0.01}{\sqrt{.000049}} = \frac{.01}{.007}$$

$$= 1.429$$

The calculated value of is less than the table value of = 1.645 (for one tail test) at 5% probability level, so our null hypothesis is accepted. Therefore, it can be concluded that the process is under control.

14.6 Testing Equality of Proportions

Let n_1 and n_2 be two large samples taken from two different population and we wish to test Ho: $P_1=P_2$, where P_1 and P_2 are population proportions of two quality characteristics. Suppose p_1 and p_2 be the corresponding sample proportions obtained from the two random samples drawn these populations. If n_1 and n_2 are sufficiently large, then under Ho:

$$Z = \frac{p_1 - p_2}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}} \sim N(0,1)$$

When q_1= 1-p_1 and q_2= 1-p_2

However if n_1 and n_2 are moderately large with define

$$p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} \ and\ q = 1 - p.$$

In this case under Ho,

$$Z = \frac{p_1 - p_2}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim N(0,1)$$

Example 2.5: A machine produced 20 defective articles in a batch of 400. After overhauling it produced 10 defectives in a batch of 300. Has the machine improved?

Ho: there is no difference in the improvement of the machine before and after overhauling.

Or Ho: p_1-p_2= 0

P_1= 20/400= .05 so q_1 = 1-p_1= 1-.05= .95

P_2 = proportion of defective articles after overhauling = 10/300= .033

So q_2= 1-p_2= 1-.033= 0.97

$$SE(p_1 - p_2) = \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$$

$$= \sqrt{\frac{.05 \times .95}{400} + \frac{.033 \times .967}{300}}$$

$$= \sqrt{0.00023} = 0.015$$

Then

$$Z = \frac{p_1 - p_2}{SE(p_1 - p_2)} = \frac{.05 - .033}{.015} = .\frac{017}{.015} = 1.134$$

The calculated value of | Ƶ | is smaller than the table value of Ƶ at 5% level of significance, so our null hypothesis Ho is accepted. Therefore, it can be concluded that the machine has not improved after overhauling

You may try the following exercise

Exercise P- 2.2: In a sample of 800 men from a certain city, 500 men are found to be smokers. In a sample of 900 from another city, 450 are found to be smokers. Do the data indicate that the two cities are significantly different with respect to prevalence of smoking habits among men?

14.7 Testing significance of standard derivation

The variance of a sample standard deviation say's is $\frac{\sigma^2}{2n}$ if a large sample of size n is drawn from a normal population of variance. It is to be noted that if parent population is not normal then this formula is not to be relied upon. Thus, for normal parent population to test Ho: $\sigma = \sigma_0$

Under Ho:

$$Z = \frac{s - \sigma_0}{\sqrt{\frac{\sigma_0^2}{2n}}} \; is\ a\ S.N.V.$$

14.8 Testing Equality of Standard Deviations

Let s_1 and s_2 be the sample standard deviation of the two large samples of sizes n_1 and n_2 taken from two normal population with variance $\sigma_1^2\ and \sigma_2^2$ respectively. For testing Ho: $\sigma_1 = \sigma_2$ under Ho, we have

$$Z = \frac{s_1 - s_2}{\sqrt{\frac{\sigma_1^2}{2n_1} + \frac{\sigma_2^2}{2n_2}}} \quad (\sigma_1, \sigma_2 known)$$

Or

$$Z = \frac{s_1 - s_2}{\sqrt{\frac{s_1^2}{2n_1} + \frac{s_2^2}{2n_2}}} \quad (\sigma_1, \sigma_2\ not\ known)$$

Or

$$Z = \frac{s_1 - s_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Where

$$S^2 = \frac{1}{n_1 + n_2}[n_1 s_1^2 + n_2 s_2^2]$$

as standard Normal Variate (S.N.V.).

Example 2.6: Random samples of sizes 300 and 270 taken from two normal population with variance $\sigma_1^2\ and \sigma_2^2$ respectively gave sample standard deviations as 240 and 202.

Test the hypothesis Ho: $\sigma_1 = \sigma_2$ at 1% level of significance.

Solution:

Here Ho: $\sigma_1 = \sigma_2$

Against H_1: $\sigma_1 \neq \sigma_2$

Under Ho,

$$Z = \frac{s_1 - s_2}{\sqrt{\frac{s_1^2}{2n_1} + \frac{s_2^2}{2n_2}}}$$

$$= \frac{240 - 202}{\sqrt{\frac{(240)^2}{2 \times 300} + \frac{(202)^2}{2 \times 270}}} = \frac{38}{13} \approx 3$$

For two sided test the tabulated value of Z at 1% level of significance is 2.58.

As calculated value falls in the critical region, Ho may be rejected at 1% level of significance and we may conclude that the difference between the standard deviations is significant at 1% level of significance.

You may try the following exercises.

15.9 Self Assessment Exercises

E-2.3 A random sample of 900 members is found to have a men of 3.4cms. Could it be regarded as a sample from a large population with mean 3.25cms. and s.d. 2.61 cm. at 5% level of significance/

E-2.4 A random sample of 6400 Englishmen has mean height of 67.85 inches with s.d. 2.56 inches while a sample of 1600 Austrians has a mean height of 68.55 inches with s.d. 2.52 inches. Do the date indicate that the Austrians are on average taller than Englishmen at 5% (or 1%) of significance?

E- 2.5 If the expectation is that 2% of men of exact age 60 years will die within the year and if out of 900 such men, 24 die within the year, can the group be regarded as a random sample of such men at 5% level of significance?

E- 2.6 In a large city A, 20% of a random sample of 900 school boys had a slight physical defect. In another city B, 18.5% of a random sample of 1600 school boys had the some defect. Test whether the difference between the two properties is significant at 5% level of significance?

15.10 Solutions and Answers

E-2.3 Cal. |Z|= 1.72

Ho is accepted at 5% level of significance and are may conclude that sample belongs to a large population with mean 3.25cms and s.d. 2.61cms.

E-2.4 Cal. |Z|= 9.21 (One sided test)

Ho is rejected at 5% (or 1%) level of significance and are may conclude that Austrian on an overage, taller than Englishmen.

E- 2.5 Cal. |Z|= 1.43

Ho is accepted at 5% level of significance and it may be concluded that the difference between the proportions is insignificant.

E- 2.6 Cal. |Z|= 0.94

Ho is accepted at 5% level of significance and it may be concluded that the difference between the proportions is insignificant.

15.11 Summary

A random sample of size n where n>30 is regarded as a large sample. In case of large samples the probability distribution of test statistic is approximately normal. The various tests have been discussed along-with illustrations and situations where they are applicable.

15.12 Further Readings

1. Fundamentals of Statistics volume I by A.N. Goon, B.D. Gupta and Dasgupta. Pub; Calcutta Publishing House, Kolkata
2. Introduction to Mathematical Statistics by Mood, Graybill and Boes. Pub: Mac. Graw Hill
3. Introduction to Mathematical statistics by Hogg and Craig.

Unit-16: Non- Parametric Tests

Structure

16.1 Introduction

In this unit we will be introduce you with non- parametric tests. In section 3.3, an attempt is made to explain the difference between parametric and non- parametric inferences highlighting the wider scope of latter to the former in terms of its use with less restrictive assumptions. In the following sections some of the non- parametric tests have been discussed. We will start with sign test discussed in section 3.4, which is two sample tests. It is an alternative to paired t-test, which is used only if sampling is form normal population. For same problem, Wilcoxon test is also developed in section 3.5. If the samples are independent, Mann Whitney- Wilcoxon Test (discussed in section 3.5) is used to test the run test has been testing the randomness of a series. It can also be used to test the equality of populations form which two independent samples are drawn. For your convenience, table are provided in the last as appendix to be used for the calculation of P-values. A summary is also provided to check whether you have achieved the objectives or not.

At the end of each sub section few unsolved problems are given. Solve these and check whether you have achieved the following objectives.

16.2 Objectives

After reading this unit you should be able to :

- Discriminate between parametric and non parametric inference
- Formulate the null and alternative non- parametric hypothesis
- Choose a non parametric test for the formulated hypothesis
- Draw inference after calculation of the test statistics

16.3 Test of significance based on Chi-square Distribution.

You have already studied the tests of significance in the earlier units. You may not that the test discussed so far are mostly concerned with testing the validity of certain statement about the parameter of the population from which the sample is drawn. For example, you have studied the tests for testing the mean/ variance of a normal population, equality of means/ variances of two normal populations. In most of these situations, we assume that the samples comes from a population having a known specific mathematical form of the distribution function (through its probability distribution function or probability mass function); the only thing unknown is the parameters (arbitrary constants involved in the distribution). Therefore, we consider for the population a family of distribution F $(X|\theta)$. A particular value of θ completely determines a unique member of the family. Since these are related to the parameter of the population, these are called parameter tests.

Often however we may not have enough information to specify a family of distribution for the population from which sample is drawn. For example we may not be able to decide whether the sample is from normal family of distribution or exponential family of distribution or from any other distribution. Naturally, a question that arises at this stage is "Can we still develop tests"? The answer is yes. We may develop tests, consider a non-parametric model for the population. By a non-parametric model, we mean a much wider class of distribution whose mathematical form is unspecified. We only make general assumptions about the distribution; for example, family of all distributions having finite mean, family of all distributions having unique mode and family of distributions symmetric about their median etc. The tests based on such assumptions about distribution of the population from which sample is drawn are called non parametric tests or distribution free tests.

The terms non-parametric and distribution free are used interchangeably. It should be noted however, that non parametric does not mean that there is no parameter and distribution free does not mean that there is no distribution involved in the test procedure.

In non parametric or distribution free test procedure, the most that we will assume is that the underlying distribution is of the continuous type. This is a mathematically convenient assumption that allows us to assume that the observations can be arranged in increasing (decreasing) order with no ties (you know that for continuous random variables P(X=Y) =0). This family is so wide that it general the non- parametric or distribution free test procedures may not be as sharp and powerful as those developed for specific parametric or distribution-free test procedures are almost as good as the parametric test procedure we must make sure that the underlying assumption about the distribution is true or at least reasonably justified. If it is felt that there is not enough information available to choose a specific parametric distribution, it is advisable to use a non parametric or distribution free test procedure.

In the following sections, we will discuss a few non parametric or distribution free test procedures; namely Sign Test, Wilcoxon Signed- Ranks Test, Mann- Whitney Test and Run Test.

16.4 Sign Test

Suppose we want to test whether a particular type of coaching helps the students to improve their performance in the examination. For this purpose a group (say, n) of students was selected randomly and two examinations were conducted for them, one before providing the coaching and the other after the coaching. The marks obtained by the i^{th} student (i= 1,2,…..,n) in the two examinations are X_i and Y_i respectively. How do we go about testing whether the coaching has improved the performance of the students in general? Without giving a deeper thought to the problem one can suggest the use of paired t-test. But you should note that the test is valid if we assume that (X_i, Yi) follows bi-variate normal distribution and the marks of the students may follow some skewed distribution. Under such a suspicion one can proceed in a very simple way as described below.

Let us consider the difference $D_i= X_i-Y_i$. If D_i is positive, it indicates that the performance of the i^{th} student has improved. On the other hand, if D_i is negative, it is an indicator of deterioration of the performance of the student. If most of the D_i is positive, it indicates that in general the performance of the students have increased. In the light of the above arguments, one can develop a test procedure as given below.

Let us consider the null hypothesis to be no significant improvement in the performance hence probability of increase in the mark is same as the probability of a decrease i.e. P (X>Y) = P (X<Y). Therefore, the null hypothesis is specified as Ho: P (X>Y) = P (X<Y). The alternative hypothesis is, naturally, that there is improvement in the performance after the coaching i.e. H_1: P (X>Y) < P (X<Y). We assume here that (X, Y) has jointly continuous distribution so that P (X=Y) = 0, that is, ties occur with zero probability. In that case

$1= P(X>Y) + P(X<Y)$

Therefore, under H_o

$P(X>Y) = ½ = P(X-Y<0)$

Similarly we may see that under H_1

$P(Y-X<0)>1/2$.

Let S is the number of pairs of (X_i, Y_i) for which $X_i > Y_i$ i.e., the difference, $D_i= Y_i-X_i$, is negative. Then under H_o: S has binomial distribution with parameter n and p =P (X>Y)= ½. Since, positive values D_i indicates improvement therefore, we may reject the null hypothesis if the values of S is small and define the test procedure as:

Reject H_o: if $S \leq C$ Where C is to be chosen such that P ($S \leq C|H_0$) is less than or equal to the prefixed level of significance (say, α). Now,

$$P\,(S \leq C|H0) = \frac{1}{2^n}\sum_{k=0}^{c}\binom{n}{k}$$

and, hence C is chosen to be the largest value such that

$$\frac{1}{2^n}\sum_{k=0}^{c}\binom{n}{k} \leq \alpha$$

Since the test statistic count the number of negative (or positive) sign of Di, the resulting test is called ***Sign test***.

It should be noted that to perform the sign test we do not need the numerical values of Di= Yi-Xi, i= 1,2,……n. We only need to know the sign of Di= Yi-Xi for each i. Also even though we have assumed that P(Yi-Xi= 0)=0, ties do occur in practice. *In case of ties, we remove the tied pairs from consideration and perform the test with remaining n observations.*

n = n- number of ties observations*

Example 3.1:

To test the effectiveness of a new medicine in reducing the weight, nine randomly selected persons were weighted before and after they took the medicine for three months. Their weights (in kilograms) before and after taking the medicine were recorded as follows:

S. No.	Weight before use of medicine (X)	Weight after use of medicine (Y)	Sign of D= Y-X
1	80	79	-
2	82	83	+
3	75	75	Tie
4	90	81	-
5	98	95	-
6	87	86	-
7	100	101	+
8	107	105	-
9	103	100	-

The null hypothesis that we wish to test here is that the medicine is ineffective that is H_o p = P (X>Y) + P (X<Y) = ½

Against the hypothesis that medicine is effective, that is

H_1: P (X>Y)> 1/2 (or P (X<Y) < ½)

Since there is one tie (S. No. 3), we drop that observation and use the remaining 8 pairs of observation only. We may also note that negative sign of D is indicator of effectiveness of the medicine. Therefore lesser the number of positive differences (Say, S), greater chances of effectiveness of the medicine and hence we may reject the null hypothesis of non effectiveness in favor of alternative hypothesis of effectiveness of the medicine if positive differences are less in number. Here the number of positive differences are less in number. Here the number of positive differences are less in number. Here the number of positive differences is 2. Further, if we decide to reject the null hypothesis for S≤C,

$$P\,(S \leq C|H0) = \frac{1}{2^8}\sum_{k=0}^{c}\binom{8}{k} = \begin{cases} 0.1446 \text{ for } C = 2 \\ 0.0352 \text{ for } C = 2 \end{cases}$$

If the level of significance α is prefixed at 0.10, the test procedure may be defined as reject Ho at 10% level of significance if S is less than or equal to 1. Here, S=2, therefore we may conclude that data do not provide enough evidence for the rejection null hypothesis. Calculating the p value, we arrive at the same conclusion because p-value, P(S≤2|Ho)= 0.1446, is greater than the prefixed level of significance.

Remark:

1. The test discussed above consider one type of one sided alternative hypothesis. Similarly for other type of one sided alternative or two sided alternative we can develop the test procedures. The test are presented below.

To test null hypothesis H_o P (X<Y) = P (X>Y) = ½ for jointly continuous random variables X and Y (based on the number S of alternative differences of D= Y-X in a sample of size n, without ties) against the alternative hypothesis.

(i) H_1: P (X<Y)> ½, reject Ho at $\propto$ % level of significance if 0≤S≤C where C is chosen such that

$$\frac{1}{2^n}\sum_{k=0}^{c}\binom{n}{k} \leq \alpha$$

(ii) H_1: P (X<Y)< ½, reject Ho at $\propto$ % level of significance if n-C≤ S≤ n where C is chosen such that

$$\frac{1}{2^n}\sum_{k=n-c}^{n}\binom{n}{k} = \frac{1}{2^n}\sum_{k=0}^{c}\binom{n}{k} \leq \alpha.$$

(iii) H_1: P (X<Y)≠ ½, reject Ho at ∝ % level of significance if 0≤S≤C or n-C ≤ S≤ n where C is chosen such that

$$\frac{1}{2^n}\sum_{k=0}^{c}\binom{n}{k}+\frac{1}{2^n}\sum_{k=n-c}^{n}\binom{n}{k}=\frac{1}{2^{n-1}}\sum_{k=0}^{c}\binom{n}{k}\leq \alpha.$$

2. In many applications X's and Y's independent. This happens, for example when we need to compare two treatments or drugs to ascertain which is better. In this case the two drugs cannot, in general be administered the first drug and choose n other patients who receive the second drug. We may select n pair of patients in such a way that patients in such are matched as closely as possible for similar characteristics. In such pair we randomly select one to receive the first drug and the other to receive the second drug. Let Xi's and Yi's i=1,2,…n are the responses who receive the first and second drug respectively. Even now, the X's and Y's are independent. Under the null hypothesis that both the drugs are equally effective, we note that P(X<Y)= P(X>Y)= 1/2 and we can use the sign test described above. The above discussion should not be misunderstood as sign test can be used for two independent samples. It is only indicated that independent sample situation can be converted in paired observation case by prober planning before taking the actual observation.

3. ***Test of location:***

Suppose the we have a sample $X_1, X_2,\ldots.X_n$ from a continuous population and wish to test that the sample comes from a population having k_p as the quantile of order p i.e. p= P(X≤ k_o). Let us denote the unknown population quantile of order p by μ_p. In other words, we wish to test the null hypothesis H_o: μ_p= k_o. The alternative hypothesis H_1 can be one sided H_1; μ_p< k_o or H_1: = μ_p> ko or it can be two sided alternative H_1: $\mu_p \neq k_o$. Let S be the number of negatives in Di= Xi- k_o. Following the arguments given above we may reject Ho at α% level of significance against the alternative hypothesis.

(i) H_1: μ_p> k_o if 0≤S≤C where C is chosen such that

$$\sum_{k=0}^{c}\binom{n}{k}q^kp^{n-k}\leq \alpha$$

(ii) H_1: μ_p< k_o if n-C≤ S≤ n where C is chosen such that

$$\sum_{k=n-c}^{c}\binom{n}{k}q^kp^{n-k}=\sum_{k=n-c}^{c}\binom{n}{k}q^kp^{n-k}\leq \alpha.$$

(iii) H_1: $\mu_p \neq k_o$ if 0≤S≤C or n-C ≤ S≤ n where C is chosen such that

$$\sum_{k=n-c}^{c}\binom{n}{k}q^kp^{n-k}+\sum_{k=n-c}^{c}\binom{n}{k}q^kp^{n-k}\leq \alpha.$$

Or

$$2\sum_{k=0}^{c}\binom{n}{k} q^k p^{n-k} \leq \alpha.$$

It may be noted here that for testing H_0: $\mu_p \geq k_o$ against H_1: $\mu_p > k_o$ the test (i) and for testing H_0: $\mu_p \leq k_o$ against H_1: $\mu_p < k_o$ the test (ii) give above to be used.

Example 3.2:

A bank manager claims that the average waiting time in getting a demand draft prepared is 20 minutes. In order to verify his claim a random sample of 12 customers showed waiting times 25, 25, 19, 16, 21, 24, 28, 18, 24, 28, 15, and 11 minutes. Does the data support the claim?

Here the null hypothesis to be tested is that the average waiting time (μ, say) is 20 minutes i.e. Ho: μ = 20 and the alternative hypothesis is H1: $\mu \neq 0$. Let us assume that the waiting time X has a symmetric distribution about μ . Therefore the null hypothesis can equivalently be expressed as P(X<20) = P(X>20)= ½. There are seven negative signs in X_i-20, i= 1,2,…12 that is S=7. For given level of significance the α say, .10., the rejection region for null hypothesis is 0 ≤ S ≤ C or n-C ≤ S ≤ n where C is chosen such that

$$2\sum_{k=0}^{c}\binom{12}{k}\left(\frac{1}{2}\right)^{12} \leq .10$$

It follows that C=2 and the actual size of the test is 0.0386. For C=3, the size of the test become 0.2534. Since the calculated value of S= 7 is not in the rejection region we conclude that the data do not provide evidence for the rejection of the null hypothesis at 10% level of significance.

The same conclusion is arrived at through P-value calculation also because under Ho,

$$P(S \geq 7) = \sum_{k=7}^{12}\binom{12}{k}\left(\frac{1}{2}\right)^{12} = 0.3871$$

And hence the P-value is 2(.3871)= 0.7742.

Now you can solve the following problem on your own. Check your answers given in the section 3.9.

Ex- 3.1: Ten chickens are fed an experimental diet for four weeks and their weight gains in grams are recorded as given below:

80,75,65,84,40,60,49,50,38,39.

Test the manufacturer's claim that average weight increase is 48 grams or more.

Ex-3.2 A course was taught by two methods: the classroom instruction method A and the self study method B. Ten students each are selected randomly and carefully matched according to their background and ability prior to instruction. At the end of the course ten pairs of scores are recorded as follows:

Pair No.	Score	
	Method A	Method B
1	96	99
2	112	110
3	115	111
4	98	103
5	95	90
6	110	95
7	98	100
8	111	92
9	97	99
10	98	93

Do the data indicate any significant different between the two methods at 10% level of significance?

E- 3.3 Eleven students are given a course designed to improve their I.Q. on a standard test. Their scores before and after the course are recorded as follows:

I.Q. before	95	110	120	98	99	95
I.Q. after	98	114	124	93	99	106
I.Q. before	89	93	95	116	95	
I.Q. after	106	95	103	124	111	

What can you say about the effectiveness of the course?

The sign test discussed above is a test of median or for comparing two population based on paired observation. It is in fact rather crude in the sense that it used only the sign of the difference but ignores the magnitude of these differences. Thus it provides equal weight to the differences. Let us recall the example of improvement in the performance of the student after coaching. It looks illogical to give equal weight to greater gains and marginal loss and vice-a versa. An alternative test to sign test is Wilcoxon test, which takes into account both the magnitude and the sign of the differences.

16.5 Wilcoxon Signed- Rank Test

Let $X_1, X_2, \ldots .. X_n$ be a random sample from a continuous symmetric population and m (unknown) is the population median. We wish to test the hypothesis Ho: $m = m_o$ (assumed median). Consider the differences $D_i = X_i - m_o$. Clearly under null hypothesis Di has continuous and symmetric distribution bout zero. Consequently, under null hypothesis, we expect negative and positive differences to be about evenly spread out and hence number of positive and negative differences should be more our less equal as expected in median test also. In addition to this positive and negative differences of equal absolute magnitude should occur with equal probability. Since Xi's have a continuous distribution $P(D_i=0)=0$ and we assume that $|D_i|>0$ for i=1,2,……,n and $D_i \neq D_j$ for $i \neq j$. Now consider the absolute value of the differences and rank them from. 1 (for smallest) to n (for the largest) keeping track of the original sign. Now, find sum of the ranks of the positive differences W+ and sum of the ranks of the negative differences W-. Obviously

$$W_+ + W_- = \sum_{i=1}^{n} i = \frac{n(n+1)}{2}$$

So that W+ and W- offer equivalent test statistics. Under null hypothesis we expect W+ and W- to be more or less same. A large value of W+ indicate that most of the larger ranks are assigned to the positive Di's hence it follows that large values of W+ support the true median m is greater than the assumed median m_o. Therefore if we have to test the null hypothesis H_o: $m=m_o$ against the alternative hypothesis H_1: $m>m_o$, we may reject H_o if W+ is large. Let W_o be the observed value of W+ then p-value (probability that $W_+ \geq W_o$ under the null hypothesis) can be obtained from the table of Cumulative Right Tail Probability For Wilcoxon Signed Rank Test' given in the appendix as Table I. A similar argument applies to the alternative hypothesis H_1: $m<m_o$ and H_1: $m\neq m_o$ to get the corresponding region of rejection of null hypothesis and associated p-values. The various region of rejection with corresponding null hypothesis and associated Ho, alternative hypothesis Hi and p-value is given in following table:

Wilcoxon Signed Rank test

Null Ho	Alternative H_1	Reject H_o if	p-value
$m= m_o$	$m> m_o$	W+ is large	$P(W_+ \geq W_o(c)\ H_o)$
$m= m_o$	$m< m_o$	W- is large	$P(W_- \geq W_o(c)\ H_o)$
$m= m_o$	$m\neq m_o$	W+ or W- is large	2 (smaller tail probability

The table mentioned above provides the cumulative right tail probability $P(W\geq W_0\ |H_o)$ for values of $W_0 \geq n(n+1)/4$ and n=2,3,…..15. Here W is either W+ or W- as the case may be. For large n we use the normal approximation. Under H0, the common distribution of W+ and W- is symmetric about the mean n(n+1)/4 with variance n(n+1)(2n+1)/24, therefore

$$Z = \frac{W - n(n+1)/4}{\sqrt{n(n+1)(2n+1)/24}}$$

Follows standard normal distribution and p-values can be obtained. For example

$P(W_+ \geq W_0\ |H_o) = P(W\geq W_0 - 1/2\ |H_o)$ (continuity correction)

$$= P\left(Z \geq \frac{W_{+} - .5 - n(n+1)/4}{\sqrt{n(n+1)(2n+1)/24}}\right)$$

Problem of Ties and Zeroes:

Even though we assume that the distribution is continuous, ties and zero do occur in practice. As in the case of sign test, the recommended procedure is to drop the zeros and use the reduced sample size. If there are ties i.e. if some of the Di's are equal, assign to each such Di a rank equal to the simple arithmetic average of the ranks, which these Di's would have received if they were not equal. For example an ordered observations 2,2,5,5,7,8,8,8,10 are assigned respectively the ranks 1.5, 1.5, 3.5, 3.5, 5, 7,7,7, 9.

Example 3.3:

A random sample of 15 infants shows the following pulse rate (beats per minute):

119, 120, 125, 122, 118, 117, 126, 114, 115, 123, 121, 120, 124, 127, 126

Does the data confirm that the median pulse rate of infants is different from 120 beats per minute?

We assume that the distribution of pulse rate is symmetric and continuous. The null hypothesis to be tested here is H_0: m=120 against the alternative H_1; m $\neq$120.Let us compute the $D_i = X_i$ -120 $|D_i|$ and rank of $|D_i|$.

Di: -1, 0, 5, 2, -2, -3, 6, -6, -5, 3, 1, 0, 4, 7, 6

|Di| : 1, 0, 5, 2, 2, 3, 6, 6, 5, 3, 1, 0, 4, 7, 6

Excluding the two O's we arranged the observations in increasing order as

1, 1, 2, 2, 3, 3, 4, 5, 5, 6, 6, 6, 7

With ranks

1.4, 1.5, 3.5, 5.5, 5.5, 7, 8.5, 11, 11, 11, 13

Hence

W- = 1.5+3.5+5.5+8.5+11= 30

So that

W_+= 13(13+1)/2-30= 61

For the reduced sample size 13, from the table mentioned above we get $P(W_+ > W_o |Ho|) = .153$ so that the associated p- value is 2(.153) = .306. Thus under Ho the chance of observing as large a value of W+ as 61 is .306 and hence we can hardly reject H_o.

It may be interesting to see what would have happened if we would have used sign test. The number of negative signs are 5 so that p- value under H_o: p = ½ is

$2P(S \leq 5) = 2(.295) = .59$

We note that the result remains the same but sign test looks H_o: μ= 120 against H_1; $\mu \neq 120$ using two tailed t- test and compare the result.

Remarks:

1. In general, the sign test performs worse than the t-test and Wilcoxon signed- rank test, at least for large samples. For large samples, Wilcoxon signed rank test performs relatively well in comparison with the t-test (even when the underlying distribution is normal). For small samples from non-normal populations either the sign test or Wilcoxon signed rank test may be more powerful than t-test. However the Wilcoxon signed rank test should be preferred to sign test in the situations when the underlying distribution is symmetric.
2. The Wilcoxon signed rank test can also be used as a test for symmetry. Suppose X_1, X_2,….X_n is a random sample from a continuous population is symmetric about m. we can test either
   H_o: distribution is symmetric about m (known)
   Or
   H_o: distribution is symmetric and m= m_o
   By using W_+ or W_- exactly in the same way as described above.
3. The Wilcoxon signed rank test can be used for paired (X_i, Y_i), i=1,2,….,n. The null hypothesis to be tested in this case is Ho: median of D= (X-Y) = mo. The alternative hypothesis may one sided or two sided as per need. In this case the assumption is that D has a continuous and symmetric distribution about their mean m. The test is performed exactly as above by taking D_i= X_i-Y_i- m_o. In most of the cases m_o= o. To clear this point let us consider the following example.

Example 3.4:

In order to determine if smoking results in increased heart activity a random sample of 20 people was taken. Their pulse rate before and after smoking was taken. The results thus obtain is given below:

S. No	Pulse rate	
	Before	After
1	70	69
2	69	72
3	72	71
4	74	74
5	66	68

6	68	67
7	69	72
8	70	72
9	71	72
10	69	70
11	73	75
12	72	73
13	68	71
14	72	72
15	67	69
16	70	71
17	68	72
18	69	70
19	70	71
20	71	71

Define $D_i = X_i - Y_i$ i= 1,2,3,…..20, assuming that D_i have continuous symmetric distribution. If the pulse rates do not increase the population median m should be zero. Hence the null hypothesis in this case should be H_o: m=0 and it is to be H_1: m<o. We can use Wilcoxon signed rank test for this purpose. Ignoring the three observation where $D_i = 0$, D_i, $|Di|$ and rank of $|D_i|$ in increasing order of magnitude is given in the following table:

D_i	$\|D_i\|$	Rank of $\|D_i\|$
1	1	5
-3	3	15
1	1	5
-2	2	11.5
1	1	5
-3	3	15
-2	2	11.5
-1	1	5
-1	1	5
-2	2	11.5
-1	1	5
-3	3	15
-2	2	11.5
-1	1	5
-4	4	17
-1	1	5
-1	1	5

Thus W+ = 5+5+5+=15

So that

W-= 17(17+1)/2-15= 153-15= 138

Since n>15, we use normal approximation. We have under the null hypothesis

E(W-) = 17(17+1)/4 = 76.5

And V (W-) = 17(18) (35)/24 = 446.25

Hence

P(W≥138| Ho) = $P\left(Z > \frac{138-.5-76.5}{\sqrt{446.25}}\right)$

= P (Z>2.89) = .0019

Since p-value is highly significant, it is reasonable to conclude that smoking does increase the pulse rate.

The students are advised to check that if we use sign test the p-value comes to be .0064 if we use t-test the p-value < .005.

Now you can solve the following problems on your own. Check your answer given in the section 3.3.8.

E 3.4 A private technical college brochure claims that the average amount needed for boarding and lodging is Rs. 75/- per day. A random sample of nine students from the college showed the following daily expenditure:

75, 92, 80, 84, 73, 60, 84, 91, 78

Is there evidence to suggest that the college estimate is not correct? Assume that the daily expenditures are normally distributed, analyze the same data using t-test.

E-3.5 In order to keep the track of the inflation, a Network News program visited 10 selected supermarkets on 1.10.2006 and 1.10.2007 and purchage 30 pre selected items. The total cost data were as follows:

Super Market	Total Cost (in Rs.) on	
	1.10.2006	1.10.2007
1	5016	5318
2	5284	5694
3	4938	5242
4	4985	4982
5	5547	5486

6	5351	5642
7	5149	5100
8	4920	5210
9	5310	5580
10	5070	5475

Is there evidence to suggest that the average market cost remained the same over the year? Analyze the data using t-test also.

E-3.6 A group of 10 students is given a task to perform. Each student is then watched T.V. for an hour and asked to repeat the task. The time taken to perform the task before and after watching the T.V. is as follows:

Student's S. No.	Time taken to perform the task (min)	
	Before watching T.V.	After watching T.V.
1	5.0	8.5
2	3.5	4.0
3	6.5	6.0
4	8.5	8.0
5	4.5	6.0
6	5.0	5.0
7	2.5	4.0
8	3.0	5.5
9	4.5	4.0
10	3.5	6.0

How strong is the evidence that watching T.V. adversely effect the performance of the students?

E 3.7 The following observations were taken from a table of random numbers from a distribution F with median 0:

0.464	0.137	2.455	-.323	-0.068
0.906	-0.513	-0.525	0.595	0.881
-0.482	1.678	-0.057	-1.229	-0.486
-1.787	-0.261	1.237	1.046	

Is it reasonable to conclude that F is symmetric distribution?

The sign and the Wilcoxon signed rank test are in fact one sample tests (note that the paired observation case discussed above are actually converted into one sample problem by defining the difference) that are the non-parametric analogs of one one sample t-test. We will now discuss some non-parametric analogs of two sample t-test.

16.6 Mann- Whitney- U- Test

Let $X_1, X_2, \ldots X_m$ and $Y_1, Y_2, \ldots Y_n$

Be two independent samples from continuous distribution functions F and G respectively. We want to test H_o: f(x) = G(x) for all x ∈R against one or two sided alternative hypothesis. Let us define.

$$Z_{ij} = \begin{cases} 1, & if\ Y_j < X_i \\ 0, & if\ Y_j < X_i \end{cases}$$

Hence, $\sum_{i=1}^{m} \sum_{j=1}^{n} Z_{ij}$ is the number of Yj's < X_i and the statistic U_x is defined as

$$U_X = \sum_{i=1}^{m} \sum_{j=1}^{n} Z_{ij}$$

Since rank of Xi is the number of Y's and X's less than Xi

R (Xi) = (number of Y_j's < X_i) + rank of Xi in the sample of X's

Therefore we have

$$T_X = \sum_{i=1}^{m} R(X_i) = U_X + \sum_{i=1}^{m} i = U_X + \frac{m(m+1)}{2}$$

It follows that

$$U_X = T_X - \frac{m(m+1)}{2}$$

Example 3.5: Consider the example 3.4. It may be noted that there is no Y less than X_1 and X_2 three values of Y less than X_3 and two values of Y less than X_4. Therefore U_X = 0+0+3+2= 5 which is equal to $T_x - m(m+1)/2$ = 15-4(5)/2= 15-10= 5.

Thus U_X and T_x are equivalent test statistics. Therefore we can also define Mann- Whitney Wilcoxon test based on U_X as follows:

Mann-Whitney- Wilcoxon test for testing

H_o: f(x) = G(x)

Alternative H_1	Reject Ho if	p-value
f(x) ≥ G(x)	$U_x \leq c_1$ (U_x is too small)	$P(U_x \leq u_x \mid H_0)$
f(x) ≤ G(x)	$U_x \geq c_1$ (U_x is too large)	$P(U_x \geq u_x \mid H_0)$
f(x) ≠G(x)	$U_x \geq c_3$ and $Ux \leq c_4$ (U_x is either too large or too small)	2 (smaller tail probability)

In the above table u_x is the observed value of U_x.

In order to compute p-values (or to find the critical region if size is given), one needs the distribution T_x (or U_x) under null hypothesis. For the distribution of the test statistics you may refer Mann and Whitney (1947) and note that under H_o. Tx is m n (m+n+1)/2. Moreover, table for right tail probability $P(Tx \geq tx |H_o)$ are available and can be for the calculation of p-values. The left tail probabilities are obtained from the relation

$P(T_x \leq t_x |H_o) = P (T_x \geq m(m+n+1) – t_x |H_o)$

It may be noted that leveling of X and Y can be interchanged, we assume the sample than X's (ranks of X vary from 1 to m) and largest Y's are smaller than X's (ranks of X vary from n+1 to m+n). Thus,

M (m+1)$\leq$ Tx $\leq \sum_{i=1}^{m}(n+i) = m\,(m+2n+1)/2.$

For large values of m or n,

$$Z = \frac{T_x - \dfrac{m\,(m+n+1)}{2}}{\sqrt{\dfrac{mn(m+n+1)}{12}}}$$

Follows approximately standard normal distribution. Since Tx is an integer the application of continuity correction (subtracting 0.5 form the numerator) results in an improved approximation.

Example 3.6:

Seventeen students were randomly selected participate in an educational research project. A group of eight students was asked to attend a traditional lecture course for four weeks. The remaining nine students were provided self instructional material on videocassettes. At the end of four weeks all the students took the same test with the following results:

Lecture: 75 82 28 82 94 78 76 94

Self – instruction: 78 95 63 37 48 74 65 77 63

The null hypothesis is this case is that there is no difference between the two methods of instruction. Thus if $F_{(x)}$ and $G_{(x)}$ are the distribution functions of the scores of the students taking lecture and self instruction courses respectivel, Ho: $F_{(x)} \neq G_{(x)}$ against the alternative hypothesis H_1; $F \neq G$. Represent the lecture scores by X and the self-instruction scores by y, so that m = 8 and n=9, m<n. Combining the scores arranging in increasing order assigning average rank to tired observing we have.

Scores	Rank	X or Y
28	1	X
37	2	Y

48	3	Y
63	4.5	Y
63	4.5	Y
64	6	X
65	7	Y
74	8	Y
75	9	X
76	10	X
77	11	Y
78	12.5	X
78	12.5	Y
82	14.5	X
82	14.5	X
94	16	X
95	17	Y

The calculated value of T_x is

T_x= 1+6+9+10+12.5+14.5+16= 83.5

Since the mean is 8(8+9+1)/2= 72, we see that t_x is in the right tail and from the table we see that $P(T_x \geq 84 |H_o) = .161$ and $P(T_x \geq 84 |H_o) = .138$. Therefore we may estimate $P(T_x \geq 84 |H_o)$ as the average of $P (T_x \geq 84 |H_o)$ and $P (T_x \geq 84/H_o)$ to get $P (T_x \geq 84 |H_o) = (.161+.138)/2$. Therefore P-value comes out to be 0.299. Since the p-value shows that there is about 30% chance of observing as large a value of T_x as 83.5 in random sampling under H_o, we may conclude that that data do not provide enough information for the rejection of the hypothesis.

Example 3.7: In order to check that the breading strength of copper wire of brand Y is more than that of brand X, 6 measurements were taken for each with the following results:

X: 3.7 2.8 7.1 8.4 6.2 2.7

Y: 6.4 6.8 9.1 7.4 6.9 6.8

Let us test null hypothesis of equality of median strength of the two brands i.e., $H_o : m_x = m_y$ against and alternative hypothesis $H_1: m_x < m_y$.

Combining the two samples and arranging them in increasing order of magnitude we have

Observation	Rank	X or Y
2.7	1	X
2.8	2	X
3.7	3	X
6.2	4	X

6.4	5	Y
6.8	6.5	Y
6.8	6.5	Y
6.9	8	Y
7.1	9	X
7.4	10	Y
8.4	11	X
9.1	12	Y

Thus, t_x = 1+2+3+4+9+11=30 with p-value $P(T_x \leq 30/Ho) = P(T_x \geq 48 \mid Ho) = .09$. We may conclude that the data do not substantiate the claim that $m_x < m_y$ at 5% level they do not substantiate it at 10% level of significance.

Example 3.8: The failure times (in hundred hours) of a certain type of light bulb manufactured by two different companies, X and Y, are given below:

X: 1.7 1.8 1.9 1.1 0.7 0.9 2.1

Y: 2.1 2.7 1.6 1.8 1.7 1.8 1.6

X: 1.6 1.7 1.3

Y: 2.2 2.4 1.3 1.9 1.8 2.0

Do the data indicate a significant different between live of light bulbs manufactured by two different companies?

Here m = 10 and n = 13. If we denote the distribution of the life of the bulbs produced by the company X as F and that by company Y by G, the null hypothesis to be tested is H_o: F =G and the alternative hypothesis is H_1: F $\neq$ G. Combine the data and arrange in increasing order of magnitude, keeping the trace that observation is related to which company. It gives the following:

Observation	Rank	X or Y
0.7	1	X
0.9	2	X
1.1	3	X
1.3	4.5	X
1.3	4.5	Y
1.6	7	X
1.6	7	Y
1.6	7	Y
1.7	10	X
1.7	10	X
1.7	10	Y

1.8	13.5	X
1.8	13.5	Y
1.8	13.5	Y
1.8	13.5	Y
1.9	16.5	X
1.9	16.5	Y
2.0	18	Y
2.1	19.5	X
2.1	19.5	Y
2.2	21	Y
2.4	22	Y
2.7	23	Y

Thus T_x = 1+2+3+4.5+7+10+10+13.5+16.5+19.5= 87

Here, we use the normal approximation. We may note that under Ho

$E(T_x)$ = m (m+n+1/2) = 10 (10+13+1) = 120,

and

$V(T_x)$ = m n (m+n+1)/2 = 10 (13) (24)/2 = 260.

Since T_x < 120, the calculated value of the statistic lies in the left tail, Hence the p- value is

$2P(T_x \leq 87/H_o) = 2P(T_x \geq 87.5/ H_o)$ (for continuity correction)

$$= 2P\left(Z \leq \frac{87.5 - 120}{\sqrt{260}}\right)$$

$$2P\,(Z \leq\ -2.02) = 2(.0217) =\ .0434$$

Now, if the level of significance is fixed at α= .05, we conclude the null hypothesis Ho: F = G may be rejected and we say with 5% level of significance that lifetime of the bulbs produced by the two companies can not be regarded as similar. On the other hand, if α = .01, we have to conclude that data do not provide enough information for rejection of the null hypothesis.

Now you can solve the following problems on your own. Check your answers given in the section 3.3.8.

E- 3.8: Two laboratory cultures are to be compared for the difference in bacteria counts. Independent random samples of six from culture A and eight from culture B are taken. The number of bacteria per unit of volume is recorded as follows:

Culture A : 32 29 34 47 33 27

Culture B :	38	36	33	42	34	40	39	32

Is there any evidence to conclude that bacteria counts for the two populations are not same?

E-3.9 The order in which the test questions are asked, affect the student's ability to answer them correctly and hence affect the student's total grade. In order to check this proposition, two were made. Test A had questions set in increasing order of differently; in test B the order was reversed. A random sample of 20 students was selected in such a way that 10 pairs of students were matched in ability. From each pair, one student was assigned randomly to take test A and other test B. The following scores were obtained:

Test A:	83	82	95	92	91	60	89	69	70	72
Test B: 76	62	70	74	52	63	48	80	76	74	

Is there any evidence to indicate that the score on test B are lower than on test A? Use both Mann-Whitney – Wilcoxon and t-tests.

E- 3.10 Nineteen pieces of flint were collected, nine from area A and ten from area B. The object of the study was to determine if the pieces of the flint were of equal hardness. For the purposes of the study, nineteen pieces of flints of equal hardness from a third area were brought in. Each of the sample pieces was then rubbed against a pieced from the third area. The nineteen sample pieces were then ordered according to the amount of damage sustained from the softest (most damaged) to the hardest (least damaged):

A	A	B	B	A	A	B	B	B	B
A	A	B	A	A	A	B	B	B	

Is there any evidence to suggest that the flints from area A and B are of equal hardness?

The equality of two distribution from which the samples has been drawn can also be tested using "run" as a test statistics.

Now we will consider tests based on runs.

16.7 Run Test

Let us see first explain the meaning of "run" and "length of a run".

Run: If there is a sequence of two types of symbols a run means one or more identical symbols preceded and followed by a different symbol (or no symbol).

Length of a run: It is number of like symbols in a run.

Example 3.9:

The sexes of 15 children in order of their birth in hospital are recorded with the following result:

G G B B B G B B G G G G B G G

Where G stand for girl and B stand for boy. We see that at the beginning there are two 'G' followed by 'B' and preceded by none; thus if forms a run of length 2. Then there are three 'B' preceded and followed by 'G' thus is also forms a run and the length of this run in 3. Similarly other runs are underlined above. We may also note that there are in all 7 runs in this sequence.

The tests based on total number of runs can be used for testing the equality of the distributions from which the samples are drawn. Besides this, runs are also used for testing the randomness of a series of observations. These tests discussed below:

Wald- Wolfowith Run test for testing of equality of two distributions:

Let $X_1, X_2, \ldots, X_m$ and $Y_1, Y_2, \ldots Y_n$ be independent random samples with respective continuous distribution F and G respectively. If we combine the sample and arranged the observation in order of increasing magnitude writing X for the m observations on X and Y for Y- observations, we get a sequences of X and Y symbols. As explained above we may count the total number of runs in this sequence (m + n) symbols of X and Y. We assume for the time being that there are no ties. Under the null hypothesis H_o: F(x) = G(x) for all x, we expect the symbols X and Y to be well mixed giving rise to large number of runs. On the other hand if X's tend to be larger than Y's (i.e., F(x) > G(x)) then most of the Y's preceded the X's and therefore the total number of runs will lesser than expected under Ho. Similarly lesser number of runs is expected if Y's tend to be larger X's (i.e., F (x) > G(x)). Therefore we see that smaller values of total number of runs only indicate that F(x) and G(x) are not equal; it cannot be an indicator of F(x) > G(x) [or F(x) < G(x)]. In other words, the above discussion clearly establishes that run test can only be used for two sided alternative.

The run test for testing the null hypothesis H_o: F(x) = G(x) for all x, against the alternative hypothesis H_1: $F(x) \neq G(x)$ for at least one x, based on the total number of runs R of X and Y in the combined ordered sample is to reject Ho if the calculated number of runs r is small i.e., $r \leq C$ where C is to be chosen such that $P(R \leq C \mid H_o)$ is equal to or less than the prefixed level of significance α . Alternative we may reject Ho, if the P-value $P(R \leq r \mid Ho)$ is less than α .

In order to perform the test (or to compute the P-value), we need the distribution of R under the null hypothesis. Tables are available for left tail cumulative probabilities (i.e. $P(R \leq r/ H_o)$) for all values of $m \leq n$ with m + n 20. For larger values of m or n, we use the normal approximation. We have, under the null hypothesis H_o.

$$E(R) = 1 + [2mn / (m + n)]$$

and

$$V(R) = \frac{2mn\,(2mn - m - n)}{(m+n-1)\,(m+n)^2}$$

Therefore, the statistic

$$Z = \frac{R + 0.5 - \left(1 + \frac{2mn}{(m+n)}\right)}{\sqrt{\frac{2mn\,(2mn - m - n)}{(m+n-1)\,(m+n)^2}}}$$

Is approximately distributed as standard normal variable. Note that for continuity, .5 is added in the numerator of the above statistic because we are considering here the left tail critical region.

Problem of ties:

In case of ties between X's and Y's conservative procedure is to break the ties in all possible ways. For each such resolution of ties, compute the values of R. then choose that value of R, which is largest. Naturally it will give the largest p-value. For example if one of X and on of Y are tied, there are two possible resolutions:

….XY…

and

….YX….

Each may give different value of R. We take the larger of the two R-values.

Example 3.10:

Let us consider example no. 3.6, where we used Mann – Whitney- Wilcoxon test to compare two methods of instruction. You may note from the table showing the ordered arranged of the observations that 63, 78 and 82 are the repeated observations. But both the 63's are observations for self instruction method (Y) and both the 82's correspond to lecture method (X); therefore these do not effect the number of runs and the problem of ties do not arise because of these. However one to the 78's corresponds to X and the other to Y there is one tie between X and Y values at 78. Thus we can have two ways of breaking the tie. Taking the X value corresponding to 78 first, we have the sequence of X, Y in the combined sample as follows:

$$\underline{X}\ \ \underline{Y\,Y\,Y\,Y}\ \ \underline{X}\ \ \underline{Y\,Y}\ \ \underline{X\,X}\ \ \underline{Y}\ \ \underline{X}\ \ \underline{Y}\ \ \underline{X\,X\,X}\ \ \underline{Y}$$

So that R = 10. Taking Y value corresponding to 78 first, we have

$$\underline{X}\ \underline{YYYY}\ \underline{X}\ \underline{YY}\ \underline{XX}\ \underline{Y}\ \underline{X}\ \underline{Y}\ \underline{XXX}\ \underline{Y}$$

to give R = 8. Thus, the larger of the two, i.e., R = 10 is to be taken for testing of the hypothesis. The p-value is therefore, $P(R \leq 10'H_o) = .702$. You may note that it is considerably larger than the corresponding p-values obtained in using the Mann-Whitney-Wilcoxon test. In any case the data do not provide enough evidence for the rejection of the null hypothesis and therefore we may accept that there is no difference between the two methods of instruction.

Run test for testing of randomness:

Suppose that we wish to ascertain that the sex of the children in order of their births in random. The data for this purpose may be collected from a hospital by noting the sex of the children in order of their births. Now suppose that we get the data as given below for ten consecutive births:

MMMMMFFFFF

Or

MFMFMFMFMF

We notice that in the first data set male and female births are clustered at one place giving total number of runs as 2. On the other hand in the text data set male and female birth is alternate and total number of runs is 10. But, neither of the above series supports the view that sex of the child in order of birth is random because pattern follows in both the data set and in its turn the number of runs are either too large or to small to accept the hypothesis of randomness. Note that in each case the number of male and female births is equal to 5 and hence the lack of randomness could not have been noted by the use of chi-square or binomial tests on the frequencies. Thus it is only a run test, focusing on the order of the event, which reveals the striking feature of lack of randomness.

From the above discussions you may have got an idea that the statistic R can be used for testing whether an ordered sequences of two types of symbols is random arrangement or not. The null hypothesis to be tested here is

H_o: The sequences is random.

It is to be tested against the alternative hypothesis

H_1: sequences is non random.

As discussed above a large value of R or a small value of R indicates non randomness of the sequence and thus critical region is two sided. The hull hypothesis is rejected if the observed numbers of runs r_o form the data (consisting of m symbols of one type and n symbols of other type) is either too large or too small and the P-value is therefore, 2P (smaller tail probability).

Example 3.11:

Students were asked to toss a fair coin 20 times and report the result. A student reported the following sequences of head (H) and tail (T):

HH T H T H TH TT H T H TTT HH T H

Does this sequence of 10 heads and 10 tails indicate departure from randomness?

We observe that there here m= 10, n= 10 and there are 15 runs in all in the sequence which is greater than

$$E\ (R) = 1+2\ m\ n/\ (m + n)$$

$$= 1+2\ (10)\ (10)\ (10)$$

$$=11$$

Hence p-value is 2P $(R \geq 15/H_o)$. From the table we get

$P\ (R \geq 15/H_o) = 1-P\ (R \leq 14/\ H_o) = 1-.949 = .051$

$= P\ (R \leq 7/\ H_o)$

So that the P-value is 2 (.051) = .102 and hence at 5 % level of significance, we may conclude that data do not provide the evidence against the randomness.

Now you can solve the following problems on your own. Check your answers given in the section 3.3.8.

E- 3.11 Use run test to analyze the data given in E. 10

E- 3.12: In a study to test the equipotentiality theory, the learning (in a brightness – discrimination task) of 21 normal rats was compared with the relearning of 8 postoperative rats with cortical lesions. That is, the number of trials to relearning required postoperative by the 8 E rates was compared with the number of trials to learning required by the 21 C rates.

Trials required to learning/ relearning by E and C rats:

E rats: 20 55 29 24 75 56 31 45

C rats: 23, 8, 24, 15, 8, 6, 15, 15, 21, 23, 16, 15, 24, 15, 21, 15, 18, 14, 22, 15, 14

Is there no difference between normal rats and postoperative rats with cortical lesions with respect to rate of learning/ relearning in the brightness- discrimination task?

E- 3.13 Seventeen items emerging from a production line are tested and classified as defective (D) or non defective (N). The following sequence is obtained:

D D N D N N N N D D N N N N D N D

How strong is the evidence that there is lack of randomness in the series?

16.8 Summary

In this unit we have described four non parametric tests. The sign test is a test of median or quintiles is one sample case and for testing the equality of two population based on paired observations. However, the sigh test only utilizes the sign of the differences and ignores the

The number of positive signs is 7, the number of negative signs is 3 and n= 10.

P- value is 0.1718

Data do not provide enough evidence for rejection of H_o even at 10% level of significance.

E- 3.2 Ho: P(X>Y) = P(X<Y)

H_1: P(X>Y) $\neq$ P(X<Y)

The number of positive signs is 6, the number of negative signs is 4 and n= 10.

P- value is 0.7538

Data do not provide enough evidence for rejection of H_o even at 10% level of significance.

E- 3.3 Ho: P(X>Y) = P(X<Y)

H_1: P(X>Y) $\neq$ P(X<Y)

The number of positive signs is 1, the number of negative signs is 9 and n= 10 (one dite).

P- value is 0.0056

Data provide enough evidence for rejection of H_o even at 1% level of significance.

E- 3.4 The null hypothesis is H_o: μ = 75. It is to be tested against H_1: $\mu \neq 75$.

$W_+ = 29$ calculated t= 1.424

E- 3.5 The null hypothesis is H_o: $m = 0$. It is to be tested against H_1: $m \neq 0$.

$W- = 49$ calculated t= -3.76

E- 3.6 The null hypothesis is H_o: $m = 0$. It is to be tested against H_1: $m < 0$.

$W_- = 37.5.$

E- 3.7 The null hypothesis is H_o: F is symmetric. It is to be tested against H_1: F is not symmetric

Data provide evidence for rejection of H_o even at 1% level of significance.

E- 3.4 The null hypothesis is H_o: $\mu = 75$. It is to be tested against H_1: $\mu \neq 75$.

$W_+ = 29$ calculated $t= 1.424$

E- 3.5 The null hypothesis is H_o: $m = 0$. It is to be tested against H_1: $m \neq 0$.

$W_- = 49$ calculated $t= -3.76$

E- 3.6 The null hypothesis is H_o: $m = 0$. It is to be tested against H_1: $m < 0$.

$W_- = 37.5.$

E- 3.7 The null hypothesis is H_o: F is symmetric. It is to be tested against H_1: F is not symmetric

$W_+ = 111.$

E-3.8 The null hypothesis is H_o: $F(x) = G(x)$. It is to be tested against H_1: $F(x) \neq G(x)$

$m = 6$ and $n = 8$

$t_x = 33.5$

The t_x is in the left tail. Therefore P-value is $3P\,(T_x \leq 33.5\ |Ho)] = 2P\,(Tx \geq 56.5 \mid H_o) = [P\,(T_x \geq 56\ |H_o) + P\,(T_x \geq 57 \mid H_o)] = 0.162$.

E-3.9 The null hypothesis is H_o: $F(x) = G(x)$. It is to be tested against H_1: $F(x) \geq G(x)$

$m = 10$ and $n = 10$

$t_x = 130.5$

P-value = .029

E-3.10 The null hypothesis is H_o: $F(x) = G(x)$. It is to be tested against H_1: $F(x) \neq G(x)$

$m = 9$ and $n = 10$

$t_x = 82$

P-value = .604

m = 7 and n = 10

R = 9

P-value = .549

16.9 Appendix

Table 3.1: Table for Right Tail probabilities of Mann Whitney- Wilcoxon Test Statistic.

Right tail probabilities are given for $t_x \geq m\ (m+n+1)/2$ for $m \leq n$.

Left tail probabilities are obtained from the relation $P(T_x \leq t_x) = P\ (T_x \geq m\ (m+n+1)/2 - t_x)$

----

n	t_x	$P(T_x \geq t_x)$	n	t_x	$P(T_x \geq t_x)$	n	t_x	$P(T_x \geq t_x)$
	m=1			m=2			m=3	
1	2	.500		7	.400		15	.291
2	2	.667		8	.200		16	.218
	3	.333		9	.100		17	.164
3	3	.500	4	7	.600		18	.109
	4	.250		8	.400		19	.073
4	3	.600		9	267		20	.036
	4	.400		10	.133		21	.018
	5	.200		11	.067	10	13	.545
5	4	.500	5	8	.571		14	.455
	5	.333		9	.429		15	.379
	6	.167		10	.286		16	.303
6	4	.571		11	.190		17	.242
	5	.429		12	.095		18	.182
	6	.286		13	.048		19	.136
	7	.143	6	9	.571		20	.091
	4	.250		8	.400		19	.073
4	3	.600		9	.267		20	.036
	4	.400		10	.133		21	.018
	5	.200		11	.067	10	13	.545
5	4	.500	5	8	.571		14	.455
	5	.333		9	.429		15	.379
	6	.167		10	.286		16	.303

	4	.571		11	.190		17	.242
	5	.429		12	.095		18	.182
	6	.286		13	.048		19	.136
	7	.143	6	9	.571		20	.091
7	5	.500		10	.429		21	.061
	6	.375		11	.321		22	.030
	7	.250		12	.214		23	.015
	8	.125		13	.143			
8	5	.556		14	.071	M=3		
	6	.444		15	.036	3	11	500
	7	.333	7	10	.556		12	.350
	8	.222		11	.444		13	.200
	9	.111		12	.333		14	.100
9	6	.500		13	.250		15	.050
	7	.400		14	.167	4	12	.578
		.300		15	.111		13	.429
	9	.200		16	.056		14	.314
	10	.100		17	.028		15	.200
10	6	.545	8	11	.556		16	.114
	7	.455		12	.444		17	.057
	8	.364		13	.356		18	.029
	9	.273		14	.267	5	14	.500
	10	.182		15	.200		15	.393
	11	.091		16	.133		16	.286
				17	.089		17	.196
	M=2			18	.044		18	.125
2	5	.667		19	.022		19	.071
	6	.333	9	12	.545		20	.036
	7	.167		13	.455		21	.018
3	6	.600		14	.364	6	15	.548
	M=3	.		M=3			M=5	
	16	.452		31	.018		25	
	17	.357		32	.009		26	
	18	.274		33	.005		27	
	19	.190	10	21	.531		28	
	20	.131		22	.469		29	

	21	.083		23	.406		30
	22	.048		24	.346		31
	23	.024		25	.287		32
	24	.012		26	.234		33
7	17	.500		27	.185		34
	18	.417		28	.143	7	24
	19	.333		29	.108		25
	20	.258		30	.080		26
	21	.192		31	.056		27
	22	.133		32	.038		28
	23	.092		33	.024		29
	24	.058		34	.014		30
	25	.033		35	.007		31
	26	.017		36	.003		32
	27	.008					33
8	18	.539			M=4		34
	19	.461	4	18	.557		35
	20	.388		19	443		36
	21	.315		20	.343		37
	22	.248		21	.243		38
	23	.188		22	.171	8	26
	24	.139		23	.100		27
	25	.097		24	.057		28
	26	.067		25	.029		29
	27	.042		26	.014		30
	28	.024	5	20	.548		31
	29	.012		21	.452		32
	30	.006		22	.365		33
9	20	.500		23	.278		34
	21	.432		24	.206		35
	22	.364		25	.143		36
	23	.300		26	.095		37
	24	.241		27	.056		38
	25	.186		28	.032		39
	26	.141		29	.016		40
	27	.105		30	.008		41

	28	.073	6	22	.543		42	
	29	.050		23	.457	9	28	
	30	.032		24	.381		29	
	M=4			M=5			M=5	
	30	.413		32	.210		36	.472
	31	.355		33	.155		37	.416
	32	.302		34	.111		38	.362
	33	.252		35	.075		39	311
	34	.207		36	.048		40	.262
	35	.165		31	.028		41	.218
	36	.130		38	.016		42	.177
	37	.099		39	.008		43	.142
	38	.014		40	.004		44	.111
	39	.053	6	30	.535		45	085
	40	.038		31	.465		46	.064
	41	.025		32	.396		47	.047
	42	.017		33	.331		48	.033
	43	.010		34	.268		49	.023
	44	.006		35	.214		50	.015
	45	.003		36	.165		51	.009
	46	.001		37	.123		52	.005
10	30	.527		38	.089		53	.003
	31	.473		39	.063		54	.002
	32	.420		40	.041		55	.001
	33	.367		41	.026	9	38	.500
	34	.318		42	.015		39	.449
	35	.270		43	.009		40	.399
	36	.227		44	.004		41	.350
	31	.187		45	.002		42	.303
	38	.152		33	.500		43	.259
	39	.120		34	.438		44	.219
	40	.094		35	378		45	.182
	41	.071		36	.319		46	.149
	42	.053		37	.265		47	.120
	43	.038		38	.216		48	.095
	44	.027		39	.172		49	.073

	45	.018		40	.134		50	.056
	46	.012		41	.101		51	.041
	47	.007		42	.074		52	.030
	48	.004		43	.053		53	.021
	49	.002		44	.037		54	.014
	50	.001		45	.024		55	.009
				46	.015		56	.006
	M=5			47	.009		57	.003
5	28	.500		48	.005		58	.002
	29	.421		49	.003		59	.001
	30	.345		50	.001		60	.000
	31	.274	8	35	.528		40	.523
	M=5			M=6			M=6	
	41	.477		57	.001		67	.001
	42	.430	7	42	.527		68	.001
	43	.384		43	.473		69	.000
	44	.339		44	.418	9	48	.523
	45	.297		45	.365		49	.477
	46	.257		46	.314		50	.432
	47	.220		47	.267		51	.388
	48	.185		48	.223		52	.344
	49	.155		49	.183		53	.303
	50	.127		50	.147		54	.264
	51	.103		51	.117		55	.228
	52	.082		52	.090		56	.194
	53	.065		53	.069		57	.164
	54	.050		54	.051		58	.136
	55	.038		55	.037		59	.112
	56	.028		56	.026		60	.091
	57	.020		57	.017		61	.072
	58	.014		58	.011		62	.057
	59	.010		59	.007		63	.044
	60	.006		60	.004		64	.033
	61	.004		61	.002		65	.025
	62	.002		62	.001		66	.018
	63	.001		63	.001		67	.013

	64	.001	8	45	.525		68	.009
	65	.000		46	.475		69	.006
				47	.426		70	.004
	M=6			48	.377		71	.002
6	39	.531		49	.331		72	.001
	40	.469		50	.286		73	.001
	41	.409		51	.245		74	.000
	42	.350		52	.207		75	.000
	43	.294		53	.172	10	51	.521
	44	.242		54	.141		52	.479
	45	.197		55	.114		53	.437
	46	.155		56	.091		54	.396
	47	.120		57	.071		55	.356
	48	.090		58	.054		56	.318
	49	.066		59	.041		57	.281
	50	.047		60	.030		58	.246
	51	.032		61	.021		59	.214
	52	.021		62	.015		60	.184
	53	.013		63	.010		61	.157
	54	.008		64	.006		62	.132
	55	.004		65	.004	10	63	.110
	56	.002		66	.002		64	.090
	M=6			M=7			M=7	
	65	.074		57	.478		77	.036
	66	.059		58	.433		78	.027
	67	.047		59	.389		79	.021
	68	.036	9	60	.347		80	.016
	69	.028		61	.306		81	.011
	70	.021		62	.368		82	.008
	71	.016		63	.232		83	.006
	72	.011		64	.198		84	004
	73	.008		65	.268		85	.003
	74	.005		66	.240		86	.002
	75	.004		67	.116		87	.001
	76	.002		68	.095		88	.001
	77	.001		69	.076		89	.000

	78	.001		70	.060		90	.000
	79	.000		71	.047		91	.000
	80	.000		72	.036	10	63	.519
	81	000		73	.027		64	.481
				74	.020		65	.443
	M=7			75	.014		66	.406
7	53	.500		76	.010		67	.370
	54	.451		77	.007		68	.335
	55	.402		78	.005		69	.300
	56	.355		79	.003		70	.268
	57	.310		80	.002		71	.237
	58	.267		81	.001		72	.209
	59	.228		82	.001		73	.182
	60	.191		83	.000		74	.157
	61	.159		84	.000		75	.135
	62	.130		60	.500		76	.115
	63	.104		61	.459		77	.097
	64	.082		62	.419		78	.081
	65	.064		63	.379		79	.067
	66	.049		64	.340		80	.054
	67	.036		65	.303		81	.044
	68	.027		66	.268		82	.035
	69	.019		67	.235		83	.028
	70	.013		68	.204		84	.022
	71	.009		69	.176		85	.017
	72	.006		70	.150		86	.012
	73	.003		71	.126		87	.009
	74	.002		72	.105		88	.007
	75	.001		73	.087		89	.005
	76	.001		74	.071		90	.003
	77	.000		75	.057		91	.002
8	56	.522		76	.045		92	.002
	M=7			M=8			M=8	
	93	.001		76	.371		88	.158
	94	.001		77	.336		89	.137
	95	.000		78	.030		90	.118

	96	.000		79	.271		91	.102
	97	.000		80	.240		92	.086
	98	.000		81	.212		93	.073
				82	.185		94	.061
				83	.161		95	.051
8	68	.520		84	.138		96	.042
	69	.480		85	.118		97	.034
	70	.439		86	.100		98	.027
	71	.399		87	.084		99	.022
	72	.307		88	.069		100	.017
	73	.323		89	.057		101	.013
	74	.287		90	.046		102	.010
	75	.253		91	.037		103	.008
	76	.221		92	.030		104	.006
	77	.191		93	.023		105	.004
	78	.164		94	.018		106	.003
	79	.139		95	.014		107	.002
	80	.117		96	.010		108	.002
	81	.097		97	.008		109	.001
	82	.080		98	.006		110	.001
	83	.065		99	.004		111	.000
	84	.052		100	.003		112	.000
	85	.041		101	.002		113	.000
	86	.032		102	.001		114	.000
	87	.025		103	.001		115	.000
	88	.019		104	.000		116	.000
	89	.014		105	.000			
	90	.010		106	.000		M=9	
	91	.007		107	.000	9	86	.500
	92	.005		108	.000		87	.466
	93	.003	10	76	.517		88	.432
	94	.002		77	.483		89	.398
	95	.001		78	.448		90	.365
	96	.001		79	.414		91	.333
	97	.001		80	.381		92	.302
	98	.000		81	.348		93	.273

	99	.000	82	.317	94	.245
	100	.000	83	.286	95	.218
9	72	.519	84	.257	96	.193
	73	.481	85	.230	97	.170
	74	.444	86	.204	98	.149
	75	.407	87	.180	99	.129
	M=9		M=9		M=10	
	100	.111	105	.121	114	.264
	101	.095	106	.106	115	.241
	102	.081	107	.091	116	.218
	103	.068	108	.078	117	.197
	104	.057	109	.067	118	.176
	105	.047	110	.056	119	.157
	106	.039	111	.047	120	.140
	107	.031	112	.039	121	.124
	108	.025	113	.033	122	.109
	109	.020	114	.027	123	.095
	110	.016	115	.022	124	.083
	111	.012	116	.017	125	.072
	112	.009	117	.014	126	.062
	113	.007	118	.011	127	.053
	114	.005	119	.009	128	.045
	115	.004	120	.007	129	.038
	116	.003	121	.005	130	.032
	117	.002	122	.004	131	.026
	118	.001	123	.003	132	.022
	119	.001	124	.002	133	.018
	120	.001	125	.001	134	.014
	121	.000	126	.001	135	.012
	122	.000	127	.001	136	.009
	123	.000	128	.000	137	.007
	124	.000	129	.000	138	.006
	125	.000	130	.000	139	.004
	126	.000	131	.000	140	.003
10	90	.516	132	.000	141	.003
	91	.484	133	.000	142	.002

92	.452	134	.000	143	.001
93	.421	135	.000	144	.001
94	.390			145	.001
95	.360	M=10		146	.001
96	.330	105	.515	147	.000
97	.302	106	.485	148	.000
98	.274	107	.456	149	.000
99	.248	108	.427	150	.000
100	.223	109	.398	151	.000
101	.200	110	.370	152	.000
102	.178	111	.342	153	.000
103	.158	112	.315	154	.000
104	.139	113	.279	155	.000

Table 3.2: Right Tail probabilities for Wilcoxon Signed Rank Statistic.

Right tail probabilities $P(W \geq W_o/H_o)$ are given for $W_o \geq n(n+1)/4$. Here W is interpreted as either as either W+ or W-. Left tail probabilities are obtained from the relation $P(W \leq W_o/H_o) = P(W_o \geq n(n+1)/2 - W_o/H_o)$

n	W_o	$(W \geq W_o/H_o)$	n	W_o	$(W \geq W_o/H_o)$	n	W_o	$(W \geq W_o/H_o)$
2	2	.500		26	.023		44	.004
	3	.250		27	.016		45	.002
3	3	.625		28	.008	10	28	.500
	4	.375	8	18	.527		29	.461
	5	.250		19	.473		30	.423
	6	.125		20	.422		31	.385
4	5	.562		21	.371		32	.348
	6	.438		22	.320		33	.312
	7	.312		23	.273		34	.278
	8	.188		24	.230		35	.246
	9	.125		25	.191		36	.216
	10	.062		26	.156		37	.188
5	8	.500		27	.125		38	.161
	9	.406		28	.098		39	.138
	10	.312		29	.074		40	.116
	11	.219		30	.055		41	.097
	12	.156		31	.039		42	.080

	13	.094
	14	.062
	15	.031
6	11	.500
	12	.422
	13	.344
	14	.281
	15	.219
	16	.156
	17	.109
	18	.078
	19	.047
	20	.031
	21	.016
7	14	.531
	15	.469
	16	.406
	17	.344
	18	.289
	19	.234
	20	.188
	21	.148
	22	.109
	23	.078
	24	.055
	25	.039
44	46	.139
	47	.120
	48	.103
	49	.087
	50	.074
	51	.062
	52	.051
	53	.042
	54	.034
	55	.027

	32	.027
	33	.020
	34	.012
	35	.008
	36	.004
9	23	.500
	24	.455
	25	.410
	26	.367
	27	.326
	28	.285
	29	.248
	30	.213
	31	.180
	32	.150
	33	.125
	34	.102
	35	.082
	36	.064
	37	.049
	38	.037
	39	.027
	40	.020
	41	.014
	42	.010
	43	.006
12	64	.026
	65	.021
	66	.017
	67	.013
	68	.010
	69	.008
	70	.006
	71	.005
	72	.003
	73	.002

	43	.065
	44	.053
	45	.042
	46	.032
	47	.024
	48	.019
	49	.014
	50	.010
	51	.007
	52	.005
	53	.003
	54	.002
	55	.001
11	33	.517
	34	.483
	35	.449
	36	.416
	37	.382
	38	.350
	39	.319
	40	.289
	41	.260
	42	.232
	43	.207
	44	.183
	45	.160
13	77	.013
	78	.011
	79	.009
	80	.007
	81	.005
	82	.004
	83	.003
	84	.002
	85	.002
	86	.001

	56	.021		74	.002		87	.001
	57	.016		75	.001		88	.001
	58	.012		76	.001		89	.000
	59	.009		77	.000		90	.000
	60	.007		78	.000		91	.000
	61	.005	13	46	.500	14	53	.500
	62	.003		47	.473		54	.476
	63	.002		48	.446		55	.452
	64	.001		49	.420		56	.428
	65	.001		50	.393		57	.404
	66	.000		51	.368		58	.380
12	39	.515		52	.342		59	.357
	40	.485		53	.318		60	.335
	41	.455		54	.294		61	.313
	42	.425		55	.271		62	.292
	43	.396		56	.249		63	.271
	44	.367		57	.227		64	.251
	45	.339		58	.207		65	.232
	46	.311		59	.188		66	.213
	47	.285		60	.170		67	.196
	48	.259		61	.153		68	.179
	49	.235		62	.137		69	.163
	50	.212		63	.122		70	.148
	51	.190		64	.108		71	.134
	52	.170		65	.095		72	.121
	53	.151		66	.084		73	.108
	54	.133		67	.073		74	.097
	55	.117		68	.064		75	.086
	56	.102		69	.055		76	.077
	57	.088		70	.047		77	.068
	58	.076		71	.040		78	.059
	59	.065		72	.034		79	.052
	60	.055		73	.029		80	.045
	61	.046		74	.024		81	.039
	62	.039		75	.020		82	.034
	63	.032		76	.016		83	.029

14	84	.025	15	66	.381	15	94	.028
	85	.021		67	.360		95	.024
	86	.018		68	.339		96	.021
	87	.015		69	.319		97	.018
	88	.012		70	.300		98	.015
	89	.010		71	.281		99	.013
	90	.008		72	.262		100	.011
	91	.007		73	.244		101	.009
	92	.005		74	.227		102	.008
	93	.004		75	.211		103	.006
	94	.003		76	.195		104	.005
	95	.003		77	.180		105	.004
	96	.002		78	.165		106	.003
	97	.002		79	.151		107	.003
	98	.001		80	.138		108	.002
	99	.001		81	.126		109	.002
	100	.011		82	.115		110	.001
	101	.000		83	.104		111	.001
	102	.000		84	.094		112	.001
	103	.000		85	.084		113	.001
	104	.000		86	.076		114	.000
	105	.000		87	.068		115	.000
15	60	.511		88	.060		116	.000
	61	.489		89	.053		117	.000
	62	.467		90	.047		118	.000
	63	.445		91	.042		119	.000
	64	.423		92	.036		120	.000
	65	.402		93	.032			

Table 3.3 Left Probabilities, R ($R \leq r_o$ Ж H_o), For Total number of runs

	r_o																	
(m, n)	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**
(2,3)	.200	.500	.900	1.000														
(2,4)	.133	.400	.800	1.000														
(2,5)	.095	.333	.714	1.000														
(2,6)	.071	.286	.643	1.000														
(2,7)	.056	.250	.583	1.000														
(2,8)	044	.222	.533	1.000														
(2,9)	.036	.200	.491	1.000														
(2,10)	.030	.182	.455	1.000														
(3,3)	.100	.300	.700	.900	1.000													
(3,4)	.057	.200	.543	.800	.971	1.000												
(3,5)	.036	.143	.429	.714	.929	1.000												
(3,6)	.024	.107	.345	.643	.881	1.000												
(3,7)	.017	.083	.283	.583	.833	1.000												
(3,8)	.012	.067	.236	.533	.788	1.000												
(3,9)	.009	.055	.200	.491	.745	1.000												
(3,10)	.007	.045	.171	.455	.706	1.000												
(4,4)	.029	.114	.371	.629	.886	.971	1.000											
(4,5)	.016	.071	.262	.500	.786	.929	.992	1.000										
(4,6)	.010	.048	.190	.405	.690	.881	.976	1.000										
(4,7)	.006	.033	.142	.333	.606	.833	.954	1.000										
(4,8)	.004	.024	.109	.279	.533	.788	.929	1.000										
(4,9)	.003	.018	.085	.236	.471	.745	.902	1.000										
(4,10)	.002	.014	.068	.203	.419	.706	.874	1.000										
(5,5)	.008	.040	.167	.357	.643	.833	.960	.992	1.000									
(5,6)	.004	.024	.110	.262	.522	.738	.911	.976	.998	1.000								
(5,7)	.003	.015	.076	.197	.424	.652	.854	.955	.992	1.000								
(5,8)	.002	.010	.054	.152	.347	.576	.793	.929	.984	1.000								

(5,9)	.001	.007	.039	.119	.287	.510	.734	.902	.972	1.000								
(5,10)	.001	.005	.029	.095	.239	.455	.678	.874	.958	1.000								
(6,6)	.002	.013	.067	.175	.392	.608	.825	.933	.987	.998	1.000							
(6,7)	.001	.008	.043	.121	.296	.500	.733	.879	.966	.992	.999	1.000						
(6,8)	.001	.005	.028	.086	.226	.413	.646	.821	.937	.984	.998	1.000						
(6,9)	.000	.003	.019	.063	.175	.343	.566	.762	.902	.972	.994	1.000						
(6,10)	.000	.002	.013	.047	.137	.288	.497	.706	.864	.958	.990	1.000						
(7,7)	.001	.004	.025	.078	.209	.383	.617	.791	.922	.975	.996	.999	1.000					
(7,8)	.000	.002	.015	.051	.149	.296	.514	.704	.867	.949	.988	.998	1.000					
(7,9)	.000	.001	.010	.035	.108	.231	.427	.622	.806	.916	.975	.994	.999	1.000				
(7,10)	.000	.001	.006	.024	.080	.182	.355	.549	.743	.879	.957	.990	.998	1.000				
(8,8)	.000	.001	.009	.032	.100	.214	.405	.595	.786	.900	.968	.991	.999	1.000				
(8,9)	.000	.001	.005	.020	.069	.157	.319	.500	.702	.843	.939	.980	.996	.999	1.000			
(8,10)	.000	.000	.003	.013	.048	.117	.251	.419	.621	.782	.903	.964	.990	.998	1.000			
(9,9)	.000	.000	.003	.012	.044	.109	.238	.399	.601	.762	.891	.956	.988	.997	1.000			
(9,10)	.000	.000	.002	.008	.029	.077	.179	.319	.510	.681	.834	.923	.974	.992	.999	1.000		
(10,10)	.000	.000	.001	.004	.019	.051	.128	.242	.410	.586	.798	.872	.949	.981	.996	.999	1.000	

16.11 Further Readings

1. Fundamentals of Statistics volume I by A.N. Goon, B.D. Gupta and Dasgupta. Pub; Calcutta Publishing House, Kolkata
2. Introduction to Mathematical Statistics by Mood, Graybill and Boes. Pub: Mac. Graw Hill
3. Introduction to Mathematical statistics by Hogg and Craig.