Statistics

DMTH404

LOVELY
PROFESSIONAL
UNIVERSITY

STATISTICS

Produced & Printed by
EXCEL BOOKS PRIVATE LIMITED
A-45, Naraina, Phase-I,
New Delhi-110028
for
Lovely Professional University
Phagwara

SYLLABUS

Statistics

Objectives:

- To understand the value of Statistics in acquiring knowledge and making decisions in today's society.
- To learn about the basic theory of Probability, random variable, moments generating function, Probability distribution, reliability theory, laws of large numbers, correlation and regression, sampling theory, theory of estimation and testing of hypotheses.

Sr. No.	Content
1	The sample space, Events, Basic notions of probability, Methods of enumeration of Probability, conditional probability and independence, Baye's theorem
2	General notion of a variable, Discrete random variables, Continuous random variables, Functions of random Variables, Two dimensional random variables, Marginal and conditional probability distributions, Independent random variables, Distribution of product and quotient of independent random variables, n-dimensional random variables
3	Expected value of a random variable, Expectation of a function of a random variable, Properties of expected value, Variance of a random variable and their properties, Approximate expressions for expectations and variance, Chebyshev inequality
4	The Moment Generating Function: Examples of moment generating functions, Properties of moment generating function, Reproductive properties, Discrete Distributions : Binomial, Poison, Geometric, Pascal Distributions, Continuous Distributions :Uniform, Normal, Exponential
5	Basic concepts, The normal failure law, The exponential failure law, Weibul failure law, Reliability of systems
6	Weak Law of Large Numbers, Strong Law of Large Number, Central Limit Theorem, Confidence Intervals
7	The correlation coefficient, Conditional expectation, Regression of the mean
8	Samples, Sample Statistics, Sampling Distribution of Sample Mean and Sample Variance, t-distribution , Chi Square distribution, F- distribution
9	Estimation of Parameters: Criteria for estimates, Maximum likelihood estimates, Method of least squares
10	t-test, chi square Godness of fit, Z-test with examples

CONTENTS

Unit 1:	Sample Space of A Random Experiment	1
Unit 2:	Methods of Enumeration of Probability	13
Unit 3:	Conditional Probability and Independence Baye's Theorem	27
Unit 4:	Random Variable	48
Unit 5:	Functions of Random Variables	54
Unit 6:	Probability	69
Unit 7:	Modern Approach to Probability	85
Unit 8:	Expected Value with Perfect Information (EVPI)	117
Unit 9:	Variance of a Random Variable and their Properties	127
Unit 10:	Approximate Expressions for Expectations and Variance	132
Unit 11:	Chebyshev's Inequality	150
Unit 12:	The Moment Generating Functions	158
Unit 13:	Moment Generating Function of a Random Vector	168
Unit 14:	Theoretical Probability Distributions	176
Unit 15:	Exponential Distribution and Normal Distribution	200
Unit 16:	Reliability Theory	223
Unit 17:	System Reliability	234
Unit 18:	The Weak Law	254
Unit 19:	The Laws of Large Numbers Compared	262
Unit 20:	Control Limit Theorem	267
Unit 21:	Confidence Intervals	273
Unit 22:	Correlation	287
Unit 23:	Regression Analysis	314
Unit 24:	Sampling Distributions	341
Unit 25:	Chi - Sqaure (χ^2) Distribution	352
Unit 26:	χ^2 - Test Hypothesis	359
Unit 27:	T - Distributions	372
Unit 28:	F-distribution	378
Unit 29:	Estimation of Parameters: Criteria for Estimates	383
Unit 30:	Method of Least Square	399
Unit 31:	Hypothesis Testing	413
Unit 32:	Hypothesis Concerning Standard Deviation	433

Unit 1: Sample Space of A Random Experiment

Notes

CONTENTS

Objectives

Introduction

1.1 Random Experiments

1.2 Sample Space

1.3 Events

1.4 Algebra of Events

1.5 Summary

1.6 Keywords

1.7 Self Assessment

1.8 Review Questions

1.9 Further Readings

Objectives

After studying this unit, you will be able to:

- Discuss random and non-random experiments,
- Explain the sample space of a random experiment and classify it as discrete or continuous,
- Describe events with subsets ofthe sample space,
- Discuss and identify relations between events,

Introduction

Many situations arise in our everyday life as well as in scientific, administrative or organisational work, where we cannot predict'the outcome of our actions or of the experiment we are conducting. Such experiments, whose outcome cannot be predicted, are called random experiments. We give a wide variety of examples in Sec. 5.2 to explain the concept of a random experiment. The set of all possible outcomes of an experiment is called its sample space. We have illustrated the different types of sample spaces that we generally come across in Sec. 5.3. Section 5.4 deals with the study of events associated with a random experiment whose sample space is either finite or countably infinite. In Sec. 5.5 we discuss methods of combining events to generate new events. Here is a list of what you should be able to do by the end of this unit.

Notes

1.1 Random Experiments

We give below some examples of a random experiment :

- A physicist performs an experiment to discover laws governing the flow of an electrical current or the propagation of sound, heat or light etc.
- A chemist studies the reactions of chemicals and tries to understand the chemical properties of matter.
- A physician compares two or more drugs to find out the most effective one by trying them out on experimental animals or on patients.
- To describe the relationship between the price of a commodity and its demand and supply, an economist observes the values assumed by these variables by conducting a market survey over a period of time.

With a little imagination, we can construct many more examples of such experiments.

Experimentation is not necessarily restricted to a laboratory or to a university or a college. It forms an important part of our everyday life. When you buy a dress or a shirt, when you vote for a candidate at an electron, which you inspect a few grains of rice to decide whether the rice is cooked or not, when you decide to register for this course, you are performing an experiment. Thus, experimentation constitutes an integral part of our lives as well as our learning processes. In this unit we shall develop methods of describing the results of an experiment. Once we can describe the results we'll be able io talk about the chances of their occurrence.

Consider the following simple experiments :

Experiment 1 : A stone is allowed to fall freely from height and we observe whether or not the stone hits the ground.

Experiment 2 : Water in a pot is heated for a sufficiently long time to a temperature greater than 100°C. We observe whether the water turns into steam.

In these experiments, we have no doubt about the final outcome. The stone will eventually hit the ground. The water in the pot will ultimately turn into steam. These experiments have only one possible outcome. Even if these experiments are repeated again and again, every such repetition will yield the same result.

On the other hand, in the following experiments there are two or more possible results.

Experiment 3 : A coin is tossed to deoide which of the two teams A and B would bat first in a game of cricket. The coin may tum up a head or a tail.

Experiment 4 : A person coming out of a polling centre is requested to disclose the name of the candidate in whose favour he/she has voted. Helshe may refuse to tell us or give the name of any one of the candidate.

Experiment 5 : Three consecutive items produced by a machine are inspected and classified as good or bad (defective). We may get 0, 1,2, or 3 defective items as a result of this inspection.

Experiment 6 : A newly invented vaccine against a disease is given to 30 healthy people. These thirty people as well as another group of 20 similar people who are not vaccinated, are watched over the next six months to see whether they develop the disease. The total number of affected people may vary between 0 and 50.

Experiment 7 : A small town has 100 telephones. The number of busy telephones between 9 and 10 a.m. is noted for each day of a week. The number of busy telephones may be any number between 0 to 100.

Experiment 8 : A group of ten persons is classified according to their blood groups 0, A, B and AB. The number of persons in each group may vary between 0 and 10, subject to the frequencies of all four classes adding up to 10.

Experiment 9 : The number of accidents along the Bombay-Bangalore national highway during the month is noted.

Experiment 10 : A radio-active substance emits particles called a-particles. The number of such particles reaching an observation screen during one hour is noted.

Experiment 11 : Thirteen cards are selected without replacement from a well-shuffled pack of 52 playing cards.

The nine experiments, 3-11, have two common features.

(i) Each of these experiments hve more than one possibie outcome.

(ii) It is impossible to predict the outcome of the experiment.

For example, we cannot predict whether a coin, when it is tossed, will turn up a head or a tail (Experiment 3). Can we predict without error the number of busy telephones (Experiment 7)? It is impossible to predict the 13 cards we shall obtain from a well-shuffled pack (Experiment 11).

Do you agree that all the experiments 3-1 1 have the above-mentioned features (i) and (ii)? Go through them carefully again, and convince yourself.

This discussion leads us to the following definition.

Definition 1 : An experiment with more than one possible outcome and whose result cannot be predicted, is called a random experiment. Experiment

So, Experiments 3 to 11 are random experiments, while in Experiments 1 and 2 the outcome of the experiment can be predicted. Therefore, Experiments 1 and 2 do not qualify as random experiments. You will meet many more illustrations of random experiments in this and subsequent units.

Note In the dictionary you will find that something that is random, happens or is chosen without a definite plan. patteron or purpose.

1.2 Sample Space

In the previous section you have seen a number of examples of random experiments. The first step we take in the study of such experiments is to specify the set of all possible outcomes of the experiment under consideration.

When a coin is tossed (Experiment 3), either a head turns up or a tail turns up. We do not consider the possibility of the coin standing on its edge or that of its rolling away out of I sight. Thus, the set SZ of all possible outcomes consists of two elemends, Head and Tail. Therefore, we write SZ = (Head, Tail) or, more simply, SZ = (H, T).

Note Ω is the Greek letter capital 'omega'

Notes

In Experiment 4, the person coming out of the polling centre may give us the name of the candidate for whom helshe voted, or may refuse to disclose hisher choice. If there are 5 candidates C_1, C_2, C_3. C_4 and C_5, seeking election, then there are six possible outcomes, five corresponding to the five candidates and the sixth one corresponding to the refusal R of the interviewed person to disclose hisher choice. The set of all possible outcomes is thus, $\{C_1, C_2, C_3{*}\ C_4, C_5, R\}$.

Note that here we have ignored certain possibilities, like the possibility of the person not voting at all or voting in such a manner that hisher ballot paper becomes invalid.

Experiment 5 is comparatively simple, if we agree that it is possible to classify each item as Good (G) or Bad (B) without error. Then R = {GGG, GGB, GBG, BGG, BBG, BGB, GBB, BBB} where. for example, GBG denotes the outcome when the first and third units are good and the second one is bad.

The situation in Experiment 6 is a little more complicated. To test the efficacy of the vaccine, we will have to look at the number of vaccinated persons who were affected (x) q the number of non-vaccinated ones who were affected (y). Here x can be any integer between 0 and 30 and y can be any integer between 0 and 20. The set Ω of all possible outcomes is

$$\Omega = \{(x,y) \mid x = 0, 1, ..., 30, y = 0, l, 2,..., 20\}.$$

This specification of Ω is valid only if we assume that we are able to observe all the 50 persons for the entire period of six months. In particular, we assume that none of them becomes untraceable because of hisher leaving the town or because of hisher death due to some other cause.

In the illustrations discussed so far, do you notice that the number of points in Ω is finite in each case? It is 2 for Experiment 3,6 for ExIjeriment 4,31 x 21 = 651 for Experiment 6. But this is not always true.

Consider, for example, Experiments 9 and 10. The number of accidents along the Bombay-Bangalore highway during the month of observation can be zero, one, two, . . . or some other positive integer. Similarly, the number of a-particles emitted by the radio-active substance can be any positive integer. Can we say that the number of accidents or a-particles would not exceed a specified limit? No. Because of this, and also in order to simplify our mathematics, we usually postulate that in both these examples the set of all possible outcomes is R = {0, 1,2, ...}, i.e., it is the set of all non-negative integers.

We are now in a position to introduce certain terms in a formal manner.

Definition 2 : The set Ω of all possible outcomes of an experiment E is called the sample space of the experiment. Each individual outcome of E is called a point, a sample point or an element of Ω.

You would also notice that in every experiment th,at was discussed, we made certain assumptions like the coin not being able to stand on its edge or not rolling away, all the fifty persons being available for the entire period of six months for observation, etc. Such assumptions are necessary to simplify our problems as well as our mathematics.

In all the examples discussed so far, the sample space is either a finite set, i.e., a set containing a finite number of points or is an infinite set whose elements can be arranged in an unending sequence, i.e., has a countable infinity of elements. We have a special name for such spaces.

Definition 3 : A sample space containing a finite number of points or a countable infinity of points is called a discrete sample space.

In this block we shall be concerned only with discrete sample spaces. However, there are mhy situations where we have to deal with sample spaces which are not discrete. For example, consider the age of a person. Although there are limitations to the accuracy with which we can measure the age of a person, in the idealised situation we can think of age being any number between 0 and ∞. Of course, no one has met a person. with infinite age of for that matter who is more than 150 years old. Nevertheless, most of the actuarial and demographic studies are carried out assuming that there is no upper bound on age. Thus, we may say that the sample space of the experiment of finding out the age of an arbitrarily selected person is the interval]0, ∞[. Since the elements of the interval]10, ∞ [cannot be arranged in a sequence, such a sample space is not a discrete sample space.

Some other examples where non-discrete sample spaces are appropriate are (i) the price of wheat, (ii) the amount of ozone in a volume of space, (iii) the length of a telephone conversation, (iv) the duration one spends in a queue, (v) the yield of rice in our country in one year.

In all these examples, it is necessary to deal with non-discrete sample spaces. However, we'll defer the study of probability theory for such experiments to the next block.

1.3 Events

We have described a number of random experiments till now. We have also identified the sample spaces associated with them. In the study of random experiments, we are interested not only in the individual outcomes but also in certain events. As you will see later, events are subsets of the sample space. In this section we shall formalise the intuitive concept of an event associated with a random experiment which has a discrete sample space. We shall also study methods of generating new events from specified ones and study their inter-relationships.

Consider the experiment of inspecting three items (Experiment 5). The sample space has the eight points,

GGG, GGB, GBG, BGG, BBG, BGB, GBB, BBB.

We label these points $\omega_1, \omega_2, \ldots, \omega_8$, respectively.

Suppose we are interested in those outcomes which correspond to the event of obtaining exactly one good item in the three inspected items. The corresponding'sample points are ω_5 = BBG, ω_6 = BGB and ω_7 = GBB. Thus, the subset $\{ \omega_5, \omega_6, \omega_7 \}$ of the sample space corresponds to the "event" A that only one of the inspected items is good.

On the other hand, consider the subset C = $\{ \omega_5, \omega_6, \omega_7, \omega_8 \}$ consisting of the points BBG, BGB, GBB, BBB. We can identify the subset C with the event "There are at least two bad items."

This discussion suggests that we can associate a subset of the sample space with an event and an event with a subset. This leads us to the following definition.

Definition 4 : When the sample space of an experiment is discrete, any subset of the sample space is called an event.

Thus, we alsq consider the empty set as an event.

You will soon find that the two extreme events, φ and ω, consisting, respectively, of no points and all the points of **R** are most uninteresting. But we need them to complete our description of the class of all events. In fact, φ is called the impossible event and Ω is called the sure event, for reasons which will be obvious in the next unit. Also, note that an individual outcome ω, when identified with the singleton (ω), constitutes an event.

The following example will help you in understanding events.

Notes

Examples: Suppose we toss a coin twice. The sample space of this experiment is Ω = (HH, HT, TH, TT}, where HT stands for a head followed by a tail, and other points are similarly defined. Let's list all the events associated with this experiment. There are 16 such events. These are :

φ,{HH}, {HT}, {TH}, {TT}

{HH, HT}, {HH, TH}, {HH, TT}, {HT, TH}

{HH, TT}, {TH, TT}, {HH, HT, TH}, {HH, TH, TT},

{HH, HT, TT}, {HT, TH, TT}, Ω

Since we have identified an event with a subset of Ω, the class of all events is the class of all the subsets of Ω. If Ω has N points, for a fixed r, we can form $\binom{N}{r}$ sets consisting of r points, where r = 0, 1, ... , N. The total number of events is, thirefore,

$$\binom{N}{0} + \binom{N}{1} + \ldots + \binom{N}{N} = (1+1)^N = 2^N.$$

Notes By binomial theorem

$$(1+x)^N = \binom{N}{0} + \binom{N}{1}x + \ldots + \binom{N}{N}x^N.$$

In Example 1, N = 4. Therefore, we have 2^4 = 16 events. If N = 10, we shall 2^{10} = 1024 events. The number of events thus increases rapidly with N. It is infinite if the sample space is infinite.

Let us now clarify the meaning of the phrase "The event A has occurred."

We continue with Experiment 5. Let A denote the event $\{\omega_5, \omega_5, \omega_7)$ = {BBG, BGB, GBB}. If, after performing the experiment, our outcome is ω_5 = BBG, which is a point of the set A, we say that the event A has occurred. If, on the other hand, the outcome is ω_8 = BBB, which is not a point of A, then we say that A has not occurred. In other words, given the outcome ω of the experiment, we say that A has occurred if $\omega \in$ A and that A has not occurred if $\omega \notin$ A.

On the other hand, if we only know that A has occured, all we know is that the outcome of the experiment is one of the points of A. It A then not possible to decide which individual outcome has resulted unless A is a singleton.

In the next section we shall talk about some ways of combining events.

1.4 Algebra of Events

In this section we shall study different ways in which we can combine two or more events. We shall also study relations ktween them. Since we are dealing with discrete sample spaces and since any subset of the sample space is an event, we shall use the terms event and subset interchangeable.

In what follows, events and sets are denoted by capital letters A, B, C, ... , with or without suffixes. We shall assume that they all consist of points chosen from the same sample space Ω.

Let Ω = { GGG, GGB, GBG, BGG, BBG, BGB, GBB, BBB } be the sample space correspondintog Experiment 5. Let A = {BBG, BGB, GBB} be the event that only one of the three inspected items is good. Here the point BGB is an element of the set A and the point BBB is not an element of A. We express this by writing BGB ∈ A and BBB ∉ A.

> *Notes* $A^c = \{w \in Q\} \mid w \notin A\}$. Then $\phi^c = \Omega$ and $\Omega^c = \phi$. Fig. 1 shows a Venn diagram representing the sets A and A^c.

Suppose, now, that the outcome of the experiment is BBB. Obviously, the event A has not occurred. But, we may say the event "not A" hasaoccurred. In probability theory, the event "not A" is called the event complementary to A and is denoted by A^c.

Let's try to understand this concept by looking back at Experiments 3-11.

Examples:

(i) For Experiment 5, if A = {BBG, BGB, GBB} , then

A^C = {GGG, GGB, BGG, GBG, BBB}.

(ii) In Experiment 6, let A denote the event that the number of infected persons is at most 40. Then

$A^c = \{(x, y) \mid x + y > 40, x = 0, 1, ..., 30, y = 0, 1, ..., 20\}$.

(iii) In Experiment 11, if B denotes the event that none of the 13 cards is a spade, B^c consists of all hands of 13 cards, each one of which has at least one spade.

Suppose now that A_1 and A_2 are two events associated with an experiment. We can get two new events, $A_1 \cap A_2$ (A_1 intersection A_2) and $A_1 \cup A_2$ (A_1 union A_2) from these two. With your knowledge of set theofy (MTE-04). you would expect the event $A_1 \cap A_2$ to correspond to the set whose elements belong to both A_1 and A_2. Thus,

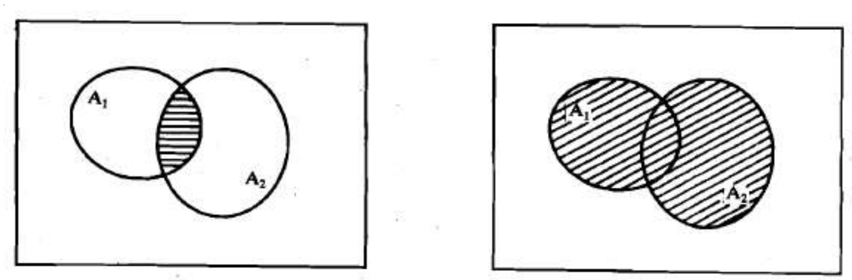
$$A_1 \cap A_2 = \{ \omega \mid \omega \in A_1 \text{ and } \omega \in A_2\}.$$

Similarly, the event $A_1 \cup A_2$ corresponds to the set whose elements belong to at least one of A_1 and A_2.

$$A_1 \cup A_2 = \{ \omega \mid \omega \in A_1 \text{ or } \omega \in A_2\}.$$

Fig. 2 (a) and (b) show the Venn diagrams representing $A_1 \cap A_2$ and $A_1 \cup A_2$.

Figure 1.1: The Shaded Region Represents the set $A_1 \cap A_2$ and $A_1 \cup A_2$.

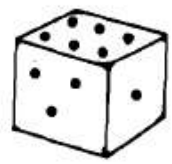


We'll try to clarify this concept with some examples.

Notes

Examples 3: In many games of chance, a small'cube (or die) with equal sides, bearing numbers 1, 2, 3, 4, 5, 6, or dots 1-6 on its six faces (Fig. 1.2), is used. When such a symmetric , die is thrown, one of its six faces would be uppermost. The number (or number of dots) on the uppermost faces is called the score obtained on the throw or roll of a die. The appropriate sample space for the experiment of throwing a die is then R = {1, 2, 3, 4, 5, 6}.

Figure 1.2



Let A_1 be the event that the score exceeds three and A_2 be the event that the score is even.

Then

$A_1 = \{4, 5, 6\}$, $A_2 = \{2, 4, 6\}$

Therefore, $A_1 \cap A_2 = \{4, 6\}$ and

$A_1 \cup A_2 = \{2, 4, 5, 6\}$.

Suppose now that the score is 6. We can say that A_1 has occurred. But then A_2 has also occurred. In other words, both A_1 and A_2 have occurred. Thus, the simultaneous occurrence of A_1 and A_2 corresponds to the occurrence of the event $A_1 \cap A_2$.

When the outcome is 5, A_1 has occurred but A_2 has not occurred. Further, when the outcome is 2, A_2 has occurred and A_1 has not. When the outcome is 4, both A_1 and A_2 have occurred. In case of each of these outcomes, 2,5 or 4, we notice that at least one of A_1 and A_2 has occurred. Note, further, that $A_1 \cup A_2$ has also occurred. Thus, the occurrence of at least one of the two events A_1 and A_2 corresponds to the occurrence of $A_1 \cup A_2$.

Examples 4: Suppose the die in Example 3 is thrown twice. Then Ω is the set $\{ (x, y) | x, y = 1, 2, 3, \ldots, 6 \}$ consisting of thirty-six points (x, y), where x is the score obtained on the first throw and y, that obtained on the second throw. If B_1 is the event that the score on the first throw is six and B_2 the event that the sum of the two scores is at least 11, then

$B_1 = \{ (6, 1), (6, 2), (6, 3), (6, 4), (6, 5), (6, 6) \}$

and

$B_2 = \{ (5, 6), (6, 5), (6, 6) \}$.

What are $B_1 \cap B_2$ and $B_1 \cup B_2$? You can check that

$B_1 \cap B_2 = \{ (6, 5), (6, 6) \}$

and

$B_1 \cup B_2 = \{ (5, 6), (6, 1), (6, 2), (6, 3), (6, 4), (6, 5), (6, 6) \}$.

Notes

The union and intersection of two sets can be utilised to define union and intersection of three or more sets.

So, if $A_1, A_2, \ldots, A_n$ are n events, then we define

$$\bigcap_{j=1}^{n} A_j = \{ \omega \mid \omega \in A_j \text{ for every } j = l, \ldots, n \}.$$

and

$$\bigcap_{j=1}^{n} A_j = \{ \omega \mid \omega \in A_j \text{ for at least one } j = 1, \ldots, n \}.$$

Note that the occurrence of $\bigcap_{j=1}^{n} A_j$ corresponds to the simultaneous occurrence of all the n events and the occurrence of $\bigcap_{j=1}^{n} A_j$ corresponds to that of at least one of the n events $A_1, \ldots, A_n$. We can similarly define the union and intersection of an infinite number of events, $A_1, A_2, \ldots, A_n, \ldots$

Another set operation with which you are familiar is a combination of complementation and intersection. Let A and B be two sets. Then the set $A \cap B^C$ is usually called the difference of A and B and is denotedby A – B. It consists of all points which belong to A but not to B.

Thus, in Example 4,

$$B_1 - B_2 = \{ (6, 1), (6, 2), (6, 3), (6, 4) \}$$

and

$B_2 - B_1 = \{ (5,6) \}$

In this notation, A^C is the set Ω – A. You can see the Venn diagram for A – B in Fig. 4.

Now, suppose A_1, A_2 and A_3 are three arbitmy events. What does the occurrence of $A_1 \cap A_2^c \cap A_3^c$ signify?

This event occurs iff only A_1 out of A_1, A_2 and A_3 occurs, that is, iff Al occurs but neither A_2 nor A_3 occur.

If you have followed this, you should be able to do this exercise quite easily.

Task If A_1, A_2 and A_3 are three arbitrary events, what does the occurrence of the following events signify?

(a) $E_1 = A_1 \cap A_2 \cap A_3$

(b) $E_2 = A_1^c \cap A_2^c \cap A_3^c$

(c) $E_3 = (A_1 \cap A_2 \cap A_3^c) \cup (A_1 \cap A_3 \cap A_2^c) \cup (A_2 \cap A_3 \cap A_1^c)$

(d) $E_1 \cup E_3$

Notes

The set operations like formation of intersection, union and complementation of two or more sets that we have listed above and their combinations are sufficient for constructing new events out of old ones. However, we need to express in a precise way commonly used expressions like (i) if the event A has occurred, B could not have occurred and (ii) the occurrence of A implies that of B. We'll explain this by taking an example first.

Examples: Let us consider the following experiments.

(i) In the experiment of tossing a die twice, let A be the event that the total score is 8 and B that the absolute difference of the two scores is 3. Then

$A = | (x, y) \mid x + y = 8, x, y = 1, 2, 3, ..., 6 \}$

$= \{ (2, 6), (3, 5), (4, 4), (5, 3), (6, 2) \}$

and $B = \{ (x,y) \mid | x - y | = 3, x, y = 1, 2, 3,..., 6 \}$

$= \{ (1, 4) \mid | x - y | = 3, x, y = 1, 2, 3, ..., 6 \}$

(ii) Consider Experiment 11, where we select 13 cards without replacement from a pack of cards. Let

event A : all the 13 cards are black and

event B : there are 6 diamonds and 7 hearts.

Note that in both the-cases there is no point which is common to both A and B. Or in other words, A n B is the empty set. Therefore, in both i) and ii) we conclude that if A occurs, B cannot occur and conversely, if B occurs A cannot occur. .

Now let us find an example for the sifuation : the occurrence of A implies that of B.

Take the experiment of tossing a die twice. Let $A = \{ (x, y) \mid x + y = 12 \}$ be the event that the total score is 12, and $B = \{ (x, y) \mid x - y = 0 \}$ be the event of having the same score on both the throws. Then

$A= \{ (6,6) \}$ and

$B = \{ (1, 1), (2, 2), (3, 3), (4, 4), (5, 5), (6, 6);$

so that whenever A occurs, B does. Note that $A \subset B$.

You were already familiar with the various operations on sets. In Sec. 5.4 we had . Sample Space of a Random identified events with subsets of the sample space. What we have done in. this section Experiment is to apply set operations to events, and to interpret the combined events.

1.5 Summary

In this introductory unit to the study of probability, we have made the following points:

- There are many situations in real life as well as in scientific work which can be regarded as experiments having more than one possible oatcome. We cannot predict the outcome that we will obtain at the conclusion of the experiment. Such experiments are called random experiments.
- The study of random experiments begins with a specification of its all possible outcomes. In this specification, we have to make certain assumptions to avoid complexities. The set of all possible outcome is called the sample space of the experiment. A sample space with a finite number or a countable infinity of points is a discrete sample space. /

- When we are dealing with a discrete sample space, we can identify events with sets of points in the sample space. Thus, an event can be formally regarded as a subset of the sample space. This definition works only when the sample space is discrete.
- We can use operations like complementation, intersection,, union and difference to generate new events.
- Some complex events can be described in terms of simpler events by using the above-mentioned set operations.

1.6 Keywords

Events: An event is a set of outcomes (a subset of the sample space) to which probability assigned.

Sample space: The sample space of ten denoted by S, W, of an experiment or random trial is the set of all possible outcomes.

Set: A set is a collection of well defined and distance object considered as an object of its own right.

Union: Two sets can be added together. It is denoted by $\cup$.

1.7 Self Assessment

1. Flipping of two coin then it is possible to get 0 heads, 1 head, 2 heads. Then sample space will be

   (a) {1, 2, 3} (b) {0, 1, 2}

   (c) {2, –1, 0} (d) {0, 1, 3}

2. An event is a set of outcomes (a subset of the sample space) to which probability assigned.

   (a) Events (b) Sample space

   (c) Set (d) Union

3. The sample space of ten denoted by S, W, of an experiment or random trial is the set of all possible outcomes.

   (a) Events (b) Sample space

   (c) Set (d) Union

4. A set is a collection of well defined and distance object considered as an object of its own right.

   (a) Events (b) Sample space

   (c) Set (d) Union

5. Two sets can be added together. It is denoted by $\cup$.

   (a) Events (b) Sample space

   (c) Set (d) Union

6. Often rolling two dice. The sum all {2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12}. However, each of these aren't equally likely. The only way to get a sum 2 is to roll a 1 on both dice, but you can get a sum of 4 by rolling a ....

   (a) 1 – 3, 2 – 2, 2 – 5, 2 – 3 (b) 1 – 3, 2 – 2, 2 – 1, 2 – 0

   (c) 3 – 1, 1 – 3, 2 – 2, 4 – 0 (d) 1 – 3, 2 – 2, 3 – 1

Notes

1.8 Review Questions

1. Classify the experiments described below as random or non-random experiments.

   (a) A spark of electricity is introduced in a cylinder containing a mixture of hydrogen and oxygen. The end product is observed.

   (b) A lake contains two types of fish. Ten fish are caught and the number of fish of each type is noted.

   (c) The time taken by a powerful radio impulse to travel from the earth to the moon and for its echo to return to the sender is observed.

   (d) Two cards are drawn from a well-shuffled pack of 52 playing cards and the suits (Club, Diamond, Heart and Spade) to which they belong are noted.

2. Write down the sample spaces of all those experiments from 3 to 11 which we have not discussed earlier. Indicate in each case the assumptions made by you.

3. Let A_1, A_2, A_3 and A_4 be arbitrary events. Find expressions for the events that correspond to occurrence of

   (a) only A_1 and A_2,

   (b) none of A_1, A_2, A_3 and A_4,

   (c) one and only one of A_1, A_2, A_3, A_4,

   (d) not more than one of A_1, A_2, A_3, A_4

   (e) at least two of A_1, A_2, A_3, A_4.

4. Express in words the following events:

   (a) $A_1^c \cap A_2 \cap A_3$

   (b) $(A_1^c \cap A_2^c \cap A_3^c) \cup (A_1 \cap A_2^c \cap A_3^c)$

   (c) $(A_1 \cup A_2) - (A_3 \cup A_4)$

   (d) $(A_1 \cup A_2) \cap A_3$

   (e) $(A_1 \cap A_2) \cap (A_2 \cap A_3) \cup (A_3 \cap A_1)$

Answers: Self Assessment

1. (b) 2. (a) 3. (b) 4. (c) 5. (d) 6. (d)

1.9 Further Readings

Books

Introductory Probability and Statistical Applications by P.L. Meyer

Introduction to Mathematical Statistics by Hogg and Craig

Fundamentals of Mathematical Statistics by S.C. Gupta and V.K. Kapoor

Unit 2: Methods of Enumeration of Probability

Notes

CONTENTS

Objectives

Introduction

2.1 Probability : Axiomatic Approach

2.1.1 Probability of an Event : Definition

2.1.2 ProbabJlity of an Event : Properties

2.2 Classical Definition of Probability

2.3 Summary

2.4 Keywords

2.5 Self Assessment

2.6 Review Questions

2.7 Further Readings

Objectives

After studying this unit, you will be able to:

- Discuss probabilities to the outcomes of a random experiment with discrete sample space,
- Explain properties of probabilities of events, I
- Describe the probability of an event,
- Explain conditional probabilities and establish Bayes theorem,

Introduction

In this unit, we shall introduce you to some simple properties of the probability of an event associated with a discrete sample space. Our definitions require you to first specify the probabilities to be attached to each individual outcome of the random experiment.

Therefore, we need to answer the question : How does one assign probabilities to each and every individual outcome? This question was answered very simply by the classical probabilists (like Jacob Bernoulli). They assumed that all outcomes are equally likely.

Therefore, for them, when a random experiment has a finite number N of outcomes, the probability of each outcome would be 1/N. Based on this assumption they developed a probability theory, which we shall briefly describe in Sec. 6.4. However, this approach has a number of logical difficulties. One of them is to find a reasonable way of specifying "equally likely outcomes."

However, one possible way out of this difficulty is to relate the probability of an event to the relative frequency with which it occurs. To illustrate this point, we consider the experiment of tossing a coin a large number of times and noting the number of times "Head" appears.

Notes

In fact, the famous mathematician, Karl Pearson, performed this experiment 24000 times. He found that the relative frequency, which is the number of heads divided by the total nuinber of tosses, approaches 112 as moFe and more repetitions of the experiment are performed. This is the same figure which the classical probabilists would assign to the probability of obtaining a head on the toss of a balanced coin.

Thus, it appears that the probability of an event could be interpreted as the long range relative frequency with which it occurs. This is called the statistical interpretation or the, 'frequentist approach to the interpretation of the probability of an event. This approach has its own difficulties. We'll not discuss these here. Apart from these two, there are a few other approaches to the interpretation of probability. These issues are full of philosophical controversies, which are still not settled.

We, shall adopt the axiomatic approach formulated by Kolmogorov and treat probabilities as numbers satisfying certain basic rules. This approach is introduced.

We deal with properties of probabilities of events and their computation. We discuss the important concept of conditional probability of an event given that another event has occurred. It also includes the celebrated Bayes' theorem. We discuss the definition and consequences of the independence of two or more events. Finally, we talk about the probabilistic structure of independent repetitions of experiments. After getting familiar with the computation of probabilities in this,unit, we shall take up the study of probability distributions in the next one.

2.1 Probability : Axiomatic Approach

We have considered a number of examples of random experiments in the last unit. The outcomes of such experiments cannot be predicted in advance. Nevertheless, we frequently make vague statements about the chances or probabilities associated with outcomes of random experiments, Cohsider the following examples of such vague statements :

(i) It is very likely that it would rain today.

(ii) The chance that the Indian team will win this match is very small. '

(iii) A person who smokes more than 10 cigarettes a day will most probably developing lung cancer.

(iv) The chances of my whning the first prize in a lottery are negligible.

(v) The price of sugar would most probably increase next week.

Probability theory attempts to quantify such vague statements about the chances being good or bad, small or large. To give you an idea of such quantification, we describe two simple random experiments and associate probabilities with their outcomes.

Example 1:

(i) A balanced coin is tossed. The two possible outcomes are head (H) and tail (T). We associate probability P{H} = 1/2 to the outcome H and probability P{T} = 1/2 to T.

(ii) A person is selected from a large group of persons and his blood group is determined. It can be one of the four blood groups O, A, B and AB. One possible assignment of probabilities to these outcomes is given below

Blood group	0	A	3	AB
Probability	0.34	0.27	0.31	0.08

Now look carefully at the probabilities attached to the sample points in Example 1 (i) and (ii). Did you notice that

(i) these are number's between 0 and 1, and

(ii) the sum of the probabilities of all the sample points is one ?

This is not true of this example alone. In general, we have the following rule or axiom about the assignment of probabilities to the points of a discrete sample space.

Axiom : Let Ω be a discrete samplk space containing the points $\omega_1, \omega_2, \ldots$; i.e.,

$$\Omega = \{\omega_1, \omega_2, \ldots\}.$$

To each point ω_j of Ω, assign a number $P\{\omega_j\}$, $0 \le P\{\omega_j\} \le 1$, such that

$$P\{\omega_1\} + P\{\omega_2\} + \ldots = 1. \qquad \ldots (1)$$

We call $P\{\omega_j\}$, the probability of ω_j.

Now see if you can do the following exercise on the basis of this axiom.

If you have done El, you would have noticed that it is possible to have more than one valid assignment of probabilities to the same sample space. If the discrete sample space is not finite, the left side of Equation (1) should be interpreted as an infinite series. For example, suppose $\Omega = \{ \omega_1, \omega_2, \ldots \}$ and

$P\{\omega_j\} = 1 / 2^j$, $\forall$ j =1 , 2, ...

Then this assignment is valid because, $0 \le P\{ \omega_j \} \le 1$, and

$$P\{\omega_1\} + P\{\omega_2\} + \ldots = \frac{1}{2} + \left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^3 + \ldots$$

$$= \frac{1}{2}\left(1 + \frac{1}{2} + \frac{1}{2^2} + \ldots\right)$$

$$= 1$$

So far we have not explained what the probability $P\{\omega_j\}$ assigned to the point oj signifies. We have just said that they are all arbitrary numbers between 0 and 1, except for the requirement that they add up to 1. In fact, we have not even tried to clarify the nature of the sample space except to assert that it be a discrete sample space. Such an approach is consistent with the usual procedure of beginning the study of a mathematical discipline with a few undefined notions and axioms and then building a theory based on the laws of logic (Remember the axioms of geometry?). It is for this reason that this approach to the specification of probabilities to discrete sample spaces is called the axiomatic approach. It was introduced by the Russian mathematician A.N. Kolmogorov in 1933. This approach is mathematically precise and is now universally accepted. But when we try to use the mathematical theory of probability to solve some real life problems, that we have to interpret the significance of statements like "The probability of an event **A** is 0.6."

We now define the probability of an event **A** for a discrete sample space.

2.1.1 Probability of an Event : Definition

Let Ω be a discrete sample space consisting of the points $\omega_1, \omega_2, \ldots$, finite or infinite in number. Let $P\{\omega_1\}, P\{\omega_2\}, \ldots$ be the probabilities assigned to the points $\omega_1, \omega_2, \ldots$

Notes

Definition 1 : The probability P(.4) of an event **A** is the sum of the Probabilities of the points in A. More formally,

$$P(A) = \sum_{\omega_j \in A} P\{\omega_j\} \quad ....(2)$$

where $\sum_{\omega_j \in A}$ stands for the fact that the sum is taken over all the points $\omega_j \in A$, A is, of course, a subset of Ω. By convention, we assign probability zero to the empty set. Thus, $P(\Phi_j) = 0$.

The following example should help in clarifying this concept.

Example 2: Let a be the sample space corresponding to three tosses of a coin with the following assignment of probabilities.

Sample point	HHH	HHT	HTH	THH	TTH	THT	HTT	TIT
Probability	1/8	1/8	118	1/8	1/8	1/8	1/8	1/8

Let's find the probabilities of the events A and B, where

A : There is exactly one head in three tosses, and

B : All the three tosses yield the same result

Now A = (HTF, THT, TITH)

Therefore,

P(A) = 1/8 + 1/8 + 1/8 = 318.

Further, B = {HHH, TTT). Therefore, P(B) = $\frac{1}{8} + \frac{1}{8} = \frac{1}{4}$.

Proceeding along these lines you should be able to do this exercise.

A word about our notation and nomenclature is necessary at this stage. Although we say that $P\{\omega_j\}$ is the probability assigned to the point wj of the sample space, it can be also interpreted as the probability of the singleton event $\{\omega_j\}$.

In fact, it would be useful to remember that probabilities are defined only for events and that $P\{\omega_j\}$ is the probability of the singleton event $\{\omega_j\}$. This type of distinction will be all the more necessary when you proceed to study probability theory for non-discrete sample spaces in Block 3.

Now let us look at some. of the probabilities of events.

2.1.2 ProbabJlity of an Event : Properties

By now you know that the probability P(A) of an event A associated with a discrete sample space is the sum of the probabilities assigned to the sample points in A. In this section we discuss the properties of the probabilities of events.

P1: For every event A, $0 \le P(A) \le 1$.

Proof: This is a straightforward consequence of the definition of P(A). Since it is the sum of non-negative numbers, $P(A) \ge 0$. Since the sum of the probabilities assigned to all the points in the sample space is one and since A is a subset of R, the sum of the probabilities assigned to the points in A cannot exceed P(R), which is one. In other words, whatever may be the event A, $0 \le P(A) \le 1$.

Now here is an important remark.

Remark 1 : If A = ϕ, $P(\phi_j) = 0$. However, P(A) = 0 does not, in general, imply that A is the empty set. For example, consider the assignment (i) of E1. You must have already shown that it is valid. If $A = \{\omega_6, \omega_7\}$. P(A) = 0 but A is not empty.

Similarly $P(\Omega) = 1$. But if P(B) = 1, does it follow that $B = \Omega$? No. Can you think of a Probability on a Discrete sample counter example ? What about El) i) again ? If we take $B = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5\}$, then P(B) = 1 but $B \neq \Omega$. In this connection, recall that the empty set and the whole space Ω were called the impossible event and the sure event respectively. In future, an event A with probability P(A) = 0 will be called a null event and an event B of probability one, will be called an almost sure event.

This remark brings out the fact that the impossible event is a null event but that a null event is not the impossible event. Similarly, the sure event is an almost sure event but an almost sure event is not necessarily the sure event.

Let us take up another property now.

P2 : $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Proof : Recall that according to the definition, $P(A \cup B)$ is the sum of the probabilities attached to the points of $A \cup B$, each point being considered only once. However, when we compute P(A) + P(B), a point in $A \cap B$ is included once in the computation of P(A) and once in the computation of P(B). Thus, the probabilities of points in A n B get added twice in the computation of P(A) + P(B). If we sdbtract the probabilities of all points in $A \cap B$, from P(A) + P(B), then we shall be left with $P(A \cup B)$, i.e.,

$$P(A \cup B) = P(A) + P(B) - \sum_{\omega_j \in A \cap B} P\{\omega_j\}$$

The last term in the above relation is, by definition, $P(A \cap B)$. Hence we have proved P2. We now list some properties which follow from P1 and P2.

P3 : If A and B are disjoint events, then

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

P4 : $P(A^c) = 1 - P(A)$

P5 : $P(A \cup B) \leq P(A) + P(B)$

Why don't you try to prove these yourself? That's what we suggest in the following exercise.

We continue with the list of properties.

P6 : If $A \subset B$, then $P(A) \leq P(B)$.

Proof: If $A \subset B$, A and B – A are disjoint events and their union, $A \cup (B - A)$ is B. Also see Fig. 1. Hence by P3, .

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

Since by P1, $P(B - A) \geq 0$, P6 follows from the above equation.

Now let us take a look at P5 again.

The inequality $P(A \cup B) \leq P(A) + P(B)$ in P5 is sometimes called Boole's inequality. We claim that equality holds in Boole's inequality if $A \cap B$ is a null event. Do you agree?

An easy induction argument leads to the following generalisation of P5.

Boole's inequality : If $A_1, A_2, \ldots, A_N$ are N events, then

$$P\left(\bigcup_{J=1}^{N} A_j\right) \leq \sum_{j=1}^{N} P(A_j)$$

Proof : By P5, the result is true for N = 2. Assume that it is true for $N \leq r$, and observe that $A_1 \cup A_2 \cup \ldots \cup A_{r+1}$ is the same as $B \cup A_{r+1}$, where $B = A_1 \cup A_2 \cup \ldots \cup Ar$. Then by P5,

$$P\left(\bigcup_{J=1}^{r+1} A_j\right) = P(B \cup A_{r+1}) \leq P(B) + P(A_{r+1}) \leq \sum_{j=1}^{r} P(A_j) + P(A_{r+1}),$$

where the last inequality is a consequence of the induction hypothesis. Hence, if Boole's inequality holds for $N \leq r$, it holds for N = r + 1 and hence for all $N \geq 2$.

A similar induction argument yields

P7 : If $A_1, A_2, \ldots, A_n$ are pair wise disjoint events, i.e., if $A_i \cap A_j = \phi$, $i \neq j$, then

$$P\left(\bigcup_{J=1}^{n} A_j\right) = P(A_1) + P(A_2) + \ldots + P(A_n) \qquad \ldots(3)$$

We sometimes refer to the relation (3) as the Property of finite additivity.

We can generalise P7 to apply to an'infinite sequence of events.

P8 : If $(A_n, n \geq 1)$ is a sequence of pair wise disjoint events, then

$$P\left(\bigcup_{J=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} P(A_j)$$

P8 is called the σ-additivity pioperty.

In the general theory of probability, which covers non-discrete sample spaces as well, σ-additivity and therefore finite additivity is included as an axiom to be satisfied by probabilities of events.

We now discuss some examples based on the above properties.

Example 3: Let us check whether the probabilities P(A) and P(B) are consistently defined in the following cases.

(i) P(A) = 0.3 P(B) = 0.4, $P(A \cap B) = 0.4$

(ii) P(A) = 0.3 P(B) = 0.4, $P(A \cap B) = 0.8$

Here we have to see whether P1, P2, P3, P5 and P6 are satisfied or not. P4, P7 and P8 do not apply here since we are considering only two sets. In both the cases P(A) and P(B) are not consistently defined. Since $A \cap B \subset A$, by P6. $P(A \cap B) \leq P(A)$. In case (i), $P(A \cap B) = 0.4 > 0.3 = P(A)$, which is impossible. Similar is the'situatmn with case (ii). Moreover, note that case (ii) also violates P1 and P2. Recall that by P2,

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

but $P(A) + P(B) - P(A \cap B) = 0.3 + 0.4 - 0.8 = -0.1$ which is impossible.

Example 4: We can extend the property P2 to the case of three events, i.e., we can show that

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(B \cap C) - (C \cap A) + P(A \cap B \cap C) \quad \text{... (5)}$$

Denote $B \cup C$ by H. Then $A \cup B \cup C = A \cup H$ and by P2, $P(A \cup B \cup C)$

$$= P(A \text{ U } H) = P(A) + P(H) - P(A \cap H). \quad \text{... (6)}$$

But $\quad P(H) = P(B \cup C) = P(B) + P(C) - P(B \cap C) \quad \text{... (7)}$

and $\quad P(A \cap H) = P(A \cap (B \cup C))$

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

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

$$= P(A \cap B) + P(A \cap C) - P(A \cap B \cap C) \quad \text{... (8)}$$

Substituting from (7) and (8) in (6) we get the required result. Also see Fig. 2.

Here are some simple exercises which you can solve by using P1-P7

2.2 Classical Definition of Probability

In the early stages, probability theory was mainly concerned with its applications to games of chance. The sample space for these games consisted of a finite number of outcomes. These simple situations led to a definition of probability which is now called the classical definiqon. It has may limitations. For example, it cannot be applied to infinite sample space. However, it is useful in understanding the concept of randomness so essential in the planning of experiments, small and large-scale sample surveys, as well as in solving some interesting problems. We shall motivate the classical definition with some examples. We shall then formulate the classical definition and apply it to solve some simple problems.

Suppose we toss a coin. This experiment has only two possible outcomes : Head (H) and Tail (T). If the coin is a balanced coin and is symmetric, there is no particular reason to expect that H is more likely than Tor that T is more likely than H. In other words, we may assume that the two outcomes H and T have the same probability or that they are equally likely. If they have the same probability, and if the sum of the.two probabilities P(H) and P(T) is to be one, we inust have P(H) = P(T) = 1/2.

Similarly, if we roll a symmetric, balanced die once, we should assign the same probability, viz. 116 to each of the six possible outcomes 1, 2, . . . ,6.

The same type of argument, when used for assigning probabilities to the results of drawing a card from a well-shuffled pack of 52 playing cards leads us to say that the probability of drawing any specified card is 1/52.

In general, we have the following :

Definition 2 : Suppose a sample space Ω has a finite number n of points $\omega_1, \omega_2, \ldots, \omega_n$. The classical definition assigns the probability l/n to each of these points, i.e.,

$$P\{\omega_j\} = \frac{1}{n}, j = 1, ..., n.$$

Notes

The above assignment is also referred to as the assignment in case of equally likely outcomes. You can check that in this case, the total of the probabilities of all then points is $n \times \frac{1}{n} = 1.$ In fact, this is a valid assignment even from the axiomatic pht'of view.

Now suppose that an event A contains m points. Then under the classical assignment, the probability P(A) of A is m/n. The early probabilists called m; the number of cases favourable to A and n, the total number of cases. Thus, according to the classical definition,

$$P(A) = \frac{\text{Number of cases favourable to A}}{\text{Total number of cases}}$$

We have already mentioned that this is a valid assignment consistent with the Axiom in Sec. 6.2. Therefore, it follows that the probabilities of events, defined in this manner, possess the properties P1 – P7.

We now give some examyies based on this definition.

Example 5: Two identical symmetric dice are thrown. Let us find the probability of obtaining a total score of 8.

The total number of possible outcomes is 6 × 6 = 36. There are 5 sample points, (2,6), (3,5), (4,4), (5,3), (6,2), which are favourable to the event A of getting a total score of 8. Hence the required probability is 5/36.

Example 6: If each card of an ordinary deck of 52 playing cards has the same probability of being drawn, let us find the probability of drawing.

(i) a red king or a black ace

(ii) a3, 4, 5, 6 or 8?

Let's tackle these one by one

(i) Since there are two red kings (diamond and heart) and two black aces (spade and club), the number of favourable cases is 4. The required probability is 4/52 = 1/13.

(ii) There are 4 cards of each of the 5 denominations 3, 4, 5, 6 and 8. Thus, the total number of favourable cases is 20 and the required probability is 20/52 = 5/13.

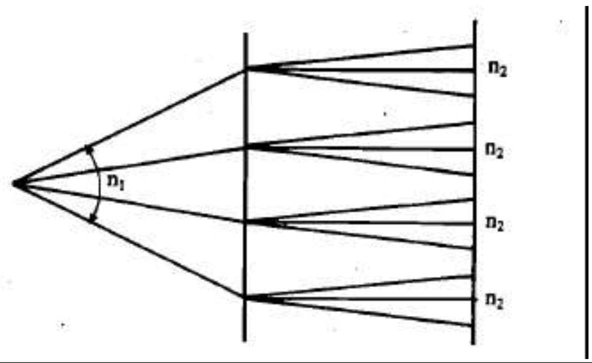
You must have realised by this time that in order to apply the ctassical definition of probability, you must be able to count the number of points favourable to an event A as well as the total number of sample points. This is not always easy. We can, however, use the theory of permutations and combinations for this purpose.

To refresh your memory, here we give two important rules which are used in counting.

1) ***Multiplication Rule :*** If an operation is performed in n_1 ways and for each of these n_1 ways, a second operation can be performed in n_2 ways, then the two operations can be performed together in n_1n_2 ways. See Fig.

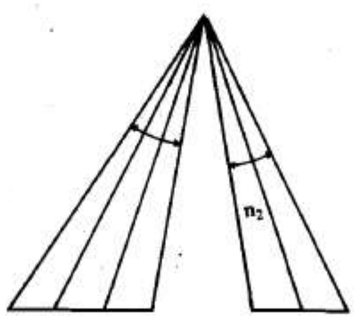
Notes

Figure 2.1



2) ***Addition Rule :*** Suppose an operation can be performed in n, ways and a second operation can be performed in n_2 ways. Suppose, further that it is not possible to perform both together. Then the number of ways in which we can perform the first or the second operation in $n_1 + n_2$. See Fig. 4.

Figure 2.2



We now illustrate the use of this theory in calculating probabilities by considering some examples. We assume that all outcomes in each of these examples are equally likely. Under this assumption, the classical definition of probability is applicable.

Example 7: We first select a digit out of the ten digits, 0, 1,2,3, ..., 9. Then we select another digit out of the renlaining nine. What will be the probability that both these digits are odd?

We can select the first digit in 10 ways and for each of these ways we can select the second digit in 9 ways. Therefore, the total number of points in the sample space is 10 × 9 = 90. The first digit, can be odd in 5 ways (1,3,5,7.9). and then the second digit can be odd in 4 ways. Thus, the total number of ways in which both the digits can be odd is 5 × 4 = 20. The required probability is therefore $\frac{20}{90} = \frac{2}{9}$.

Remark 2 : In Example 7, every digit had the same chance of being selected. This is sometimes expressed by saying that the digits were selected **at random** (with equal probability). Sele~tion~raatn dom is generally taken to be synonymous with the assignment of the same probability to all the sample points, unless stated otherwise.

Notes

We now give a number of examples to show how to calculate the probabilities of events in a variety of situations. Please go through these examples carefully. If you understand them. you will have no difficulty in doing the exercises later.

Example 8: A box contains ninety good and ten defective screws. Let us find theprobability that 5 screws selected at random out of this box are all good.

Let A be the event that the 5 selected screws are all good.

Now we can choose 5 screws out of 100 screws in ways. If the selected 5 screws are to be good, they will have to be selected out of the 90 good screws. This can be done in ways. This is the number of sample points favourable to A. Hence the probability of A

Example 9: A government prints 10 lakh lottery tickets of value of Rs. 2 each. We would like to know the number of tickets that must be bought to have a chance of 0.5 or more to win the first prize of 2 lakhs.

The prize-winning ticket can be mndomly selected out of the 10 lakh tickets in 10^6 ways.

Now, let m denote the number of tickets that we must buy. Then m is the number of points favourable to our winning the first prize. Therefore, the probability of our winning the first prize, is, $\frac{m}{10^6}$.

Since we want that $\frac{m}{10^6} \geq \frac{1}{2}$, therefore $m \leq \frac{10^6}{2}$. This means that we must buy at least $\frac{10^6}{2}$ = 500,000 tickets, at a cost of at least Rs. 10 lakhs ! Not a profitable proposition at all !

Example 10: In a study centre batch of 100 students, 54 opted for MTE-06,69 opted for MTE - 11 and 35 opted for both MTE-06 and MTE-11. If one of these students is selected at random, let us find the probability that the student has opted for MTE-06 or MTE- 11.

Let M denote the event that the randomly selected student has opted for MTE-06 and S the event that helshe has opted for MTE- 11. We want to know P(M U S). According to the classical definition. $P(M) = \frac{54}{100}$, $P(S) = \frac{69}{100}$ and $P(M \cap S) = \frac{35}{100}$. Thus

$P(M \cup S) = P(M) + P(S) - P(M \cap S)$

Suppose now we want to know the probability that the randomly selected student has opted for neither MTE-06 nor MTE-11. This means that we want to know $P[M^C \cap S^C]$.

Now,

$$M^C \cap S^C = (M \cap S)^C$$

Therefore,

$$P(M^C \cap S^C) = 1 - P[M \cup S] = 1 - 0.88 = 0.12$$

Lastly, to obtain the probability that the student has opted for MTE-06 but not for MTE-11, i.e., to obtain $P(M \cap S^C)$, observe that $M = (M \cap S) \cup (M \cap S^C)$ and that $M \cap S$ and $M \cap S^C$ are disjoint events. Thus.

$$P(M) = P(M \cap S) + P(M \cap SC)$$

Notes

or $P(M \cap SC) = P(M) - P(M \cap S)$

$$= \frac{54}{100} - \frac{35}{100} = 0.19.$$

Example 11: A throws six unbiased dice and wins if he has at least one six. B throws twelve unbiased dice and wins if he has at least two sixes. Who do you think is more likely to win?

We would urge you- to make a guess first and then go through the following computations. Check if your intuition was correct.

The total number of outcomes for A is $n_A = 6^6$ and that for B is $n_B = 6^{12}$, We will first calculate the probabilities qA and qB that A and B, respectively, loose their games. Then the I probabilities of their winning are $P_A = 1 - q_A$ and $P_B = 1 - q_B$, respectively. We do this because q_A and q_B are easier to compute.

Now A loses if he does not have a six on any of the 6 dice he rolls. This can happen in 5^6 different ways, since he can have nu sis on each die in 5 ways. Hence $q_A = 5^6/6^6$ and therefore, $P_A = 1 - (5/6)^6 \equiv 0.665$.

In order to calculate q_B, observe that B loses if he has no six or exactly one six. The probability that he has no six is $5^{12}/6^{12} = (5/6)^{12}$. Now the single six can occur on any one of the 12 dice, i.e., in $\binom{12}{1}$ ways. Then all the remaining 11 dice have to have a score other than six. This can happen in 5^{11} ways.

Therefore, the total number of ways of obtaining one six is $\binom{12}{1}5^{11}$. Hence the probability that B has exactly one six is $\frac{12 \times 5^{11}}{6^{12}}$.

The events of "no six" and "one six" in the throwing of 12 dice are disjoint events. Hence the probability

$$q_B = (5/6)^{12} + 12\frac{5^{11}}{6^{12}} \equiv 0.381$$

Thus, $P_B \equiv 1 - 0.38\,1 = 0.619$.

Comparing PA and PB, we can conclude that A has a greater probability of winning.

Now here are some exercises which you should try to solve.

So far we have seen various examples of assigning probabilities to sample points and have also discussed some properties of probabilities of events. In the next section we shall talk about the concept of conditional probability.

Notes

2.3 Summary

- When a random experiment has a finite number N of outcomes, the probability of each outcome would be 1/N. Based on this assumption they developed a probability theory, which we shall briefly describe in Sec. 6.4. However, this approach has a number of logical difficulties. One of them is to find a reasonable way of specifying "equally likely outcomes."
- Probability theory attempts to quantify such vague statements about the chances being good or bad, small or large. To give you an idea of such quantification, we describe two simple random experiments and associate probabilities with their outcomes.
- Let Ω be a discrete sample space consisting of the points $\omega_1, \omega_2, \ldots$, finite or infinite in number. Let $P\{\omega_1\}, P\{\omega_2\}, \ldots$ be the probabilities assigned to the points $\omega_1, \omega_2, \ldots$
- A word about our notation and nomenclature is necessary at this stage. Although we say that $P\{\omega_j\}$ is the probability assigned to the point wj of the sample space, it can be also interpreted as the probability of the singleton event $\{\omega_j\}$.

  In fact, it would be useful to remember that probabilities are defined only for events and that $P\{\omega_j\}$ is the probability of the singleton event $\{\omega_j\}$. This type of distinction will be all the more necessary when you proceed to study probability theory for non-discrete sample spaces in Block 3.

2.4 Keywords

Multiplication Rule : If an operation is performed in n_1 ways and for each of these n_1 ways, a second operation can be performed in n_2 ways, then the two operations can be performed together in n_1n_2 ways.

Addition Rule : Suppose an operation can be performed in n, ways and a second operation can be performed in n_2 ways. Suppose, further that it is not possible to perform both together. Then the number of ways in which we can perform the first or the second operation in $n_1 + n_2$.

2.5 Self Assessment

1. If P(A) = 0.3, P(B) = 0.4, $P(A \cap B) = 0.4$. Then find $P(A \cup B)$

   (a) –0.1 (b) 0.3

   (b) 0.4 (d) 0.2

2. If P(A) = 0.5, P(B) = 0.7, $P(A \cap B) = 0.4$. Then find $P(A \cup B)$

   (a) 0.8 (b) 0.1

   (c) 0.3 (d) 0.4

3. If two identical symmetric dice are thrown. Find the probabilities of obtaining 9 total score of 8.

   (a) 5/36 (b) 2/4

   (c) 4/36 (d) 6/36

4. If each card on an orindary deck of 52 playing cards has the same probability a being drawn, then find a red king or a black ace?
   (a) 1/13 (b) 2/13
   (c) 4/52 (d) 4/43
5. If P(A) and P(B) is given then $P(A \cup B)$ is equal to
   (a) $P(A) + P(B) + P(A \cap B)$ (b) $P(A) + P(B) - P(A \cap B)$
   (c) $P(A) - P(B) + P(A \cap B)$ (d) $P(A) - P(B) - P(A \cap B)$

2.6 Review Questions

1. Prove the following : Space
   (a) If P(A) = P(B) = 1; then $P(A \cup B) = P(A \cap B) = 1$.
   (b) If P(A) = P(B) = P(C) = 0, then $P(A \cup B \cup C) = 0$.
   (c) We have mentioned that by convention we take $P(\phi) = 0$.
   But see if you can prove it by using P4.
2. Fill in the blanks in the following table :

P(A)	P(B)	$P(A \cup B)$	$P(A \cap B)$
0.4	0.8		0.3
	0.5	0.6	0.25

3. Explain why each one of the following statements is incorrect.
   (a) The probability that a student will pass an examination is 0.65 and that he would fail is 0.45.
   (b) The probability that team A would win a match is 0.75, that the game will end in , a draw is 0.15 and that team A will not loose the game is 0.95.
   (c) The following is the table of probabilities for printing mistakes in a book.

No. of printing mistakes	0	1	2	3	4	5 or more
Probability	0.12	0.25	0.36	0.14	0.09	0.07

   (d) The probabilities that a bank will get 0, 1, 2, or more than 2 bad cheques on a given day are 0.08, 0.21, 0.29 and 0.40, respectively.
4. There are two assistants Seema (S) and Wilson (W) in an office. The probability that Seema will be absent on any given day is 0.05 and that Wilson will be absent on any given day is 0.10. The probability that both will be absent on the same day is 0.02. Find the probability that on a given day,
   (a) both Seema and Wilson would be present,
   (b) at least one of them would be present, and
   (c) only one of them will be absent.

Notes

5. A large office has three Xerox machines M_1, M_2 and M_3. The probability that on a given day

   M_1 works is 0.60

   M_2 works is 0.75

   M_3 works is 0.80

   both M_1 and M_2 work is 0.50

   both M_1 and M_3 work is 0.40

   both M_2 and M_3 work is 0.70

   all of them work is 0.25.

   Find the probability that on a given day at least one of the three machines works.

Answers: Self Assessment

1. (b) 2. (a) 3. (a) 4. (a) 5. (b)

2.7 Further Readings

Books

Introductory Probability and Statistical Applications by P.L. Meyer

Introduction to Mathematical Statistics by Hogg and Craig

Fundamentals of Mathematical Statistics by S.C. Gupta and V.K. Kapoor

Unit 3: Conditional Probability and Independence Baye's Theorem

Notes

CONTENTS

Objectives

Introduction

3.1 Conditional Probability

3.2 Baye's Theorem

3.3 Independence of Events

3.4 Repeated Experiments and Trials

3.5 Summary

3.6 Keywords

3.7 Self Assesment

3.8 Review Questions

3.9 Further Readings

Objectives

After studying this unit, you will be able to:

- Discuss probabilities to the outcomes of a random experiment with discrete sample space,
- Explain properties of probabilities of events, I
- Describe the probability of an event,
- Explain conditional probabilities and establish Bayes theorem,

Introduction

Suppose that two series of tickets are issued for a lottery. Let 1,2,3,4,5 be the numbers on the 5 tickets in series I and let 6,7, 8,9, be the numbers on the. 4 tickets in series 11. I hold the ticket, bearing number 3. Suppose the first prize in the lottery is decided by selecting one of the 5 + 4 = 9 numbers at random. The probability that I will win the prize is 1/9. Does this probability change if it is known that the prize-winning ticket is from series I? Ineffect, we want to know the probability of my winning the prize, conditional on the knqwledge that the prize-winning ticket is from series I.

In order to answer this question, observe that the given information reduces our sample-space from the set { 1,2,3,.4,5,6,7,8,9 } to its subset { 1.2.3.4.5 } containing 5 points. In fact, this subset { 1,2,3,4,5 } corresponds to the event H that the prize winning ticket. belongs to series I. If the prize winning ticket is selected by choosing one of these 5 numbers at random, the probability that I will win the prize is 115. Therefore, it seems logical to say that the conditional probability of the event **A** of my winning the prize, given that the prize-winning number is from series I, is

P(A | H) = 1/5.

Notes

Here P(A | H) is read as the conditional probability of A given the event H. Note that we can write

$$P(A \mid H) = \frac{1/9}{5/9} = \frac{P(A \cap H)}{P(H)}$$

This discussion enables us to introduce the following formal definition. In what follows we assume that we are given a random experiment with discrete sample space R, and all relevant events are subsets of R.

3.1 Conditional Probability

Definition 3 : Let H be an event of positive probability, that is, P(H) > 0. The conditional probability P(A | H) of an event A, given the event H, is

$$P(A \mid H) = \frac{P(A \cap H)}{P(H)} \qquad ...(9)$$

Notice that we have not put any restriction on the event A except that A and H be subsets of the same sample space R and that P(H) > 0.

Now we give two examples to help clarify this concept.

Example 12: In a small town of 1000 people there are 400 females and 200 colour-blind persons. It is known that ten per.cent, i.e. 40, of the 400 females are colour-blind. Let us find the probability that a randomly chosen person is colour-blind, given that the selected person is a female.

Now suppose we denote by A the event that the randomly chosen person is colour-blind and by H the event that the randomly chosen person is a female. You can see that

$P(A \cap H) = 40/1000 = 0.04$ and that

$P(H) = 400/1000 = 0.4$.

Then

$$P(A \mid H) = \frac{P(A \cap H)}{P(H)} = \frac{0.04}{0.40} = 0.1.$$

Now can you find the probability that a randomly chosen person is colour-blind, given that the selgcted person is a male?

If you denote by M the event that the selected person is a male, then

$$P(M) = \frac{600}{1000} = 0.6 \text{ and}$$

$$P(A \cap M) = \frac{600}{1000} = 0.16.$$

Therefore, $P(A \mid M) = \frac{0.16}{0.6} = 0.266.$

You must have noticed that P(A | M) > P(A | H). So there are greater chances of a man being colour-blind as compared to a woman.

Example 13: A manufacturer of automobile parts knows from past experience that the probability that an order will be completed on time is 0.75. The probability that an order is completed and delivered on time is 0.60. Can ygu help him to find the probability that an order will be delivered on time given that it is completed ?

Let A be the event that an order is delivered on time and H the event that it is completed on time. Then P(H) = 0.75 and $P(A \cap H) = 0.60$. We need $P(A \mid H)$.

$$P(A \mid H) = \frac{P(A \cap H)}{P(H)} = \frac{0.60}{0.75} = 0.8.$$

Have you understood the definition of conditional probability? You can find out for yourself by doing these simple exercises.

Task If A is the event that a person suffers from high blood pressure and B is the event that he is a smoker, explain in words what the following probabilities represent.

(a) $P(A \mid B)$

(b) $P(Ac \mid B)$

(c) $P(A \mid B^c)$

(d) $P(Ac \mid Bc)$.

Two unbiased dice are rolled. They both show the same score. What is the probability that their common score is 6?

We now state some of the properties of $P(A \mid H)$.

P1 : For any setA, $0 \leq P(A \mid H) \leq 1$.

Recall that since $A \cap H \subset H$, $P(A \cap H) \leq P(H)$. The required property follows immediately.

P2 : $P(A \mid H) = 0$ if and only if $A \cap H$ is a null set. In particular, $P(\phi \mid H) = 0$ and $P(A \mid H) = 0$ if A and H are disjoint events.

P3 : $P(A \mid H) = 1$ if and only if $P(A \cap H) = P(H)$.

In particular,

$P(\Omega \mid H)= 1$ $P(H \mid H) = 1$

P'4: $P(A \cup B \mid H) = P(A \mid H) + P(B \mid H) - P(A \cap B \mid H)$.

How do we get P'4 ? Well, since

$(A \cup B) \cap H = (A \cap H) \cup (B \cap H)$,

P2 gives us

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

Now use the definition of the conditional probability to obtain P4.

Using P'4 and P3 and P4 of Sec. 6.2.2, we get

P5 : If A and B are disjoint events, $P(A \cup B \mid H) = P(A \cap H) + P(B \mid H)$

and $P(A^C \mid H) = 1 - P(A \mid H)$

Compare $P'_1 - P'_5$ with the properties of (unconditiond) probabilities given in Sec. 6.2.2. You will find that the conditional probabilities, given the event H, have all the properties of unconditional probabilities, which are sometimes called the absolute properties.

We can use the conditional probatii:itics to compute the unconditional probabilities of events by employing the following obvious fact,

$$P(A \cap H) = P(H)\, P(A \cap H) \qquad \text{... (10)}$$

obtained from Definition 3 of $P(A \cap H)$.

Here is an important remark related to (10).

Notes See E 14 for the interpretations of $P(A \mid H)$ and $P(A^C \mid H)$

Remark 3 : Relation (10) holds even if $P(H) = 0$, provided we interpret $P(A \mid H) = 0$ if $P(H) = 0$. In words, this means that if the probability of -occurrence of H is zero, we say that the probability of occurrence of A, given that H has occurred, is also zero. This is so, because $P(H) = 0$ implies $P(A \cap H) = 0$, $(A \cap H)$ being a subset of H,

We now give an example to illustrate the use of Relation (10).

Example 14: Two cards are drawn at random and without ieplacement from a pack of 52 playing cards. Let us find the probability that both the cards are red.

Let A_1 and A_2 denote, respectively the events that cards drawn on the first and second draw are red. Then by the classical definition, $P(A_1) = 26/52$, since there are 26 red cards. If the first card is red, we are left with 25 red cards in the pack of 51 cards. Hence $P(A_2 \mid A_1) = 25/51$. Thus, the probability $P(A_1 \cap A_2)$ of both cards being red is

$$P(A_1 \cap A_2) = P(A_1)P(A_2 \mid A_1)$$

$$= \frac{26}{52}, \frac{25}{51} = 0.245.$$

Relation (10) specifies the probability of $A \cap H$ in terms of $P(H)$ and $P(A/H)$. We can extend this relation to obtain the probability, $P(A_1 \cap A_2 \cap A_3)$ in terms of $P(A_1)$, $P(A_2 \mid A_1)$ and $P(A_3 \mid A_1 \cap A_2)$. We, of course, assume that $P(A_1)$ and $P(A_1 \cap A_2)$ are both positive. Can you guess what this relation could be? Suppose we write

$$P(A_1 \cap A_2 \cap A_3) = P(A_1)\frac{P(A_1 \cap A_2)}{P(A_1)}.\frac{P(A_1 \cap A_2 \cap A_3)}{P(A_1 \cap A_2)}$$

Does this give you any clue? This gives us,

$$P(A_1 \cap A_2 \cap A_3) = P(A_1) \,.\, P(A_2 \mid A_1) \,.\, P(A_3 \mid A_1 \cap A_2).$$

Now let us use this to compute some probabilities.

Example 15: A box of mangoes is inspected by examining three randomly selected mangoes drawn without replacement. If all the three mangoes are good, the box is sent to the market, otherwise it is rejected. Let us calculate the probability that a box of 100 mangoes containing 90 good mangoes and 10 bad ones will pass the inspection.

Let A_1, A_2 and A_3, respectively denote the events that the first, second and third mangoes are good. Then $P(A_1) = 90/100$, $P(A_2 \mid A_1) = 89/99$, and $P(A_3 \mid A_1 \cap AZ) = 88/98$ according to the classichl definition. Thus.

$$P(A_1 \cap A_2 \cap A_3) = \frac{90}{100}.\frac{89}{99}.\frac{88}{98} = 0.727.$$

We end this section with a derivation of a well-known theorem in probability theory, called the Bayes' theorem.

Consider an event B and its complementary event B^C. The pair (B, B^C) is called a partition of Ω, since they satisfy $B \cap B^C = \phi$, and $B \cup B^C$ is the whole sample space Ω. Observe.that for any event A,

$$A = A \cap \Omega = A \cap (B \cup B^C) = (A \cap B) \cup (A \cap B^C).$$

Since $A \cap B$ and $A \cap B^C$ are subsets of the disjoint sets B and B^C, respectively, they themselves are disjoint. As a consequence, $P(A) = P(A \cap B) + P(A \cap B^C)$.

Now using Relation (10), we have

Here we do not insist that P(B) and $P(B^C)$ be positive and follow the convention stated in Remark 3.

It is now possible to extend Equation (11) to the case when we have a partition of Ω consisting of more than two sets. More specifically, we say that the n sets $B_1, B_2, \ldots, B_n$ constitute a partition of Ω if any two of them are disjoint, i.e., 1

$$B_i \cap B_j = \phi, \ i \neq j, \ i, j = 1, \ldots, n$$

and their union is Ω, i.e.,

$$\bigcup_{j=1}^{n} B_j = \Omega.$$

We can now write for any event A,

$$A = A \cap \Omega = A \cap \left(\bigcup_{j=1}^{n} B_j\right) = \bigcup_{j=1}^{n} (A \cap B_j).$$

Since $A \cap B_i$ and $A \cap B_j$ are respectively subsets of B_i and B_j, $i \neq j$, they are disjoint. Consequently by P7,

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

$$\text{or } P(A) = \sum_{j=1}^{n} P(B_j) P(A \cap B_j). \qquad \ldots (12)$$

which is obtained by using (10). This result (12) leads to the celebrated Bayes' theorem, which we now state.

Notes

3.2 Baye's Theorem

Theorem 1 (Bayes' Theorem) : If $B_1, B_2, \ldots, B_n$ are n events which constitute a partition of Ω and A is an event of positive probability, then

$$P(B_r \mid A) = \frac{P(B_r)P(A \mid B_r)}{\sum_{j=1}^{n} P(B_j)P(A \mid B_j)}$$

foranyr, $1 \le r \le n$.

Proof: Observe that by definition,

$$P(Br \mid A) = \frac{P(A \cap B_r)}{P(A)}$$

$$= \frac{P(B_r)P(A \mid B_r)}{P(A)}, \qquad \text{by (10)}$$

$$= \frac{P(B_r)P(A \mid B_r)}{\sum_{j=1}^{n} P(B_j)P(A \mid B_j)}, \qquad \text{by (12)}$$

The proof is complete.

In the examples that follow, you will see a variety of situations in which Bayes' theorem is useful.

Example 16: It is known that 25 per cent of the people in a community suffer from TB. A test to diagnose this'disease is such that the probability is 0.99 that a person suffering from it will show a positive result indicating its presence. The same test has probability 0.20 that a person not suffering from TB has a positive test result. If a randomly selected person from the community has positive test result, let us find the probability that he has TB.

Let B_1 denote the event that a randomly selected person has TB. Let $B_2 = B_j^c$. Then from the given information, $P(B_1) = 0.25$, $P(B_2) = 0.75$. Let A denote the event that the test for the randomly selected person yields a positive result. Then $P(A \mid B_1) = 0.99$ and $P(A \mid B_2) = 0.20$. We need to obtain $P(B1 \mid A)$. By applying Bayes' theorem we get

$$P(B1 \mid A) = \frac{P(B_1)P(A \mid B_1)}{P(B_1)P(A \mid B_1) + P(B_2)P(A \mid B_2)}$$

Example 17: We have three boxes, each containing two covered compartments. The first box has a gold coin in each compartment. The second box has a gold coin in one compartment and a silver coin in the other. The third box has a silver coin in each of its compartments. We choose a box at random and open a drawer at random. It contains a gold coin. We would like to know the probability that the other compartment also has a gold coin.

Let B_1, B_2, B_3, respectively, denote the events that Box 1, Box 2 and Box 3 are selected. It is easy to see that B_1, B_2, B_3 constitute a partition of the sample space of the experiment.

Since the boxes are selected at random, we have

$P(B_1) = P(B_2) = P(B_3) = 1/3$.

Notes

Let A denote the event that a gold coin is located. The composition of the boxes implies that

$P(A \mid B_1) = 1, P(A \mid B_2) = 1/2, P(A \mid B_3) = 0.$

Since one gold coin is observed, we' will have a gold coin in the other unobserved compartment of the box only. if we have selected Box 1. Thus, we need to obtain P(B 1 A).

Now by Bayes Theorem

$$P(B1 \mid A) = \frac{P(B_1)P(A|B_1)}{P(B_1)P(A|B_1) + P(B_2)P(A|B_2) + P(B_3)P(A|B_3)}$$

Do you feel confident enough to try and solve these exercises now? In each of them, the crucial step is to define the relevant events properly. Once you do that, the actual calculation of probabilities is child's play.

Notes This is an example of a Markov chain, named after the Russian mathematician. A, Markov (1856-1922) who initiated their study. This procedure is called Poly's urn scheme.

Example 1: A manufacturing firm purchases a certain component, for its manufacturing process, from three sub-contractors A, B and C. These supply 60%, 30% and 10% of the firm's requirements, respectively. It is known that 2%, 5% and 8% of the items supplied by the respective suppliers are defective. On a particular day, a normal shipment arrives from each of the three suppliers and the contents get mixed. A component is chosen at random from the day's shipment:

(a) What is the probability that it is defective?

(b) If this component is found to be defective, what is the probability that it was supplied by (i) A, (ii) B, (iii) C ?

Solution.

Let A be the event that the item is supplied by A. Similarly, B and C denote the events that the item is supplied by B and C respectively. Further, let D be the event that the item is defective. It is given that :

$$P(A) = 0.6, P(B) = 0.3, P(C) = 0.1, P(D/A) = 0.02$$

$$P(D/B) = 0.05, P(D/C) = 0.08.$$

(a) We have to find P(D)

From equation (1), we can write

$$P(D) = P(A \cap D) + P(B \cap D) + P(C \cap D)$$

$$= P(A)P(D/A) + P(B)P(D/B) + P(C)P(D/C)$$

$$= 0.6 \times 0.02 + 0.3 \times 0.05 + 0.1 \times 0.08 = 0.035$$

(b) (i) We have to find P(A/D)

$$P(A/D) = \frac{P(A)P(D/A)}{P(D)} = \frac{0.6 \times 0.02}{0.035} = 0.343$$

Notes

Similarly, (ii) $P(B/D) = \dfrac{P(B)P(D/B)}{P(D)} = \dfrac{0.3 \times 0.05}{0.035} = 0.429$

and (iii) $P(C/D) = \dfrac{P(C)P(D/C)}{P(D)} = \dfrac{0.1 \times 0.08}{0.035} = 0.228$

Alternative Method :

The above problem can also be attempted by writing various probabilities in the form of following table :

	A	B	C	Total
D	$P(A \cap D) = 0.012$	$P(B \cap D) = 0.015$	$P(C \cap D) = 0.008$	0.035
$\bar{D}$	$P(A \cap \bar{D}) = 0.588$	$P(B \cap \bar{D}) = 0.285$	$P(C \cap \bar{D}) = 0.092$	0.965
Total	0.600	0.300	0.100	1.000

Thus $P(A/D) = \dfrac{0.012}{0.035}$ etc.

Example 2: A box contains 4 identical dice out of which three are fair and the fourth is loaded in such a way that the face marked as 5 appears in 60% of the tosses. A die is selected at random from the box and tossed. If it shows 5, what is the probability that it was a loaded die?

Solution.

Let A be the event that a fair die is selected and B be the event that the loaded die is selected from the box.

Then, we have $P(A) = \dfrac{3}{4}$ and $P(B) = \dfrac{1}{4}$.

Further, let D be the event that 5 is obtained on the die, then

$$P(D/A) = \frac{1}{6} \text{ and } P(D/B) = \frac{6}{10}$$

Thus, $P(D) = P(A).P(D/A) + P(B).P(D/B) = \dfrac{3}{4} \times \dfrac{1}{6} + \dfrac{1}{4} \times \dfrac{6}{10} = \dfrac{11}{40}$

We want to find P(B/D), which is given by

$$P(B/D) = \frac{P(B \cap D)}{P(D)} = \frac{1}{4} \times \frac{6}{10} \times \frac{40}{11} = \frac{6}{11}$$

Example 3: A bag contains 6 red and 4 white balls. Another bag contains 3 red and 5 white balls. A fair die is tossed for the selection of bag. If the die shows 1 or 2, the first bag is selected otherwise the second bag is selected. A ball is drawn from the selected bag and is found to be red. What is the probability that the first bag was selected?

Solution.

Notes

Let A be the event that first bag is selected, B be the event that second bag is selected and D be the event of drawing a red ball.

Then, we can write

$$P(A)=\frac{1}{3},\ P(B)=\frac{2}{3},\ P(D/A)=\frac{6}{10},\ P(D/B)=\frac{3}{8}$$

Further, $P(D)=\frac{1}{3}\times\frac{6}{10}+\frac{2}{3}\times\frac{3}{8}=\frac{9}{20}$.

$$\therefore\quad P(A/D)=\frac{P(A\cap D)}{P(D)}=\frac{1}{3}\times\frac{6}{10}\times\frac{20}{9}=\frac{4}{9}$$

Example 4: In a certain recruitment test there are multiple-choice questions. There are 4 possible answers to each questio n out of which only one is correct. An intelligent student knows 90% of the answers while a weak student knows only 20% of the answers.

(i) An intelligent student gets the correct answer, what is the probability that he was guessing?

(ii) A weak student gets the correct answer, what is the probability that he was guessing?

Solution.

Let A be the event that an intelligent student knows the answer, B be the event that the weak student knows the answer and C be the event that the student gets a correct answer.

(i) We have to find $P(\bar{A}/C)$. We can write

$$P(\bar{A}/C)=\frac{P(\bar{A}\cap C)}{P(C)}=\frac{P(\bar{A})P(C/\bar{A})}{P(\bar{A})P(C/\bar{A})+P(A)P(C/A)} \quad \text{.... (1)}$$

It is given that P(A) = 0.90, $P(C/\bar{A})=\frac{1}{4}=0.25$ and $P(C/A)=1.0$

From the above, we can also write $P(\bar{A})=0.10$

Substituting these values, we get

$$P(\bar{A}/C)=\frac{0.10\times0.25}{0.10\times0.25+0.90\times1.0}=\frac{0.025}{0.925}=0.027$$

(ii) We have to find $P(\bar{B}/C)$. Replacing $\bar{A}$ by $\bar{B}$, in equation (1), we can get this probability.

It is given that P(B) = 0.20, $P(C/\bar{B})=0.25$ and $P(C/B)=1.0$

From the above, we can also write $P(\bar{B})=0.80$

Thus, we get $P(\bar{B}/C)=\frac{0.80\times0.25}{0.80\times0.25+0.20\times1.0}=\frac{0.20}{0.40}=0.50$

Notes

Example 5: An electronic manufacturer has two lines A and B assembling identical electronic units. 5% of the units assembled on line A and 10%of those assembled on line B are defective. All defective units must be reworked at a significant increase in cost. During the last eight-hour shift, line A produced 200 units while the line B produced 300 units. One unit is selected at random from the 500 units produced and is found to be defective. What is the probability that it was assembled (i) on line A, (ii) on line B?

Answer the above questions if the selected unit was found to be non-defective.

Solution.

Let A be the event that the unit is assembled on line A, B be the event that it is assembled on line B and D be the event that it is defective.

Thus, we can write

$$P(A) = \frac{2}{5},\ P(B) = \frac{3}{5},\ P(D/A) = \frac{5}{100} \text{ and } P(D/B) = \frac{10}{100}$$

Further, we have

$$P(A \cap D) = \frac{2}{5} \times \frac{5}{100} = \frac{1}{50} \text{ and } P(B \cap D) = \frac{3}{5} \times \frac{10}{100} = \frac{3}{50}$$

The required probabilities are computed form the following table:

	A	B	Total
D	$\frac{1}{50}$	$\frac{3}{50}$	$\frac{4}{50}$
$\bar{D}$	$\frac{19}{50}$	$\frac{27}{50}$	$\frac{46}{50}$
Total	$\frac{20}{50}$	$\frac{30}{50}$	1

From the above table, we can write

$$P(A/D) = \frac{1}{50} \times \frac{50}{4} = \frac{1}{4},\ P(B/D) = \frac{3}{50} \times \frac{50}{4} = \frac{3}{4}$$

$$P(A/\bar{D}) = \frac{19}{50} \times \frac{50}{46} = \frac{19}{46},\ P(B/\bar{D}) = \frac{27}{50} \times \frac{50}{46} = \frac{27}{46}$$

3.3 Independence of Events

From the examples discussed in the previous section you know that the conditional probability P(A 1 H) is, in general, not the same as the unconditional probability P(A). Thus, the knowledge of H affects the chances of occurrence of A. The following example illustrates this fact more explicitly.

Example 18: A box has 4 tickets numbered 1, 2,3 and 4. One of these tickets is drawn at random. Let A = { 1, 2 } be the event that the randomly selected ticket bears the number 1 or 2. Similarly define B = { 1 }. Then

$P(A) = 1/2$, $P(B) = 1/4$ and $P(A \cap B) = 1/4$.

Notes

Therefore, P(B | A) = (1/4) 1 (1/2) = 1/2.

So we have P(B | A) > P(B).

On the other hand, if C= { 1,2,3 } and D = {1,2,4}, then P(C) = P(D) = 3/4 and $P(C \cap D)$ = 112.

Thus,

$P(D \mid C) = \frac{1/2}{3/4} = 2/3$, and in this case,

P(D | C) < P(D)

This example illustrates that additional information (about the occurrence of an event) can increase or decrease the probability of occurrence of another event: We would be interested in those situations which correspond to the cases when P(B | A) = P(B), as in the following example.

Example 19: We continue with the previous example. But now define H = {1,2} and K = {1 , 3}. Then

P(H) = 1/2, P(K) = 1/2 and $P(H \cap K)$ = 114.

Hence

In this example, knowledge of the occurrence of H does not alter the probability of occurrence of K. We call such events, independent events.

Thus, two events A and B are independent, if

$$P(B \mid A) = P(B). \qquad ...(13)$$

However, in this definition, we need to have P(A) > 0. Using the definition of P(B | A), we can rewrite (13) as

$$P(A \cap B) = P(A)\,P(B) \qquad ...(14)$$

which does not require that P(A) or P(B) be positive. We shall now use (14) to define independence of two events.

Definition 4 : Let A and B be two events associated with the same random experiment. They are said to be stochastically independent or simply independent if

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

So the events A and B in Example 18 are not independent. Similarly, events C and D are also not independent. But events K and H in Example 19 are independent.

See if you can apply Definition 4 and solve this exercise.

Task Two unbiased dice are rolled. Let

A_1 be the event "odd face with the first die"

A_2 be the event "odd face with the second die"

B_1 be the event that the score on the first die is 1

B_2 be the event that the total score is at most 3.

Check the independence of the events

(a) A_1 and A_2

(b) B_1 and B_2

Notes

We now proceed to study some implications of independence of two events A_1 and A_2.

Recall that

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

Then

$$P(A_1 \cap A_2^c) = P(A_1) - P(A_1 \cap A_2)$$

Now, if A_1 and A_2 are independent, we get

$$P(A_1 \cap A_2^c) = P(A_1)\{1 - P(A_2)\}$$

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

Thus, the independence of A_1 andA_2 implies that of A_1 and A_2^c. Now interchange the roles of A_1 and A_2 What do you get? We get that if A_1 and A_2 are independent, then so are A_2^c and A_2. The independence of A_1^c and A_2 then implies the independence of A_1^c and A_2^c too.

Now here is an interesting fact.

If A is an almost sure event, then A and another event B are independent.

Let us see how. Since A is an almost sure event, $P(A) = 1$. Hence $P(A^C) = 0$ and therefore, $P(A^C \cap B) = 0$. In particular,

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

One consequence of this is that

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

which implies that A and B are independent.

Can you prove a similar result for a null event ? You can check that if A is a null event, then A and any other event B are independent.

Now, can we extend the definition of independence of two events to that of the independence of three events? The obvious way seems to be to call A_1, A_2, A_3, independent if $P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3)$. But this does not work. Because if 3 events are independgnt, we would expect any two of them also to be independent. But this is not ensured by the condition above. To appreciate this, consider the case when $A_1 = A_2 = A$, $0 < P(A) < 1$, and $P(A_3) = 0$. Then $P(A_1 \cap A_2) = P(A) \neq P(A_1)\, P(A_2) = P[(A)]^2$.

Thus, A_1 and A_2 are not independent, but $P(A_1 \cap A_2 \cap A_3) = P(A_1)\, P(A_2)\, P(A_3)$ is satisfied. So, to get around this problem we add some more conditions and get the following definition

Definition 5 : Three events A_1, A_2 and A_3 corresponding to the same random experiment are said to be stochastically or mutually independent if

$P(A_1 \cap A_2) = p(A_1)\, P(A_2)$

$P(A_2 \cap A_3) = P(A_2)\, P(A_3)$ ...(15)

and $P(A_1 \cap A_2 \cap A_3) = P(A_1)\, P(A_2)\, P(A_3)$.

Let's try to understand this through an example.

Example 20: An unbiased coin is tossed thx& times. Let A, denote the event that a head turns up on the j-th toss, j = 1,2,3. Let's see if A_1, A_2 and A_3 are independent.

Since the coin is unbiased, we assign the same probability, 1/8, to each of the eight possible outcomes.

Check that

$P(A_1) = P(A_2) = P(A_3) = 1/2$

$P(A_1 \cap A_2) = P(A_2 \cap A_3) = P(A_3 \cap A_1) = 1/4$, and

$P(A_1 \cap A_2 \cap A_3) = 1/8$.

Thus, all the four equations in (15) are satisfied and the events A_1, A_2, A_3 are mutually independent.

We have seen that the last condition in (15) alone is not enough, since it does not guarantee the independence of pairs of events.

Similarly, the first three equations of (15) alone are not sufficient to guarantee that all the four conditions required for mutual independence would be satisfied. To see this, consider the following example.

Example 21: An unbiased die is rolled twice. Let A_1 denote the event "odd face on the first roll", A_2 denote the event "odd face on the second roll" and A_3 denote the event that the total score is odd. With the classical assignment of probability 1/36 to each of the sample points, you can easily check that

$P(A_1) = P(A_2) = P(A_3) = 18/36 = 1/2$, and that

$P(A_1 \cap A_2) = P(A_2 \cap A_3) = P(A_3 \cap A_1) = 9/36 = 1/4$.

Thus, the first three equations in (15) are satisfied. But the last one is not valid. The reason for it is that $P(A_1 \cap A_2 \cap A_3)$ is zero (Do you agree ?), and $P(A_1)$, $P(A_2)$, $P(A_3)$ are all positive.

If the first three equationsof (15) are satisfied, we say that A_1, A_2 and A_3 are pairwise independent. Example 21 shows that pairwise independence does not guarantee mutual independence.

Now we are sure you can define the concept of independence of n events. Does your definition agree with Definition 6?

Definition 6 : The n events $A_1, A_2, , \ldots , A_n$ corresponding to the same random experiment are mutually independent if for all r = 2. . . ., n, $1 \le i_j < i_2 < \ldots < i_r \le n$, the product rule holds.

$$P(A_{i_1} \cap \ldots \cap A_{i_r}) \underset{j=1}{\overset{r}{\pi}} P(A_{i_j}) \qquad \ldots(17)$$

Since r of the n events can be chosen in $\binom{n}{r}$ ways, (17) represents

$$\binom{n}{2} + \binom{n}{3} + \ldots + \binom{n}{n} = 2^n - n - 1$$

conditions.

Try to write Definition 6 for n = 3 and see if it matches Definition 5.

Notes

We have already seen that if A, and'A2 are independent, then

A_1^c and A_2 or A_1 and A_2^c or A_1^c and A_2^c are independent. We now give a similar remark about n independent events.

Remark 4 : If $A_1, A_2, \ldots . A_n$ are n independent events, then we may replace some or all of them by their complements without losing independence. In particular, when $A_1, A_2, \ldots, A_n$ are independent, the product rule (17) holds even with some or all of $A_{i_1}, \ldots, A_{i_r}$ are replaced by their complements.

We shall not prove this assertion, but shall use it in the following examples.

Example 22: Suppose A_1, A_2, A_3 are three independent events, with $P(A_j) = P_j$ and we want to obtain the probability that at least one of them occurs.

We want to find $P(A_1 \cup A_2 \cup A_3)$. Recall that (Example 8)

$$\begin{aligned} P(A_1 \cup A_2 \cup A_3) &= P(A_1) + P(A_2) + P(A_3) - P(A_1 \cup A_2) - P(A_2 \cup A_3) - P(A_3 \cup A) + P(A_1 \cup A_2 \cup A_3) \\ &= P_1 + P_2 + P_3 - P_1P_2 - P_2P_3 - P_3P_1 + P_1P_2P_3 \\ &= 1 - (1 - P_1)(1 - P_2)(1 - P3). \end{aligned}$$

We could have amved at this expression more easily by using Remark 4. This is how we can proceed.

$$\begin{aligned} (A_1 \cup A_2 \cup A_3) &= 1 - P((A_1 \cup A_2 \cup A_3)^c) \\ &= 1 - P(A_1^c \cap A_2^c \cap A_3^c) \\ &= 1 - P(A_1^c)P(A_2^c)P(A_3^c) \end{aligned}$$

Example 23: If A_1, A_2 and A_3 are independent events, then can we say that $A_1 \cup A_2$ and A_3 are independent? Let's see.

We have

$$\begin{aligned} P(A_1 \cup A_2) &= P(A_1) + P(A_2) - P(A_1 \cap A_2) \\ &= P(A_1) + P(A_2) - P(A_1)\,P(A_2) \end{aligned}$$

and

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

implying the independence of $A_1 \cup A_2$ and A_3.

Example 24: An automatic machine produces bolts. Each bolt has probability 1/10 of being defective. Assuming that a bolt is defective independently of all other bolts, let's find

(i) the probability that a good bolt is followed by two defective ones.

(ii) the probability of getting one good and two defective bolts, not necessarily in that order.

Let A_j denote the event that the j-th inspected bolt is defective, j = 1,2,3. The assumption of independence implies that A_1, A_2 and A_3 are independent.

Notes

(i) We want $P(A_1^c \cap A_2 \cap A_3)$. By Remark 4, we can write

$$P(A_1^c \cap A_2 \cap A_3) = P(A_1^c)P(A_2)P(A_3)$$

$$= \frac{9}{10}.\frac{1}{10}.\frac{1}{10} = 0.009.$$

(ii) We want to find,the probability of

$$(A_1^c \cap A_2 \cap A_3) \cup (A_1 \cap A_2^c \cap A_3) \cup (A_1 \ A_2 \ A_3^c).$$

Notice that these events are disjoint and that each has the probability 0.009 (see (i)). Hence, the required probability is

Example 25: The probability that a person A will be alive 20 years hence is 0.7 and the probability that another person B will be alive 20 years hence is 0.5. Assuming independence, let's find the probability that neither of them will be alive after 20 years.

The probability that A dies before twenty years have elapsed is 0.3 and the corresponding probability for B is 0.5. Hence the probability that neither of them will be alive 20 years hence is

$$0.3 \times 0.5 = 0.15,$$

by virtue of independence.

We now give you some exercises based on the concept of independence.

3.4 Repeated Experiments and Trials

We must mention that we have earlier discussed rolls of two dice or three or more tosses of a coin without bringingin the concept of repeated trials. The following discussion is only an elementary introduction to the topic of repeated trials.

To fix ideas, consider the simple experiment of tossing a coin twice. The sample space corresponding to the first toss is S_1 = {H, T} say, where H = Head, T = Tail. Similarly the sample space S_2 for the second toss is also {H, T}. Now observe that the sample space for two tosses is Ω = {(H, H), (H, T), (T, H), (T, T)}, where (H, H) stands for head on first toss followed by a head on the second toss. The pairs (H, T), etc. are also similarly defined. Note that Ω consists of all ordered pairs that can be formed by choosing a point from S_1 followed by a point from S_2. Mathematically we say that Ω is the Cartesian product $S_1 \times S_2$ (read, S_1 cross S_2) of S_1 and S_2.

Now consider an experiment of tossing a coin and then rolling a die. The sample space corresponding to toss of the coin is S_1 = (H, T) and that corresponding to the roll of the die is S_2 = (1, 2, 3, 4, 5, 6). The sample space of the combined experiment is

W = {(H, 1), (H, 2), (H, 3), (H, 4), (H, 5), (H, 6),

(T, 1), (T, 2), (T, 3), (T, 4), (T, 5), (T, 6) = $S_1 \times S_2$.

Taking a cue from these two examples we can say that if S_1 and S_2 are the sample spaces for two random experiments $\in_1$ and $\in_2$ then the Cartesian product $S_1 \times S_2$ is the sample space of the experiment consistihg of both $\in_1$ and $\in_2$.

Sometimes we refer to $S_1 \times S_2$ as the product space of the two experiments.

Notes

We are sure that you will be able to do this simple exercise.

> *Task* Find the sample spaces of the following experiments
>
> (a) Rolling two dice
>
> (b) Drawing two cards from a pack of 52 playing cards, with replacement.

Do you remember the definition of the Cartesian product of $n(n \geq 3)$ sets? We say that the Cartesian product

$$S_1 \times S_2 \times ... \times S_n = \{(x_1 ...., x_n) \mid x_j \in S_j, j = 1 ,... n\}.$$

Now, if $S_1, S_2, ... S_n$ represent the sample spaces corresponding to repetitions $\in_1, \in_2, . . . , \in_n$ of the same experiment E, then the Cartesian product $S_1 \times S_2 \times . . . \times S_n$ represents the sample space for n repetitions or n trials of the experiment $\in$.

[(H, T), (T, H), (T, T)] which is the Cartesian product of {H, T} with itself. Suppose the Probability on a Discrete Sample coin is unbiased so that P{H} = P{T} = ID for both the first and the second toss. Since the coin is unbiased, we may regard the four points in Ω as equally likely and assign probability 1/4 to each one of them. However, another way of looking at this assignment is to assume that the results in the two tosses are independent. Mor specifically, we may consider specifying P{(H, H)} say, by the multiplication rul Jav ailable to us under independence, i.e., we may take

$$P\{H, H\} = P\{H\} . P\{H\} = \frac{1}{2}.\frac{1}{2} = \frac{1}{4}.$$

and make similar calculations for other pomts.

When such a situation holds, we say that the two tosses or the two trials of tossing the coin are independent. This is equivalent to saying that the events Head on first toss and Head on second toss are independent and that we may make similar statements about the other points also. The following example illustrates the method of defining probabilities on the product spaces when we are unable (or unwilling) to assume equally likely outcomes.

Example 26: Suppqe this successive units manufactured by a machine are such that each unit has probability p of being defective (D) and (1 – p) of being good (G). We examine three units manufactured by this machine. The sample space for this experiment is the Cartesian product $S_1 \times S_2 \times S_3$, where $S_1 = S_2 = S_3 = \{D, G\}$, i.e.,

Ω = [(D, D, D), (D, D, G), (D, G, D), (G, D, D), (G, G, D), (G, D, G), (D, G, G), (G, G, G)].

The statement that "the successive units are independent of each other'' is interpreted by assigning probabilities to points of Ω by the product rule. In particular,

$$P\{(D, D, D)\} = P(D)\ P(D)\ P(D) = p^3,$$

$$P\{(D, D, G)\} = P(D)\ P(D)\ P(G) = p2q$$

$$= P\{(D, G, D)\} = P\{(G, D, D)\},$$

$$P\{(G, G, D)\} = P(G)\ P(G)\ P(D) = (1 - p)^2 p$$

and lastly,

$$P((G, G, G)) = P(G)\ P(G)\ P(G) = (1 - p)^3.$$

Notice that the sum of the probabilities of the eight points in Ω is

which is as it should be.

Summarising the discussion so far, consider two random experiments $\in_1$ and $\in_2$ with sample spaces S_1 and S_2, respectively. Let u_1, u_2 . . . be the points of S_1 and let v_1, v_2 . . ., be the points of S_2. Suppose p_1, p_2, . . ,a nd q_l, q_2,. . . are the associated probabilities, i.e., $P(u_i) = p_i$ and $P(v_j) = q_j$, with $p_i, q_j \geq 0 \ \sum_i p_i = 1, \sum_j q_j = 1.$ We say that $\in_1$ and $\in_2$ are independent experiments if the events "first outcome is u_i" and the event "second outcome is v_j", are independent,

i.e., if the assignment of probabilities on the product space S1 **x** S2 is such that

$$P\{(u_j, v_j)\} = P(u_i)\, P(v_j) = p_i q_j$$

This assignment is a valid assignment because $P((u_i, v_j)) \geq 0$ and

$$\sum_i \sum_j P(u_i, v_j) = \sum_i \sum_j p_i q_i$$

$$= \sum_i p_i \sum_j q_j = 1.$$

where the sums are taken over all values of i and j.

Can we extend these concepts to the case of n (n > 2) random experiments?

Let us denote the n random experiments by $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$. Let S_1. S_2, . . . , S_n be the corresponding sample spaces. Let $P(x_j)$ denote the probability assigned to the outcome x_j of the random experiment ε_j. We say that $\varepsilon_1 \ldots, \varepsilon_n$ are independent experiments, if the assignment of probabilities on the product space $S_1 \times S_2 \times \ldots \times S_n$ is such that

$$P\{(x_1, x_2, \ldots, x_n)\} = P(x_1)\, P(x_2) \ldots P(x_n).$$

The random experiments $\varepsilon_1 \ldots, \varepsilon_n$ are said to be repeated independent trials of an experiment ε if the sample space of $\varepsilon_1, \ldots \varepsilon_n$ are all identical and so are the assignment of probabilities, it is in this sense that the experiment discussed in Example 26 corresponds to 3 independent repetitions of the experiment of inspecting a unit, where the probability of a unit being defective is P.

Before we conclude ow discussion of product spaces and repeated aials, let us revert to the case of two independent experiments ε_1 and ε_2 with sample spaces S_1 and S_2

Suppose

$S_1 = (u_1, u_2, \ldots)$, $P(u_j) = p_i$, $i \geq 1$

$S_2 = (v_1, v_2, \ldots)$, $P(v_j) = q_j$, $j \geq 1$

Let $A_1 = (u_{i_1}, u_{i_2}, \ldots.)$ and $A_2 = (v_{j_1}, v_{j_2}, \ldots)$ be two events in S_1 and S_2. Then $A_1 \times A_2$ is event in $S_1 \times S_2$ and

$$A_1 \times A_2 = \{(u_{i_r}, v_{j_s}) \mid r, s = 1, 2, \ldots\}.$$

Under the assumption that E~ and E~ are independent, we can write

$$P(A1 \times A2) = \sum_r \sum_s P\{(u_{i_r}, v_{j_s})\}$$

$$= \sum p_i\, q_j$$

Notes

$$= \sum_r p_{i_r} \sum_s q_{r_s}$$

$$= P(A_1) \times P(A_2).$$

Thus, the multiplication rule 'is valid not only for individual sample points of $S_1 \times S_2$ but also for events in the component sample spaces S_1 and S_2 also. Here we have talked about events related to two experiments. But we can easily extend this fact to events related to three or more experiments.

The independent Bernoulli trials provide the simplest example of repeated independent trials. Here each trial has only two possible outcomes, usually called success (S) and failure (F). We further assume that the probability of success is the same in each trial, and therefore, the the probability of failure is also the same for each trial. Usually we denote the probability of success by p and that of failure by q = 1 – p.

Suppose, we consider three independent Bernoulli trials. The sample space is the Cartesian product (S, F) × (S, F) × (S, F). It, therefore, consists of the eight points

SSS, SSF, SFS, FSS, FFS, FSF, SFF, FFF.

In view of independence, the corresponding probabilities are

$p^3, p^2q, p^2q, p^2q, pq^2, pq^2, pq^2, q^3$.

Do they add up to one? Yes.

In general, the sample space corresponding to n independent Bernoulli trials consists of 2^n points. **A** generic point in this sample space consists of the sequence of n letters, j of which are S and n – j are F, j = 0.1, . . ., n. Each such point carries the probability $p^i q^{n-j}$, probability of successes in n independent Bernoulli trials. We first note that there are $\binom{n}{j}$ points with j successes and (n – j) failures (we ask you to prove this in E27). Since each such point carries the probability $p^i q^{n-j}$, the probability of j successes, denoted by b(j, n, p) is

$$b(j, n, p) = \binom{n}{j} p^j q^{n-j}, 0, 1, ...n.$$

These are called binomial probabilities and we shall return to a discussion of this topic when we discuss the binomial distribution in Unit 8.

Task Prove that there are $\binom{n}{j}$ pints with j successes and (n – j) failures in n Bernoulli trials.

Now we bring this unit to a closk. But before that let's briefly recall the important concepts that we studied in it.

3.5 Summary

- We have acquainted you with the concept of conditional probability P(A | B) of a given the event B.

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}, P(B) > 0$$

Notes

- We have stated and proved Bayes' theorem :

  If $B_1, B_2, \ldots, B_n$, are n events which constitute a partition of Ω, and A is an event of positive probability, then

$$P(B_r | A) = \frac{P(B_r)P(A | B_r)}{\sum_{1}^{n} P(B_j)P(AIB_j)}$$

  for any r, $1 \le r \le n$.

- We have defined and discussed the consequences of independence of two or more events.
- We have seen the method of assignment of probabilities when dealing with independent repetitions of an experiment.

3.6 Keywords

Conditional probability: The concept of conditional probability P(A | B) of a given the event B.

$$P(A | B) = \frac{P(A \cap B)}{P(B)}, P(B) > 0$$

Bayes' theorem: If $B_1, B_2, \ldots, B_n$, are n events which constitute a partition of W, and A is an event of positive probability, then

$$P(B_r | A) = \frac{P(B_r)P(A | B_r)}{\sum_{1}^{n} P(B_j)P(AIB_j)}$$

for any r, $1 \le r \le n$.

3.7 Self Assesment

1. There are 1000 people. There 400 females and 200 colour-blind person. Find the probability of females.

   (a) 0.4 (b) 0.6

   (c) 0.16 (d) 0.21

2. There are $\begin{pmatrix} n \\ j \end{pmatrix}$ points with j successes and (n – j) failures. Since each such point carries the probability $p^i q^{n-j}$, the probability of j successes, denoted by b(j, n, p) is $b(j, n, p) = \begin{pmatrix} n \\ j \end{pmatrix} p^j q^{n-j}, 0, 1, \ldots n.$

   (a) Binomial probabilities (b) Conditional probability

   (c) Bayer's theorem (d) Clanical probability

   These are called binomial probabilities.

3. the concept of conditional probability P(A | B) of a given the event B.

$$P(A | B) = \frac{P(A \cap B)}{P(B)}, P(B) > 0$$

Notes

(a) Binomial probabilities (b) Conditional probability

(c) Bayer's theorem (d) Clanical probability

4. If $B_1, B_2, \ldots, B_n$, are n events which constitute a partition of Ω, and A is an event of positive probability, then

$$P(B_r \mid A) = \frac{P(B_r)P(A \mid B_r)}{\sum_{1}^{n} P(B_j)P(AIB_j)}$$

for any r, 1 ≤ r ≤ n.

(a) Binomial probabilities (b) Conditional probability

(c) Bayer's theorem (d) Clanical probability

5. There are 1000 people. There 400 females and 600 males then find the probability of males.

(a) 0.4 (b) 0.6

(c) 0.16 (d) 2.01

3.8 Review Questions

1. In a city the weather changes frequently. It is known from past experience that a rainy day is followed by a sunny day with probability 0.4 and that sa sunny day is followed by a rainy day with probability 0.7. Assume that the weather on any given day depends only on the weather of the previous day. Find the probability that
   (a) a rainy day is followed by a rainy day
   (b) it would rain on Saturday and Sunday when Friday was rainy
   (c) the entire period from Monday to Friday is rainy given, that the previous Sunday was sunny.
2. An urn contains 4 white and 4 black balls. A ball is drawn at random, its colour is noted and is returned to the urn. Moreover, 2 additional balls of the colour drawn are put in the urn and then a ball is drawn at random. What is the probability that the second ball is black?
3. In a community 2 per cent of the people suffer from cancer. The probability that a doctor is able to correctly diagnose a person with cancer as suffering from cancer is 0.80. The doctor wrongly diagnoses a person without cancer as having cancer with probability 0.05. What is the probability that a randomly selected person diagnosed as having cancer is really suffering from cancer?
4. An explosion in a factory manufacturing explosives can occur because of (i) leakage of electricity, (ii) defects in machinery, (iii) carelessness of worker or (iv) sabotage. The probability that
   (a) there is a leakage of electricity is 0.20
   (b) the machinery is defective is 0.30
   (c) the workers are careless is 0.40
   (d) there is sabotage is 0.10

Answers: Self Assessment

1. (a) 2. (a) 3. (b) 4. (c) 5. (b)

3.9 Further Readings

Books

Introductory Probability and Statistical Applications by P.L. Meyer

Introduction to Mathematical Statistics by Hogg and Craig

Fundamentals of Mathematical Statistics by S.C. Gupta and V.K. Kapoor

Notes

Unit 4: Random Variable

CONTENTS

Objectives

Introduction

4.1 Definition of a Random Variable

4.2 Probability Distribution of a Random Variable

4.2.1 Discrete and Continuous Probability Distributions

4.2.2 Cumulative Probability Function or Distribution Function

4.3 Summary

4.4 Keywords

4.5 Self Assessment

4.6 Review Questions

4.7 Further Readings

Objectives

After studying this unit, you will be able to:

- Define random variables
- Define probability distributions of random variable

Introduction

In order to discuss the applications of probability to practical situations, it is necessary to associate some numerical characteristics with each possible outcome of the random experiment. This numerical characteristic is termed as random variable.

4.1 Definition of a Random Variable

A random variable X is a real valued function of the elements of sample space S, i.e., different values of the random variable are obtained by associating a real number with each element of the sample space. A random variable is also known as a stochastic or chance variable.

Mathematically, we can write $X = F(e)$, where e ÎS and X is a real number. We can note here that the domain of this function is the set S and the range is a set or subset of real numbers.

Example 1: Three coins are tossed simultaneously. Write down the sample space of the random experiment. What are the possible values of the random variable X, if it denotes the number of heads obtained?

Solution.

Notes

The sample space of the experiment can be written as

S = {(H,H,H), (H,H,T), (H,T,H), (T,H,H), (H,T,T), (T,H,T), (T,T,H), (T,T,T)}

We note that the first element of the sample space denotes 3 heads, therefore, the corresponding value of the random variable will be 3. Similarly, the value of the random variable corresponding to each of the second, third and fourth element will be 2 and it will be 1 for each of the fifth, sixth and seventh element and 0 for the last element. Thus, the random variable X, defined above can take four possible values, i.e., 0, 1, 2 and 3.

It may be pointed out here that it is possible to define another random variable on the above sample space.

4.2 Probability Distribution of a Random Variable

Given any random variable, corresponding to a sample space, it is possible to associate probabilities to each of its possible values. For example, in the toss of 3 coins, assuming that they are unbiased, the probabilities of various values of the random variable X, defined in example 1 above, can be written as :

$$P(X=0)=\frac{1}{8},\ P(X=1)=\frac{3}{8},\ P(X=2)=\frac{3}{8} \text{ and } P(X=3)=\frac{1}{8}.$$

The set of all possible values of the random variable X along with their respective probabilities is termed as Probability Distribution of X. The probability distribution of X, defined in example 1 above, can be written in a tabular form as given below :

X	:	0	1	2	3	*Total*
$p(X)$	:	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$	1

Note that the total probability is equal to unity.

In general, the set of n possible values of a random variable X, i.e., $\{X_1, X_2, ... X_n\}$ along with their respective probabilities $p(X_1)$, $p(X_2)$, ... $p(X_n)$, where $\sum_{i=1}^{n} p(X_i) = 1$, is called a probability distribution of X. The expression p(X) is called the probability function of X.

4.2.1 Discrete and Continuous Probability Distributions

Like any other variable, a random variable X can be discrete or continuous. If X can take only finite or countably infinite set of values, it is termed as a discrete random variable. On the other hand, if X can take an uncountable set of infinite values, it is called a continuous random variable.

The random variable defined in example 1 is a discrete random variable. However, if X denotes the measurement of heights of persons or the time interval of arrival of a specified number of calls at a telephone desk, etc., it would be termed as a continuous random variable.

The distribution of a discrete random variable is called the Discrete Probability Distribution and the corresponding probability function p(X) is called a Probability Mass Function. In order that any discrete function p(X) may serve as probability function of a discrete random variable X, the following conditions must be satisfied:

Notes

(i) $p(X_i) \geq 0 \; \forall \; i = 1, 2, \ldots\ldots n$ and

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

In a similar way, the distribution of a continuous random variable is called a Continuous Probability Distribution and the corresponding probability function p(X) is termed as the Probability Density Function. The conditions for any function of a continuous variable to serve as a probability density function are:

(i) $p(X) \geq 0 \; \forall$ real values of X, and

(ii) $\int_{-\infty}^{\infty} p(X)dX = 1$

Remarks:

1. When X is a continuous random variable, there are an infinite number of points in the sample space and thus, the probability that X takes a particular value is always defined to be zero even though the event is not regarded as impossible. Hence, we always measure the probability of a continuous random variable lying in an interval.

2. The concept of a probability distribution is not new. In fact it is another way of representing a frequency distribution. Using statistical definition, we can treat the relative frequencies of various values of the random variable as the probabilities.

Example 2: Two unbiased die are thrown. Let the random variable X denote the sum of points obtained. Construct the probability distribution of X.

Solution.

The possible values of the random variable are:

2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12

The probabilities of various values of X are shown in the following table:

Probability Distribution of X

X	2	3	4	5	6	7	8	9	10	11	12	*Total*
$p(X)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	1

Example 3: Three marbles are drawn at random from a bag containing 4 red and 2 white marbles. If the random variable X denotes the number of red marbles drawn, construct the probability distribution of X.

Solution.

The given random variable can take 3 possible values, i.e., 1, 2 and 3. Thus, we can compute the probabilities of various values of the random variable as given below:

P(X = 1, i.e., 1R and 2 W marbles are drawn) $= \frac{{}^4C_1 \times {}^2C_2}{{}^6C_3} = \frac{4}{20}$

P(X = 2, i.e., 2R and 1W marbles are drawn) $= \frac{{}^4C_2 \times {}^2C_1}{{}^6C_3} = \frac{12}{20}$

Notes

P(X = 3, i.e., 3R marbles are drawn) $= \frac{{}^{4}C_{3}}{{}^{6}C_{3}} = \frac{4}{20}$

> *Notes* In the event of white balls being greater than 2, the possible values of the random variable would have been 0, 1, 2 and 3.

4.2.2 Cumulative Probability Function or Distribution Function

This concept is similar to the concept of cumulative frequency. The distribution function is denoted by F(x).

For a discrete random variable X, the distribution function or the cumulative probability function is given by F(x) = P(X £ x).

If X is a random variable that can take values, say 0, 1, 2, ..., then

F(1) = P(X = 0) + P(X =1), F(2) = P(X = 0) + P(X =1) +P(X = 2), etc.

Similarly, if X is a continuous random variable, the distribution function or cumulative probability density function is given by

$$F(x) = P(X \le x) = \int_{-\infty}^{x} p(X)dX$$

4.3 Summary

- A random variable X is a real valued function of the elements of sample space S, i.e., different values of the random variable are obtained by associating a real number with each element of the sample space. A random variable is also known as a stochastic or chance variable.

  Mathematically, we can write X = *F*(e), where e ÎS and X is a real number. We can note here that the domain of this function is the set S and the range is a set or subset of real numbers.

- The random variable defined in example 1 is a discrete random variable. However, if X denotes the measurement of heights of persons or the time interval of arrival of a specified number of calls at a telephone desk, etc., it would be termed as a continuous random variable.

- When X is a continuous random variable, there are an infinite number of points in the sample space and thus, the probability that X takes a particular value is always defined to be zero even though the event is not regarded as impossible. Hence, we always measure the probability of a continuous random variable lying in an interval.

- The concept of a probability distribution is not new. In fact it is another way of representing a frequency distribution. Using statistical definition, we can treat the relative frequencies of various values of the random variable as the probabilities.

4.4 Keywords

Random variable: A random variable X is a real valued function of the elements of sample space S, i.e., different values of the random variable are obtained by associating a real number with each element of the sample space.

Notes

Discrete Probability Distribution: The distribution of a discrete random variable is called the Discrete Probability Distribution.

Continuous Probability Distribution: The distribution of a continuous random variable is called a Continuous Probability Distribution and the corresponding probability function p(X) is termed as the Probability Density Function.

4.5 Self Assessment

State whether the following statements are true or false:

(a) A random variable takes a value corresponding to every element of the sample space.

(b) The probability of a given value of the discrete random variable is obtained by its probability density function.

(c) Distribution function is another name of cumulative probability function.

(d) Any function of a random variable is also a random variable.

4.6 Review Questions

1. An urn contains 4 white and 3 black balls. 3 balls are drawn at random. Write down the probability distribution of the number of white balls. Find mean and variance of the distribution.

2. A consignment is offered to two firms A and B for Rs 50,000. The following table shows the probability at which the firm will be able to sell it at different prices :

SellingPrice(in Rs)	40,000	45,000	55,000	70,000
Prob. of A	0.3	0.4	0.2	0.1
Prob. of B	0.1	0.2	0.4	03

Which of the two firms will be more inclined towards the offer?

3. If the probability that the value of a certain stock will remain same is 0.46, the probabilities that its value will increase by Re. 0.50 or Re. 1.00 per share are respectively 0.17 and 0.23 and the probability that its value will decrease by Re. 0.25 per share is 0.14, what is the expected gain per share?

4. In a college fete a stall is run where on buying a ticket a person is allowed one throw of two dice. If this gives a double six, 10 times the ticket money is refunded and in other cases nothing is refunded. Will it be profitable to run such a stall? What is the expectation of the player? State clearly the assumptions if any, for your answer.

5. The proprietor of a food stall has introduced a new item of food. The cost of making it is Rs 4 per piece and because of its novelty, it would be sold for Rs 8 per piece. It is, however, perishable and pieces remaining unsold at the end of the day are a dead loss. He expects the daily demand to be variable and has drawn up the following probability distribution expressing his estimates:

No. of pieces demanded	:	50	51	52	53	54	55
Probability	:	0.05	0.07	0.20	0.35	0.25	0.08

Compute his expected profit or loss if he prepares 53 pieces on a particular day.

Answers: Self Assessment

1. T 2. F 3. T 4. T

4.7 Further Readings

Notes

Books

Introductory Probability and Statistical Applications by P.L. Meyer

Introduction to Mathematical Statistics by Hogg and Craig

Fundamentals of Mathematical Statistics by S.C. Gupta and V.K. Kapoor

Notes

Unit 5: Functions of Random Variables

CONTENTS

Objectives

Introduction

5.1 Function of Two Random Variables

5.1.1 Direct Method

5.1.2 Transformation Approach

5.2 Functions of More than Two Random Variables

5.3 Summary

5.4 Keywords

5.5 Self Assessment

5.6 Review Questions

5.7 Further Readings

Objectives

After studying this unit, you will be able to:

- Discuss the functions of random variables
- Describe the transformation approach

Introduction

In the last unit, we have introduced bivariate distributions and multivariate distributions. Most of the times we would like to know the probabilistic behaviour of a function g(X,Y) of the random vector (X,Y). The function g could be either the sum X+Y or the max (X,Y) or some other function depending on the phenomenon under study. We give two approaches for obtaining the distribution function of a function of two random variables. These distributions can be considered as distributions of functions of independent standard normal random variables. Properties of these distributions are investigated in detail.

5.1 Function of Two Random Variables

We shall talk about functions of two random variables and discuss methods for obtaining their distribution functions. Some of the important functions X which we shall consider are X + Y, XY, $\frac{X}{Y}$, max (X,Y), |X-Y|.

Let us start with a random vector (X,Y).B y definition X and Y are random variables defined on the sample space S of some experiment and each of which assigns a real number to every $s \in S$. Let g(x,y) be a real-valued function defined on R x R. Then the composite function Z = g(X,Y) defined by

$$Z(s) = g\,[X(s), Y(s)],\ s \in S$$

assignes to every outcome $s \in S$ a real number. Z is called a function of the random vector (X,Y).

For example, if g(x, y) = x + y, then we get Z = X + Y and if g(x,y) =xy, then we get Z = XY and so on.

Now let us see how do we find the distribution function of Z. As in the univariate case, we shall restrict ourselves to the continuous case. Here we shall discuss two methods for obtaining distribution functions - Direct Method and Transformation Method. We shall first discuss Direct Method.

5.1.1 Direct Method

Let (X,Y) be a random vector with the joint density function fx, y (x,y). Let g(x,y) be a real-valued function defined on R x R. For z E 2, define

$$D_z = \{(x, Y) : g(x, y) \le z\}$$

Then the distribution function of Z is defined as

$$P\,[Z \le z] = \iint_{D_z} f_{X,Y}\ (X,Y)\,dx\,dy \qquad ...(1)$$

Theoretically it is not difficult to write down the distribution function using (1). But in actual practise it is sometimes difficult to evaluate the double integral.

We shall now illustrate the computation of distribution functions in the following examples.

Example 1: Suppose (X,Y) has the uniform distribution on [0,1] × [0,1] the unit square. Then the joint density of (X,Y) is

$$fX, y(x, y) = \begin{cases} 1 & \text{if } 0 < x, y < 1 \\ 0 & \text{otherwise} \end{cases}$$

Let us find the distribution function of Z = g(X,Y) = XY.

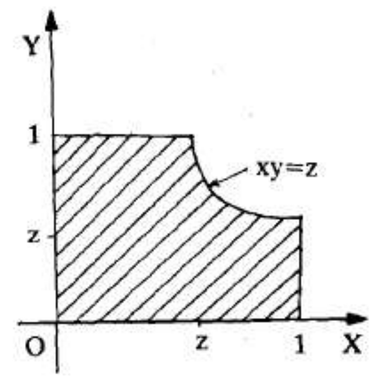
From the definition of a distrihution function of Z, we have

Fz (z) = P [XY Z]

$$= \iint_{D_z} fX, Y(x, y)dxdy \quad \text{if } 0 < z < 1$$

Notes

Figure 5.1



where Dz = {(x,y) : xy ≤ z}

$$= \int\int_{D_z^1} dxdy \quad \text{if } 0 < z < 1.$$

where D_z^1 = {(x,y) : xy ≤ z, 0 < x < 1,0 < y < 1}

In order to evaluate the last integral, let us look at the set of all points (x,y) such that x y ≤ z when 0 < x < l and 0 < y < l (See Fig. 1).

If 0 < x < z, then for any 0 < y < 1, the product xy ≤ z and if x > z, then xy ≤ z only when 0 < y < z/x. This is the region shaded in Fig. 1. Hence for 0 < z < 1

$$Fz(z) = P[Z \le z]$$

$$= \int_0^z \left\{ \int_0^1 dy \right\} dx + \int_z^1 \left\{ \int_0^{z/x} dx \right\} dx$$

$$= \int_0^z dx + \int_z^1 \frac{z}{x} dx$$

$$= z + z\,[\ln x]_z^1 = z - z \ln z.$$

Therefore

$$F_z(z) = \begin{cases} 0 & \text{if } z = 0 \\ z - z\ln z & \text{if } 0 < z < 1 \\ 1 & \text{if } z \ge 1 \end{cases}$$

is the distribution function of Z. The density function fz (z) of Z is obtained by differentiating Fz (z) with respect to z. Then you can check that

$$f_z(z) = 0 \quad \text{if } z \le 0 \text{ or } z \ge 1$$

$$= -\ln z \quad \text{if } 0 < z < 1$$

Notes

Example 2: Suppose X and Y are independent exponential random variables with the density function

$$f(x) = \lambda e^{-\lambda x},\ x > 0 \text{ and } f(y) = \lambda e^{-\lambda y},\ y > 0$$
$$= 0,\quad x \le 0. \qquad = 0,\ y \le 0$$

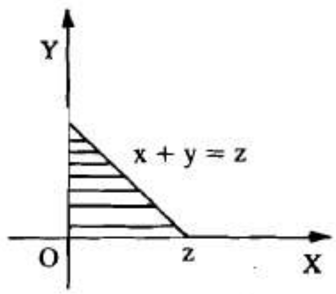
Define Z = X + Y and let us find the distribution function of Z.

From the definition of Z,

$$Fz(z) = P[Z \le z] = 0 \text{ if } z < 0$$

and, for z > 0

Figure 5.2



$$Fz(z) = P[Z \le z]$$

$$= \iint_{\{(x,y) : x + y \le z\}} f_{x,y}(x,y)\, dx\, dy$$

where $f_{x,y}(x, y)$ is the joint density of (X,Y). Since X and Y are independent random variables, the joint density of (X,Y) is given by

$$f_{x,y}(x, y) = f_x(x)\, f_y(Y)$$

where fx (x)and fy (y) are the marginal density functions

i.e. $$f_{x,y}(x, y) = \lambda e^{-\lambda x} \times \lambda e^{-\lambda y},\ x > 0,\ y > 0$$
$$= 0,\ \text{otherwise}$$

Now for z > 0, the set { (x, y) : x + y ≤ z, x > 0, y > 0} is the region shaded in Fig. 2.

Hence, for z > 0,

$$Fz(z) = \int_0^z \left[\int_0^{z-x} \lambda e^{-\lambda y}\, dy \right] l e^{-\lambda x}\, dx$$

$$= \int_0^z \left[-e^{-\lambda x} \right]_0^{z-x} \lambda e^{-\lambda x} dx$$

$$= \int_0^z \left[1 - e^{-\lambda(z-x)} \right] \lambda e^{-\lambda x} dx$$

Notes

$$= \int_0^z \left[\lambda e^{-\lambda x} - \lambda e^{-\lambda x}\right] dx$$

$$= \left[-e^{-\lambda x}\right]_0^z - z\lambda e^{-\lambda z}$$

$$= 1 - e^{-\lambda z} - \lambda z e^{-\lambda z}$$

Now we leave it to you to check that the density function of Z is

$$f_z(z) = \lambda^2 z\, e^{-\lambda z} \text{ for } z > 0$$

$$= 0 \text{ otherwise}$$

In this density function familiar to you? Recall that this function is the gamma density function you have studied in Unit 11. Hence Example 2 says that the sum of two independent exponential random variables has gamma distribution.

Let us consider another example.

Example 3: Suppose X and Y are independent random varihles with the same density function I (x) and the distribution function F(x). Define Z = max(X,Y). Let us determine the distribution function of *2.*

By definition, the distribution function F_z is given by

$$F_z(z) = P[Z \le z]$$

$$= P[\max(X, Y) \le z]$$

$$= P[X \le z, Y \le z]$$

$$= P[X \le z]\, P[Y \le Z] = [F(z)]^2$$

by the independence of X and Y and the fact that

$$P[X \le z] = P[Y \le z] = F(z).$$

Since F is diflerentiable almost everywhere and the density corresponding to F is fit follows that Z has a probability density function fz and

$$fz(z) = 2F(z)\, f(z), -\infty < z < \infty.$$

To get more practise why don't you try some exercises now.

The examples and exercises discussed above deal with the method of obtaining the distribution function of Z= g(X,Y) directly. This method is applicable even when (X,Y) does not have a density function.

Next we shall discuss another method for obtaining the distribution and density functions.

5.1.2 Transformation Approach

Suppose (X_1, X_2) is a bivariate random vector with the density function $fx_1, x_2\,(x_1, x_2)$ and we would like to determine the distribution function of the density function of $Z_1 = g_1(X_1, X_2)$. To determine this, let us suppose that we can find another function $Z_2 = g_2(X_1, X_2)$ such that the transformation from (X_1, X_2) to (Z_1, Z_2) is one-to-one. In other words to every point (x_1, x_2) in R^2,

Notes

there corresponds a point (x_1, x_2) in R^2 given by the above transformation and conversely to every point (z_1, z_2) there corresponds a unique point (x_1, x_2) such that

$$z_1 = g_1(x_1, x_2)$$
$$z_2 = g_2(x_1, x_2)$$

For example suppose that $Z_1 = X_1 + X_2$. Then we can choose $Z_2 = X_1 - X_2$. You can easily see that the transformation $(x_1, x_2) \to (z_1, z_2)$ from R^2 to R^2 is one-one and in this case we have

$$X_1 = \frac{Z_1 - Z_2}{2} \text{ and } X_2 = \frac{Z_1 + Z_2}{2}$$

So, in general, one can assume that we can express (X_1, X_2) in terms of (Z_1, Z_2) uniquely.

That means that there is exist real valued functions h_1 and h_2 such that

$$X_1 = h_1(Z_1, Z_2)$$
$$X_2 = h_2(Z_1, Z_2)$$

Let us further assume that h_1 and h_2 have continuous partial derivatives with respect to Z_1, Z_2. Consider the Jacobian of the tranformation $(Z_1, Z_2) \to (X_1, X_2)$

$$\begin{vmatrix} \frac{\partial h_1}{\partial z_1} & \frac{\partial h_2}{\partial z_2} \\ \frac{\partial h_2}{\partial z_2} & \frac{\partial h_2}{\partial z_2} \end{vmatrix} = \frac{\partial h_1}{\partial z_1}\frac{\partial h_2}{\partial z_2} - \frac{\partial h_2}{\partial z_2}\frac{\partial h_2}{\partial z_2}$$

Recall that you. have seen 'Jacobians' in Unit 9, Block 3 of MTE - 07 we denote this Jacobian by $J = \frac{\partial(x_1, x_2)}{\partial(z_1, z_2)}$. Assume that J is not zero for all (z_1, z_2). Then, it can be shown, by using the change of variable formula for double integrals [see MTE-07, Unit 11, Block 41 we can show that the random vector (Z_1, Z_2) has a density and the density function $\phi(z_1, z_2)$ of (Z_1, Z_2) is

$$\phi(z_1, z_2) = f[h_1(z_1, z_2), h_2(z_1, z_2)] \mid J \mid \text{ if } (z_1, z_2) \text{ E } B \quad ...(2)$$
$$= 0 \qquad \text{otherwise}$$

where $B = \{ (z_1, z_2) : z_1 = g_1(x_1, x_2), z_2 = g(x_1, x_2) \text{ for some } (x_1, x_2) \}$.

From the joint density of (Z_1, Z_2) obtained above, the marginal density of Z_1 can be derived and it is given by

$$\phi_2(z_1) = \int_{-\infty}^{\infty} f(z_1, z_2)\, dz_2$$

Let us now compute the density function of $Z_1 = X_1 + X_2$ where X_1 and X_2 are independent and identically distributed standard normal variables. We have seen that in this cask $Z_2 - X_1 + X_2$ and we can write

$X_1 = \frac{Z_1 + Z_2}{2}$ and $X_2 = \frac{Z_1 - Z_2}{2}$. Let us now calculate the Jacobian of the transformation. It is given by

$$\begin{vmatrix} \frac{\partial x_1}{\partial z_1} & \frac{\partial x_1}{\partial z_2} \\ \frac{\partial x_2}{\partial z_1} & \frac{\partial x_2}{\partial z_2} \end{vmatrix} = \begin{vmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{vmatrix} = -\frac{1}{2}.$$

Notes

Now since X_1 and X_2 are independent, we have

$$f_{X_1, X_2}(x_1, x_2) = \frac{1}{\sqrt{2\pi}} e - \frac{1}{2} x_1^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x_2^2}, -\infty < x_1, x_2 < \infty.$$

Hence by (1) the joint probability density function of (Z_1, Z_2) is

$$\phi(z_1, z_2) = f\left[\frac{z_1 + z_2}{2}, \frac{z_1 - z_2}{2}\right] | J |, -\infty < z_1, z_2 < \infty$$

$$= \frac{1}{4\pi} \exp\left\{-\frac{1}{2}\left[\frac{z_1 + z_2}{2}\right] - \frac{1}{2}\left[\frac{z_1 - z_2}{2}\right]^2\right\}$$

$$= \frac{1}{4\pi} \exp\left\{-\left[\frac{{z_1}^2}{4} + \frac{{z_2}^2}{4}\right]\right\}, -\infty < z_1, z_2 < \infty$$

Then the marginal density of Z_1 is given by

$$\phi z_1(z_1) = \int_{-\infty}^{+\infty} f(z1, z2) \partial z_2 \frac{1}{\sqrt{4\pi}} e^{\frac{z_1^2}{4}}, -\infty < z_2 < \infty.$$

Note that we can calculate the marginal density of Z2 also. It is given by

$$\phi z_2(z_2) = \int_{-\infty}^{+\infty} \phi(z_1, z_2) \partial z_2 \frac{1}{\sqrt{4\pi}} e^{\frac{z_1^2}{4}}, -\infty < z_2 < \infty.$$

In other words Z_1 has N(0, 2) and Z_2 has N(0, 2) as their distribution functions. In fact Z_1 and Z_2 are independent random variables since

$$\phi(z_1, z_2) = \phi z_1(z_1)\, \phi z_2(z_2)$$

for all z_1 and z_2.

We shall illustrate this method with one more example.

Example 4 : Suppose X_1 and X_2 are independent random variables with common density function

$$f(x) = \frac{1}{2} e^{-x/2} \text{ for } 0 < x < \infty$$

$$= 0 \qquad \text{otherwise.}$$

Let us find the distribution function of $Z_1 = \frac{1}{2}(X_1 - X_2)$.

Here it is convenient to choose $Z_2 = X_2$. Note that the transformation

$(X_1, X_2) \rightarrow (Z_1, Z_2)$ gives a one-to-one mapping from the set A = $\{ (x_1, x_2) : 0 < x_1 < w, 0 < x_2 < \infty\}$ onto the set

B = $\{ (z_1, z_2) : z_2 > 0, -\infty < x_2 < \infty$ and $z_2 > -2z_1\}$. The inverse transformation is

$$X_1 = 2Z_1 + Z_2$$

and

$$X_2 = Z_2.$$

Notes

Since $x_1 > 0$, it follows that $2z_1 + z_2 > 0$, that is, $z_2 > -2z1$.

Since $x_2 > 0$, it follows that $z_2 > 0$. Obviously, $-\infty < z_1 < \infty$.

Further more you can check that the Jacobian of the transformation is equal to 2.

Now the joint density function is

$$fx_1,x_2, (x_1, x_2) = \frac{1}{4}e^{\frac{-(x+y)}{2}}$$

Therefore from (2) the joint density function of (Z_1, Z_2) is

$$4\,[z_1, z_2] = fx_1, x_2, [2z_1 + z_2, z_2]\ |\ 1\ | \qquad \text{provided } (z_1, z_2) \in B$$

$$= 0 \qquad \text{, otherwise.}$$

Therefore

$$f[z_1, z_2] = \frac{1}{2}e^{-z_1-z_2} \qquad \text{if } (z_1, z_2) \in B$$

$$= 0 \qquad \text{otherwise ...*}$$

and the marginal density of Z_1 is

$$\phi z_1\,(z_1) = \int_{-2z_1}^{\infty} \frac{1}{2}e^{-z_1-z_2}dz_2 \quad \text{if } -\infty < z_1 < 0$$

$$= \int_{0}^{\infty} \frac{1}{2}e^{-z_1-z_2}dz_2 \quad \text{if } 0 \le z_1 < \infty$$

This shows that

$$\phi z_1(z_1) = \frac{1}{2}e^{-|z_1|}, -\infty < z_1 < \infty$$

The distribution with the density function given by * is known as double exponential distribution.

An important application of the transformation approach is to determine distribution of the sum of two independent random variables not necessarily identically distributed. Let us now look at this problem.

Suppose X_1 and X_2 are independent random variables with the density functions f1 (x_1) and $f_2\,(x_2)$ respectively and we want to determine the distribution function of $X_1 + X_2$. Let $Z_1 = X_1 + X_2$. We apply transformation method here. Set $Z_2 = X_2$. Then the transformation $(X_1, X_2) \to (Z_1, Z_2)$ is invertible and

$X_1 = Z_1 - Z_2$

$X_2 = Z_2$.

The Jacobian of the transformation is equal to unity. Since the joint density of (X_1, X_2) is $f_1(x_1)\,f_2(x_2)$, it follows from (2) that the joint density of (Z_1, Z_2) is given by

$\phi\,(z_1, z_2) = f_1\,(z_1 - z_2)\,f_2\,(z_2), -\infty < z_1, z_2 < \infty$.

Notes

Hence the density of Z_1 is given by 4

$$\phi z_1(z_1) = \int_{-\infty}^{\infty} f(z_1, z_2) dz_2$$

$$\int_{-\infty}^{\infty} f_1(z_1 - z_2) f_2(z_2) dz_2, -\infty < z_1, z_2, < \infty \qquad ...(3)$$

This formula giving the density function of Z1 is known as the convolution formula. This is called the convolution formula because the density function is the convolution product of the density functions of X_1 and X_2.

Let us now dlculate the distribution function of Z_1. We denote the distribution function of Z_1 by ϕz_1. Then we have

$$\phi z_1(z) = \int_{-\infty}^{z} \phi_1(z_1) dz_1$$

$$= \int_{-\infty}^{z} \left[\int_{-\infty}^{\infty} f_1(z_1 - z_2) f_2(z_2) dz_2 \right] dz_1$$

$$= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} f_1(z_1 - z_2) dz_1 \right] f_2(z_2) dz_2$$

$$= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{z-z_2} f_1(u) du \right] f_2(z_2) dz_2$$

(by the transformation $u = z_1 - z_2$)

$$= \int_{-\infty}^{\infty} F_1(z - z_2) f_2(z_2) dz_2$$

where F_1 is the distribution function of X_1.

Therefore the distribution function of Z_1 is the convolution product of the distribution function of X_1 and the density function of X_2.

The above relation gives an explicit formula for the distribution function of Z_1.

Let us see an ,example.

Example 5: Suppose X_1 and X_2 are independent random variables with the gamma distributions having parameters (α_1, λ) and (α_2, λ) respectively. Let us find the density function of the sum $Z = X_1 + X_2$ using the convolution formula.

The density of X_1 is

$$Fx_i(x_i) = \frac{\lambda^{\alpha_i} x_i^{\alpha-1} e^{-1x_i}}{\Gamma(\alpha_i)}, \qquad x_i > 0$$

$$= 0 \qquad \text{otherwise}$$

Notes

for i = 1, 2. We use Formula (3) to mmpute the density function of 2. For z > 0, we have

$$\phi z(z) = \int_{-\infty}^{\infty} fx_1(z-u)fx_2(u)du$$

$$= \int_{0}^{\infty} fx_1(z-u)fx_2(u)du$$

$$= \int_{0}^{z} \frac{\lambda^{\alpha_1} e^{-\lambda(z-u)}}{\Gamma(\alpha_1)}(z-u)^{\alpha_1-1} \frac{\lambda^{\alpha_2} e^{-\lambda u}}{\Gamma(\alpha_2)} u^{\alpha_2-1} du.$$

$$= \frac{\lambda^{\alpha_1+\alpha 2} e^{-\lambda z}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \int_{0}^{z} \{(z-u)^{\alpha_1-1} u^{\alpha_2-1} du\}$$

$$= \frac{\lambda^{\alpha_1+\alpha 2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} e^{-\lambda z} z^{\alpha_1+\alpha_2-1} \left\{ \int_{0}^{1} (1-v)^{\alpha_1-1} v^{\alpha_2-1} dv \right\}$$

(by the transformation $v = \frac{u}{z}$)

$$= \frac{\lambda^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} e^{-\lambda z} z^{\alpha_1+\alpha_2-1} B(\alpha_2, \alpha_1)$$

\ $$\phi z(z) = \frac{\lambda^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1+\alpha_2)} e^{-\lambda z} z^{\alpha_1+\alpha_2-1}, 0 < z < \infty$$

$$= 0 \quad , \qquad z < 0.$$

The last equality follows from the properties of beta function and gamma function.

This example shows that the convolution of gamma distributions with parameters (α_1, λ) and (α_2, λ) is a gamma distribution with parameter ($\alpha_1 + \alpha_2$, λ).

Next we shall consider another example in which we illustrate another method called Moment Generating Function approach. This method is useful for finding the distribution functions of sums or linear combinations of independent random variables.

Example 6: Suppose X_1 and X_2 are independent random variables with distributions $N[\mu_1, \sigma_1^2]$ and $N[\mu_2, \sigma_2^2]$ respectively. Define $Z = X_1 + X_2$. Then the m.g.f. of Z is

$$Mz(t) = E[e^{t(X_1 + X_2)}]$$

$$= E[e^{tX_1} e^{tX_2}]$$

$$= E[e^{tX_1}] E[e^{tX_2}]$$

The last relation follows from the fact that e^{tX_1} and e^{lX_2} are independent random variables when XI and X2 are independent. But we have proved earlier that

Notes

Hence

$$Mz(t) = \exp\left\{t[\mu_1 + \mu_2] + \frac{1}{2}t^2\left[\sigma_1^{\ 2} + \sigma_2^{\ 2}\right]\right\}, \quad -\infty < t < \infty.$$

But this function is the m.g.f. of N [$\mu_1 + \mu_2$, $\sigma_1^2 + \sigma_2^2$]. From the uniqueness property (Theorem 1 of Unit 10), it follows that Z has

$$N[\ \mu_1 + \mu_2,\ \sigma_1^2 + \sigma_2^2].$$

In the next section we shall talk about functions of more than two random variables.

5.2 Functions of More than Two Random Variables

Suppose we haven random variables X_1, ..., X_n not necessarily independent and we are interested in finding the distribution function of a function $Z_1 = g_1$ (X_1, ..., X_n) or the joint distribution function of $Z_j = g_i$ (X_1 ,.........., X_n), $1 \leq i \leq r$, where r is any positive integer $1 \leq r \leq n$. The methods described in the previous section can be extended to this general case. We will not go into detailed description or extension of the methods. We will illustrate by a few examples.

Example 7: Suppose X_1, X_2 ....X_n, is a random sample of size n, from a certain population. We shall discuss this concept of random sampling in greater detail in Block 4 Unit 15. In the present context it will suffice to record that the above statement is a convenient alternative way of expressing the fact that X_1, X_2, ... X_n are independent and identically distributed n random variables with a common distribution function F(x) which coincide with the population distribution function (see Unit 15, Block 4). Define Z_1 = min (X_1, ...,X_n) and Z_n = max (X_1, ...,X_n). Let us find the joint distribution of (Z_1, Z_n).

We first note that $Z_1 \leq Z_n$. Let us compute the distribution function G_{z_i}, z_n of (Z_1, Z_n). Let (z_1, z_n) be a fixed pair where $-\infty < z_1 \leq z_n <$ m. We first consider the case $z_1 = z_n$. Then

$$\begin{aligned}
G_{z_1}, z_n(z_1, z_n) &= P\ [Z_1 \leq z_1, Z_n \leq Z_n] \\
&= P[Z_n \leq z_n]\text{ , since the event } [Z_n \leq z_n] \\
&\qquad \text{implies the event } [Z_1 \leq z_1], \\
&= P[X_i \leq Z_n, 1 \leq i \leq n]\ z_1 \text{ and } z_n \text{ being equal.} \\
&= \prod_{i=1}^{n} P[X_i \leq z_n], \text{ since } X_i\text{'s are independent} \\
&= F\ (z_n)^n.
\end{aligned}$$

Now, suppose that $z_1 < z_n$. Then we have

$$\begin{aligned}
G_{z_1}, z_n\ (z_1, z_n) &= P\ [Z_1 \leq z_1, Z_n, \leq z_n\ ] \\
&= P\ [Z_n \leq z_n\ ] - P\ [Z_n < z_n, Z_1 > z_1\ ] \\
&= P[Z_n \leq Z_n] - P[z_1 < Z_1 \leq Z_n \leq z_n] \\
&= P[Z_n \leq z_n] - P(z_1 < X_i \leq z_n \text{ for } 1 \leq i\ £\ n) \\
&= P\ (Z_n \leq z_n) - \prod_{i=1}^{n} P[z_1 < X_i \leq z_n], \text{ since Xi's are independent}
\end{aligned}$$

Notes

$$= P\,[X_i \le z_n \text{ for } 1 \le i \le n\,] - \prod_{i=1}^{n} P[z_1 < X_i \le z_n]$$

$$= \prod_{i=1}^{n} P[X_i \le z_n] - \prod_{i=1}^{n} P[z_1 < X_i \le z_n]$$

$$= [F(z_n)]^n - [F(z_n) - F(z_1)]^n.$$

Therefore if $-\infty < z_1 \le z_n < \infty$, we get the distribution function as

$$G_{Z_1, Z_n}(Z_1, Z_n) = F(Z_n)^n - [F(Z_n) - F(Z_1)]^n \qquad ...(4)$$

The joint probability density function of (Z_1, Z_n) is obtained by the relation

$$G_{Z_1}, Z_n(Z_1, Z_n) = \frac{\partial^2 G_{Z_1}, Z_n(Z_1, Z_n)}{\partial Z_1 \partial Z_n}$$

Then from (4), we have

$Gz1, z2(z1, zn) = n(n-1)\,[F(z_1) - F(z_1)]^{n-2} f(z_1)\, f(z_n)$ if $-\infty < z_1 < z_n < \infty$

$= 0$ otherwise.

The quantity $Z_n - Z_1$ is called the range. Infact, Range is the difference between the largest and the smallest observations. We shall now find the distribution of the range $W_1 = Z_n - Z_1$ for the observations given in Example 7

Example 8: Let $X_1, X_2, ..., X_n$ and Z_1, Z_n are as given Example 7. Let us find the distribution of WI = Zn - Zi.

Here we make use of the transformation method.

Set $W_2 = Z_1$

Now you can check that the transformation $(Z_1, Z_n) \to (W_1, W_2)$ is one-to-one and the inverse transformation is given by Z1 = W_2, $Z_n = W_1 + W_2$. The Jacobian of this transformation is equal to -1. Hence the joint density of (W_1, W_2) is given by

$$G(w_1, w_2) = g_{z1}, z_n\,(w_2, w_2 + w_1),\ 0 < w_1 < \infty,\ -\infty < w_2 < \infty$$

where $g_{z1}, z_n,$ is the joint density of $(Z_1, Z_n,)$ which we have calculated in Example 7.

Then we have

$$G\{w_l, w_2) = n(n-1)\,[(F(w_2+w_l) - F(w_2)]^{n-2} f(w_2)f(w_2+w_1) \text{ if } 0 < w_1 < \infty \text{ and } -\infty < w_2 < \infty$$

$= 0$, otherwise

and the marginal density function of W_1 is

$$\phi_{w_1}(w_1) = \int_{-\infty}^{\infty} \phi(w_1, w_2)\,dw_2$$

$= 0$, otherwise

Notes

Let us consider a special case of the above problem when $X_1, ..., X_n$ are independent and identically distributed with uniform distribution on [0, I]. Then

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } 0 < x < 1 \\ 1 & \text{if } x \geq 1 \end{cases}$$

and

$$f(x) = \begin{cases} 1 & \text{if } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

In this case

$$\phi_{w_1}(w_1) = n(n-1) \int_0^{1-w_1} w_1^{n-2} dw_2, \text{ if } 0 < w_1 < 1$$

$$= n(n-1)\ w_1^{n-2}(1-w_1), \text{ if } 0 < w_1 < 1$$

$$= 0 \qquad \text{, otherwise}$$

Now for a short exercise

In the next three sections we shall discuss three standard distributions each of which appear as the distribution of a certain function of standard normal variable. We shall make use of the different approaches discussed in this unit to obtain their distribution functions. All these distributions play an important role in statistical inference which will be discussed in Block 4.

5.3 Summary

- Let (X,Y) be a random vector with the joint density function fx, y (x,y). Let g(x,y) be a real-valued function defined on R x R. For z E 2, define

  $$D_z = \{(x, Y) : g(x, y) \leq z\}$$

  Then the distribution function of Z is defined as

  $$P[Z \leq z] = \iint_{D_z} f_{X,Y}(X,Y)\, dx\, dy$$

  Theoretically it is not difficult to write down the distribution function using (1). But in actual practise it is sometimes difficult to evaluate the double integral.

- Suppose (X_1, X_2) is a bivariate random vector with the density function $fx_1, x_2 (x_1, x_2)$ and we would like to determine the distribution function of the density function of $Z_1 = g_1 (X_1, X_2)$. To determine this, let us suppose that we can find another function $Z_2 = g_2 (X_1, X_2)$ such that the transformation from (X_1, X_2) to (Z_1, Z_2) is one-to-one. In other words to every point (x_1, x_2) in R^2, there corresponds a point (x_1, x_2) in R^2 given by the above transformation and conversely to every point (z_1, z_2) there corresponds a unique point (x_1, x_2) such that

  $$z_1 = g_1(x_1, x_2)$$

  $$z_2 = g_2(x_1, x_2)$$

5.4 Keywords

Double exponential distribution: The distribution with the density function given by * is known as double exponential distribution.

Range: It is the difference between the largest and the smallest observations.

5.5 Self Assessment

Fill in the blanks:

1. Let (X,Y) be a random vector with the joint .... fx, y (x,y). Let g(x,y) be a real-valued function defined on R x R. For z E 2, define

$$D_z = \{(x, Y) : g(x, y) \leq z\}$$

2. The distribution with the density function given by * is known as ....
3. An important application of the .... approach is to determine distribution of the sum of two independent random variables not necessarily identically distributed.
4. The .... of the transformation is equal to unity.
5. .... is the difference between the largest and the smallest observations.

5.6 Review Questions

1. Suppose X and Y are independent random variables, each having uniform distribution on (0, 1). Determine the density function of Z = X + Y.
2. Suppose (X,Y) has the joint probability density function

$$f(x,y) = x + y, \text{ if } 0 < x, y < 1$$
$$= 0, \text{ otherwise}$$

   Find the density function of Z = XY.
3. Suppose X and Y are independent r. vs with density function f(x) and distribution function F(x). Find the density function of Z = min (X,Y).
4. Suppose X_1 and X_2 are independent random variables with gamma densities $f_i(x_i)$ given by

$$f_i(X_i) = \frac{1}{\Gamma(\alpha_i)} x_i^{\alpha_i - 1} e^{-x_i}, \ 0 < x_i < \infty$$
$$= 0 \text{ otherwise}$$

   for i = 1, 2. Let $Z_1 = X_1 + X_2$ and $Z_2 = \frac{X_1}{X_1 + X_2}$. Show that Z_1 and Z_2 are independent random variables. Find the distribution functions of Z_2 and Z_1.
5. (Box - Muller transformation) Let X_1 and X_2 be independent random variables uniformly distributed on [0, 1]. Define

$$Z_1 = (-2 \log X_1)^{1/2} \cos(2\pi X_2),$$
$$Z_2 = (-2 \log X_1)^{1/2} \sin(2\pi X_2)$$

   Show that Z_1 and Z_2 are independent standard normal random variables.

Notes

6. Suppose (X_1, X_2) have the joint density function

$$f(x_1, x_2) = 4x_1 x_2 \text{ if } 0 < x_1, x_2 < 1$$
$$= 0 \qquad \text{otherwise.}$$

Define $Z_1 = \dfrac{X_1}{X_2}$ and $Z_2 = X_1 X_2$. Determine the joint density function of (Z_1, Z_2)

7. Suppose $X_1, \ldots, X_n$ are n independent random variables with the same distribution $N(\mu, \sigma^2)$. Define

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$\bar{X}$ is called the sample mean. Extending the m.g.f. approach for more than two random variables, show that $\bar{X}$ has the distribution $N\left[\mu, \dfrac{\sigma^2}{n}\right]$.

Answers: Self Assessment

1. density function 2. double exponential distribution 3. transformation

4. Jacobian 5. Range

5.7 Further Readings

Books

Introductory Probability and Statistical Applications by P.L. Meyer

Introduction to Mathematical Statistics by Hogg and Craig

Fundamentals of Mathematical Statistics by S.C. Gupta and V.K. Kapoor

Unit 6: Probability

Notes

CONTENTS

Objectives

Introduction

6.1 Classical Definition

6.2 Counting Techniques

6.2.1 Fundamental Principle of Counting

6.2.2 Permutation

6.2.3 Combination

6.2.4 Ordered Partitions

6.3 Statistical or Empirical definition of Probability

6.3.1 Definition (R. Von Mises)

6.7 Summary of Formulae

6.8 Keywords

6.9 Self Assessment

6.10 Review Questions

6.11 Further Readings

Objectives

After studying this unit, you will be able to:

- Define Classical Probability
- Discuss Counting Techniques
- Discuss Statistical or Empirical definition of Probability

Introduction

The concept of probability originated from the analysis of the games of chance in the 17th century. Now the subject has been developed to the extent that it is very difficult to imagine a discipline, be it from social or natural sciences, that can do without it. The theory of probability is a study of Statistical or Random Experiments. It is the backbone of Statistical Inference and Decision Theory that are essential tools of the analysis of most of the modern business and economic problems.

Often, in our day-to-day life, we hear sentences like 'it may rain today', 'Mr X has fifty-fifty chances of passing the examination', 'India may win the forthcoming cricket match against Sri lanka', 'the chances of making profits by investing in shares of company A are very bright', etc. Each of the above sentences involves an element of uncertainty.

Notes

6.1 Classical Definition

This definition, also known as the mathematical definition of probability, was given by J. Bernoulli. With the use of this definition, the probabilities associated with the occurrence of various events are determined by specifying the conditions of a random experiment. It is because of this that the classical definition is also known as 'a priori' definition of probability.

Definition

If n is the number of equally likely, mutually exclusive and exhaustive outcomes of a random experiment out of which m outcomes are favourable to the occurrence of an event A, then the probability that A occurs, denoted by P(A), is given by :

$$P(A) = \frac{\text{Number of outcomes favourable to } A}{\text{Number of exhaustive outcomes}} = \frac{m}{n}$$

Various terms used in the above definition are explained below :

1. ***Equally likely outcomes:*** The outcomes of random experiment are said to be equally likely or equally probable if the occurrence of none of them is expected in preference to others. For example, if an unbiased coin is tossed, the two possible outcomes, a head or a tail are equally likely.

2. ***Mutually exclusive outcomes:*** Two or more outcomes of an experiment are said to be mutually exclusive if the occurrence of one of them precludes the occurrence of all others in the same trial i.e. they cannot occur jointly. For example, the two possible outcomes of toss of a coin are mutually exclusive. Similarly, the occurrences of the numbers 1, 2, 3, 4, 5, 6 in the roll of a six faced die are mutually exclusive.

3. ***Exhaustive outcomes:*** It is the totality of all possible outcomes of a random experiment. The number of exhaustive outcomes in the roll of a die are six. Similarly, there are 52 exhaustive outcomes in the experiment of drawing a card from a pack of 52 cards.

4. ***Event:*** The occurrence or non-occurrence of a phenomenon is called an event. For example, in the toss of two coins, there are four exhaustive outcomes, viz. (H, H), (H, T), (T, H), (T, T). The events associated with this experiment can be defined in a number of ways. For example, (i) the event of occurrence of head on both the coins, (ii) the event of occurrence of head on at least one of the two coins, (iii) the event of non-occurrence of head on the two coins, etc.

   An event can be simple or composite depending upon whether it corresponds to a single outcome of the experiment or not. In the example, given above, the event defined by (i) is simple, while those defined by (ii) and (iii) are composite events.

Example 1: What is the probability of obtaining a head in the toss of an unbiased coin?

Solution.

This experiment has two possible outcomes, i.e., occurrence of a head or tail. These two outcomes are mutually exclusive and exhaustive. Since the coin is given to be unbiased, the two outcomes are equally likely. Thus, all the conditions of the classical definition are satisfied.

No. of cases favourable to the occurrence of head = 1

No. of exhaustive cases = 2

$\therefore$ Probability of obtaining head $P(H) = \frac{1}{2}$.

Notes

Example 2: What is the probability of obtaining at least one head in the simultaneous toss of two unbiased coins?

Solution.

The equally likely, mutually exclusive and exhaustive outcomes of the experiment are (H, H), (H, T), (T, H) and (T, T), where H denotes a head and T denotes a tail. Thus, n = 4.

Let A be the event that at least one head occurs. This event corresponds the first three outcomes of the random experiment. Therefore, m = 3.

Hence, probability that A occurs, i.e., $P(A) = \frac{3}{4}$.

Example 3: Find the probability of obtaining an odd number in the roll of an unbiased die.

Solution.

The number of equally likely, mutually exclusive and exhaustive outcomes, i.e., n = 6. There are three odd numbers out of the numbers 1, 2, 3, 4, 5 and 6. Therefore, m = 3.

Thus, probability of occurrence of an odd number $= \frac{3}{6} = \frac{1}{2}$.

Example 4: What is the chance of drawing a face card in a draw from a pack of 52 well-shuffled cards?

Solution.

Total possible outcomes n = 52.

Since the pack is well-shuffled, these outcomes are equally likely. Further, since only one card is to be drawn, the outcomes are mutually exclusive.

There are 12 face cards, $\therefore$ m = 12.

Thus, probability of drawing a face card $= \frac{12}{52} = \frac{3}{13}$.

Example 5: What is the probability that a leap year selected at random will contain 53 Sundays?

Solution.

A leap year has 366 days. It contains 52 complete weeks, i.e, 52 Sundays. The remaining two days of the year could be anyone of the following pairs:

(Monday, Tuesday), (Tuesday, Wednesday), (Wednesday, Thursday), (Thursday, Friday), (Friday, Saturday), (Saturday, Sunday), (Sunday, Monday). Thus, there are seven possibilities out of which last two are favourable to the occurrence of 53rd Sunday.

Hence, the required probability $= \frac{2}{7}$.

Notes

Example 6: Find the probability of throwing a total of six in a single throw with two unbiased dice.

Solution.

The number of exhaustive cases n = 36, because with two dice all the possible outcomes are :

(1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6),

(2, 1), (2, 2), (2, 3), (2, 4), (2, 5), (2, 6),

(3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6),

(4, 1), (4, 2), (4, 3), (4, 4), (4, 5), (4, 6),

(5, 1), (5, 2), (5, 3), (5, 4), (5, 5), (5, 6),

(6, 1), (6, 2), (6, 3), (6, 4), (6, 5), (6, 6).

Out of these outcomes the number of cases favourable to the event A of getting 6 are : (1, 5), (2, 4), (3, 3), (4, 2), (5, 1). Thus, we have m = 5.

$$\therefore \quad P(A) = \frac{5}{36}.$$

Example 7: A bag contains 15 tickets marked with numbers 1 to 15. One ticket is drawn at random. Find the probability that

(i) the number on it is greater than 10,

(ii) the number on it is even,

(iii) the number on it is a multiple of 2 or 5.

Solution.

Number of exhaustive cases n = 15

(i) Tickets with number greater than 10 are 11, 12, 13, 14 and 15. Therefore, m = 5 and hence the required probability $= \frac{5}{15} = \frac{1}{3}$.

(ii) Number of even numbered tickets m = 7

$$\therefore \text{ Required probability } = \frac{7}{15}.$$

(iii) The multiple of 2 are : 2, 4, 6, 8, 10, 12, 14 and the multiple of 5 are : 5, 10, 15. $\therefore$ m = 9 (note that 10 is repeated in both multiples will be counted only once).

$$\text{Thus, the required probability } = \frac{9}{15} = \frac{3}{5}.$$

Notes

6.2 Counting Techniques

Counting techniques or combinatorial methods are often helpful in the enumeration of total number of outcomes of a random experiment and the number of cases favourable to the occurrence of an event.

6.2.1 Fundamental Principle of Counting

If the first operation can be performed in any one of the m ways and then a second operation can be performed in any one of the n ways, then both can be performed together in any one of the m $\times$ n ways.

This rule can be generalised. If first operation can be performed in any one of the n_1 ways, second operation in any one of the n_2 ways, ... kth operation in any one of the n_k ways, then together these can be performed in any one of the n_1 ´ n_2 × ...... × n_k ways.

6.2.2 Permutation

A permutation is an arrangement of a given set of objects in a definite order. Thus composition and order both are important in a permutation.

(a) *Permutations of n objects*

The total number of permutations of n distinct objects is n!. Using symbols, we can write ^nP_n = n!, (where n denotes the permutations of n objects, all taken together).

Let us assume there are n persons to be seated on n chairs. The first chair can be occupied by any one of the n persons and hence, there are n ways in which it can be occupied. Similarly, the second chair can be occupied in n - 1 ways and so on. Using the fundamental principle of counting, the total number of ways in which n chairs can be occupied by n persons or the permutations of n objects taking all at a time is given by:

^nP_n = n(n - 1)(n - 2) ...... 3.2.1 = n!

(b) *Permutations of n objects taking r at a time*

In terms of the example, considered above, now we have n persons to be seated on r chairs, where r £ n.

Thus, ^nP_r = n(n - 1)(n - 2) ...... [n - (r - 1)] = n(n - 1)(n - 2) ...... (n - r + 1).

On multiplication and division of the R.H.S. by (n - r)!, we get

$$^nP_r = \frac{n(n-1)(n-2)\ ....\ (n-r+1)(n-r)!}{(n-r)!} = \frac{n!}{(n-r)!}$$

(c) *Permutations of n objects taking r at a time when any object may be repeated any number of times*

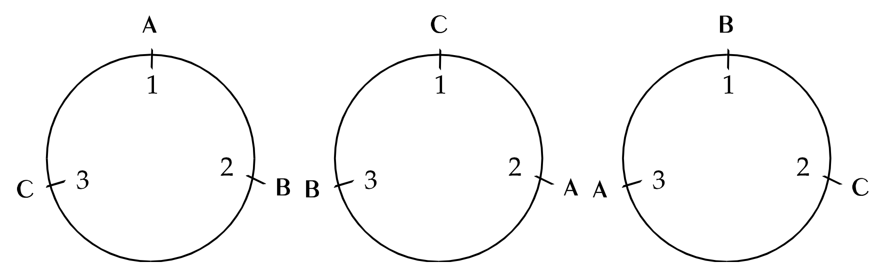
Here, each of the r places can be filled in n ways. Therefore, total number of permutations is n^r.

(d) *Permutations of n objects in a circular order*

Suppose that there are three persons A, B and C, to be seated on the three chairs 1, 2 and 3, in a circular order. Then, the following three arrangements are identical:

Notes

Figure 7.1



Similarly, if n objects are seated in a circle, there will be n identical arrangements of the above type. Thus, in order to obtain distinct permutation of n objects in circular order we divide ${}^{n}P_{n}$ by n, where ${}^{n}P_{n}$ denotes number of permutations in a row. Hence, the number of permutations in a circular order $\frac{n!}{n} = (n-1)!$

(e) *Permutations with restrictions*

If out of n objects n_1 are alike of one kind, n_2 are alike of another kind, ... n_k are alike, the number of permutations are $\frac{n!}{n_1!n_2! \text{ .... } n_k!}$

Since permutation of n_i objects, which are alike, is only one (i = 1, 2, ... k). Therefore, n! is to be divided by $n_1!$, $n_2!$ .... $n_k!$, to get the required permutations.

Example 8: What is the total number of ways of simultaneous throwing of (i) 3 coins, (ii) 2 dice and (iii) 2 coins and a die ?

Solution.

(i) Each coin can be thrown in any one of the two ways, i.e, a head or a tail, therefore, the number of ways of simultaneous throwing of 3 coins = $2^3 = 8$.

(ii) Similarly, the total number of ways of simultaneous throwing of two dice is equal to $6^2 = 36$ and

(iii) the total number of ways of simultaneous throwing of 2 coins and a die is equal to $2^2 \times 6 = 24$.

Example 9: A person can go from Delhi to Port-Blair via Allahabad and Calcutta using following mode of transport:

Delhi to Allahabad	*Allahabad to Calcutta*	*Calcutta to Port-Blair*
By Rail	*By Rail*	*By Air*
By Bus	*By Bus*	*By Ship*
By Car	*By Car*	
By Air	*By Air*	

In how many different ways the journey can be planned?

Notes

Solution.

The journey from Delhi to Port-Blair can be treated as three operations; From Delhi to Allahabad, from Allahabad to Calcutta and from Calcutta to Port-Blair. Using the fundamental principle of counting, the journey can be planned in $4 \times 4 \times 2 = 32$ ways.

Example 10: In how many ways the first, second and third prize can be given to 10 competitors?

Solution.

There are 10 ways of giving first prize, nine ways of giving second prize and eight ways of giving third prize. Therefore, total no. ways is $10 \times 9 \times 8 = 720$.

Alternative method :

Here n = 10 and r = 3, $\therefore \; {}^{10}P_3 = \dfrac{10!}{(10-3)!} = 720$

Example 11:

(a) There are 5 doors in a room. In how many ways can three persons enter the room using different doors?

(b) A lady is asked to rank 5 types of washing powders according to her preference. Calculate the total number of possible rankings.

(c) In how many ways 6 passengers can be seated on 15 available seats.

(d) If there are six different trains available for journey between Delhi to Kanpur, calculate the number of ways in which a person can complete his return journey by using a different train in each direction.

(e) In how many ways President, Vice-President, Secretary and Treasurer of an association can be nominated at random out of 130 members?

Solution.

(a) The first person can use any of the 5 doors and hence can enter the room in 5 ways. Similarly, the second person can enter in 4 ways and third person can enter in 3 ways.

Thus, the total number of ways is ${}^5P_3 = \dfrac{5!}{2!} = 60$.

(b) Total number of rankings are ${}^5P_5 = \dfrac{5!}{0!} = 120$. (Note that 0! = 1)

(c) Total number of ways of seating 6 passengers on 15 seats are

$${}^{15}P_6 = \frac{15!}{9!} = 36{,}03{,}600.$$

(d) Total number of ways of performing return journey, using different train in each direction are $6 \times 5 = 30$, which is also equal to ${}^6P_2$.

(e) Total number of ways of nominating for the 4 post of association are

$${}^{130}P_4 = \frac{130!}{126!} = 27{,}26{,}13{,}120 .$$

Notes

Example 12: Three prizes are awarded each for getting more than 80% marks, 98% attendance and good behaviour in the college. In how many ways the prizes can be awarded if 15 students of the college are eligible for the three prizes?

Solution.

Note that all the three prizes can be awarded to the same student. The prize for getting more than 80% marks can be awarded in 15 ways, prize for 90% attendance can be awarded in 15 ways and prize for good behaviour can also be awarded in 15 ways.

Thus, the total number of ways is $n^r = 15^3 = 3{,}375$.

Example 13:

(a) In how many ways can the letters of the word EDUCATION be arranged?

(b) In how many ways can the letters of the word STATISTICS be arranged?

(c) In how many ways can 20 students be allotted to 4 tutorial groups of 4, 5, 5 and 6 students respectively?

(d) In how many ways 10 members of a committee can be seated at a round table if (i) they can sit anywhere (ii) president and secretary must not sit next to each other?

Solution.

(a) The given word EDUCATION has 9 letters. Therefore, number of permutations of 9 letters is 9! = 3,62,880.

(b) The word STATISTICS has 10 letters in which there are $3S^{'s}$, $3T^{'s}$, $2I^{'s}$, 1A and 1C. Thus, the required number of permutations $\frac{10!}{3!3!2!1!1!} = 50{,}400$.

(c) Required number of permutations $\frac{20!}{4!5!5!6!} = 9{,}77{,}72{,}87{,}522$

(d) (i) Number of permutations when they can sit anywhere = (10-1)!= 9! = 3,62,880.

(ii) We first find the number of permutations when president and secretary must sit together. For this we consider president and secretary as one person. Thus, the number of permutations of 9 persons at round table = 8! = 40,320.

∴ The number of permutations when president and secretary must not sit together = 3,62,880 - 40,320 = 3,22,560.

Example 14:

(a) In how many ways 4 men and 3 women can be seated in a row such that women occupy the even places?

(b) In how many ways 4 men and 4 women can be seated such that men and women occupy alternative places?

Solution.

(a) 4 men can be seated in 4! ways and 3 women can be seated in 3! ways. Since each arrangement of men is associated with each arrangement of women, therefore, the required number of permutations = 4! 3! = 144.

Notes

(b) There are two ways in which 4 men and 4 women can be seated

MWMWMWMWMWMW or WMWMWMWMWMWM

∴ The required number of permutations = 2 .4! 4! = 1,152

Example 15: There are 3 different books of economics, 4 different books of commerce and 5 different books of statistics. In how many ways these can be arranged on a shelf when

(a) all the books are arranged at random,

(b) books of each subject are arranged together,

(c) books of only statistics are arranged together, and

(d) books of statistics and books of other subjects are arranged together?

Solution.

(a) The required number of permutations = 12!

(b) The economics books can be arranged in 3! ways, commerce books in 4! ways and statistics book in 5! ways. Further, the three groups can be arranged in 3! ways. ∴ The required number of permutations = 3! 4! 5! 3! =1,03,680.

(c) Consider 5 books of statistics as one book. Then 8 books can be arranged in 8! ways and 5 books of statistics can be arranged among themselves in 5! ways.

∴ The required number of permutations = 8! 5! = 48,38,400.

(d) There are two groups which can be arranged in 2! ways. The books of other subjects can be arranged in 7! ways and books of statistics can be arranged in 5! ways. Thus, the required number of ways = 2! 7! 5! = 12,09,600.

6.2.3 Combination

When no attention is given to the order of arrangement of the selected objects, we get a combination. We know that the number of permutations of n objects taking r at a time is ${}^nP_r$. Since r objects can be arranged in r! ways, therefore, there are r! permutations corresponding to one combination. Thus, the number of combinations of n objects taking r at a time, denoted by

${}^nC_r$, can be obtained by dividing ${}^nP_r$ by r!, i.e., ${}^nC_r = \frac{{}^nP_r}{r!} = \frac{n!}{r!(n-r)!}$.

Note: (a) Since ${}^nC_r = {}^nC_{n-r}$, therefore, ${}^nC_r$ is also equal to the combinations of n objects taking (n - r) at a time.

(b) The total number of combinations of n distinct objects taking 1, 2, ... n respectively, at a time is ${}^nC_1 + {}^nC_2 + \ldots\ldots + {}^nC_n = 2^n - 1$.

Example 16:

(a) In how many ways two balls can be selected from 8 balls?

(b) In how many ways a group of 12 persons can be divided into two groups of 7 and 5 persons respectively?

Notes

(c) A committee of 8 teachers is to be formed out of 6 science, 8 arts teachers and a physical instructor. In how many ways the committee can be formed if

1. Any teacher can be included in the committee.

2. There should be 3 science and 4 arts teachers on the committee such that (i) any science teacher and any arts teacher can be included, (ii) one particular science teacher must be on the committee, (iii) three particular arts teachers must not be on the committee?

Solution.

(a) 2 balls can be selected from 8 balls in $^8C_2 = \frac{8!}{2!6!} = 28$ ways.

(b) Since $^nC_r = {}^nC_{n-r}$, therefore, the number of groups of 7 persons out of 12 is also equal to the number of groups of 5 persons out of 12. Hence, the required number of groups is

$$^{12}C_7 = \frac{12!}{7!5!} = 792 \cdot$$

Alternative Method. We may regard 7 persons of one type and remaining 5 persons of another type. The required number of groups are equal to the number of permutations of 12 persons where 7 are alike of one type and 5 are alike of another type.

(c) 1. 8 teachers can be selected out of 15 in $^{15}C_8 = \frac{15!}{8!7!} = 6,435$ ways.

2. (i) 3 science teachers can be selected out of 6 teachers in $^6C_3$ ways and 4 arts teachers can be selected out of 8 in $^8C_4$ ways and the physical instructor can be selected in $^1C_1$ way. Therefore, the required number of ways = $^6C_3 \times {}^8C_4 \times {}^1C_1$ = $20 \times 70 \times 1 = 1400$.

(ii) 2 additional science teachers can be selected in $^5C_2$ ways. The number of selections of other teachers is same as in (i) above. Thus, the required number of ways = $^5C_2 \times {}^8C_4 \times {}^1C_1 = 10 \times 70 \times 1 = 700$.

(iii) 3 science teachers can be selected in $^6C_3$ ways and 4 arts teachers out of remaining 5 arts teachers can be selected in $^5C_4$ ways.

$\therefore$ The required number of ways = $^6C_3 \times {}^5C_4 = 20 \times 5 = 100$.

6.2.4 Ordered Partitions

1. Ordered Partitions (distinguishable objects)

(a) The total number of ways of putting n distinct objects into r compartments which are marked as 1, 2, ... r is equal to r^n.

Since first object can be put in any of the r compartments in r ways, second can be put in any of the r compartments in r ways and so on.

(b) The number of ways in which n objects can be put into r compartments such that the first compartment contains n_1 objects, second contains n_2 objects and so on the rth compartment contains n_r objects, where $n_1 + n_2 + ... + n_r = n$, is given by

$$\frac{n!}{n_1!n_2! \; ...... \; n_r!}.$$

To illustrate this, let r = 3. Then n_1 objects in the first compartment can be put in ${}^{n}C_{n_1}$ ways. Out of the remaining n - n_1 objects, n_2 objects can be put in the second compartment in ${}^{n-n_1}C_{n_2}$ ways. Finally the remaining n - n_1 - n_2 = n_3 objects can be put in the third compartment in one way. Thus, the required number of ways is

$${}^{n}C_{n_1} \times {}^{n-n_1}C_{n_2} = \frac{n!}{n_1!n_2!n_3!}.$$

2. *Ordered Partitions (identical objects)*

(a) The total number of ways of putting n identical objects into r compartments marked as 1, 2, ... r, is ${}^{n+r-1}C_{r-1}$, where each compartment may have none or any number of objects.

We can think of n objects being placed in a row and partitioned by the (r - 1) vertical lines into r compartments. This is equivalent to permutations of (n + r - 1) objects out of which n are of one type and (r - 1) of another type. The required number of permutations are $\frac{(n+r-1)!}{n!(r-1)!}$, which is equal to ${}^{(n+r-1)}C_n$ or ${}^{(n+r-1)}C_{(r-1)}$.

(b) The total number of ways of putting n identical objects into r compartments is ${}^{(n-r)+(r-1)}C_{(r-1)}$ or ${}^{(n-1)}C_{(r-1)}$, where each compartment must have at least one object.

In order that each compartment must have at least one object, we first put one object in each of the r compartments. Then the remaining (n - r) objects can be placed as in (a) above.

(c) The formula, given in (b) above, can be generalised. If each compartment is supposed to have at least k objects, the total number of ways is ${}^{(n-kr)+(r-1)}C_{(r-1)}$, where k = 0, 1, 2, ... etc. such that $k < \frac{n}{r}$.

Example 17: 4 couples occupy eight seats in a row at random. What is the probability that all the ladies are sitting next to each other?

Solution.

Eight persons can be seated in a row in 8! ways.

We can treat 4 ladies as one person. Then, five persons can be seated in a row in 5! ways. Further, 4 ladies can be seated among themselves in 4! ways.

$\therefore$ The required probability $= \frac{5!4!}{8!} = \frac{1}{14}$.

Example 18: 12 persons are seated at random (i) in a row, (ii) in a ring. Find the probabilities that three particular persons are sitting together.

Solution.

(i) The required probability $= \frac{10!3!}{12!} = \frac{1}{22}$.

(ii) The required probability $= \frac{9!3!}{11!} = \frac{3}{55}$.

Notes

Example 19: 5 red and 2 black balls, each of different sizes, are randomly laid down in a row. Find the probability that

(i) the two end balls are black,

(ii) there are three red balls between two black balls and

(iii) the two black balls are placed side by side.

Solution.

The seven balls can be placed in a row in 7! ways.

(i) The black can be placed at the ends in 2! ways and, in-between them, 5 red balls can be placed in 5! ways.

$\therefore$ The required probability $= \frac{2!5!}{7!} = \frac{1}{21}$.

(ii) We can treat BRRRB as one ball. Therefore, this ball along with the remaining two balls can be arranged in 3! ways. The sequence BRRRB can be arranged in 2! 3! ways and the three red balls of the sequence can be obtained from 5 balls in ${}^5C_3$ ways.

$\therefore$ The required probability $= \frac{3!2!3!}{7!} \times {}^5C_3 = \frac{1}{7}$.

(iii) The 2 black balls can be treated as one and, therefore, this ball along with 5 red balls can be arranged in 6! ways. Further, 2 black ball can be arranged in 2! ways.

$\therefore$ The required probability $= \frac{6!2!}{7!} = \frac{2}{7}$.

Example 20: Each of the two players, A and B, get 26 cards at random. Find the probability that each player has an equal number of red and black cards.

Solution.

Each player can get 26 cards at random in ${}^{52}C_{26}$ ways.

In order that a player gets an equal number of red and black cards, he should have 13 cards of each colour, note that there are 26 red cards and 26 black cards in a pack of playing cards. This can be done in ${}^{26}C_{13} \times {}^{26}C_{13}$ ways. Hence, the required probability $= \frac{{}^{26}C_{13} \times {}^{26}C_{13}}{{}^{52}C_{26}}$.

Example 21: 8 distinguishable marbles are distributed at random into 3 boxes marked as 1, 2 and 3. Find the probability that they contain 3, 4 and 1 marbles respectively.

Solution.

Since the first, second .... 8th marble, each, can go to any of the three boxes in 3 ways, the total number of ways of putting 8 distinguishable marbles into three boxes is 3^8.

The number of ways of putting the marbles, so that the first box contains 3 marbles, second contains 4 and the third contains 1, are $\frac{8!}{3!4!1!}$

Notes

$\therefore$ The required probability $= \frac{8!}{3!4!1!} \times \frac{1}{3^8} = \frac{280}{6561}$.

Example 22: 12 'one rupee' coins are distributed at random among 5 beggars A, B, C, D and E. Find the probability that :

(i) They get 4, 2, 0, 5 and 1 coins respectively.

(ii) Each beggar gets at least two coins.

(iii) None of them goes empty handed.

Solution.

The total number of ways of distributing 12 one rupee coins among 5 beggars are ${}^{12+5-1}C_{5-1} = {}^{16}C_4 = 1820$.

(i) Since the distribution 4, 2, 0, 5, 1 is one way out of 1820 ways, the required probability $= \frac{1}{1820}$.

(ii) After distributing two coins to each of the five beggars, we are left with two coins, which can be distributed among five beggars in ${}^{2+5-1}C_{5-1} = {}^{6}C_4 = 15$ ways.

$\therefore$ The required probability $= \frac{15}{1820} = \frac{3}{364}$.

(iii) No beggar goes empty handed if each gets at least one coin. 7 coins, that are left after giving one coin to each of the five beggars, can be distributed among five beggars in ${}^{7+5-1}C_{5-1} = {}^{11}C_4 = 330$ ways.

$\therefore$ The required probability $= \frac{330}{1820} = \frac{33}{182}$.

6.3 Statistical or Empirical definition of Probability

The scope of the classical definition was found to be very limited as it failed to determine the probabilities of certain events in the following circumstances :

(i) When n, the exhaustive outcomes of a random experiment is infinite.

(ii) When actual value of n is not known.

(iii) When various outcomes of a random experiment are not equally likely.

In addition to the above this definition doesn't lead to any mathematical treatment of probability.

In view of the above shortcomings of the classical definition, an attempt was made to establish a correspondence between relative frequency and the probability of an event when the total number of trials become sufficiently large.

6.3.1 Definition (R. Von Mises)

If an experiment is repeated n times, under essentially the identical conditions and, if, out of these trials, an event A occurs m times, then the probability that A occurs is given by P(A) = $\lim_{n\to\infty} \frac{m}{n}$, provided the limit exists.

Notes

This definition of probability is also termed as the empirical definition because the probability of an event is obtained by actual experimentation.

Although, it is seldom possible to obtain the limit of the relative frequency, the ratio $\frac{m}{n}$ can be regarded as a good approximation of the probability of an event for large values of n.

This definition also suffers from the following shortcomings :

(i) The conditions of the experiment may not remain identical, particularly when the number of trials is sufficiently large.

(ii) The relative frequency, $\frac{m}{n}$, may not attain a unique value no matter how large is the total number of trials.

(iii) It may not be possible to repeat an experiment a large number of times.

(iv) Like the classical definition, this definition doesn't lead to any mathematical treatment of probability.

6.7 Summary of Formulae

1. (a) The number of permutations of n objects taking n at a time are n!

   (b) The number of permutations of n objects taking r at a time, are ${}^nP_r = \frac{n!}{(n-r)!}$

   (c) The number of permutations of n objects in a circular order are (n - 1)!

   (d) The number of permutations of n objects out of which n_1 are alike, n_2 are alike, ... n_k are alike, are $\frac{n!}{n_1!n_2! \ldots n_k!}$

   (e) The number of combinations of n objects taking r at a time are ${}^nC_r = \frac{n!}{r!(n-r)!}$

2. (a) The probability of occurrence of at least one of the two events A and B is given by : $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 1 - P(\bar{A} \cap \bar{B})$.

   (b) The probability of occurrence of exactly one of the events A or B is given by : $P(A \cap \bar{B}) + P(\bar{A} \cap B)$ or $P(A \cup B) - P(A \cap B)$

3. (a) The probability of simultaneous occurrence of the two events A and B is given by: $P(A \cap B) = P(A).P(B/A)$ or $= P(B).P(A/B)$

   (b) If A and B are independent $P(A \cap B) = P(A).P(B)$.

6.8 Keywords

Classical: If n is the number of equally likely, mutually exclusive and exhaustive outcomes of a random experiment out of which m outcomes are favourable to the occurrence of an event A, then the probability that A occurs, denoted by P(A), is given by :

$$P(A) = \frac{\text{Number of outcomes favourable to } A}{\text{Number of exhaustive outcomes}} = \frac{m}{n}$$

Notes

Equally likely outcomes: The outcomes of random experiment are said to be equally likely or equally probable if the occurrence of none of them is expected in preference to others. For example, if an unbiased coin is tossed, the two possible outcomes, a head or a tail are equally likely.

6.9 Self Assessment

Choose the appropriate answer:

1. If A and B are any two events of a sample space S, then $P(A \cup B)$ $+P(A \cap B)$ equals

   (a) $P(A)+P(B)$

   (b) $1-P(\overline{A} \cap \overline{B})$

   (c) $1-P(\overline{A} \cup \overline{B})$

   (d) none of the above.

2. If A and B are independent and mutually exclusive events, then

   (a) $P(A) = P(A/B)$

   (b) $P(B) = P(B/A)$

   (c) either P(A) or P(B) or both must be zero.

   (d) none of the above.

3. If A and B are independent events, then $P(A \cap B)$ equals

   (a) $P(A) + P(B)$

   (b) $P(A).P(B/A)$

   (c) $P(B).P(A/B)$

   (d) $P(A).P(B)$

4. If A and B are independent events, then $P(A \cup B)$ equals

   (a) $P(A).P(B) + P(B)$

   (b) $P(A).P(\overline{B}) + P(B)$

   (c) $P(\overline{A}).P(\overline{B}) + P(A)$

   (d) none of the above.

5. If A and B are two events such that $P(A \cup B) = \frac{5}{6}$, $P(A \cap B) = \frac{1}{3}$, $P(\overline{A}) = \frac{1}{3}$, the events are

   (a) dependent

   (b) independent

   (c) mutually exclusive

   (d) none of the above.

6. Four dice and six coins are tossed simultaneously. The number of elements in the sample space are

   (a) $4^6 \times 6^2$ (b) $2^6 \times 6^2$ (c) $6^4 \times 2^6$ (d) none of these.

Notes

6.10 Review Questions

1. Define the term 'probability' by (a) The Classical Approach, (b) The Statistical Approach. What are the main limitations of these approaches?
2. Discuss the axiomatic approach to probability. In what way it is an improvement over classical and statistical approaches?
3. Distinguish between objective probability and subjective probability. Give one example of each concept.
4. State and prove theorem of addition of probabilities for two events when (a) they are not independent, (b) they are independent.
5. Explain the meaning of conditional probability. State and prove the multiplication rule of probability of two events when (a) they are not independent, (b) they are independent.
6. Explain the concept of independence and mutually exclusiveness of two events A and B. If A and B are independent events, then prove that $\overline{A}$ and $\overline{B}$ are also independent.

   (b) For two events A and B it is given that

   $$P(A) = 0.4, \quad P(B) = p, \quad P(A \cup B) = 0.6$$

   (i) Find the value of p so that A and B are independent.

   (ii) Find the value of p so that A and B are mutually exclusive.
7. Explain the meaning of a statistical experiment and corresponding sample space. Write down the sample space of an experiment of simultaneous toss of two coins and a die.

Answers: Self Assessment

1. (a) 2. (c) 3. (d) 4. (b) 5. (b) 6. (c)

6.11 Further Readings

Books

Introductory Probability and Statistical Applications by P.L. Meyer

Introduction to Mathematical Statistics by Hogg and Craig

Fundamentals of Mathematical Statistics by S.C. Gupta and V.K. Kapoor

Unit 7: Modern Approach to Probability

CONTENTS

Objectives

Introduction

7.1 Axiomatic or Modern Approach to Probability

7.1.1 Definition of Probability (Modern Approach)

7.2 Theorems on Probability

7.3 Theorems on Probability

7.4 Summary of Formulae

7.5 Keywords

7.6 Self Assessment

7.7 Review Questions

7.8 Further Readings

Objectives

After studying this unit, you will be able to:

- Explain Axiomatic or Modern Approach to Probability
- Describe Theorems on Probability
- Discuss Theorems on Probability (Contd.)

Introduction

In last unit, you have studied about probability. A phenomenon or an experiment which can result into more than one possible outcome, is called a random phenomenon or random experiment or statistical experiment. Although, we may be aware of all the possible outcomes of a random experiment, it is not possible to predetermine the outcome associated with a particular experimentation or trial.

Consider, for example, the toss of a coin. The result of a toss can be a head or a tail, therefore, it is a random experiment. Here we know that either a head or a tail would occur as a result of the toss, however, it is not possible to predetermine the outcome. With the use of probability theory, it is possible to assign a quantitative measure, to express the extent of uncertainty, associated with the occurrence of each possible outcome of a random experiment.

7.1 Axiomatic or Modern Approach to Probability

This approach was introduced by the Russian mathematician, A. Kolmogorov in 1930s. In his book, 'Foundations of Probability' published in 1933, he introduced probability as a function of the outcomes of an experiment, under certain restrictions. These restrictions are known as Postulates or Axioms of probability theory. Before discussing the above approach to probability, we shall explain certain concepts that are necessary for its understanding.

Notes

Sample Space

It is the set of all possible outcomes of a random experiment. Each element of the set is called a sample point or a simple event or an elementary event. The sample space of a random experiment is denoted by S and its element are denoted by e_i, where i = 1, 2, ... n. Thus, a sample space having n elements can be written as :

$$S = \{e_1, e_2, ..., e_n\}.$$

If a random experiment consists of rolling a six faced die, the corresponding sample space consists of 6 elementary events. Thus, S = {1, 2, 3, 4, 5, 6}.

Similarly, in the toss of a coin S = {H, T}.

The elements of S can either be single elements or ordered pairs. For example, if two coins are tossed, each element of the sample space would consist of the set of ordered pairs, as shown below :

$$S = \{(H, H), (H, T), (T, H), (T, T)\}$$

Finite and Infinite Sample Space

A sample space consisting of finite number of elements is called a finite sample space, while if the number of elements is infinite, it is called an infinite sample space. The sample spaces discussed so far are examples of finite sample spaces. As an example of infinite sample space, consider repeated toss of a coin till a head appears. Various elements of the sample space would be :

$$S = \{(H), (T, H), (T, T, H), ... \}.$$

Discrete and Continuous Sample Space

A discrete sample space consists of finite or countably infinite number of elements. The sample spaces, discussed so far, are some examples of discrete sample spaces. Contrary to this, a continuous sample space consists of an uncountable number of elements. This type of sample space is obtained when the result of an experiment is a measurement on continuous scale like measurements of weight, height, area, volume, time, etc.

Event

An event is any subset of a sample space. In the experiment of roll of a die, the sample space is S = {1, 2, 3, 4, 5, 6}. It is possible to define various events on this sample space, as shown below :

Let A be the event that an odd number appears on the die. Then A = {1, 3, 5} is a subset of S. Further, let B be the event of getting a number greater than 4. Then B = {5, 6} is another subset of S. Similarly, if C denotes an event of getting a number 3 on the die, then C = {3}.

It should be noted here that the events A and B are composite while C is a simple or elementary event.

Occurrence of an Event

An event is said to have occurred whenever the outcome of the experiment is an element of its set. For example, if we throw a die and obtain 5, then both the events A and B, defined above, are said to have occurred.

It should be noted here that the sample space is certain to occur since the outcome of the experiment must always be one of its elements.

7.1.1 Definition of Probability (Modern Approach)

Notes

Let S be a sample space of an experiment and A be any event of this sample space. The probability of A, denoted by P(A), is defined as a real value set function which associates a real value corresponding to a subset A of the sample space S. In order that P(A) denotes a probability function, the following rules, popularly known as axioms or postulates of probability, must be satisfied.

Axiom I : For any event A in sample space S, we have $0 \le P(A) \le 1$.

Axiom II : P(S) = 1.

Axiom III : If $A_1, A_2, \ldots\ldots A_k$ are k mutually exclusive events (i.e., $A_i \underset{i \neq j}{\cap} A_j = \phi$, where ϕ denotes a null set) of the sample space S, then

$$P(A_1 \cup A_2 \ldots\ldots \cup A_k) = \sum_{i=1}^{k} P(A_i)$$

The first axiom implies that the probability of an event is a non-negative number less than or equal to unity. The second axiom implies that the probability of an event that is certain to occur must be equal to unity. Axiom III gives a basic rule of addition of probabilities when events are mutually exclusive.

The above axioms provide a set of basic rules that can be used to find the probability of any event of a sample space.

Probability of an Event

Let there be a sample space consisting of n elements, i.e., $S = \{e_1, e_2, \ldots\ldots e_n\}$. Since the elementary events $e_1, e_2, \ldots\ldots e_n$ are mutually exclusive, we have, according to axiom III, $P(S) = \sum_{i=1}^{n} P(e_i)$. Similarly, if $A = \{e_1, e_2, \ldots\ldots e_m\}$ is any subset of S consisting of m elements, where m £ n, then $P(A) = \sum_{i=1}^{m} P(e_i)$. Thus, the probability of a sample space or an event is equal to the sum of probabilities of its elementary events.

It is obvious from the above that the probability of an event can be determined if the probabilities of elementary events, belonging to it, are known.

The Assignment of Probabilities to various Elementary Events

The assignment of probabilities to various elementary events of a sample space can be done in any one of the following three ways :

1. ***Using Classical Definition***

   We know that various elementary events of a random experiment, under the classical definition, are equally likely and, therefore, can be assigned equal probabilities. Thus, if there are n elementary events in the sample space of an experiment and in view of the fact that $P(S) = \sum_{i=1}^{n} P(e_i) = 1$ (from axiom II), we can assign a probability equal to $\frac{1}{n}$ to every elementary event or, using symbols, we can write $P(e_i) = \frac{1}{n}$ for i = 1, 2, ... n.

Notes

Further, if there are m elementary events in an event A, we have,

$$P(A)=\frac{1}{n}+\frac{1}{n}+\ ......\ +\frac{1}{n}\ (m \text{ times})=\frac{m}{n}=\frac{n(A), \textit{i.e., number of elements in A}}{n(S), \textit{i.e., number of elements in S}}$$

We note that the above expression is similar to the formula obtained under classical definition.

2. ***Using Statistical Definition***

   Using this definition, the assignment of probabilities to various elementary events of a sample space can be done by repeating an experiment a large number of times or by using the past records.

3. ***Subjective Assignment***

   The assignment of probabilities on the basis of the statistical and the classical definitions is objective. Contrary to this, it is also possible to have subjective assignment of probabilities. Under the subjective assignment, the probabilities to various elementary events are assigned on the basis of the expectations or the degree of belief of the statistician. These probabilities, also known as personal probabilities, are very useful in the analysis of various business and economic problems where it is neither possible to repeat the experiment nor the outcomes are equally likely.

   It is obvious from the above that the Modern Definition of probability is a general one which includes the classical and the statistical definitions as its particular cases. Besides this, it provides a set of mathematical rules that are useful for further mathematical treatment of the subject of probability.

7.2 Theorems on Probability

Theorem 1.

$P(\phi)=0$, where ϕ is a null set.

Proof.

For a sample space S of an experiment, we can write $S\cup\phi=S$.

Taking probability of both sides, we have $P(S\cup\phi)=P(S)$.

Since S and ϕ are mutually exclusive, using axiom III, we can write

$$P(S)+P(\phi)=P(S). \text{ Hence, } P(\phi)=0.$$

Theorem 2.

$P(\overline{A})=1-P(A)$, where $\overline{A}$ is compliment of A.

Proof.

Let A be any event in the sample space S. We can write

$$A\cup\overline{A}=S \text{ or } P(A\cup\overline{A})=P(S)$$

Since A and $\overline{A}$ are mutually exclusive, we can write

$$P(A)+P(\overline{A})=P(S)=1. \text{ Hence, } P(\overline{A})=1-P(A).$$

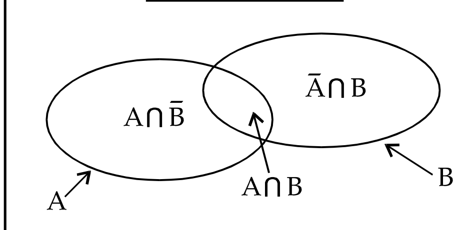
Notes

Theorem 3.

For any two events A and B in a sample space S

$$P(\bar{A} \cap B) = P(B) - P(A \cap B)$$

Figure 7.2



Venn Diagram

Proof.

From the Venn diagram, we can write

$B = (\bar{A} \cap B) \cup (A \cap B)$ or $P(B) = P[(\bar{A} \cap B) \cup (A \cap B)]$

Since $(\bar{A} \cap B)$ and $(A \cap B)$ are mutually exclusive, we have

$P(B) = P(\bar{A} \cap B) + P(A \cap B)$

or $P(\bar{A} \cap B) = P(B) - P(A \cap B)$.

Similarly, it can be shown that

$P(A \cap \bar{B}) = P(A) - P(A \cap B)$

Additive Laws

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

Proof.

From the Venn diagram, given above, we can write

$$A \cup B = A \cup (\bar{A} \cap B) \text{ or } P(A \cup B) = P[A \cup (\bar{A} \cap B)]$$

Since A and $(\bar{A} \cap B)$ are mutually exclusive, we can write

$$P(A \cup B) = P(A) + P(\bar{A} \cap B)$$

Substituting the value of $P(\bar{A} \cap B)$ from theorem 3, we get

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

Remarks:

1. If A and B are mutually exclusive, i.e., $A \cap B = \phi$, then according to theorem 1, we have $P(A \cap B) = 0$. The addition rule, in this case, becomes $P(A \cup B) = P(A) + P(B)$, which is in conformity with axiom III.

2. The event $A \cup B$ (i.e. A or B) denotes the occurrence of either A or B or both. Alternatively, it implies the occurrence of at least one of the two events.

3. The event $A \cap B$ (i.e. A and B) is a compound (or joint) event that denotes the simultaneous occurrence of the two events.

4. Alternatively, the event $A \cup B$ is also denoted by A + B and the event $A \cap B$ by AB.

Notes

Corollaries:

1. From the Venn diagram, we can write $P(A \cup B) = 1 - P(\overline{A} \cap \overline{B})$, where $P(\overline{A} \cap \overline{B})$ is the probability that none of the events A and B occur simultaneously.

2. $P(\text{exactly one of } A \text{ and } B \text{ occurs}) = P\left[\left(A \cap \overline{B}\right) \cup \left(\overline{A} \cap B\right)\right]$

   $= P\left(A \cap \overline{B}\right) + P\left(\overline{A} \cap B\right)$ $\left[\text{Since } \left(A \cap \overline{B}\right) \cap \left(\overline{A} \cap B\right) = \phi\right]$

   $= P(A) - P(A \cap B) + P(B) - P(A \cap B)$ (using theorem 3)

   $= P(A \cup B) - P(A \cap B)$ (using theorem 4)

   = P(at least one of the two events occur) – P(the two events occur jointly)

3. The addition theorem can be generalised for more than two events. If A, B and C are three events of a sample space S, then the probability of occurrence of at least one of them is given by

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

   Applying theorem 4 on the second and third term, we get

   $$= P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C) \quad .... (1)$$

   Alternatively, the probability of occurrence of at least one of the three events can also be written as

   $$P(A \cup B \cup C) = 1 - P\left(\overline{A} \cap \overline{B} \cap \overline{C}\right) \quad .... (2)$$

   If A, B and C are mutually exclusive, then equation (1) can be written as

   $$P(A \cup B \cup C) = P(A) + P(B) + P(C) \quad .... (3)$$

   If $A_1, A_2, ... A_n$ are n events of a sample space S, the respective equations (1), (2) and (3) can be modified as

   $$P(A_1 \cup A_2 \ \dots \ \cup A_n) = \sum P(A_i) - \sum\sum P\left(A_i \cap A_j\right) + \sum\sum\sum P\left(A_i \cap A_j \cap A_k\right)$$
   $$+(-1)^n P(A_1 \cap A_2 \cap \ \dots \cap A_n) \ (i \neq j \neq k, \text{ etc.}) \quad .... (4)$$

   $$P(A_1 \cup A_2 \cup \ \dots \ \cup A_n) = 1 - P\left(\overline{A}_1 \cap \overline{A}_2 \cap \ \dots \ \cap \overline{A}_n\right) \quad .... (5)$$

   $$P(A_1 \cup A_2 \cup \ \dots \ \cup A_n) = \sum_{i=1}^{n} P(A_i) \quad .... (6)$$

   (if the events are mutually exclusive)

4. The probability of occurrence of at least two of the three events can be written as

   $$P[(A \cap B) \cup (B \cap C) \cup (A \cap C)] = P(A \cap B) + P(B \cap C) + P(A \cap C) -$$
   $$3P(A \cap B \cap C) + P(A \cap B \cap C)$$
   $$= P(A \cap B) + P(B \cap C) + P(A \cap C) - 2P(A \cap B \cap C)$$

5. The probability of occurrence of exactly two of the three events can be written as

   $$P\left[\left(A \cap B \cap \overline{C}\right) \cup \left(A \cap \overline{B} \cap C\right) \cup \left(\overline{A} \cap B \cap C\right)\right] = P[(A \cap B) \cup (B \cap C) \cup (A \cap C)]$$
   $$-P(A \cap B \cap C) \quad \text{(using corollary 2)}$$

   = P(occurrence of at least two events) – P(joint occurrence of three events)

   $$= P(A \cap B) + P(B \cap C) + P(A \cap C) - 3P(A \cap B \cap C) \quad \text{(using corollary 4)}$$

6. The probability of occurrence of exactly one of the three events can be written as

$P\left[\left(A\cap\overline{B}\cap\overline{C}\right)\cup\left(\overline{A}\cap B\cap\overline{C}\right)\cup\left(\overline{A}\cap\overline{B}\cap C\right)\right]$ = P(at least one of the three events occur) - P(at least two of the three events occur).

$$= P(A)+P(B)+P(C)-2P(A\cap B)-3P(B\cap C)-2P(A\cap C)+3P(A\cap B\cap C).$$

Example 23: In a group of 1,000 persons, there are 650 who can speak Hindi, 400 can speak English and 150 can speak both Hindi and English. If a person is selected at random, what is the probability that he speaks (i) Hindi only, (ii) English only, (iii) only one of the two languages, (iv) at least one of the two languages?

Solution.

Let A denote the event that a person selected at random speaks Hindi and B denotes the event that he speaks English.

Thus, we have n(A) = 650, n(B) = 400, $n(A\cap B)=150$ and n(S) = 1000, where n(A), n(B), etc. denote the number of persons belonging to the respective event.

(i) The probability that a person selected at random speaks Hindi only, is given by

$$P\left(A\cap\overline{B}\right)=\frac{n(A)}{n(S)}-\frac{n(A\cap B)}{n(S)}=\frac{650}{1000}-\frac{150}{1000}=\frac{1}{2}$$

(ii) The probability that a person selected at random speaks English only, is given by

$$P\left(\overline{A}\cap B\right)=\frac{n(B)}{n(S)}-\frac{n(A\cap B)}{n(S)}=\frac{400}{1000}-\frac{150}{1000}=\frac{1}{4}$$

(iii) The probability that a person selected at random speaks only one of the languages, is given by

$$P\left[\left(A\cap\overline{B}\right)\cup\left(\overline{A}\cap B\right)\right]=P(A)+P(B)-2P(A\cap B) \quad \text{(see corollary 2)}$$

$$=\frac{n(A)+n(B)-2n(A\cap B)}{n(S)}=\frac{650+400-300}{1000}=\frac{3}{4}$$

(iv) The probability that a person selected at random speaks at least one of the languages, is given by

$$P(A\cup B)=\frac{650+400-150}{1000}=\frac{9}{10}$$

Alternative Method

The above probabilities can easily be computed by the following nine-square table :

	B	$\overline{B}$	*Total*
A	150	500	650
$\overline{A}$	250	100	350
Total	400	600	1000

Notes

From the above table, we can write

(i) $P\left(A \cap \overline{B}\right) = \frac{500}{1000} = \frac{1}{2}$

(ii) $P\left(\overline{A} \cap B\right) = \frac{250}{1000} = \frac{1}{4}$

(iii) $P\left[\left(A \cap \overline{B}\right) \cup \left(\overline{A} \cap B\right)\right] = \frac{500+250}{1000} = \frac{3}{4}$

(iv) $P\left(A \cup B\right) = \frac{150+500+250}{1000} = \frac{9}{10}$

This can, alternatively, be written as $P\left(A \cup B\right) = 1 - P\left(\overline{A} \cap \overline{B}\right) = 1 - \frac{100}{1000} = \frac{9}{10}$.

Example 24: What is the probability of drawing a black card or a king from a well-shuffled pack of playing cards?

Solution.

There are 52 cards in a pack, $\therefore$ n(S) = 52.

Let A be the event that the drawn card is black and B be the event that it is a king. We have to find $P\left(A \cup B\right)$.

Since there are 26 black cards, 4 kings and two black kings in a pack, we have n(A) = 26, n(B) = 4 and $n\left(A \cap \mathrm{B}\right) = 2$ Thus, $P\left(A \cup B\right) = \frac{26+4-2}{52} = \frac{7}{13}$

Alternative Method

The given information can be written in the form of the following table:

	B	$\overline{B}$	*Total*
A	2	24	26
$\overline{A}$	2	24	26
Total	4	48	52

From the above, we can write

$$P\left(A \cup B\right) = 1 - P\left(\overline{A} \cap \overline{B}\right) = 1 - \frac{24}{52} = \frac{7}{13}$$

Example 25: A pair of unbiased dice is thrown. Find the probability that (i) the sum of spots is either 5 or 10, (ii) either there is a doublet or a sum less than 6.

Solution.

Since the first die can be thrown in 6 ways and the second also in 6 ways, therefore, both can be thrown in 36 ways (fundamental principle of counting). Since both the dice are given to be unbiased, 36 elementary outcomes are equally likely.

Notes

(i) Let A be the event that the sum of spots is 5 and B be the event that their sum is 10. Thus, we can write

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

We note that $(A \cap B) = \phi$, i.e. A and B are mutually exclusive.

$\therefore$ By addition theorem, we have $P(A \cup B) = P(A) + P(B) = \frac{4}{36} + \frac{3}{36} = \frac{7}{36}$.

(ii) Let C be the event that there is a doublet and D be the event that the sum is less than 6. Thus, we can write

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

D = {(1, 1), (1, 2), (1, 3), (1, 4), (2, 1), (2, 2), (2, 3), (3, 1), (3, 2), (4, 1)}

Further, $(C \cap D)$ = {(1, 1), (2, 2)}

By addition theorem, we have $P(C \cup D) = \frac{6}{36} + \frac{10}{36} - \frac{2}{36} = \frac{7}{18}$.

Alternative Methods:

(i) It is given that n(A) = 4, n(B) = 3 and n(S) = 36. Also $n(A \cap B) = 0$. Thus, the corresponding nine-square table can be written as follows:

	B	$\bar{B}$	*Total*
A	0	4	4
$\bar{A}$	3	29	32
Total	3	33	36

From the above table, we have $P(A \cup B) = 1 - \frac{29}{36} = \frac{7}{36}$.

(ii) Here n(C) = 6, n(D) = 10, $n(C \cap D) = 2$ and n(S) = 36. Thus, we have

	C	$\bar{C}$	*Total*
D	2	8	10
$\bar{D}$	4	22	26
Total	6	30	36

Thus, $P(C \cup D) = 1 - P(\bar{C} \cap \bar{D}) = 1 - \frac{22}{36} = \frac{7}{18}$.

Example 26: Two unbiased coins are tossed. Let A_1 be the event that the first coin shows a tail and A_2 be the event that the second coin shows a head. Are A_1 and A_2 mutually exclusive? Obtain $P(A_1 \cap A_2)$ and $P(A_1 \cup A_2)$. Further, let A_1 be the event that both coins show heads and A_2 be the event that both show tails. Are A_1 and A_2 mutually exclusive? Find $P(A_1 \cap A_2)$ and $P(A_1 \cup A_2)$.

Notes

Solution.

The sample space of the experiment is S = {(H, H), (H, T), (T, H), (T, T)}

(i) A_1 = {(T, H), (T, T)} and A_2 = {(H, H), (T, H)}

Also $(A_1 \cap A_2)$ = {(T, H)}, Since $A_1 \cap A_2 \neq \phi$, A_1 and A_2 are not mutually exclusive. Further, the coins are given to be unbiased, therefore, all the elementary events are equally likely.

$\therefore \ P(A_1) = \frac{2}{4} = \frac{1}{2},\ P(A_2) = \frac{2}{4} = \frac{1}{2},\ P(A_1 \cap A_2) = \frac{1}{4}$

Thus, $P(A_1 \cup A_2) = \frac{1}{2} + \frac{1}{2} - \frac{1}{4} = \frac{3}{4}$.

(ii) When both the coins show heads; A_1 = {(H, H)}

When both the coins show tails; A_2 = {(T, T)}

Here $A_1 \cap A_2 = \phi$, $\therefore$ A_1 and A_2 are mutually exclusive.

Thus, $P(A_1 \cup A_2) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$.

Alternatively, the problem can also be attempted by making the following nine-square tables for the two cases :

(*i*)	A_2	$\bar{A}_2$	*Total*
A_1	1	1	2
$\bar{A}_1$	1	1	2
Total	2	2	4

(*ii*)	A_2	$\bar{A}_2$	*Total*
A_1	0	1	1
$\bar{A}_1$	1	2	3
Total	1	3	4

Theorem 5. Multiplication or Compound Probability Theorem

A compound event is the result of the simultaneous occurrence of two or more events. For convenience, we assume that there are two events, however, the results can be easily generalised. The probability of the compound event would depend upon whether the events are independent or not. Thus, we shall discuss two theorems; (a) Conditional Probability Theorem, and (b) Multiplicative Theorem for Independent Events.

(a) Conditional Probability Theorem

For any two events A and B in a sample space S, the probability of their simultaneous occurrence, is given by

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

or equivalently

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

Here, P(B/A) is the conditional probability of B given that A has already occurred. Similar interpretation can be given to the term P(A/B).

Proof.

Let all the outcomes of the random experiment be equally likely. Therefore,

$$P(A \cap B) = \frac{n(A \cap B)}{n(S)} = \frac{\text{no. of elements in } (A \cap B)}{\text{no. of elements in sample space}}$$

For the event B/A, the sample space is the set of elements in A and out of these the number of cases favourable to B is given by $n(A \cap B)$.

$$\therefore \quad P(B/A) = \frac{n(A \cap B)}{n(A)}.$$

If we multiply the numerator and denominator of the above expression by n(S), we get

$$P(B/A) = \frac{n(A \cap B)}{n(A)} \times \frac{n(S)}{n(S)} = \frac{P(A \cap B)}{P(A)}$$

$$\text{or} \quad P(A \cap B) = P(A).P(B/A).$$

The other result can also be shown in a similar way.

> *Notes* To avoid mathematical complications, we have assumed that the elementary events are equally likely. However, the above results will hold true even for the cases where the elementary events are not equally likely.

(b) Multiplicative Theorem for Independent Events

If A and B are independent, the probability of their simultaneous occurrence is given by

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

Proof.

We can write $A = (A \cap B) \cup (A \cap \overline{B})$.

Since $(A \cap B)$ and $(A \cap \overline{B})$ are mutually exclusive, we have

$$P(A) = P(A \cap B) + P(A \cap \overline{B}) \quad \text{(by axiom III)}$$

$$= P(B).P(A/B) + P(\overline{B}).P(A/\overline{B})$$

If A and B are independent, then proportion of A's in B is equal to proportion of A's in $\overline{B}$'s, i.e., $P(A/B) = P(A/\overline{B})$.

Thus, the above equation can be written as

$$P(A) = P(A/B)\left[P(B) + P(\overline{B})\right] = P(A/B)$$

Substituting this value in the formula of conditional probability theorem, we get

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

Remarks:

The addition theorem is used to find the probability of A or B i.e. $P(A \cup B)$, where as multiplicative theorem is used to find the probability of A and B i.e. $P(A \cap B)$.

Corollaries :

1. (i) If A and B are mutually exclusive and P(A).P(B) > 0, then they cannot be independent since $P(A \cap B) = 0$.

   (ii) If A and B are independent and P(A).P(B) > 0, then they cannot be mutually exclusive since $P(A \cap B) > 0$.

Notes

2. Generalisation of Multiplicative Theorem :

   If A, B and C are three events, then

   $$P(A \cap B \cap C) = P(A).P(B/A).P[C/(A \cap B)]$$

   Similarly, for n events A_1, A_2, ... A_n, we can write

   $$P(A_1 \cap A_2 \cap \ ... \ \cap A_n) = P(A_1).P(A_2/A_1).P[A_3/(A_1 \cap A_2)] \\ ... \ P[A_n/(A_1 \cap A_2 \cap \ ... \ \cap A_{n-1})]$$

   Further, if A_1, A_2, ... A_n are independent, we have

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

3. If A and B are independent, then A and $\overline{B}$, $\overline{A}$ and B, $\overline{A}$ and $\overline{B}$ are also independent.

   We can write $P(A \cap \overline{B}) = P(A) - P(A \cap B)$ (by theorem 3)

   $= P(A) - P(A).P(B) = P(A)[1 - P(B)] = P(A).P(\overline{B})$, which shows that A and $\overline{B}$ are independent. The other results can also be shown in a similar way.

4. The probability of occurrence of at least one of the events A_1, A_2, ... A_n, is given by

   $$P(A_1 \cup A_2 \cup \ .... \ \cup A_n) = 1 - P(\overline{A}_1 \cap \overline{A}_2 \cap \ .... \ \cap \overline{A}_n).$$

   If A_1, A_2, ... A_n are independent then their compliments will also be independent, therefore, the above result can be modified as

   $$P(A_1 \cup A_2 \cup \ .... \ \cup A_n) = 1 - P(\overline{A}_1).P(\overline{A}_2) \ .... \ P(\overline{A}_n).$$

Pair-wise and Mutual Independence

Three events A, B and C are said to be mutually independent if the following conditions are simultaneously satisfied :

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

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

If the last condition is not satisfied, the events are said to be pair-wise independent.

From the above we note that mutually independent events will always be pair-wise independent but not vice-versa.

Example 27: Among 1,000 applicants for admission to M.A. economics course in a University, 600 were economics graduates and 400 were non-economics graduates; 30% of economics graduate applicants and 5% of non-economics graduate applicants obtained admission. If an applicant selected at random is found to have been given admission, what is the probability that he/she is an economics graduate?

Solution.

Let A be the event that the applicant selected at random is an economics graduate and B be the event that he/she is given admission.

Notes

We are given n(S) = 1000, n(A) = 600, $n(\bar{A}) = 400$

Also, $n(B) = \frac{600 \times 30}{100} + \frac{400 \times 5}{100} = 200$ and $n(A \cap B) = \frac{600 \times 30}{100} = 180$

Thus, the required probability is given by $P(A/B) = \frac{n(A \cap B)}{n(B)} = \frac{180}{200} = \frac{9}{10}$

Alternative Method :

Writing the given information in a nine-square table, we have:

	B	$\bar{B}$	*Total*
A	180	420	600
$\bar{A}$	20	380	400
Total	200	800	1000

From the above table we can write $P(A/B) = \frac{180}{200} = \frac{9}{10}$

Example 28: A bag contains 2 black and 3 white balls. Two balls are drawn at random one after the other without replacement. Obtain the probability that (a) Second ball is black given that the first is white, (b) First ball is white given that the second is black.

Solution.

First ball can be drawn in any one of the 5 ways and then a second ball can be drawn in any one of the 4 ways. Therefore, two balls can be drawn in 5 × 4 = 20 ways. Thus, n(S) = 20.

(a) Let A_1 be the event that first ball is white and A_2 be the event that second is black. We want to find $P(A_2 / A_1)$.

First white ball can be drawn in any of the 3 ways and then a second ball can be drawn in any of the 4 ways, ∴ $n(A_1) = 3 \times 4 = 12$.

Further, first white ball can be drawn in any of the 3 ways and then a black ball can be drawn in any of the 2 ways, ∴ $n(A_1 \cap A_2) = 3 \times 2 = 6$.

Thus, $P(A_2 / A_1) = \frac{n(A_1 \cap A_2)}{n(A_1)} = \frac{6}{12} = \frac{1}{2}$.

(b) Here we have to find $P(A_1 / A_2)$.

The second black ball can be drawn in the following two mutually exclusive ways :

(i) First ball is white and second is black or

(ii) both the balls are black.

Thus, $n(A_2) = 3 \times 2 + 2 \times 1 = 8$, ∴ $P(A_1 / A_2) = \frac{n(A_1 \cap A_2)}{n(A_2)} = \frac{6}{8} = \frac{3}{4}$.

Notes

Alternative Method :

The given problem can be summarised into the following nine-square table:

	A_2	$\overline{A}_2$	*Total*
A_1	6	6	12
$\overline{A}_1$	2	6	8
Total	8	12	20

The required probabilities can be directly written from the above table.

Example 29: Two unbiased dice are tossed. Let w denote the number on the first die and r denote the number on the second die. Let A be the event that w + r ≤ 4 and B be the event that w + r ≤ 3. Are A and B independent?

Solution.

The sample space of this experiment consists of 36 elements, i.e., n(S) = 36. Also, A = {(1, 1), (1, 2), (1, 3), (2, 1), (2, 2), (3, 1)} and B = {(1, 1), (1, 2), (2, 1)}.

From the above, we can write

$$P(A) = \frac{6}{36} = \frac{1}{6},\ P(B) = \frac{3}{36} = \frac{1}{12}$$

Also $(A \cap B) = \{(1,1),(1,2),(2,1)\}$ $\therefore$ $P(A \cap B) = \frac{3}{36} = \frac{1}{12}$

Since $P(A \cap B) \neq P(A)P(B)$, A and B are not independent.

Example 30: It is known that 40% of the students in a certain college are girls and 50% of the students are above the median height. If 2/3 of the boys are above median height, what is the probability that a randomly selected student who is below the median height is a girl?

Solution.

Let A be the event that a randomly selected student is a girl and B be the event that he/she is above median height. The given information can be summarised into the following table :

	B	$\overline{B}$	*Total*
A	10	30	40
$\overline{A}$	40	20	60
Total	50	50	100

From the above table, we can write $P(A/\overline{B}) = \frac{30}{50} = 0.6$.

Example 31: A problem in statistics is given to three students A, B and C, whose chances of solving it independently are $\frac{1}{2}, \frac{1}{3}$ and $\frac{1}{4}$ respectively. Find the probability that

(a) the problem is solved.

(b) at least two of them are able to solve the problem.

(c) exactly two of them are able to solve the problem.

(d) exactly one of them is able to solve the problem.

Solution.

Let A be the event that student A solves the problem. Similarly, we can define the events B and C. Further, A, B and C are given to be independent.

(a) The problem is solved if at least one of them is able to solve it. This probability is given by

$$P(A \cup B \cup C) = 1 - P(\bar{A}).P(\bar{B}).P(\bar{C}) = 1 - \frac{1}{2} \times \frac{2}{3} \times \frac{3}{4} = \frac{3}{4}$$

(b) Here we have to find $P[(A \cap B) \cup (B \cap C) \cup (A \cap C)]$

$$P[(A \cap B) \cup (B \cap C) \cup (A \cap C)] = P(A)P(B) + P(B)P(C) + P(A)P(C) - 2P(A)P(B)P(C)$$

$$= \frac{1}{2} \times \frac{1}{3} + \frac{1}{3} \times \frac{1}{4} + \frac{1}{2} \times \frac{1}{4} - 2.\frac{1}{2} \times \frac{1}{3} \times \frac{1}{4} = \frac{7}{24}$$

(c) The required probability is given by $P\left[\left(A \cap B \cap \bar{C}\right) \cup \left(A \cap \bar{B} \cap C\right) \cup \left(\bar{A} \cap B \cap C\right)\right]$

$$= P(A).P(B) + P(B).P(C) + P(A).P(C) - 3P(A).P(B).P(C)$$

$$= \frac{1}{6} + \frac{1}{12} + \frac{1}{8} - \frac{1}{8} = \frac{1}{4}.$$

(d) The required probability is given by $P\left[\left(A \cap \bar{B} \cap \bar{C}\right) \cup \left(\bar{A} \cap B \cap \bar{C}\right) \cup \left(\bar{A} \cap \bar{B} \cap C\right)\right]$

$$= P(A) + P(B) + P(C) - 2P(A).P(B) - 2P(B).P(C) - 2P(A).P(C) + 3P(A).P(B).P(C)$$

$$= \frac{1}{2} + \frac{1}{3} + \frac{1}{4} - \frac{1}{3} - \frac{1}{6} - \frac{1}{4} + \frac{1}{8} = \frac{11}{24}.$$

Note that the formulae used in (a), (b), (c) and (d) above are the modified forms of corollaries (following theorem 4) 3, 4, 5 and 6 respectively.

Example 32: A bag contains 2 red and 1 black ball and another bag contains 2 red and 2 black balls. One ball is selected at random from each bag. Find the probability of drawing (a) at least a red ball, (b) a black ball from the second bag given that ball from the first is red; (c) show that the event of drawing a red ball from the first bag and the event of drawing a red ball from the second bag are independent.

Solution.

Let A_1 be the event of drawing a red ball from the first bag and A_2 be the event of drawing a red ball from the second bag. Thus, we can write:

$$n(A_1 \cap A_2) = 2 \times 2 = 4, \quad n(A_1 \cap \bar{A}_2) = 2 \times 2 = 4,$$

$$n(\bar{A}_1 \cap A_2) = 1 \times 2 = 2, \quad n(\bar{A}_1 \cap \bar{A}_2) = 1 \times 2 = 2$$

Notes

Also, $n(S) = n(A_1 \cap A_2) + n(A_1 \cap \overline{A}_2) + n(\overline{A}_1 \cap A_2) + n(\overline{A}_1 \cap \overline{A}_2) = 12$

Writing the given information in the form of a nine-square table, we get

	A_2	$\overline{A}_2$	*Total*
A_1	4	4	8
$\overline{A}_1$	2	2	4
Total	6	6	12

(a) The probability of drawing at least a red ball is given by

$$P(A_1 \cup A_2) = 1 - \frac{n(\overline{A}_1 \cap \overline{A}_2)}{n(S)} = 1 - \frac{2}{12} = \frac{5}{6}$$

(b) We have to find $P(\overline{A}_2 / A_1)$

$$P(\overline{A}_2 / A_1) = \frac{n(A_1 \cap \overline{A}_2)}{n(A_1)} = \frac{4}{8} = \frac{1}{2}$$

(c) A_1 and A_2 will be independent if $P(A_1 \cap A_2) = P(A_1).P(A_2)$

Now $P(A_1 \cap A_2) = \frac{n(A_1 \cap A_2)}{n(S)} = \frac{4}{12} = \frac{1}{3}$

$$P(A_1).P(A_2) = \frac{n(A_1)}{n(S)}.\frac{n(A_2)}{n(S)} = \frac{8}{12} \times \frac{6}{12} = \frac{1}{3}$$

Hence, A_1 and A_2 are independent.

Example 33: An urn contains 3 red and 2 white balls. 2 balls are drawn at random. Find the probability that either both of them are red or both are white.

Solution.

Let A be the event that both the balls are red and B be the event that both the balls are white. Thus, we can write

$n(S) = {}^5C_2 = 10,\ n(A) = {}^3C_2 = 3,\ n(B) = {}^2C_2 = 1,$ also $n(A \cap B) = 0$

$\therefore$ The required probability is $P(A \cup B) = \frac{n(A) + n(B)}{n(S)} = \frac{3+1}{10} = \frac{2}{5}$

Example 34: A bag contains 10 red and 8 black balls. Two balls are drawn at random. Find the probability that (a) both of them are red, (b) one is red and the other is black.

Solution.

Let A be the event that both the balls are red and B be the event that one is red and the other is black.

Two balls can be drawn from 18 balls in ${}^{18}C_2$ equally likely ways.

$$\therefore \quad n(S) = {}^{18}C_2 = \frac{18!}{2!16!} = 153$$

Notes

(a) Two red balls can be drawn from 10 red balls in ${}^{10}C_2$ ways.

$$\therefore\ n(A) = {}^{10}C_2 = \frac{10!}{2!8!} = 45$$

$$\text{Thus, } P(A) = \frac{n(A)}{n(S)} = \frac{45}{153} = \frac{5}{17}$$

(b) One red ball can be drawn in ${}^{10}C_1$ ways and one black ball can be drawn in ${}^{8}C_1$ ways.

$$\therefore\ n(B) = {}^{10}C_1 \times {}^{8}C_1 = 10 \times 8 = 80 \quad \text{Thus, } P(B) = \frac{80}{153}$$

Example 35:

Five cards are drawn in succession and without replacement from an ordinary deck of 52 well-shuffled cards:

(a) What is the probability that there will be no ace among the five cards?

(b) What is the probability that first three cards are aces and the last two cards are kings?

(c) What is the probability that only first three cards are aces?

(d) What is the probability that an ace will appear only on the fifth draw?

Solution.

(a) $$P(\text{there is no ace}) = \frac{48 \times 47 \times 46 \times 45 \times 44}{52 \times 51 \times 50 \times 49 \times 48} = 0.66$$

(b) $$P\begin{pmatrix}\text{first three card are aces and} \\ \text{the last two are kings}\end{pmatrix} = \frac{4 \times 3 \times 2 \times 4 \times 3}{52 \times 51 \times 50 \times 49 \times 48} = 0.0000009$$

(c) $$P(\text{only first three card are aces}) = \frac{4 \times 3 \times 2 \times 48 \times 47}{52 \times 51 \times 50 \times 49 \times 48} = 0.00017$$

(d) $$P\begin{pmatrix}\text{an ace appears only} \\ \text{on the fifth draw}\end{pmatrix} = \frac{48 \times 47 \times 46 \times 45 \times 4}{52 \times 51 \times 50 \times 49 \times 48} = 0.059$$

Example 36:

Two cards are drawn in succession from a pack of 52 well-shuffled cards. Find the probability that:

(a) Only first card is a king.

(b) First card is jack of diamond or a king.

(c) At least one card is a picture card.

(d) Not more than one card is a picture card.

(e) Cards are not of the same suit.

Notes

(f) Second card is not a spade.

(g) Second card is not a spade given that first is a spade.

(h) The cards are aces or diamonds or both.

Solution.

(a) $P(\text{only first card is a king}) = \dfrac{4 \times 48}{52 \times 51} = \dfrac{16}{221}.$

(b) $P\begin{pmatrix}\text{first card is a jack of} \\ \text{diamond or a king}\end{pmatrix} = \dfrac{5 \times 51}{52 \times 51} = \dfrac{5}{52}.$

(c) $P\begin{pmatrix}\text{at least one card is} \\ \text{a picture card}\end{pmatrix} = 1 - \dfrac{40 \times 39}{52 \times 51} = \dfrac{7}{17}.$

(d) $P\begin{pmatrix}\text{not more than one card} \\ \text{is a picture card}\end{pmatrix} = \dfrac{40 \times 39}{52 \times 51} + \dfrac{12 \times 40}{52 \times 51} + \dfrac{40 \times 12}{52 \times 51} = \dfrac{210}{221}.$

(e) $P(\text{cards are not of the same suit}) = \dfrac{52 \times 39}{52 \times 51} = \dfrac{13}{17}.$

(f) $P(\text{second card is not a spade}) = \dfrac{13 \times 39}{52 \times 51} + \dfrac{39 \times 38}{52 \times 51} = \dfrac{3}{4}.$

(g) $P\begin{pmatrix}\text{second card is not a spade} \\ \text{given that first is spade}\end{pmatrix} = \dfrac{39}{51} = \dfrac{13}{17}.$

(h) $P\begin{pmatrix}\text{the cards are aces or} \\ \text{diamonds or both}\end{pmatrix} = \dfrac{16 \times 15}{52 \times 51} = \dfrac{20}{221}.$

Example 37: The odds are 9 : 7 against a person A, who is now 35 years of age, living till he is 65 and 3 : 2 against a person B, now 45 years of age, living till he is 75. Find the chance that at least one of these persons will be alive 30 years hence.

Solution.

Note: If a is the number of cases favourable to an event A and a *is the number of cases favourable to its compliment event (a* + a = *n), then odds in favour of A are a :* a *and odds against A are* a *: a.*

$$\text{Obviously } P(A) = \frac{a}{a+\alpha} \text{ and } P(\bar{A}) = \frac{\alpha}{a+\alpha}.$$

Let A be the event that person A will be alive 30 years hence and B be the event that person B will be alive 30 years hence.

$$\therefore \; P(A) = \frac{7}{9+7} = \frac{7}{16} \text{ and } P(B) = \frac{2}{3+2} = \frac{2}{5}$$

We have to find $P(A \cup B)$. Note that A and B are independent.

Notes

$$\therefore\ P(A\cup B)=\frac{7}{16}+\frac{2}{5}-\frac{7}{16}\times\frac{2}{5}=\frac{53}{80}$$

Alternative Method :

$$P(A\cup B)=1-\frac{9}{16}\times\frac{3}{5}=\frac{53}{80}$$

Example 38: If A and B are two events such that $P(A\cap B)=\frac{1}{3}$, find P(B), $P(A\cup B)$, P(A/B), P(B/A), $P(\overline{A}\cup B)$, $P(\overline{A}\cap\overline{B})$ and $P(\overline{B})$. Also examine whether the events A and B are : (a) Equally likely, (b) Exhaustive, (c) Mutually exclusive, and (d) Independent.

Solution.

The probabilities of various events are obtained as follows:

$$P(B)=P(\overline{A}\cap B)+P(A\cap B)=\frac{1}{6}+\frac{1}{3}=\frac{1}{2}$$

$$P(A\cup B)=\frac{2}{3}+\frac{1}{2}-\frac{1}{3}=\frac{5}{6}$$

$$P(A/B)=\frac{P(A\cap B)}{P(B)}=\frac{1}{3}\times\frac{2}{1}=\frac{2}{3}$$

$$P(B/A)=\frac{P(A\cap B)}{P(A)}=\frac{1}{3}\times\frac{3}{2}=\frac{1}{2}$$

$$P(\overline{A}\cup B)=P(\overline{A})+P(B)-P(\overline{A}\cap B)=\frac{1}{3}+\frac{1}{2}-\frac{1}{6}=\frac{2}{3}$$

$$P(\overline{A}\cap\overline{B})=1-P(A\cup B)=1-\frac{5}{6}=\frac{1}{6}$$

$$P(\overline{B})=1-P(B)=1-\frac{1}{2}=\frac{1}{2}$$

(a) Since P(A) ≠ P(B), A and B are not equally likely events.

(b) Since $P(A\cup B)\neq 1$, A and B are not exhaustive events.

(c) Since $P(A\cap B)\neq 0$, A and B are not mutually exclusive.

(d) Since $P(A)P(B)=P(A\cap B)$, A and B are independent events.

Example 39: Two players A and B toss an unbiased die alternatively. He who first throws a six wins the game. If A begins, what is the probability that B wins the game?

Solution.

Let A_i and B_i be the respective events that A and B throw a six in i th toss, i = 1, 2, ... B will win the game if any one of the following mutually exclusive events occur: $\overline{A}_1B_1$ or $\overline{A}_1\overline{B}_1\overline{A}_2B_2$ or $\overline{A}_1\overline{B}_1\overline{A}_2\overline{B}_2\overline{A}_3B_3$, etc.

Notes

Thus, $P(\text{B wins}) = \frac{5}{6} \times \frac{1}{6} + \frac{5}{6} \times \frac{5}{6} \times \frac{5}{6} \times \frac{1}{6} + \frac{5}{6} \times \frac{5}{6} \times \frac{5}{6} \times \frac{5}{6} \times \frac{5}{6} \times \frac{1}{6} + \ \ldots\ldots$

$$= \frac{5}{36}\left[1 + \left(\frac{5}{6}\right)^2 + \left(\frac{5}{6}\right)^4 + \ \ldots\ldots\ \right] = \frac{5}{36} \times \frac{1}{1 - \left(\frac{5}{6}\right)^2} = \frac{5}{11}$$

Example 40: A bag contains 5 red and 3 black balls and second bag contains 4 red and 5 black balls.

(a) If one ball is selected at random from each bag, what is the probability that both of them are of same colour?

(b) If a bag is selected at random and two balls are drawn from it, what is the probability that they are of (i) same colour, (ii) different colours?

Solution.

(a) Required Probability $= \begin{bmatrix} \text{Probability that ball} \\ \text{from both bags are red} \end{bmatrix} + \begin{bmatrix} \text{Probability that balls} \\ \text{from both bags are black} \end{bmatrix}$

$$= \frac{5}{8} \times \frac{4}{9} + \frac{3}{8} \times \frac{5}{9} = \frac{35}{72}$$

(b) Let A be the event that first bag is drawn so that $\overline{A}$ denotes the event that second bag is drawn. Since the two events are equally likely, mutually exclusive and exhaustive, we have $P(A) = P(\overline{A}) = \frac{1}{2}$.

(i) Let R be the event that two drawn balls are red and B be the event that they are black. The required probability is given by

$$= P(A)\left[P(R/A) + P(B/A)\right] + P(\overline{A})\left[P(R/\overline{A}) + P(B/\overline{A})\right]$$

$$= \frac{1}{2}\left[\frac{{}^5C_2 + {}^3C_2}{{}^8C_2}\right] + \frac{1}{2}\left[\frac{{}^4C_2 + {}^5C_2}{{}^9C_2}\right] = \frac{1}{2}\left[\frac{10+3}{28}\right] + \frac{1}{2}\left[\frac{6+10}{36}\right] = \frac{229}{504}$$

(ii) Let C denote the event that the drawn balls are of different colours. The required probability is given by

$$P(C) = P(A)P(C/A) + P(\overline{A})P(C/\overline{A})$$

$$= \frac{1}{2}\left[\frac{5 \times 3}{{}^8C_2}\right] + \frac{1}{2}\left[\frac{4 \times 5}{{}^9C_2}\right] = \frac{1}{2}\left[\frac{15}{28} + \frac{20}{36}\right] = \frac{275}{504}$$

Example 41: There are two urns U_1 and U_2. U_1 contains 9 white and 4 red balls and U_2 contains 3 white and 6 red balls. Two balls are transferred from U_1 to U_2 and then a ball is drawn from U_2. What is the probability that it is a white ball?

Solution.

Let A be the event that the two transferred balls are white, B be the event that they are red and C be the event that one is white and the other is red. Further, let W be the event that a white ball is drawn from U_2. The event W can occur with any one of the mutually exclusive events A, B and C.

$$P(W) = P(A).P(W/A) + P(B)P(W/B) + P(C)P(W/C)$$

$$= \frac{{}^{9}C_2}{{}^{13}C_2} \times \frac{5}{11} + \frac{{}^{4}C_2}{{}^{13}C_2} \times \frac{3}{11} + \frac{9 \times 4}{{}^{13}C_2} \times \frac{4}{11} = \frac{57}{143}$$

Example 42: A bag contains tickets numbered as 112, 121, 211 and 222. One ticket is drawn at random from the bag. Let E_i (i = 1, 2, 3) be the event that *i* th digit on the ticket is 2. Discuss the independence of E_1, E_2 and E_3.

Solution.

The event E_1 occurs if the number on the drawn ticket 211 or 222, therefore, $P(E_1) = \frac{1}{2}$. Similarly $P(E_2) = \frac{1}{2}$ and $P(E_3) = \frac{1}{2}$.

Now $P(E_i \cap E_j) = \frac{1}{4}$ (i, j = 1, 2, 3 and i $\neq$ j).

Since $P(E_i \cap E_j) = P(E_i)P(E_j)$ for i $\neq$ j, therefore E_1, E_2 and E_3 are pair-wise independent.

Further, $P(E_1 \cap E_2 \cap E_3) = \frac{1}{4} \neq P(E_1).P(E_2).P(E_3)$, therefore, E_1, E_2 and E_3 are not mutually independent.

Example 43: Probability that an electric bulb will last for 150 days or more is 0.7 and that it will last at the most 160 days is 0.8. Find the probability that it will last between 150 to 160 days.

Solution.

Let A be the event that the bulb will last for 150 days or more and B be the event that it will last at the most 160 days. It is given that P(A) = 0.7 and P(B) = 0.8.

The event $A \cup B$ is a certain event because at least one of A or B is bound to occur. Thus, $P(A \cup B) = 1$. We have to find $P(A \cap B)$. This probability is given by

$$P(A \cap B) = P(A) + P(B) - P(A \cup B) = 0.7 + 0.8 - 1.0 = 0.5$$

Example 44: The odds that A speaks the truth are 2 : 3 and the odds that B speaks the truth are 4 : 5. In what percentage of cases they are likely to contradict each other on an identical point?

Solution.

Let A and B denote the respective events that A and B speak truth. It is given that $P(A) = \frac{2}{5}$ and $P(B) = \frac{4}{9}$.

The event that they contradict each other on an identical point is given by $(A \cap \bar{B}) \cup (\bar{A} \cap B)$, where $(A \cap \bar{B})$ and $(\bar{A} \cap B)$ are mutually exclusive. Also A and B are independent events. Thus, we have

$$P\left[(A \cap \bar{B}) \cup (\bar{A} \cap B)\right] = P(A \cap \bar{B}) + P(\bar{A} \cap B) = P(A).P(\bar{B}) + P(\bar{A}).P(B)$$

Notes

$$= \frac{2}{5} \times \frac{5}{9} + \frac{3}{5} \times \frac{4}{9} = \frac{22}{45} = 0.49$$

Hence, A and B are likely to contradict each other in 49% of the cases.

Example 45: The probability that a student A solves a mathematics problem is $\frac{2}{5}$ and the probability that a student B solves it is $\frac{2}{3}$. What is the probability that (a) the problem is not solved, (b) the problem is solved, (c) Both A and B, working independently of each other, solve the problem?

Solution.

Let A and B be the respective events that students A and B solve the problem. We note that A and B are independent events.

$$(a)\;\; P\left(\bar{A} \cap \bar{B}\right) = P\left(\bar{A}\right).P\left(\bar{B}\right) = \frac{3}{5} \times \frac{1}{3} = \frac{1}{5}$$

$$(b)\;\; P(A \cup B) = 1 - P\left(\bar{A} \cap \bar{B}\right) = 1 - \frac{1}{5} = \frac{4}{5}$$

$$(c)\;\; P(A \cap B) = P(A)P(B) = \frac{2}{5} \times \frac{2}{3} = \frac{4}{15}$$

Example 46: A bag contains 8 red and 5 white balls. Two successive drawings of 3 balls each are made such that

(i) balls are replaced before the second trial, (ii) balls are not replaced before the second trial. Find the probability that the first drawing will give 3 white and the second 3 red balls.

Solution.

Let A be the event that all the 3 balls obtained at the first draw are white and B be the event that all the 3 balls obtained at the second draw are red.

(a) When balls are replaced before the second draw, we have

$$P(A) = \frac{{}^{5}C_3}{{}^{13}C_3} = \frac{5}{143} \text{ and } P(B) = \frac{{}^{8}C_3}{{}^{13}C_3} = \frac{28}{143}$$

The required probability is given by $P(A \cap B)$, where A and B are independent. Thus, we have

$$P(A \cap B) = P(A).P(B) = \frac{5}{143} \times \frac{28}{143} = \frac{140}{20449}$$

(b) When the balls are not replaced before the second draw

We have $P(B/A) = \frac{{}^{8}C_3}{{}^{10}C_3} = \frac{7}{15}$. Thus, we have

$$P(A \cap B) = P(A).P(B/A) = \frac{5}{143} \times \frac{7}{15} = \frac{7}{429}$$

Example 47: Computers A and B are to be marketed. A salesman who is assigned the job of finding customers for them has 60% and 40% chances respectively of succeeding in case of computer A and B. The two computers can be sold independently. Given that the salesman is able to sell at least one computer, what is the probability that computer A has been sold?

Solution.

Let A be the event that the salesman is able to sell computer A and B be the event that he is able to sell computer B. It is given that P(A) = 0.6 and P(B) = 0.4. The probability that the salesman is able to sell at least one computer, is given by

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

(note that A and B are given to be independent)

$$= 0.6 + 0.4 - 0.6 \times 0.4 = 0.76$$

Now the required probability, the probability that computer A is sold given that the salesman is able to sell at least one computer, is given by

$$P(A / A \cup B) = \frac{0.60}{0.76} = 0.789$$

Example 48: Two men M_1 and M_2 and three women W_1, W_2 and W_3, in a big industrial firm, are trying for promotion to a single post which falls vacant. Those of the same sex have equal probabilities of getting promotion but each man is twice as likely to get the promotion as any women.

(a) Find the probability that a woman gets the promotion.

(b) If M_2 and W_2 are husband and wife, find the probability that one of them gets the promotion.

Solution.

Let p be the probability that a woman gets the promotion, therefore 2p will be the probability that a man gets the promotion. Thus, we can write, $P(M_1) = P(M_2) = 2p$ and $P(W_1) = P(W_2) = P(W_3) = p$, where $P(M_i)$ denotes the probability that i th man gets the promotion (i = 1, 2) and $P(W_j)$ denotes the probability that j th woman gets the promotion.

Since the post is to be given only to one of the five persons, the events M_1, M_2 , W_1, W_2 and W_3 are mutually exclusive and exhaustive.

$$\therefore\ P(M_1 \cup M_2 \cup W_1 \cup W_2 \cup W_3) = P(M_1) + P(M_2) + P(W_1) + P(W_2) + P(W_3) = 1$$

$$\Rightarrow\ 2p + 2p + p + p + p = 1 \text{ or } p = \frac{1}{7}$$

(a) The probability that a woman gets the promotion

$$P(W_1 \cup W_2 \cup W_3) = P(W_1) + P(W_2) + P(W_3) = \frac{3}{7}$$

(b) The probability that M_2 or W_2 gets the promotion

$$P(M_2 \cup W_2) = P(M_2) + P(W_2) = \frac{3}{7}$$

Notes

Example 49: An unbiased die is thrown 8 times. What is the probability of getting a six in at least one of the throws?

Solution.

Let A_i be the event that a six is obtained in the ith throw (i = 1, 2, ... 8). Therefore, $P(A_i) = \frac{1}{6}$.

The event that a six is obtained in at least one of the throws is represented by $(A_1 \cup A_2 \cup \text{ .... } \cup A_8)$. Thus, we have

$$P(A_1 \cup A_2 \cup \text{ .... } \cup A_8) = 1 - P(\bar{A}_1 \cap \bar{A}_2 \cap \text{ .... } \cap \bar{A}_8)$$

Since A_1, A_2, ... A_8 are independent, we can write

$$P(A_1 \cup A_2 \cup \text{ .... } \cup A_8) = 1 - P(\bar{A}_1).P(\bar{A}_2). \text{ .... } P(\bar{A}_8) = 1 - \left(\frac{5}{6}\right)^8.$$

Example 50: Two students X and Y are very weak students of mathematics and their chances of solving a problem correctly are 0.11 and 0.14 respectively. If the probability of their making a common mistake is 0.081 and they get the same answer, what is the chance that their answer is correct?

Solution.

Let A be the event that both the students get a correct answer, B be the event that both get incorrect answer by making a common mistake and C be the event that both get the same answer. Thus, we have

$$P(A \cap C) = P(X \text{ gets correct answer}).P(Y \text{ gets correct answer})$$

$$= 0.11 \times 0.14 = 0.0154 \text{ (note that the two events are independent)}$$

Similarly,

$$P(B \cap C) = P(X \text{ gets incorrect answer}) \times P(Y \text{ gets incorrect answer})$$

$$\times \text{P}(X \text{ and } Y \text{ make a common mistake})$$

$$= (1 - 0.11)(1 - 0.14) \times 0.081 = 0.062$$

Further, $C = (A \cap C) \cup (B \cap C)$ or $P(C) = P(A \cap C) + P(B \cap C)$, since $(A \cap C)$ and $(B \cap C)$ are mutually exclusive. Thus, we have

$$P(C) = 0.0154 + 0.0620 = 0.0774$$

We have to find the probability that the answers of both the students are correct given that they are same, i.e.,

$$P(A/C) = \frac{P(A \cap C)}{P(C)} = \frac{0.0154}{0.0774} = 0.199$$

Notes

Example 51: Given below are the daily wages (in rupees) of six workers of a factory :

77, 105, 91, 100, 90, 83

If two of these workers are selected at random to serve as representatives, what is the probability that at least one will have a wage lower than the average?

Solution.

The average wage $\bar{X} = \dfrac{77+105+91+100+90+83}{6} = 91$

Let A be the event that two workers selected at random have their wages greater than or equal to average wage.

$$\therefore \quad P(A) = \frac{{}^3C_2}{{}^6C_2} = \frac{1}{5}$$

Thus, the probability that at least one of the workers has a wage less than the average $= 1 - \dfrac{1}{5} = \dfrac{4}{5}$

Example 52: There are two groups of subjects one of which consists of 5 science subjects and 3 engineering subjects and the other consists of 3 science subjects and 5 engineering subjects. An unbiased die is cast. If the number 3 or 5 turns up, a subject from the first group is selected at random otherwise a subject is randomly selected from the second group. Find the probability that an engineering subject is selected ultimately.

Solution.

Let A be the event that an engineering subject is selected and B be the event that 3 or 5 turns on the die. The given information can be summarised into symbols, as given below :

$$P(B) = \frac{1}{3}, \quad P(A/B) = \frac{3}{8}, \quad \text{and} \quad P(A/\bar{B}) = \frac{5}{8}$$

To find P(A), we write

$$P(A) = P(A \cap B) + P(A \cap \bar{B}) = P(B).P(A/B) + P(\bar{B}).P(A/\bar{B})$$

$$= \frac{1}{3} \times \frac{3}{8} + \frac{2}{3} \times \frac{5}{8} = \frac{13}{24}$$

Example 53: Find the probability of obtaining two heads in the toss of two unbiased coins when (a) at least one of the coins shows a head, (b) second coin shows a head.

Solution.

Let A be the event that both coins show heads, B be the event that at least one coin shows a head and C be the event that second coin shows a head. The sample space and the three events can be written as:

S = {(H, H), (H, T), (T, H), (T, T)},   A = {(H, H)},

B = {(H, H), (H, T), (T, H)} and   C = {(H, H), (T, H)}.

Further, $A \cap B = \{(H, H)\}$ and $A \cap C = \{(H, H)\}$

Notes

Since the coins are given to be unbiased, the elementary events are equally likely, therefore

$$P(A)=\frac{1}{4},\quad P(B)=\frac{3}{4},\quad P(C)=\frac{1}{2},\quad P(A\cap B)=P(A\cap C)=\frac{1}{4}$$

(a) We have to determine P (A/B)

$$P(A/B)=\frac{P(A\cap B)}{P(B)}=\frac{1}{4}\times\frac{4}{3}=\frac{1}{3}$$

(b) We have to determine P(A/C)

$$P(A/C)=\frac{P(A\cap C)}{P(C)}=\frac{1}{4}\times\frac{2}{1}=\frac{1}{2}$$

7.3 Theorems on Probability

Theorem 6. Bayes Theorem or Inverse Probability Rule

The probabilities assigned to various events on the basis of the conditions of the experiment or by actual experimentation or past experience or on the basis of personal judgement are called prior probabilities. One may like to revise these probabilities in the light of certain additional or new information. This can be done with the help of Bayes Theorem, which is based on the concept of conditional probability. The revised probabilities, thus obtained, are known as posterior or inverse probabilities. Using this theorem it is possible to revise various business decisions in the light of additional information.

Bayes Theorem

If an event D can occur only in combination with any of the n mutually exclusive and exhaustive events A_1, A_2, ... A_n and if, in an actual observation, D is found to have occurred, then the probability that it was preceded by a particular event A_k is given by

$$P(A_k/D)=\frac{P(A_k).P(D/A_k)}{\sum_{i=1}^{n}P(A_i).P(D/A_i)}$$

Proof.

Since A_1, A_2, ... A_n are n exhaustive events, therefore,
$S=A_1\cup A_2\ \ldots\ldots\ \cup A_n$.

Since D is another event that can occur in combination with any of the mutually exclusive and exhaustive events A_1, A_2, ... A_n, we can write

$$D=(A_1\cap D)\cup(A_2\cap D)\cup\ \ldots\ldots\ \cup(A_n\cap D)$$

Taking probability of both sides, we get

$$P(D)=P(A_1\cap D)+P(A_2\cap D)+\ \ldots\ldots\ +P(A_n\cap D)$$

We note that the events $(A_1\cap D),(A_2\cap D)$, etc. are mutually exclusive.

$$P(D)=\sum_{i=1}^{n}P(A_i\cap D)=\sum_{i=1}^{n}P(A_i).P(D/A_i) \qquad \text{.... (1)}$$

The conditional probability of an event A_k given that D has already occurred, is given by

Notes

$$P(A_k/D) = \frac{P(A_k \cap D)}{P(D)} = \frac{P(A_k).P(D/A_k)}{P(D)} \quad \text{.... (2)}$$

Substituting the value of P(D) from (1), we get

$$P(A_k/D) = \frac{P(A_k).P(D/A_k)}{\sum_{i=1}^{n} P(A_i).P(D/A_i)} \quad \text{.... (3)}$$

Example 54: A manufacturing firm purchases a certain component, for its manufacturing process, from three sub-contractors A, B and C. These supply 60%, 30% and 10% of the firm's requirements, respectively. It is known that 2%, 5% and 8% of the items supplied by the respective suppliers are defective. On a particular day, a normal shipment arrives from each of the three suppliers and the contents get mixed. A component is chosen at random from the day's shipment:

(a) What is the probability that it is defective?

(b) If this component is found to be defective, what is the probability that it was supplied by (i) A, (ii) B, (iii) C ?

Solution.

Let A be the event that the item is supplied by A. Similarly, B and C denote the events that the item is supplied by B and C respectively. Further, let D be the event that the item is defective. It is given that:

P(A) = 0.6, P(B) = 0.3, P(C) = 0.1, P(D/A) = 0.02

P(D/B) = 0.05, P(D/C) = 0.08.

(a) We have to find P(D)

From equation (1), we can write

$$P(D) = P(A \cap D) + P(B \cap D) + P(C \cap D)$$

$$= P(A)P(D/A) + P(B)P(D/B) + P(C)P(D/C)$$

$$= 0.6 \times 0.02 + 0.3 \times 0.05 + 0.1 \times 0.08 = 0.035$$

(b) (i) We have to find P(A/D)

$$P(A/D) = \frac{P(A)P(D/A)}{P(D)} = \frac{0.6 \times 0.02}{0.035} = 0.343$$

Similarly, (ii) $P(B/D) = \frac{P(B)P(D/B)}{P(D)} = \frac{0.3 \times 0.05}{0.035} = 0.429$

and (iii) $P(C/D) = \frac{P(C)P(D/C)}{P(D)} = \frac{0.1 \times 0.08}{0.035} = 0.228$

Notes

Alternative Method :

The above problem can also be attempted by writing various probabilities in the form of following table :

	A	B	C	Total
D	$P(A \cap D) = 0.012$	$P(B \cap D) = 0.015$	$P(C \cap D) = 0.008$	0.035
$\bar{D}$	$P(A \cap \bar{D}) = 0.588$	$P(B \cap \bar{D}) = 0.285$	$P(C \cap \bar{D}) = 0.092$	0.965
Total	0.600	0.300	0.100	1.000

Thus $P(A/D) = \dfrac{0.012}{0.035}$ etc.

Example 55: A box contains 4 identical dice out of which three are fair and the fourth is loaded in such a way that the face marked as 5 appears in 60% of the tosses. A die is selected at random from the box and tossed. If it shows 5, what is the probability that it was a loaded die?

Solution.

Let A be the event that a fair die is selected and B be the event that the loaded die is selected from the box.

Then, we have $P(A) = \dfrac{3}{4}$ and $P(B) = \dfrac{1}{4}$.

Further, let D be the event that 5 is obtained on the die, then

$$P(D/A) = \frac{1}{6} \text{ and } P(D/B) = \frac{6}{10}$$

Thus, $P(D) = P(A).P(D/A) + P(B).P(D/B) = \dfrac{3}{4} \times \dfrac{1}{6} + \dfrac{1}{4} \times \dfrac{6}{10} = \dfrac{11}{40}$

We want to find P(B/D), which is given by

$$P(B/D) = \frac{P(B \cap D)}{P(D)} = \frac{1}{4} \times \frac{6}{10} \times \frac{40}{11} = \frac{6}{11}$$

Example 56: A bag contains 6 red and 4 white balls. Another bag contains 3 red and 5 white balls. A fair die is tossed for the selection of bag. If the die shows 1 or 2, the first bag is selected otherwise the second bag is selected. A ball is drawn from the selected bag and is found to be red. What is the probability that the first bag was selected?

Solution.

Let A be the event that first bag is selected, B be the event that second bag is selected and D be the event of drawing a red ball.

Then, we can write

$$P(A)=\frac{1}{3},\ P(B)=\frac{2}{3},\ P(D/A)=\frac{6}{10},\ P(D/B)=\frac{3}{8}$$

Further, $P(D)=\frac{1}{3}\times\frac{6}{10}+\frac{2}{3}\times\frac{3}{8}=\frac{9}{20}$.

$\therefore\ P(A/D)=\frac{P(A\cap D)}{P(D)}=\frac{1}{3}\times\frac{6}{10}\times\frac{20}{9}=\frac{4}{9}$

Example 57: In a certain recruitment test there are multiple-choice questions. There are 4 possible answers to each question out of which only one is correct. An intelligent student knows 90% of the answers while a weak student knows only 20% of the answers.

(i) An intelligent student gets the correct answer, what is the probability that he was guessing?

(ii) A weak student gets the correct answer, what is the probability that he was guessing?

Solution.

Let A be the event that an intelligent student knows the answer, B be the event that the weak student knows the answer and C be the event that the student gets a correct answer.

(i) We have to find $P(\bar{A}/C)$. We can write

$$P(\bar{A}/C)=\frac{P(\bar{A}\cap C)}{P(C)}=\frac{P(\bar{A})P(C/\bar{A})}{P(\bar{A})P(C/\bar{A})+P(A)P(C/A)} \quad .... (1)$$

It is given that P(A) = 0.90, $P(C/\bar{A})=\frac{1}{4}=0.25$ and $P(C/A)=1.0$

From the above, we can also write $P(\bar{A})=0.10$

Substituting these values, we get

$$P(\bar{A}/C)=\frac{0.10\times 0.25}{0.10\times 0.25+0.90\times 1.0}=\frac{0.025}{0.925}=0.027$$

(ii) We have to find $P(\bar{B}/C)$. Replacing $\bar{A}$ by $\bar{B}$, in equation (1), we can get this probability.

It is given that P(B) = 0.20, $P(C/\bar{B})=0.25$ and $P(C/B)=1.0$

From the above, we can also write $P(\bar{B})=0.80$

Thus, we get $P(\bar{B}/C)=\frac{0.80\times 0.25}{0.80\times 0.25+0.20\times 1.0}=\frac{0.20}{0.40}=0.50$

Example 58: An electronic manufacturer has two lines A and B assembling identical electronic units. 5% of the units assembled on line A and 10%of those assembled on line B are defective. All defective units must be reworked at a significant increase in cost. During the last eight-hour shift, line A produced 200 units while the line B produced 300 units. One unit is selected at random from the 500 units produced and is found to be defective. What is the probability that it was assembled (i) on line A, (ii) on line B?

Answer the above questions if the selected unit was found to be non-defective.

Notes

Solution.

Let A be the event that the unit is assembled on line A, B be the event that it is assembled on line B and D be the event that it is defective.

Thus, we can write

$$P(A)=\frac{2}{5},\ P(B)=\frac{3}{5},\ P(D/A)=\frac{5}{100}\ \text{and}\ P(D/B)=\frac{10}{100}$$

Further, we have

$$P(A\cap D)=\frac{2}{5}\times\frac{5}{100}=\frac{1}{50}\ \text{and}\ P(B\cap D)=\frac{3}{5}\times\frac{10}{100}=\frac{3}{50}$$

The required probabilities are computed form the following table:

	A	*B*	*Total*
D	$\frac{1}{50}$	$\frac{3}{50}$	$\frac{4}{50}$
$\overline{D}$	$\frac{19}{50}$	$\frac{27}{50}$	$\frac{46}{50}$
Total	$\frac{20}{50}$	$\frac{30}{50}$	1

From the above table, we can write

$$P(A/D)=\frac{1}{50}\times\frac{50}{4}=\frac{1}{4},\ P(B/D)=\frac{3}{50}\times\frac{50}{4}=\frac{3}{4}$$

$$P(A/\overline{D})=\frac{19}{50}\times\frac{50}{46}=\frac{19}{46},\ P(B/\overline{D})=\frac{27}{50}\times\frac{50}{46}=\frac{27}{46}$$

7.4 Summary of Formulae

1. (a) The number of permutations of n objects taking n at a time are n!

   (b) The number of permutations of n objects taking r at a time, are ${}^{n}P_r=\frac{n!}{(n-r)!}$

   (c) The number of permutations of n objects in a circular order are (n - 1)!

   (d) The number of permutations of n objects out of which n_1 are alike, n_2 are alike, ... n_k are alike, are $\frac{n!}{n_1!n_2!\ ...\ n_k!}$

   (e) The number of combinations of n objects taking r at a time are ${}^{n}C_r=\frac{n!}{r!(n-r)!}$

2. (a) The probability of occurrence of at least one of the two events A and B is given by : $P(A\cup B)=P(A)+P(B)-P(A\cap B)=1-P(\overline{A}\cap\overline{B})$.

   (b) The probability of occurrence of exactly one of the events A or B is given by : $P(A\cap\overline{B})+P(\overline{A}\cap B)$ or $P(A\cup B)-P(A\cap B)$

Notes

3. (a) The probability of simultaneous occurrence of the two events A and B is given by:
$P(A \cap B) = P(A).P(B/A)$ or $= P(B).P(A/B)$

(b) If A and B are independent $P(A \cap B) = P(A).P(B)$.

4. Bayes Theorem :

$$P(A_k/D) = \frac{P(A_k).P(D/A_k)}{\sum_{i=1}^{n} P(A_i).P(D/A_i)}, \quad (k = 1, 2, \ldots\ldots n)$$

Here A_1, A_2, ... A_n are n mutually exclusive and exhaustive events.

7.5 Keywords

Mutually exclusive outcomes: Two or more outcomes of an experiment are said to be mutually exclusive if the occurrence of one of them precludes the occurrence of all others in the same trial i.e. they cannot occur jointly. For example, the two possible outcomes of toss of a coin are mutually exclusive. Similarly, the occurrences of the numbers 1, 2, 3, 4, 5, 6 in the roll of a six faced die are mutually exclusive.

Exhaustive outcomes: It is the totality of all possible outcomes of a random experiment. The number of exhaustive outcomes in the roll of a die are six. Similarly, there are 52 exhaustive outcomes in the experiment of drawing a card from a pack of 52 cards.

7.6 Self Assessment

Choose the appropriate answer:

1. Two cards are drawn successively without replacement from a well-shuffled pack of 52 cards. The probability that one of them is king and the other is queen is

   (a) $\frac{8}{13 \times 51}$ (b) $\frac{4}{13 \times 51}$ (c) $\frac{1}{13 \times 17}$ (d) none of these.

2. Two unbiased dice are rolled. The chance of obtaining an even sum is

   (a) $\frac{1}{4}$ (b) $\frac{1}{2}$ (c) $\frac{1}{3}$ (d) none of these.

3. Two unbiased dice are rolled. The chance of obtaining a six only on the second die is

   (a) $\frac{5}{6}$ (b) $\frac{1}{6}$ (c) $\frac{1}{4}$ (d) none of these.

4. If $P(A) = \frac{4}{5}$, then odds against $\overline{A}$ are

   (a) 1 : 4 (b) 5 : 4 (c) 4 : 5 (d) none of these.

5. The probability of occurrence of an event A is 0.60 and that of B is 0.25. If A and B are mutually exclusive events, then the probability of occurrence of neither of them is

   (a) 0.35 (b) 0.75 (c) 0.15 (d) none of these.

6. The probability of getting at least one head in 3 throws of an unbiased coin is

   (a) $\frac{1}{8}$ (b) $\frac{7}{8}$ (c) $\frac{3}{8}$ (d) none of these.

Notes

7.7 Review Questions

1. What is the probability of getting exactly two heads in three throws of an unbiased coin?
2. What is the probability of getting a sum of 2 or 8 or 12 in single throw of two unbiased dice?
3. Two cards are drawn at random from a pack of 52 cards. What is the probability that the first is a king and second is a queen?
4. What is the probability of successive drawing of an ace, a king, a queen and a jack from a pack of 52 well shuffled cards? The drawn cards are not replaced.
5. 5 unbiased coins with faces marked as 2 and 3 are tossed. Find the probability of getting a sum of 12.
6. If 15 chocolates are distributed at random among 5 children, what is the probability that a particular child receives 8 chocolates?
7. A and B stand in a ring with 10 other persons. If arrangement of 12 persons is at random, find the chance that there are exactly three persons between A and B.
8. Two different digits are chosen at random from the set 1, 2, 3, 4, 5, 6, 7, 8. Find the probability that sum of two digits exceeds 13.
9. From each of the four married couples one of the partner is selected at random. What is the probability that they are of the same sex?
10. A bag contains 5 red and 4 green balls. Two draws of three balls each are done with replacement of balls in the first draw. Find the probability that all the three balls are red in the first draw and green in the second draw.
11. Two dice are thrown two times. What is the probability of getting a sum 10 in the first and 11 in the second throw?
12. 4 cards are drawn successively one after the other without replacement. What is the probability of getting cards of the same denominations?

Answers: Self Assessment

1. (a) 2. (b) 3. (d) 4. (d) 5. (c) 6. (b)

7.8 Further Readings

Books

Introductory Probability and Statistical Applications by P.L. Meyer

Introduction to Mathematical Statistics by Hogg and Craig

Fundamentals of Mathematical Statistics by S.C. Gupta and V.K. Kapoor

Unit 8: Expected Value with Perfect Information (EVPI)

Notes

CONTENTS

Objectives

Introduction

8.1 Expected Value with Perfect Information (EVPI)

8.1.1 Cost of Uncertainty

8.1.2 Marginal Analysis

8.2 Use of Subjective Probabilities in Decision Making

8.3 Use of Posterior Probabilities in Decision Making

8.4 Summary

8.5 Keywords

8.6 Self Assessment

8.7 Review Questions

8.8 Further Readings

Objectives

After studying this unit, you will be able to:

- Define expected value
- Describe expected value with perfect information

Introduction

In last unit, you will studied about random variable. This unit will provide you information related to expected value with perfect information.

8.1 Expected Value with Perfect Information (EVPI)

The expected value with perfect information is the amount of profit foregone due to uncertain conditions affecting the selection of a course of action.

Given the perfect information, a decision maker is supposed to know which particular state of nature will be in effect. Thus, the procedure for the selection of an optimal course of action, for the decision problem given in example 18, will be as follows:

If the decision maker is certain that the state of nature S_1 will be in effect, he would select the course of action A_3, having maximum payoff equal to Rs 200.

Similarly, if the decision maker is certain that the state of nature S_2 will be in effect, his course of action would be A_1 and if he is certain that the state of nature S_3 will be in effect, his course of action would be A_2. The maximum payoffs associated with the actions are Rs 200 and Rs 600 respectively.

Notes

The weighted average of these payoffs with weights equal to the probabilities of respective states of nature is termed as Expected Payoff under Certainty (EPC).

Thus, $EPC = 200 \times 0.3 + 200 \times 0.4 + 600 \times 0.3 = 320$

The difference between EPC and EMV of optimal action is the amount of profit foregone due to uncertainty and is equal to EVPI.

Thus, EVPI = EPC – EMV of optimal action = 320 – 194 = 126

It is interesting to note that EVPI is also equal to EOL of the optimal action.

8.1.1 Cost of Uncertainty

This concept is similar to the concept of EVPI. Cost of uncertainty is the difference between the EOL of optimal action and the EOL under perfect information.

Given the perfect information, the decision maker would select an action with minimum opportunity loss under each state of nature. Since minimum opportunity loss under each state of nature is zero, therefore,

EOL under certainty $= 0 \times 0.3 + 0 \times 0.4 + 0 \times 0.3 = 0$.

Thus, the cost of uncertainty = EOL of optimal action = EVPI

Example 19: A group of students raise money each year by selling souvenirs outside the stadium of a cricket match between teams A and B. They can buy any of three different types of souvenirs from a supplier. Their sales are mostly dependent on which team wins the match. A conditional payoff (in Rs.) table is as under :

Type of Souvenir →	*I*	*II*	*III*
Team A wins	1200	800	300
Team B wins	250	700	1100

(i) Construct the opportunity loss table.

(ii) Which type of souvenir should the students buy if the probability of team A's winning is 0.6?

(iii) Compute the cost of uncertainty.

Solution.

(i) The Opportunity Loss Table

Actions → / *Events* ↓	*Type of Souvenir bought* *I*	*II*	*III*
Team A wins	0	400	900
Team B wins	850	400	0

(ii) EOL of buying type I Souvenir $= 0 \times 0.6 + 850 \times 0.4 = 340$

EOL of buying type II Souvenir $= 400 \times 0.6 + 400 \times 0.4 = 400.$

EOL of buying type III Souvenir $= 900 \times 0.6 + 0 \times 0.4 = 540.$

Notes

Since the EOL of buying Type I Souvenir is minimum, the optimal decision is to buy Type I Souvenir.

(iii) Cost of uncertainty = EOL of optimal action = Rs. 340

Example 20:

The following is the information concerning a product X :

(i) Per unit profit is Rs 3.

(ii) Salvage loss per unit is Rs 2.

(iii) Demand recorded over 300 days is as under:

Units demanded	:	5	6	7	8	9
No. of days	:	30	60	90	75	45

Find: (i) EMV of optimal order.

(ii) Expected profit presuming certainty of demand.

Solution.

(i) The given data can be rewritten in terms of relative frequencies, as shown below :

Units demanded	:	5	6	7	8	9
No. of days	:	0.1	0.2	0.3	0.25	0.15

From the above probability distribution, it is obvious that the optimum order would lie between and including 5 to 9.

Let A denote the number of units ordered and D denote the number of units demanded per day.

If $D \geq A$, profit per day = 3A, and if D < A, profit per day = 3D - 2(A - D) = 5D - 2A.

Thus, the profit matrix can be written as

Units Demanded	5	6	7	8	9	
Probability $\rightarrow$ *Action (units ordered)* $\downarrow$	0.10	0.20	0.30	0.25	0.15	*EMV*
5	15	15	15	15	15	15.00
6	13	18	18	18	18	17.50
7	11	16	21	21	21	19.00
8	9	14	19	24	24	19.00
9	7	12	17	22	27	17.75

From the above table, we note that the maximum EMV = 19.00, which corresponds to the order of 7 or 8 units. Since the order of the 8th unit adds nothing to the EMV, i.e., marginal EMV is zero, therefore, order of 8 units per day is optimal.

(ii) Expected profit under certainty

$$= (5 \times 0.10 + 6 \times 0.20 + 7 \times 0.30 + 8 \times 0.25 + 9 \times 0.15) \times 3 = \text{Rs } 21.45$$

Notes

Alternative Method

The work of computations of EMV's, in the above example, can be reduced considerably by the use of the concept of expected marginal profit. Let π be the marginal profit and *l* be the marginal loss of ordering an additional unit of the product. Then, the expected marginal profit of ordering the Ath unit, is given by

$$= \pi.P(D \geq A) - \lambda.P(D < A) = \pi.P(D \geq A) - \lambda.\left[1 - P(D \geq A)\right]$$

$$= (\pi + \lambda).P(D \geq A) - \lambda \qquad \text{.... (1)}$$

The computations of EMV, for alternative possible values of A, are shown in the following table:

In our example, $\pi = 3$ and $\lambda = 2$.

Thus, the expression for the expected marginal profit of the Ath unit

$$= (3+2)P(D \geq A) - 2 = 5P(D \geq A) - 2$$

Table for computations

$Action(A)$	$P(D \geq A)^*$	$EMP = 5P(D \geq A) - 2$	*Total profit or EMV*
5	1.00	$5 \times 1.00 - 2 = 3.00$	$5 \times 3.00 = 15.00$
6	0.90	$5 \times 0.90 - 2 = 2.50$	$15.00 + 2.50 = 17.50$
7	0.70	$5 \times 0.70 - 2 = 1.50$	$17.50 + 1.50 = 19.00$
8	0.40	$5 \times 0.40 - 2 = 0.00$	$19.00 + 0.00 = 19.00$
9	0.15	$5 \times 0.15 - 2 = -1.25$	$19.00 - 1.25 = 17.75$

* This column represents the 'more than type' cumulative probabilities.

Since the expected marginal profit (EMP) of the 8th unit is zero, therefore, optimal order is 8 units.

8.1.2 Marginal Analysis

Marginal analysis is used when the number of states of nature is considerably large. Using this analysis, it is possible to locate the optimal course of action without the computation of EMV's of various actions.

An order of A units is said to be optimal if the expected marginal profit of the Ath unit is non-negative and the expected marginal profit of the (A + 1)th unit is negative. Using equation (1), we can write

$$(\pi + \lambda)P(D \geq A) - \lambda \geq 0 \text{ and} \qquad \text{.... (2)}$$

$$(\pi + \lambda)P(D \geq A + 1) - \lambda < 0 \qquad \text{.... (3)}$$

From equation (2), we get

$$P(D \geq A) \geq \frac{\lambda}{\pi + \lambda} \quad \text{or} \quad 1 - P(D < A) \geq \frac{\lambda}{\pi + \lambda}$$

$$\text{or } P(D < A) \leq 1 - \frac{\lambda}{\pi + \lambda} \quad \text{or} \quad P(D \leq A - 1) \leq \frac{\pi}{\pi + \lambda} \qquad \text{.... (4)}$$

[P(D ≤ A – 1) = P(D < A), since A is an integer]

Notes

Further, equation (3) gives

$$P(D \geq A+1) < \frac{\lambda}{\pi+\lambda} \quad \text{or} \quad 1 - P(D < A+1) < \frac{\lambda}{\pi+\lambda}$$

$$\text{or} \quad P(D < A+1) > 1 - \frac{\lambda}{\pi+\lambda} \quad \text{or} \quad P(D \leq A) > \frac{\pi}{\pi+\lambda} \qquad \text{.... (5)}$$

Combining (4) and (5), we get

$$P(D \leq A-1) \leq \frac{\pi}{\pi+\lambda} < P(D \leq A).$$

Writing the probability distribution, given in example 20, in the form of less than type cumulative probabilities which is also known as the distribution function F(D), we get

Units demanded(D)	:	5	6	7	8	9
$F(D)$	:	0.1	0.3	0.6	0.85	1.00

We are given $\pi = 3$ and $\lambda = 2$, $\therefore \quad \frac{\pi}{\pi+\lambda} = \frac{3}{5} = 0.6$

Since the next cumulative probability, i.e., 0.85, corresponds to 8 units, hence, the optimal order is 8 units.

8.2 Use of Subjective Probabilities in Decision Making

When the objective probabilities of the occurrence of various states of nature are not known, the same can be assigned on the basis of the expectations or the degree of belief of the decision maker. Such probabilities are known as subjective or personal probabilities. It may be pointed out that different individuals may assign different probability values to given states of nature.

This indicates that a decision problem under uncertainty can always be converted into a decision problem under risk by the use of subjective probabilities. Such an approach is also termed as Subjectivists's Approach.

Example 21:

The conditional payoff (in Rs) for each action-event combination are as under:

Action → *Event* ↓	1	2	3	4
A	4	−2	7	8
B	0	6	3	5
C	−5	9	2	−3
D	3	1	4	5
E	6	6	3	2

(i) Which is the best action in accordance with the Maximin Criterion?

(ii) Which is the best action in accordance with the EMV Criterion, assuming that all the events are equally likely?

Notes

Solution.

(i) The minimum payoffs for various actions are :

Action 1 = - 5

Action 2 = - 2

Action 3 = 2

Action 4 = - 3

Since the payoff for action 3 is maximum, therefore, A_3 is optimal on the basis of maximin criterion.

(ii) Since there are 5 equally likely events, the probability of each of them would be $\frac{1}{5}$.

Thus, the EMV of action 1, i.e., $EMV_1 = \frac{4+0-5+3+6}{5} = \frac{8}{5} = 1.6$

Similarly, $EMV_2 = \frac{20}{5} = 4.0$, $EMV_3 = \frac{19}{5} = 3.8$ and $EMV_4 = \frac{17}{5} = 3.4$

Thus, action 2 is optimal.

8.3 Use of Posterior Probabilities in Decision Making

The probability values of various states of nature, discussed so far, were prior probabilities. Such probabilities are either computed from the past data or assigned subjectively. It is possible to revise these probabilities in the light of current information available by using the Bayes' Theorem. The revised probabilities are known as posterior probabilities.

Example 22: A manufacturer of detergent soap must determine whether or not to expand his productive capacity. His profit per month, however, depends upon the potential demand for his product which may turn out to be high or low. His payoff matrix is given below :

	Do not Expand	*Expand*
High Demand	Rs 5,000	Rs 7,500
Low Demand	Rs 5,000	Rs 2,100

On the basis of past experience, he has estimated the probability that demand for his product being high in future is only 0.4

Before taking a decision, he also conducts a market survey. From the past experience he knows that when the demand has been high, such a survey had predicted it correctly only 60% of the times and when the demand has been low, the survey predicted it correctly only 80% of the times.

If the current survey predicts that the demand of his product is going to be high in future, determine whether the manufacturer should increase his production capacity or not? What would have been his decision in the absence of survey?

Solution.

Let H be the event that the demand will be high. Therefore,

$$P(H) = 0.4 \text{ and } P(\overline{H}) = 0.6$$

Note that H and $\overline{H}$ are the only two states of nature.

Let D be the event that the survey predicts high demand. Therefore,

$P(D/H) = 0.60$ and $P(\overline{D}/\overline{H}) = 0.80$

We have to find $P(H/D)$ and $P(\overline{H}/D)$. For this, we make the following table:

	H	$\overline{H}$	*Total*
D	$0.4 \times 0.6 = 0.24$	0.12	0.36
$\overline{D}$	0.16	$0.6 \times 0.8 = 0.48$	0.64
Total	0.40	0.60	1.00

From the above table, we can write

$$P(H/D) = \frac{0.24}{0.36} = \frac{2}{3} \text{ and } P(\overline{H}/D) = \frac{0.12}{0.36} = \frac{1}{3}$$

The EMV of the act 'don't expand' $= 5000 \times \frac{2}{3} + 5000 \times \frac{1}{3} =$ Rs 5,000

and the EMV of the act 'expand' $= 7500 \times \frac{2}{3} + 2100 \times \frac{1}{3} =$ Rs 5,700

Since the EMV of the act 'expand' > the EMV of the act 'don't expand', the manufacturer should expand his production capacity.

It can be shown that, in the absence of survey the EMV of the act 'don't expand' is Rs 5,000 and the EMV of the act expand is Rs 4,260. Hence, the optimal act is 'don't expand'.

Decision Tree Approach

The decision tree diagrams are often used to understand and solve a decision problem. Using such diagrams, it is possible to describe the sequence of actions and chance events. A decision node is represented by a square and various action branches stem from it. Similarly, a chance node is represented by a circle and various event branches stem from it. Various steps in the construction of a decision tree can be summarised as follows:

(i) Show the appropriate action-event sequence beginning from left to right of the page.

(ii) Write the probabilities of various events along their respective branches stemming from each chance node.

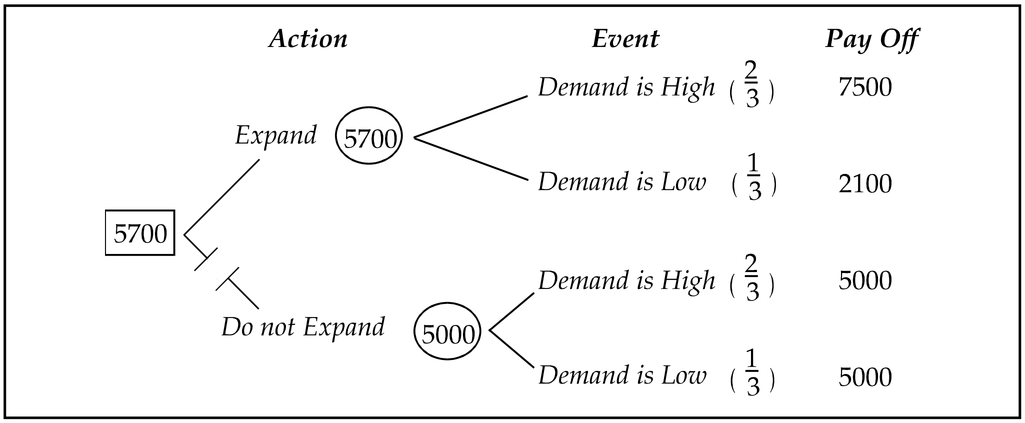
(iii) Write the payoffs at the end of each of the right-most branch.

(iv) Moving backward, from right to left, compute EMV of each chance node, wherever encountered. Enter this EMV in the chance node. When a decision node is encountered, choose the action branch having the highest EMV. Enter this EMV in the decision node and cut-off the other action branches.

Following this approach, we can describe the decision problem of the above example as given below:

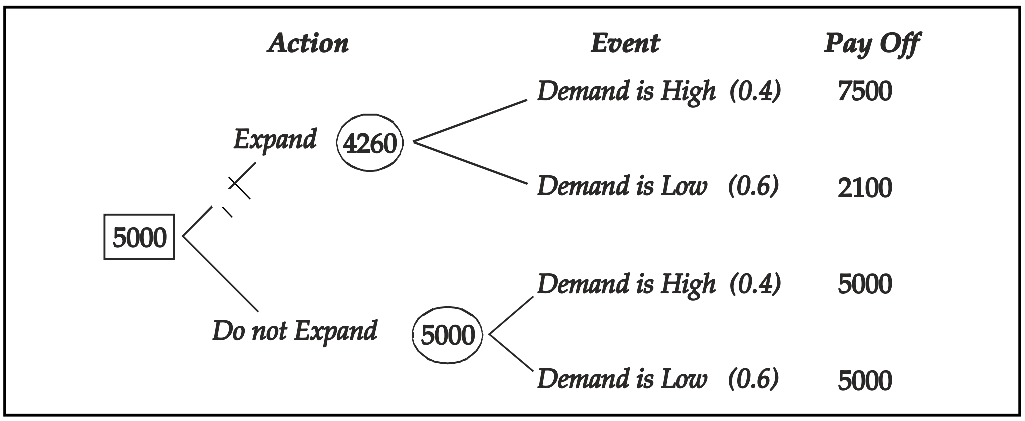
Notes

Case I : When the survey predicts that the demand is going to be high



Thus, the optimal act to expand capacity.

Case II : In the absence of survey



Thus, the optimal act is not to expand capacity.

8.4 Summary

The expected value with perfect information is the amount of profit foregone due to uncertain conditions affecting the selection of a course of action.

Given the perfect information, a decision maker is supposed to know which particular state of nature will be in effect. Thus, the procedure for the selection of an optimal course of action, for the decision problem given in example 18, will be as follows:

If the decision maker is certain that the state of nature S_1 will be in effect, he would select the course of action A_3, having maximum payoff equal to Rs 200.

When the objective probabilities of the occurrence of various states of nature are not known, the same can be assigned on the basis of the expectations or the degree of belief of the decision maker. Such probabilities are known as subjective or personal probabilities. It may be pointed out that different individuals may assign different probability values to given states of nature.

The probability values of various states of nature, discussed so far, were prior probabilities. Such probabilities are either computed from the past data or assigned subjectively. It is possible to revise these probabilities in the light of current information available by using the Bayes' Theorem.

8.5 Keywords

Cost of Uncertainty: This concept is similar to the concept of EVPI. Cost of uncertainty is the difference between the EOL of optimal action and the EOL under perfect information.

Bayes' Theorem: It is possible to revise these probabilities in the light of current information available by using the Bayes' Theorem.

8.6 Self Assessment

1. .... is used when the number of states of nature is considerably large. Using this analysis, it is possible to locate the optimal course of action without the computation of EMV's of various actions.
2. The .... of various states of nature, discussed so far, were prior probabilities.
3. The .... are often used to understand and solve a decision problem. Using such diagrams, it is possible to describe the sequence of actions and chance events.
4. A .... is represented by a circle and various event branches stem from it.

8.7 Review Questions

38. A newspaper distributor assigns probabilities to the demand for a magazine as follows:

Copies Demanded	:	1	2	3	4
Probability	:	0.4	0.3	0.2	0.1

A copy of magazine sells for Rs 7 and costs Rs 6. What can be the maximum possible expected monetary value (EMV) if the distributor can return the unsold copies for Rs 5 each? Also find EVPI.

39. A management is faced with the problem of choosing one of the three products for manufacturing. The potential demand for each product may turn out to be good, fair or poor. The probabilities for each type of demand were estimated as follows :

Demand → *Product ↓*	*Good*	*Fair*	*Poor*
A	0.75	0.15	0.10
B	0.60	0.30	0.10
C	0.50	0.30	0.20

The estimated profit or loss (in Rs) under the three states of demand in respect of each product may be taken as :

A	35,000	15,000	5,000
B	50,000	20,000	−3,000
C	60,000	30,000	20,000

Prepare the expected value table and advise the management about the choice of the product.

Notes

40. The payoffs of three acts A, B and C and the states of nature P, Q and R are given as :

	Payoffs (in Rs)		
States of Nature	*A*	*B*	*C*
P	–35	120	–100
Q	250	–350	200
R	550	650	700

The probabilities of the states of nature are 0.5, 0.1 and 0.4 respectively. Tabulate the Expected Monetary Values for the above data and state which can be chosen as the best act? Calculate expected value of perfect information also.

41. A manufacturing company is faced with the problem of choosing from four products to manufacture. The potential demand for each product may turn out to be good, satisfactory or poor. The probabilities estimated of each type of demand are given below:

	Probabilities of type of demand		
Product	*Good*	*Satisfactory*	*Poor*
A	0.60	0.20	0.20
B	0.75	0.15	0.10
C	0.60	0.25	0.15
D	0.50	0.20	0.30

The estimated profit (in Rs) under different states of demand in respect of each product may be taken as:

A	40,000	10,000	1,100
B	40,000	20,000	–7,000
C	50,000	15,000	–8,000
D	40,000	18,000	15,000

Prepare the expected value table and advise the company about the choice of product to manufacture.

Answers: Self Assessment

1. Marginal analysis 2. probability values
3. decision tree diagrams 4. chance node

8.8 Further Readings

Books

Introductory Probability and Statistical Applications by P.L. Meyer

Introduction to Mathematical Statistics by Hogg and Craig

Fundamentals of Mathematical Statistics by S.C. Gupta and V.K. Kapoor

Unit 9: Variance of a Random Variable and their Properties

CONTENTS

Objectives

Introduction

9.1 Mean and Variance of a Random Variable

9.1.1 Moments

9.2 Summary

9.3 Keywords

9.4 Self Assessment

9.5 Review Questions

9.6 Further Readings

Objectives

After studying this unit, you will be able to:

- Discuss variance of random variable
- Describe the properties random variable

Introduction

In last unit you will studied about expected random variable. This unit will provide you variance of a random variable.

9.1 Mean and Variance of a Random Variable

The mean and variance of a random variable can be computed in a manner similar to the computation of mean and variance of the variable of a frequency distribution.

Mean

If X is a discrete random variable which can take values X_1, X_2, ... X_n, with respective probabilities as $p(X_1)$, $p(X_2)$, ... $p(X_n)$, then its mean, also known as the Mathematical Expectation or Expected Value of X, is given by:

$$E(X) = X_1p(X_1) + X_2p(X_2) + ... + X_np(X_n) = \sum_{i=1}^{n} X_i p(X_i) .$$

The mean of a random variable or its probability distribution is often denoted by μ, i.e., $E(X) = \mu$.

Notes

Remarks: The mean of a frequency distribution can be written as

$X_1.\frac{f_1}{N} + X_2.\frac{f_2}{N} + \ ...... \ + X_n.\frac{f_n}{N}$, which is identical to the expression for expected value.

Variance

The concept of variance of a random variable or its probability distribution is also similar to the concept of the variance of a frequency distribution.

The variance of a frequency distribution is given by

$$\sigma^2 = \frac{1}{N}\sum f_i\left(X_i - \overline{X}\right)^2 = \sum\left(X_i - \overline{X}\right)^2.\frac{f_i}{N} = \text{Mean of } \left(X_i - \overline{X}\right)^2 \text{ values.}$$

The expression for variance of a probability distribution with mean μ can be written in a similar way, as given below :

$\sigma^2 = E(X-\mu)^2 = \sum_{i=1}^{n}(X_i - \mu)^2 p(X_i)$, where X is a discrete random variable.

Remarks: If X is a continuous random variable with probability density function p(X), then

$$E(X) = \int_{-\infty}^{\infty} X.p(X)dX$$

$$\sigma^2 = E(X-\mu)^2 = \int_{-\infty}^{\infty}(X-\mu)^2.p(X)dX$$

9.1.1 Moments

The rth moment of a discrete random variable about its mean is defined as:

$$\mu_r = E(X-\mu)^r = \sum_{i=1}^{n}(X_i - \mu)^r p(X_i)$$

Similarly, the rth moment about any arbitrary value A, can be written as

$$\mu_r' = E(X-A)^r = \sum_{i=1}^{n}(X_i - A)^r p(X_i)$$

The expressions for the central and the raw moments, when X is a continuous random variable, can be written as

$$\mu_r = E(X-\mu)^r = \int_{-\infty}^{\infty}(X-\mu)^r.p(X)dX$$

and $\mu_r' = E(X-A)^r = \int_{-\infty}^{\infty}(X-A)^r.p(X)dX$ respectively.

9.2 Summary

Notes

- The mean and variance of a random variable can be computed in a manner similar to the computation of mean and variance of the variable of a frequency distribution.
- If X is a discrete random variable which can take values X_1, X_2, ... X_n, with respective probabilities as $p(X_1)$, $p(X_2)$, ... $p(X_n)$, then its mean, also known as the Mathematical Expectation or Expected Value of X, is given by:

  $$E(X) = X_1p(X_1) + X_2p(X_2) + \ldots\ldots + X_np(X_n) = \sum_{i=1}^{n} X_i p(X_i).$$

  The mean of a random variable or its probability distribution is often denoted by μ, i.e., $E(X) = \mu$.

  Remarks: The mean of a frequency distribution can be written as

  $X_1.\frac{f_1}{N} + X_2.\frac{f_2}{N} + \ldots\ldots + X_n.\frac{f_n}{N}$, which is identical to the expression for expected value.

- The concept of variance of a random variable or its probability distribution is also similar to the concept of the variance of a frequency distribution.

  The variance of a frequency distribution is given by

  $$\sigma^2 = \frac{1}{N}\sum f_i\left(X_i - \bar{X}\right)^2 = \sum\left(X_i - \bar{X}\right)^2.\frac{f_i}{N} = \text{Mean of } \left(X_i - \bar{X}\right)^2 \text{ values.}$$

  The expression for variance of a probability distribution with mean μ can be written in a similar way, as given below :

  $\sigma^2 = E(X-\mu)^2 = \sum_{i=1}^{n}(X_i - \mu)^2 p(X_i)$, where X is a discrete random variable.

9.3 Keywords

Random variable: If X is a discrete random variable which can take values X_1, X_2, ... X_n, with respective probabilities as $p(X_1)$, $p(X_2)$, ... $p(X_n)$, then its mean, also known as the Mathematical Expectation or Expected Value of X, is given by:

$$E(X) = X_1p(X_1) + X_2p(X_2) + \ldots\ldots + X_np(X_n) = \sum_{i=1}^{n} X_i p(X_i).$$

Variance: The concept of variance of a random variable or its probability distribution is also similar to the concept of the variance of a frequency distribution.

Continuous random: If X is a continuous random variable with probability density function p(X), then

$$E(X) = \int_{-\infty}^{\infty} X.p(X)dX$$

$$\sigma^2 = E(X-\mu)^2 = \int_{-\infty}^{\infty}(X-\mu)^2.p(X)dX$$

Notes

Moment: The rth moment of a discrete random variable about its mean is defined as:

$$\mu_r = E(X-\mu)^r = \sum_{i=1}^{n}(X_i - \mu)^r p(X_i)$$

9.4 Self Assessment

1. If X is a .... variable which can take values X_1, X_2, ... X_n, with respective probabilities as $p(X_1)$, $p(X_2)$, ... $p(X_n)$, then its mean, also known as the Mathematical Expectation or Expected Value of X, is given by:

   $$E(X) = X_1p(X_1) + X_2p(X_2) + ... + X_np(X_n) = \sum_{i=1}^{n} X_i p(X_i).$$

   (a) discrete random (b) variance
   (c) continuous random (d) moment

2. The concept of .... of a random variable or its probability distribution is also similar to the concept of the variance of a frequency distribution.

   (a) discrete random (b) variance
   (c) continuous random (d) moment

3. If X is a .... variable with probability density function p(X), then

   $$E(X) = \int_{-\infty}^{\infty} X.p(X)dX$$

   $$\sigma^2 = E(X-\mu)^2 = \int_{-\infty}^{\infty} (X-\mu)^2 .p(X)dX$$

   (a) discrete random (b) variance
   (c) continuous random (d) moment

4. The rth .... of a discrete random variable about its mean is defined as:

   $$\mu_r = E(X-\mu)^r = \sum_{i=1}^{n}(X_i - \mu)^r p(X_i)$$

   (a) discrete random (b) variance
   (c) continuous random (d) moment

9.5 Review Questions

1. Obtain the probability distribution of the number of aces in simultaneous throws of two unbiased dice.
2. Explain the concept of a random variable and its probability distribution. Illustrate your answer using experiment of the toss of two unbiased coins. Find the mean and variance of the random variable defined by you for this experiment.
3. If E(X) = 1 and Var(X) = 5, find
   (i) $E[(2 + X)^2]$
   (ii) Var(4 + 3X)

4. You are told that the time to service a car at a service station is uncertain with following probability density function:

   $f(x) = 3x - 2x^2 + 1$ for 0 £ x £ 2

   = 0 otherwise.

   Examine whether this is a valid probability density function?

5. Find mean and variance of the following probability distribution :

$$\begin{array}{lcccc} X & : & -20 & -10 & 30 \\ p(X) & : & \frac{3}{10} & \frac{1}{5} & \frac{1}{2} \end{array}$$

Answers: Self Assessment

1. (a) 2. (b) 3. (c) 4. (d)

9.6 Further Readings

Books

Introductory Probability and Statistical Applications by P.L. Meyer

Introduction to Mathematical Statistics by Hogg and Craig

Fundamentals of Mathematical Statistics by S.C. Gupta and V.K. Kapoor

Notes

Unit 10: Approximate Expressions for Expectations and Variance

CONTENTS

Objectives

Introduction

10.1 Theorems on Expectation

10.1.1 Theorems on Variance

10.2 Joint Probability Distribution

10.2.1 Marginal Probability Distribution

10.2.2 Conditional Probability Distribution

10.2.3 Expectation of the Sum or Product of two Random Variables

10.2.4 Expectation of a Function of Random Variables

10.6 Summary

10.7 Keywords

10.8 Self Assessment

10.9 Review Questions

10.10 Further Readings

Objectives

After studying this unit, you will be able to:

- Discuss theorem on expectation
- Explain joint probability distribution

Introduction

In last unit you have studied about variance of random variable. This unit will explain you joint probability distribution.

10.1 Theorems on Expectation

Theorem 1.

Expected value of a constant is the constant itself, i.e., E(b) = b, where b is a constant.

Proof.

The given situation can be regarded as a probability distribution in which the random variable takes a value b with probability 1 and takes some other real value, say a, with probability 0.

Thus, we can write $E(b) = b \times 1 + a \times 0 = b$

Theorem 2.

$E(aX) = aE(X)$, where X is a random variable and a is constant.

Proof.

For a discrete random variable X with probability function p(X), we have :

$$E(aX) = aX_1.p(X_1) + aX_2.p(X_2) + ... + aX_n.p(X_n)$$

$$= a\sum_{i=1}^{n} X_i.p(X_i) = aE(X)$$

Combining the results of theorems 1 and 2, we can write

$$E(aX + b) = aE(X) + b$$

Remarks: Using the above result, we can write an alternative expression for the variance of X, as given below :

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

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

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

= Mean of Squares - Square of the Mean

We note that the above expression is identical to the expression for the variance of a frequency distribution.

10.1.1 Theorems on Variance

Theorem 1.

The variance of a constant is zero.

Proof.

Let b be the given constant. We can write the expression for the variance of b as:

$$Var(b) = E[b - E(b)]^2 = E[b - b]^2 = 0.$$

Theorem 2.

$$Var(X + b) = Var(X).$$

Proof.

We can write $Var(X + b) = E[X + b - E(X + b)]^2 = E[X + b - E(X) - b]^2$

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

Similarly, it can be shown that $Var(X - b) = Var(X)$

Remarks: The above theorem shows that variance is independent of change of origin.

Theorem 3.

$$Var(aX) = a^2Var(X)$$

Notes

Proof.

We can write Var(aX) = E[aX - E(aX)]^2 = E[aX - aE(X)]^2

= a^{2}E[X - E(X)]^2 = a^{2}Var(X).

Combining the results of theorems 2 and 3, we can write

Var(aX + b) = a^{2}Var(X).

This result shows that the variance is independent of change origin but not of change of scale.

Remarks:

1. On the basis of the theorems on expectation and variance, we can say that if X is a random variable, then its linear combination, aX + b, is also a random variable with mean aE(X) + b and Variance equal to a^{2}Var(X).
2. The above theorems can also be proved for a continuous random variable.

Example 4: Compute mean and variance of the probability distributions obtained in examples 1, 2 and 3.

Solution.

(a) The probability distribution of X in example 1 was obtained as

X	0	1	2	3
$p(X)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

From the above distribution, we can write

$$E(X) = 0 \times \frac{1}{8} + 1 \times \frac{3}{8} + 2 \times \frac{3}{8} + 3 \times \frac{1}{8} = 1.5$$

To find variance of X, we write

Var(X) = E(X^2) - [E(X)]^2, where $E(X^2) = \sum X^2 p(X)$.

Now, $E(X^2) = 0 \times \frac{1}{8} + 1 \times \frac{3}{8} + 4 \times \frac{3}{8} + 9 \times \frac{1}{8} = 3$

Thus, Var(X) = 3 - (1.5)^2 = 0.75

(b) The probability distribution of X in example 2 was obtained as

X	2	3	4	5	6	7	8	9	10	11	12	*Total*
$p(X)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	1

$$\therefore E(X) = 2 \times \frac{1}{36} + 3 \times \frac{2}{36} + 4 \times \frac{3}{36} + 5 \times \frac{4}{36} + 6 \times \frac{5}{36} + 7 \times \frac{6}{36}$$

$$+8\times\frac{5}{36}+9\times\frac{4}{36}+10\times\frac{3}{36}+11\times\frac{2}{36}+12\times\frac{1}{36}=\frac{252}{36}=7$$

Further, $E(X^2)=4\times\frac{1}{36}+9\times\frac{2}{36}+16\times\frac{3}{36}+25\times\frac{4}{36}+36\times\frac{5}{36}+49\times\frac{6}{36}$

$$+64\times\frac{5}{36}+81\times\frac{4}{36}+100\times\frac{3}{36}+121\times\frac{2}{36}+144\times\frac{1}{36}=\frac{1974}{36}=54.8$$

Thus, Var(X) = 54.8 - 49 = 5.8

(c) The probability distribution of X in example 3 was obtained as

X	1	2	3
$p(X)$	$\frac{4}{20}$	$\frac{12}{20}$	$\frac{4}{20}$

From the above, we can write

$$E(X)=1\times\frac{4}{20}+2\times\frac{12}{20}+3\times\frac{4}{20}=2$$

and $E(X^2)=1\times\frac{4}{20}+4\times\frac{12}{20}+9\times\frac{4}{20}=4.4$

$\therefore$ Var(X) = 4.4 - 4 = 0.4

Expected Monetary Value (EMV)

When a random variable is expressed in monetary units, its expected value is often termed as expected monetary value and symbolised by EMV.

Example 5: If it rains, an umbrella salesman earns Rs 100 per day. If it is fair, he loses Rs 15 per day. What is his expectation if the probability of rain is 0.3?

Solution.

Here the random variable X takes only two values, X_1 = 100 with probability 0.3 and X_2 = - 15 with probability 0.7.

Thus, the expectation of the umbrella salesman

= 100 × 0.3 - 15 × 0.7 = 19.5

The above result implies that his average earning in the long run would be Rs 19.5 per day.

Example 6: A person plays a game of throwing an unbiased die under the condition that he could get as many rupees as the number of points obtained on the die. Find the expectation and variance of his winning. How much should he pay to play in order that it is a fair game?

Solution.

Notes

The probability distribution of the number of rupees won by the person is given below :

$X(Rs)$	1	2	3	4	5	6
$p(X)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Thus, $E(X) = 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + 3 \times \frac{1}{6} + 4 \times \frac{1}{6} + 5 \times \frac{1}{6} + 6 \times \frac{1}{6} = Rs\frac{7}{2}$

and $E(X^2) = 1 \times \frac{1}{6} + 4 \times \frac{1}{6} + 9 \times \frac{1}{6} + 16 \times \frac{1}{6} + 25 \times \frac{1}{6} + 36 \times \frac{1}{6} = \frac{91}{6}$

$\therefore \quad \sigma^2 = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12} = 2.82$. Note that the unit of σ^2 will be $(Rs)^2$.

Since E(X) is positive, the player would win Rs 3.5 per game in the long run. Such a game is said to be favourable to the player. In order that the game is fair, the expectation of the player should be zero. Thus, he should pay Rs 3.5 before the start of the game so that the possible values of the random variable become 1 - 3.5 = - 2.5, 2 - 3.5 = - 1.5, 3 - 3.5 = - 0.5, 4 - 3.5 = 0.5, etc. and their expected value is zero.

Example 7: Two persons A and B throw, alternatively, a six faced die for a prize of Rs 55 which is to be won by the person who first throws 6. If A has the first throw, what are their respective expectations?

Solution.

Let A be the event that A gets a 6 and B be the event that B gets a 6. Thus, $P(A) = \frac{1}{6}$ and $P(B) = \frac{1}{6}$.

If A starts the game, the probability of his winning is given by :

$$P(A \text{ wins}) = P(A) + P(\overline{A}).P(\overline{B}).P(A) + P(\overline{A}).P(\overline{B}).P(\overline{A}).P(\overline{B}).P(A) + \ldots.$$

$$= \frac{1}{6} + \frac{5}{6} \times \frac{5}{6} \times \frac{1}{6} + \frac{5}{6} \times \frac{5}{6} \times \frac{5}{6} \times \frac{5}{6} \times \frac{1}{6} + \ldots\ldots$$

$$= \frac{1}{6}\left[1 + \left(\frac{5}{6}\right)^2 + \left(\frac{5}{6}\right)^4 + \ldots\ldots\right] = \frac{1}{6} \times \left(\frac{1}{1 - \frac{25}{36}}\right) = \frac{1}{6} \times \frac{36}{11} = \frac{6}{11}$$

Similarly, $P(B \text{ wins}) = P(\overline{A}).P(B) + P(\overline{A}).P(\overline{B}).P(\overline{A}).P(B) + \ldots.$

$$= \frac{5}{6} \times \frac{1}{6} + \frac{5}{6} \times \frac{5}{6} \times \frac{5}{6} \times \frac{1}{6} + \ldots\ldots$$

Notes

$$= \frac{5}{6} \times \frac{1}{6}\left[1 + \left(\frac{5}{6}\right)^2 + \left(\frac{5}{6}\right)^4 + \ ......\right] = \frac{5}{6} \times \frac{1}{6} \times \frac{36}{11} = \frac{5}{11}$$

Expectation of A and B for the prize of Rs 55

Since the probability that A wins is $\frac{6}{11}$, therefore, the random variable takes a value 55 with probability $\frac{6}{11}$ and value 0 with probability $\frac{5}{11}$. Hence, $E(A) = 55 \times \frac{6}{11} + 0 \times \frac{5}{11} = \text{Rs } 30$

Similarly, the expectation of B is given by $E(B) = 55 \times \frac{5}{11} + 0 \times \frac{6}{11} = Rs\ 25$

Example 8: An unbiased die is thrown until a four is obtained. Find the expected value and variance of the number of throws.

Solution.

Let X denote the number of throws required to get a four. Thus, X will take values 1, 2, 3, 4, ... with respective probabilities.

$$\frac{1}{6}, \frac{5}{6} \times \frac{1}{6}, \left(\frac{5}{6}\right)^2 \times \frac{1}{6}, \left(\frac{5}{6}\right)^3 \times \frac{1}{6} \text{...... etc.}$$

$$\therefore \quad E(X) = 1.\frac{1}{6} + 2.\frac{5}{6}.\frac{1}{6} + 3.\left(\frac{5}{6}\right)^2.\frac{1}{6} + 4.\left(\frac{5}{6}\right)^3.\frac{1}{6} ......$$

$$= \frac{1}{6}\left[1 + 2.\frac{5}{6} + 3.\left(\frac{5}{6}\right)^2 + 4.\left(\frac{5}{6}\right)^3 + ...\right]$$

Let $\quad S = 1 + 2.\frac{5}{6} + 3.\left(\frac{5}{6}\right)^2 + 4.\left(\frac{5}{6}\right)^3 + ...$

Multiplying both sides by $\frac{5}{6}$, we get

[redacted] $S = \frac{5}{6} + 2.\left(\frac{5}{6}\right)^2 + 3.\left(\frac{5}{6}\right)^3 + 4.\left(\frac{5}{6}\right)^4 + ...$

$$\therefore \quad S - \frac{5}{6}S = 1 + (2-1)\frac{5}{6} + (3-2)\left(\frac{5}{6}\right)^2 + (4-3)\left(\frac{5}{6}\right)^3 + ...$$

$$\frac{1}{6}S = 1 + \frac{5}{6} + \left(\frac{5}{6}\right)^2 + \left(\frac{5}{6}\right)^3 + ... = \frac{1}{1 - \frac{5}{6}} = 6 \qquad .... (1)$$

Notes

Thus, S = 36 and hence E(X) = $\frac{1}{6}$ × 36 = 6.

Further, to find variance, we first find E(X^2)

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

$$= \frac{1}{6}\left[1 + 2^2.\left(\frac{5}{6}\right) + 3^2.\left(\frac{5}{6}\right)^2 + 4^2.\left(\frac{5}{6}\right)^3 + ...\right]$$

Let S = $1 + 2^2.\left(\frac{5}{6}\right) + 3^2.\left(\frac{5}{6}\right)^2 + 4^2.\left(\frac{5}{6}\right)^3 + ...$

Multiply both sides by $\frac{5}{6}$ and subtract from S, to get

$$\frac{1}{6}S = 1 + \left(2^2 - 1\right)\left(\frac{5}{6}\right) + \left(3^2 - 2^2\right)\left(\frac{5}{6}\right)^2 + \left(4^2 - 3^2\right)\left(\frac{5}{6}\right)^3 + ...$$

$$= 1 + 3\left(\frac{5}{6}\right) + 5\left(\frac{5}{6}\right)^2 + 7\left(\frac{5}{6}\right)^3 + ...$$

Further, multiply both sides by $\frac{5}{6}$ and subtract

$$\frac{1}{6}S - \frac{5}{36}S = 1 + (3-1)\left(\frac{5}{6}\right) + (5-3)\left(\frac{5}{6}\right)^2 + (7-5)\left(\frac{5}{6}\right)^3 + ...$$

$$\frac{1}{36}S = 1 + 2\left(\frac{5}{6}\right)\left\{1 + \frac{5}{6} + \left(\frac{5}{6}\right)^2 + ...\right\} = 1 + \frac{5}{3} \times 6 = 11 \quad .... (2)$$

∴ S = 36 × 11 and E(X^2) = $\frac{1}{6}$ × 36 × 11 = 66

Hence, Variance = E(X^2) - $[E(X)]^2$ = 66 - 36 = 30

Generalisation:

Let p be the probability of getting 4, then from equation (1) we can write

$$pS = \frac{1}{1-q} = \frac{1}{p} \text{ or } S = \frac{1}{p^2} \text{ Therefore, } E(X) = p\left(\frac{1}{p^2}\right) = \frac{1}{p}$$

Similarly, equation (2) can be written as

$$p^2S = 1 + \frac{2q}{p} \text{ or } S = \frac{1}{p^2} + \frac{2q}{p^3} = \frac{p+2q}{p^3}$$

Therefore, $E\left(X^2\right) = p.\left(\frac{p+2q}{p^3}\right) = \frac{p+2q}{p^2}$ and Var(X) = $\frac{p+2q}{p^2} - \frac{1}{p^2} = \frac{q}{p^2}$

10.2 Joint Probability Distribution

When two or more random variables X and Y are studied simultaneously on a sample space, we get a joint probability distribution. Consider the experiment of throwing two unbiased dice. If X denotes the number on the first and Y denotes the number on the second die, then X and Y are random variables having a joint probability distribution. When the number of random variables is two, it is called a bi-variate probability distribution and if the number of random variables become more than two, the distribution is termed as a multivariate probability distribution.

Let the random variable X take values X_1, X_2, ... X_m and Y take values Y_1, Y_2, ... Y_n. Further, let p_{ij} be the joint probability that X takes the value X_i and Y takes the value Y_j, i.e., $P[X = X_i \text{ and } Y = Y_j] = p_{ij}$ (i = 1 to m and j = 1 to n). This bi-variate probability distribution can be written in a tabular form as follows:

	Y_1	Y_2	...	...	Y_n	*Marginal Probabilities of X*
X_1	p_{11}	p_{12}	...	...	p_{1n}	P_1
X_2	p_{21}	p_{22}	...	...	p_{2n}	P_2
.	.	.	...	...	.	.
.	.	.	...	...	.	.
X_m	p_{m1}	p_{m2}	...	...	p_{mn}	P_m
Marginal Probabilities of Y	P'_1	P'_2	...	...	P'_n	1

10.2.1 Marginal Probability Distribution

In the above table, the probabilities given in each row are added and shown in the last column. Similarly, the sum of probabilities of each column are shown in the last row of the table. These probabilities are termed as marginal probabilities. The last column of the table gives the marginal probabilities for various values of random variable X. The set of all possible values of the random variable X along with their respective marginal probabilities is termed as the marginal probability distribution of X. Similarly, the marginal probabilities of the random variable Y are given in the last row of the above table.

Remarks: If X and Y are independent random variables, by multiplication theorem of probability we have

$P(X = X_i \text{ and } Y = Y_j) = P(X = X_i).P(Y = Y_j)$ " i and j

Using notations, we can write $p_{ij} = P_i.P'_j$

The above relation is similar to the relation between the relative frequencies of independent attributes.

10.2.2 Conditional Probability Distribution

Each column of the above table gives the probabilities for various values of the random variable X for a given value of Y, represented by it. For example, column 1 of the table represents that $P(X_1,Y_1) = p_{11}$, $P(X_2,Y_1) = p_{21}$, ... $P(X_m,Y_1) = p_{m1}$, where $P(X_i,Y_1) = p_{i1}$ denote the probability of the event that $X = X_i$ (i = 1 to m) and $Y = Y_1$. From the conditional probability theorem, we can write

$$P\left(X = X_i / Y = Y_1\right) = \frac{\text{Joint probability of } X_i \text{ and } Y_1}{\text{Marginal probability of } Y_1} = \frac{p_{ij}}{P'_j} \text{ (for i = 1, 2, ... m).}$$

Notes

This gives us a conditional probability distribution of X given that Y = Y_1. This distribution can be written in a tabular form as shown below :

X	X_1	X_2	... ...	X_m	*Total Probability*
Probability	$\frac{p_{11}}{P_1'}$	$\frac{p_{21}}{P_1'}$	... ...	$\frac{p_{m1}}{P_1'}$	1

The conditional distribution of X given some other value of Y can be constructed in a similar way. Further, we can construct the conditional distributions of Y for various given values of X.

Remarks:

It can be shown that if the conditional distribution of a random variable is same as its marginal distribution, the two random variables are independent. Thus, if for the conditional distribution of X given Y_1 we have $\frac{p_{i1}}{P_1'} = P_i$ for " i, then X and Y are independent. It should be noted here that if one conditional distribution satisfies the condition of independence of the random variables, then all the conditional distributions would also satisfy this condition.

Example 9: Let two unbiased dice be tossed. Let a random variable X take the value 1 if first die shows 1 or 2, value 2 if first die shows 3 or 4 and value 3 if first die shows 5 or 6. Further, Let Y be a random variable which denotes the number obtained on the second die. Construct a joint probability distribution of X and Y. Also determine their marginal probability distributions and find E(X) and E(Y) respectively. Determine the conditional distribution of X given Y = 5 and of Y given X = 2. Find the expected values of these conditional distributions. Determine whether X and Y are independent?

Solution.

For the given random experiment, the random variable X takes values 1, 2 and 3 and the random variable Y takes values 1, 2, 3, 4, 5 and 6. Their joint probability distribution is shown in the following table:

$X\downarrow \backslash Y\rightarrow$	1	2	3	4	5	6	*Marginal Dist. of X*
1	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{3}$
2	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{3}$
3	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{3}$
Marginal Dist. of Y	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	1

From the above table, we can write the marginal distribution of X as given below :

X	1	2	3	*Total*
P_i	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	1

Thus, the expected value of X is $E(X) = 1.\frac{1}{3} + 2.\frac{1}{3} + 3.\frac{1}{3} = 2$

Similarly, the probability distribution of Y is

Y	1	2	3	4	5	6	*Total*
P'_j	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	1

and $E(Y) = 1.\frac{1}{6} + 2.\frac{1}{6} + 3.\frac{1}{6} + 4.\frac{1}{6} + 5.\frac{1}{6} + 6.\frac{1}{6} = \frac{21}{6} = 3.5$

The conditional distribution of X when Y = 5 is

X	1	2	3	*Total*
$P_i / Y = 5$	$\frac{1}{18} \times \frac{6}{1} = \frac{1}{3}$	$\frac{1}{18} \times \frac{6}{1} = \frac{1}{3}$	$\frac{1}{18} \times \frac{6}{1} = \frac{1}{3}$	1

$\therefore \quad E(X / Y = 5) = \frac{1}{3}(1 + 2 + 3) = 2$

The conditional distribution of Y when X = 2 is

Y	1	2	3	4	5	6	*Total*
$P'_j / X = 2$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	1

$\therefore \quad E(Y / X = 2) = \frac{1}{6}(1 + 2 + 3 + 4 + 5 + 6) = 3.5$

Since the conditional distribution of X is same as its marginal distribution (or equivalently the conditional distribution of Y is same as its marginal distribution), X and Y are independent random variables.

Example 10: Two unbiased coins are tossed. Let X be a random variable which denotes the total number of heads obtained on a toss and Y be a random variable which takes a value 1 if head occurs on first coin and takes a value 0 if tail occurs on it. Construct the joint probability distribution of X and Y. Find the conditional distribution of X when Y = 0. Are X and Y independent random variables?

Solution.

There are 4 elements in the sample space of the random experiment. The possible values that X can take are 0, 1 and 2 and the possible values of Y are 0 and 1. The joint probability distribution of X and Y can be written in a tabular form as follows:

Notes

$X \downarrow \backslash Y \rightarrow$	0	1	*Total*
0	$\frac{1}{4}$	0	$\frac{1}{4}$
1	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{2}{4}$
2	0	$\frac{1}{4}$	$\frac{1}{4}$
Total	$\frac{2}{4}$	$\frac{2}{4}$	1

The conditional distribution of X when Y = 0, is given by

X	0	1	2	*Total*
$P(X/Y=0)$	$\frac{1}{2}$	$\frac{1}{2}$	0	1

Also, the marginal distribution of X, is given by

X	0	1	2	*Total*
P_i	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	1

Since the conditional and the marginal distributions are different, X and Y are not independent random variables.

10.2.3 Expectation of the Sum or Product of two Random Variables

Theorem 1.

If X and Y are two random variables, then E(X + Y) = E(X) + E(Y).

Proof.

Let the random variable X takes values $X_1, X_2, \ldots\ldots X_m$ and the random variable Y takes values $Y_1, Y_2, \ldots\ldots Y_n$ such that $P(X = X_i \text{ and } Y = Y_j) = p_{ij}$ (i = 1 to m, j = 1 to n).

By definition of expectation, we can write

$$E(X+Y) = \sum_{i=1}^{m}\sum_{j=1}^{n}\left(X_i + Y_j\right)p_{ij} = \sum_{i=1}^{m}\sum_{j=1}^{n} X_i p_{ij} + \sum_{i=1}^{m}\sum_{j=1}^{n} Y_j p_{ij} = \sum_{i=1}^{m} X_i \sum_{j=1}^{n} p_{ij} + \sum_{i=1}^{n} Y_j \sum_{j=1}^{m} p_{ij}$$

$$= \sum_{i=1}^{m} X_i P_i + \sum_{j=1}^{n} Y_j P'_j \quad \left(\text{Here } \sum_{J=1}^{n} p_{ij} = P_i \text{ and } \sum_{i=1}^{m} p_{ij} = P'_j\right)$$

$$= E(X) + E(Y)$$

The above result can be generalised. If there are k random variables $X_1, X_2, \ldots\ldots X_k$, then $E(X_1 + X_2 + \ldots\ldots + X_k) = E(X_1) + E(X_2) + \ldots\ldots E(X_k)$.

Remarks: The above result holds irrespective of whether $X_1, X_2, \ldots\ldots X_k$ are independent or not.

Theorem 2.

Notes

If X and Y are two independent random variables, then

$$E(X.Y) = E(X).E(Y)$$

Proof.

Let the random variable X takes values X_1, X_2, ... X_m and the random variable Y takes values Y_1, Y_2, ... Y_n such that $P(X = X_i \text{ and } Y = Y_j) = p_{ij}$ (i = 1 to m, j = 1 to n).

By definition $E(XY) = \sum_{i=1}^{m}\sum_{j=1}^{n} X_i Y_j p_{ij}$

Since X and Y are independent, we have $p_{ij} = P_i . P_j'$

$$\therefore \; E(XY) = \sum_{i=1}^{m}\sum_{j=1}^{n} X_i Y_j P_i . P_j' = \sum_{i=1}^{m} X_i P_i \times \sum_{j=1}^{n} Y_j P_j'$$

$$= E(X).E(Y).$$

The above result can be generalised. If there are k independent random variables X_1, X_2, ... X_k, then

$$E(X_1 . X_2 . \ldots\ldots X_k) = E(X_1).E(X_2). \ldots\ldots E(X_k)$$

10.2.4 Expectation of a Function of Random Variables

Let $f(X,Y)$ be a function of two random variables X and Y. Then we can write $E[\phi(X,Y)] = \sum_{i=1}^{m}\sum_{j=1}^{n} \phi(X_i, Y_j) p_{ij}$

I. Expression for Covariance

As a particular case, assume that $\phi(X_i, Y_j) = (X_i - \mu_X)(Y_j - \mu_Y)$, where $E(X) = \mu_X$ and $E(Y) = \mu_Y$

Thus, $E[(X - \mu_X)(Y - \mu_Y)] = \sum_{i=1}^{m}\sum_{j=1}^{n} (X_i - \mu_X)(Y_j - \mu_Y) p_{ij}$

The above expression, which is the mean of the product of deviations of values from their respective means, is known as the Covariance of X and Y denoted as Cov(X, Y) or σ_{XY} . Thus, we can write

$$Cov(X,Y) = E[(X - \mu_X)(Y - \mu_Y)]$$

An alternative expression of Cov(X, Y)

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

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

$$= E[X.Y - X.E(Y)] = E(X.Y) - E(X).E(Y)$$

Note that E[{Y – E(Y)}] = 0, the sum of deviations of values from their arithmetic mean.

Remarks:

1. If X and Y are independent random variables, the right hand side of the above equation will be zero. Thus, covariance between independent variables is always equal to zero.
2. COV(a + bX, c + dY) = bd COV(X, Y)
3. COV(X, X) = VAR(X)

II. Mean and Variance of a Linear Combination

Let $Z = \phi(X, Y) = aX + bY$ be a linear combination of the two random variables X and Y, then using the theorem of addition of expectation, we can write

$$\mu_Z = E(Z) = E(aX + bY) = aE(X) + bE(Y) = a\mu_X + b\mu_Y$$

Further, the variance of Z is given by

$$\sigma_Z^2 = E[Z - E(Z)]^2 = E[aX + bY - a\mu_X - b\mu_Y]^2 = E\left[a(X - \mu_X) + b(Y - \mu_Y)\right]^2$$

$$= a^2E(X - \mu_X)^2 + b^2E(Y - \mu_Y)^2 + 2abE(X - \mu_X)(Y - \mu_Y)$$

$$= a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\sigma_{XY}$$

Remarks:

1. The above results indicate that any function of random variables is also a random variable.
2. If X and Y are independent, then $\sigma_{XY} = 0$, \ $\sigma_Z^2 = a^2\sigma_X^2 + b^2\sigma_Y^2$
3. If Z = aX - bY, then we can write $\sigma_Z^2 = a^2\sigma_X^2 + b^2\sigma_Y^2 - 2ab\sigma_{XY}$. However, $\sigma_Z^2 = a^2\sigma_X^2 + b^2\sigma_Y^2$, if X and Y are independent.
4. The above results can be generalised. If X_1, X_2, ... X_k are k independent random variables with means μ_1, μ_2, ... μ_k and variances σ_1^2, σ_2^2, ... σ_k^2 respectively, then

$$E(X_1 \pm X_2 \pm \; .... \; \pm X_k) = \mu_1 \pm \mu_2 \pm \; .... \pm \mu_k$$

and $$\text{Var}(X_1 \pm X_2 \pm \; .... \; \pm X_k) = \sigma_1^2 + \sigma_2^2 + \; .... + \sigma_k^2$$

Notes

1. The general result on expectation of the sum or difference will hold even if the random variables are not independent.
2. The above result can also be proved for continuous random variables.

Example 11:

A random variable X has the following probability distribution :

X	:	-2	-1	0	1	2
Probability	:	$\frac{1}{6}$	p	$\frac{1}{4}$	p	$\frac{1}{6}$

(i) Find the value of p.

(ii) Calculate E(X + 2) and $E(2X^2 + 3X + 5)$.

Solution.

Since the total probability under a probability distribution is equal to unity, the value of p should be such that $\frac{1}{6}+p+\frac{1}{4}+p+\frac{1}{6}=1$.

This condition gives $p=\frac{5}{24}$

Further, $$E(X)=-2.\frac{1}{6}-1.\frac{5}{24}+0.\frac{1}{4}+1.\frac{5}{24}+2.\frac{1}{6}=0$$

$$E(X^2)=4.\frac{1}{6}+1.\frac{5}{24}+0.\frac{1}{4}+1.\frac{5}{24}+4.\frac{1}{6}=\frac{7}{4},$$

$$E(X+2)=E(X)+2=0+2=2$$

and $$E(2X^2+3X+5)=2E(X^2)+3E(X)+5=2.\frac{7}{4}+0+5=8.5$$

Example 12:

A dealer of ceiling fans has estimated the following probability distribution of the price of a ceiling fan in the next summer season:

Price (*P*)	:	800	825	850	875	900
Probability (*p*)	:	0.15	0.25	0.30	0.20	0.10

If the demand (x) of his ceiling fans follows a linear relation x = 6000 - 4P, find expected demand of fans and expected total revenue of the dealer.

Solution.

Since P is a random variable, therefore, x = 6000 - 4P, is also a random variable. Further, Total Revenue TR = P.x = $6000P - 4P^2$ is also a random variable.

From the given probability distribution, we have

$E(P) = 800 \times 0.15 + 825 \times 0.25 + 850 \times 0.30 + 875 \times 0.20 + 900 \times 0.10$

=Rs 846.25 and

$E(P^2) = (800)^2 \times 0.15 + (825)^2 \times 0.25 + (850)^2 \times 0.30 + (875)^2 \times 0.20$

$+ (900)^2 \times 0.10 = 717031.25$

Notes

Thus, E(X) = 6000 - 4E(P) = 6000 - 4 $\times$ 846.25 = 2615 fans.

And E(TR) = 6000E(P) - 4E(P^2)

$$= 6000 \times 846.25 - 4 \times 717031.25 = \text{Rs } 22{,}09{,}375.00$$

Example 13: A person applies for equity shares of Rs 10 each to be issued at a premium of Rs 6 per share; Rs 8 per share being payable along with the application and the balance at the time of allotment. The issuing company may issue 50 or 100 shares to those who apply for 200 shares, the probability of issuing 50 shares being 0.4 and that of issuing 100 shares is 0.6. In either case, the probability of an application being selected for allotment of any shares is 0.2 The allotment usually takes three months and the market price per share is expected to be Rs 25 at the time of allotment. Find the expected rate of return of the person per month.

Solution.

Let A be the event that the application of the person is considered for allotment, B_1 be the event that he is allotted 50 shares and B_2 be the event that he is allotted 100 shares. Further, let R_1 denote the rate of return (per month) when 50 shares are allotted, R_2 be the rate of return when 100 shares are allotted and $R = R_1 + R_2$ be the combined rate of return.

We are given that P(A) = 0.2, $P(B_1/A) = 0.4$ and $P(B_2/A) = 0.6$.

(a) When 50 shares are allotted

The return on investment in 3 months = (25 - 16)50 = 450

$\therefore$ Monthly rate of return $= \frac{450}{3} = 150$

The probability that he is allotted 50 shares

$$= P(A \cap B_1) = P(A).\,P(B_1 / A) = 0.2 \times 0.4 = 0.08$$

Thus, the random variable R_1 takes a value 150 with probability 0.08 and it takes a value 0 with probability 1 - 0.08 = 0.92

$\therefore$ $E(R_1) = 150 \times 0.08 + 0 = 12.00$

(b) When 100 shares are allotted

The return on investment in 3 months = (25 - 16).100 = 900

$\therefore$ Monthly rate of return $= \frac{900}{3} = 300$

The probability that he is allotted 100 shares

$$= P(A \cap B_2) = P(A).\,P(B_2 / A) = 0.2 \times 0.6 = 0.12$$

Thus, the random variable R_2 takes a value 300 with probability 0.12 and it takes a value 0 with probability 1 - 0.12 = 0.88

$\therefore$ $E(R_2) = 300 \times 0.12 + 0 = 36$

Hence, $E(R) = E(R_1 + R_2) = E(R_1) + E(R2) = 12 + 36 = 48$

Example 14: What is the mathematical expectation of the sum of points on n unbiased dice?

Solution.

Notes

Let X_i denote the number obtained on the i th die. Therefore, the sum of points on n dice is $S = X_1 + X_2 + ... + X_n$ and

$$E(S) = E(X_1) + E(X_2) + ... + E(X_n).$$

Further, the number on the i th die, i.e., X_i follows the following distribution :

X_i	:	1	2	3	4	5	6
$p(X_i)$	:	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

$$\therefore \quad E(X_i) = \frac{1}{6}(1+2+3+4+5+6) = \frac{7}{2} \qquad (i = 1, 2, ... n)$$

Thus, $E(S) = \frac{7}{2} + \frac{7}{2} + \; .... \; + \frac{7}{2} \; (n \text{ times}) = \frac{7n}{2}$

Example 15: If X and Y are two independent random variables with means 50 and 120 and variances 10 and 12 respectively, find the mean and variance of $Z = 4X + 3Y$.

Solution.

$E(Z) = E(4X + 3Y) = 4E(X) + 3E(Y) = 4 \times 50 + 3 \times 120 = 560$

Since X and Y are independent, we can write

$Var(Z) = Var(4X + 3Y) = 16Var(X) + 9Var(Y) = 16 \times 10 + 9 \times 12 = 268$

Example 16: It costs Rs 600 to test a machine. If a defective machine is installed, it costs Rs 12,000 to repair the damage resulting to the machine. Is it more profitable to install the machine without testing if it is known that 3% of all the machines produced are defective? Show by calculations.

Solution.

Here X is a random variable which takes a value 12,000 with probability 0.03 and a value 0 with probability 0.97.

$\therefore$ $E(X) = 12000 \times 0.03 + 0 \times 0.97 =$ Rs 360.

Since E(X) is less than Rs 600, the cost of testing the machine, hence, it is more profitable to install the machine without testing.

10.6 Summary

- Expected value of a constant is the constant itself, i.e., E(b) = b, where b is a constant.
- Using the above result, we can write an alternative expression for the variance of X, as given below :

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

= Mean of Squares - Square of the Mean

Notes

- When two or more random variables X and Y are studied simultaneously on a sample space, we get a joint probability distribution. Consider the experiment of throwing two unbiased dice. If X denotes the number on the first and Y denotes the number on the second die, then X and Y are random variables having a joint probability distribution. When the number of random variables is two, it is called a bi-variate probability distribution and if the number of random variables become more than two, the distribution is termed as a multivariate probability distribution.

10.7 Keywords

Expected value: Expected value of a constant is the constant itself, i.e., E(b) = b, where b is a constant.

Variance: The variance of a constant is zero.

10.8 Self Assessment

1. When a random variable is expressed in monetary units, its expected value is often termed as expected monetary value and symbolised by ....

   (a) probability distribution (b) EMV

   (c) Covariance (d) random variables

2. The set of all possible values of the random variable X along with their respective marginal probabilities is termed as the marginal .... of X.

   (a) probability distribution (b) EMV

   (c) Covariance (d) random variables

3. If X and Y are two .... , then E(X + Y) = E(X) + E(Y).

   (a) probability distribution (b) EMV

   (c) Covariance (d) random variables

4. The mean of the product of deviations of values from their respective means, is known as the .... of X and Y denoted as Cov(X, Y) or σ_{XY} .

   (a) probability distribution (b) EMV

   (c) Covariance (d) random variables

10.9 Review Questions

1. ABC company estimates the net profit on a new product, that it is launching, to be Rs 30,00,000 if it is successful, Rs 10,00,000 if it is moderately successful and a loss of Rs 10,00,000 if it is unsuccessful. The firm assigns the following probabilities to the different possibilities : Successful 0.15, moderately successful 0.25 and unsuccessful 0.60. Find the expected value and variance of the net profits.

   Hint : See example 5.

2. There are 4 different choices available to a customer who wants to buy a transistor set. The first type costs Rs 800, the second type Rs 680, the third type Rs 880 and the fourth type Rs 760. The probabilities that the customer will buy these types are $\frac{1}{3}, \frac{1}{6}, \frac{1}{4}$ and $\frac{1}{4}$ respectively.

Notes

The retailer of these sets gets a commission @ 20%, 12%, 25% and 15% on the respective sets. What is the expected commission of the retailer?

Hint : Take commission as random variable.

3. Three cards are drawn at random successively, with replacement, from a well shuffled pack of cards. Getting a card of diamond is termed as a success. Tabulate the probability distribution of the number successes (X). Find the mean and variance of X.

   Hint : The random variable takes values 0, 1, 2 and 3.

4. A discrete random variable can take all possible integral values from 1 to k each with probability $\frac{1}{k}$. Find the mean and variance of the distribution.

   Hint : $E\left(X^2\right)=\frac{1}{k}\left(1^2+2^2+\;....\;+k^2\right)=\frac{1}{k}\left[\frac{k(k+1)(2k+1)}{6}\right].$

5. An insurance company charges, from a man aged 50, an annual premium of Rs 15 on a policy of Rs 1,000. If the death rate is 6 per thousand per year for this age group, what is the expected gain for the insurance company?

   Hint : Random variable takes values 15 and - 985.

6. On buying a ticket, a player is allowed to toss three fair coins. He is paid number of rupees equal to the number of heads appearing. What is the maximum amount the player should be willing to pay for the ticket.

   Hint : The maximum amount is equal to expected value.

7. The following is the probability distribution of the monthly demand of calculators :

Demand (x)	:	15	16	17	18	19	20
Probability $p(x)$	:	0.10	0.15	0.35	0.25	0.08	0.07

Calculate the expected demand for calculators. If the cost c of producing x calculators is given by the relation $c = 4x^2 - 15x + 200$, find expected cost.

Hint : See example 12.

Answers: Self Assessment

1. (b) 2. (a) 3. (d) 4. (c)

10.10 Further Readings

Books

Introductory Probability and Statistical Applications by P.L. Meyer

Introduction to Mathematical Statistics by Hogg and Craig

Fundamentals of Mathematical Statistics by S.C. Gupta and V.K. Kapoor

Notes

Unit 11: Chebyshev's Inequality

CONTENTS

Objectives

Introduction

11.1 Chebyshev's Inequality

11.2 Summary

11.3 Keywords

11.4 Self Assessment

11..5 Review Questions

11.6 Further Readings

Objectives

After studying this unit, you will be able to:

- Apply chebyshev's inequality
- Give example of chebyshev's inequality

Introduction

We have discussed different methods for obtaining distribution functions of random variables or random vectors. Even though it is possible to derive these distributions explicity in closed form in some special situations, in general, this is not the case. Computation of the probabilities, even when the probability distribution functions are known, is cumbersome at times. For instance, it is easy to write down the exact probabilities for a binomial distribution with parameters n = 1000 and $p = \frac{1}{50}$. However computing the individual probabilities involve factorials for integers of large order which are impossible to handle even with speed computing facilities.

In this unit, we discuss limit theorems which describe the behaviour of some distributions when the sample size n is large. The limiting distributions can be used for computation of the probabilities approximately.

Chebyshev's inequality is discussed, as an application, weak law of large numbers is derived (which describes the behaviour of the sample mean as n increases).

11.1 Chebyshev's Inequality

We prove in this section an important result known as Chebyshev's inequality. This inequality is due to the nineteenth century Russian mathematician P.L. Chebyshev.

We shall begin with a theorem.

Notes

Theorem 1: Suppose X is a random variable with mean μ and finite variance σ^2. Then for every $\varepsilon > 0$.

Proof: We shall prove the theorem for continuous r.vs. The proof in the discrete case is very similar.

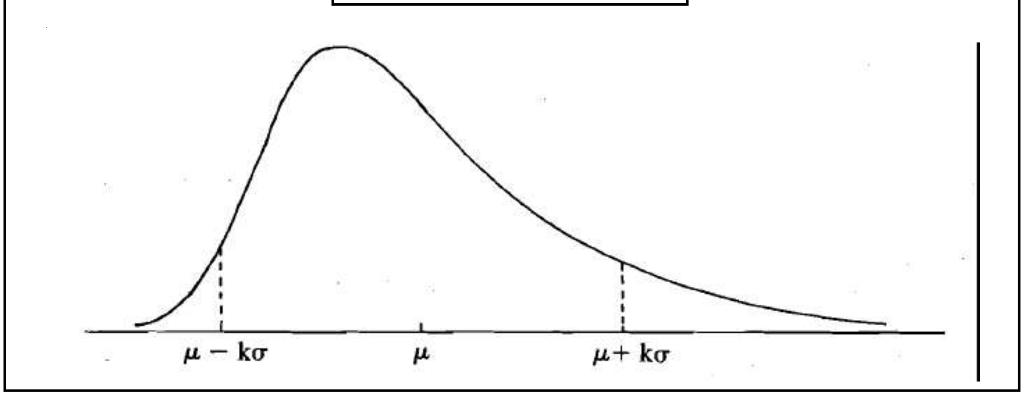
Suppose X is a random variable with probability density function f. From the definition of the variance of X, we have

$$\sigma = E[(x-\mu)^2] = \int_{-\infty}^{+\infty} (x-\mu)^2 f(x)dx.$$

Suppose $\varepsilon > 0$ is given. Put $\varepsilon_1 = \frac{\varepsilon}{\sigma}$. Now we divide the integral into three parts as shown in Fig. 1.

$$\sigma^2 = \int_{-\infty}^{\mu-\varepsilon_1\sigma} (x-\mu)^2 f(x)dx + \int_{\mu-\varepsilon_1\sigma}^{\mu-\varepsilon_1\sigma} (x-\mu)^2 f(x)dx + \int_{\mu+\varepsilon_1\sigma}^{\infty} (x-\mu)^2 f(x)dx \qquad ...(2)$$

Figure 11.1



Since the integrand (x – m)2 f(x) is non-negative, from (2) we get the inequality

$$\sigma^2 \geq \int_{-\infty}^{\mu-\varepsilon_1\sigma} (x-\mu)^2 f(x)dx + \int_{\mu+\varepsilon_1\sigma}^{\infty} (x-\mu)^2 f(x)dx \qquad ...(3)$$

Now for any $x \in]-\infty, \mu - \varepsilon_1\sigma]$, we have $x \leq \mu - \varepsilon_1\sigma$ which implies that $(x-\mu)^2 \geq \varepsilon^2\sigma^2$. Therefore we get

$$\int_{-\infty}^{\mu-\varepsilon_1\sigma} (x-\mu)^2 f(x)dx \geq \int_{-\infty}^{\mu-\varepsilon_1\sigma} \varepsilon^2\sigma^2 f(x)dx$$

$$= \varepsilon^2\sigma^2 \int_{-\infty}^{\mu-\varepsilon_1\sigma} f(x)dx.$$

Similarly for $x \in]\mu + \varepsilon_1\sigma, \infty[$ also we have $(x-\mu)^2 \geq \varepsilon_1^2\sigma^2$ and therefore

$$\int_{\mu+\varepsilon_1\sigma}^{\infty} (x-\mu)^2 f(x)dx \geq \varepsilon_1^2\sigma^2 \int_{\mu+\varepsilon_1\sigma}^{\infty} f(x)dx$$

Notes

Then by (3) we get

$$\sigma^2 \geq \varepsilon_1^2\sigma^2 \left[\int_{-\infty}^{\mu-\varepsilon_1\sigma} f(x)dx + \int_{\mu+\varepsilon_1\sigma}^{\infty} f(x)dx \right]$$

i.e.,

$$\frac{1}{\varepsilon_1^2} \geq \int_{-\infty}^{\mu-\varepsilon_1\sigma} f(x)dx + \int_{\mu+\varepsilon_1\sigma}^{\infty} f(x)dx$$

whenever $\sigma^2 \neq 0$.

Now, by applying Property (iii) of the density function given in Sec. 11.3, unit 10, we get

$$\frac{1}{\varepsilon_1^2} \geq P[X \leq \mu - \varepsilon_1\sigma] + P[X \geq \mu + \varepsilon\sigma]$$

$$= P[X - \mu \leq - \varepsilon_1\sigma] + P[X - \mu \geq \varepsilon 1\sigma]$$

$$= P[\,|X - m| \geq \varepsilon_1\sigma]$$

That is, $P[\,|X - m| \geq \varepsilon_1\sigma] \leq \frac{1}{\varepsilon_1^2}$ ...(4)

Substituting $\varepsilon_1 = \frac{\varepsilon}{\sigma}$ in (4), we gt the inequality

$$[P[\,|X - \mu| \geq \varepsilon] \leq \frac{\sigma^2}{\varepsilon^2}$$

Chebyshev's inequality also holds when the distribution of X is neither (absolutely) continuous nor discrete. We will not discuss this general case here. Now we shall make a remark.

Remark 1: The above result is very general indeed. We need to know nothing about the probability distribution of the random variable X. It could be binomial, normal, beta or gamma or any other distribution. The only restriction is that it should have finite variance. In other words the upper bound is universal in nature. The price we pay for such generality is that the upper bound is not sharp in general. If we know more about the distribution of X, then it might be possible to get a better bound. We shall illustrate this point in the following example.

Example 1: Suppose X is $N(\mu, \sigma^2)$. Then $E(X) = \mu$ and $Var(X) = \sigma^2$. Let us compute $P[\,|X - \mu| \geq 2\sigma]$.

Here $\varepsilon = 2\sigma$. By applying Chebychev's inequality we get

$$P[\,|X - \mu| \geq 2\sigma] \leq \frac{\sigma^2}{4\sigma^2} = \frac{1}{4} = .25$$

Since we know that the distribution of X is normal, we can directly compute the probability. Then we have

$$P(|X - \mu| \geq 2\sigma) = P\left[\,\left|\frac{X - \mu}{\sigma}\right| \geq 2\right]$$

Notes

Since $\frac{X-\mu}{\sigma}$ has N(0, 1) as its distribution, from the normal distribution table given in the appendix of Unit 11, we get

$$P\left(\left|\frac{X-\mu}{\sigma}\right| \geq 2\right) = 0.456$$

which is substantially small as compared to the exact value 0.25. Thus in this case we could get a better upperbound by directly using the distribution.

Let us consider another example.

Example 2: Suppose X is a random variable such that P[X = 1] = 1/2 = P[X = –1]. Let us compute an upper bound for $P[|X - \mu| > \sigma]$.

You can check that E(X) = 0 and Var(X) = 1. Hence, by Chebyshev's inequality, we get that

$$P(|X-\mu| > \sigma) \leq \frac{\sigma^2}{\sigma^2} = 1.$$

on the other hand, direct calculations show that

$$P(|X-\mu| > \sigma) = P[|X| \geq 1] = 1.$$

In this example, the upper bound obtained from Chebyshev's inequality as well as the one obtained from using the distribution of X are one and the same.

In the first example you can see an application of Chebyshev's inequality.

Example 3: Suppose a person makes 100 check transactions during a certain period. In balancing his or her check book transactions, suppose he or she rounds off the check entries to the nearest rupee instead of subtracting the exact amount he or she has used. Let us find an upper bound to the probability that the total error he or she has committed exceeds Rs. 5 after 100 transactions.

Let X_i denote the round off error in rupees made for the ith transaction. Then the total error is $X_1 + X_2 + ... + X_{100}$. We can assume that X_i, $1 \leq i \leq 100$ are independent and idelltically distributed random variables and that each X_i has uniform distribution on $\left[-\frac{1}{2}, \frac{1}{2}\right]$. We are interested in finding an upper bound for the $P[|S_{100}| > 5]$ where $S_{100} = X_1 + ... + X_{100}$.

In general, it is difficult and computationally complex to find the exact distribution. However, we can use Chebyshev's inequality to get an upper hound. It is clear that

$$E(S_{100}) = 100E(X_1) = 0$$

and

$$\text{var}(S_{100}) = 100 \text{ var } (X_i) = \frac{100}{12}.$$

since $E(X_1) = 0$ and $\text{Var}(X_1) = \frac{1}{12}$. Therefore by Chebyshev's inequality,

Notes

$$P\left(|S_{100}-0)|>5\right)\leq\frac{Var(S_{100})}{25}$$

$$=\frac{100}{12\times25}$$

$$=\frac{1}{3}.$$

Here are some exercises for you.

The above examples and exercises must have given you enough practise to apply Chebyshev's inequality. Now we shall use this inequality to establish an important result.

Suppose X_1, X_2, ..., X_n are independent and identically distributed random variables having mean μ and variance σ^2. We define

$$\overline{X}_n=\frac{1}{n}\sum_{i=1}^{n}X_i$$

Then $\overline{X}_n$ has mean μ and variance $\frac{\sigma^2}{n}$. Hence, by the Chebyshev's inequality, we get

$$P\left[|\overline{X}_n-\mu|\geq\varepsilon\right]\leq\frac{\sigma^2}{n\varepsilon^2}$$

for any $\varepsilon > 0$. If $n \to 0$, then $\frac{\sigma^2}{n\varepsilon^2}\to 0$ and therefore

$$P\left(|\overline{X}_n-\mu|\geq\varepsilon\right)\to 0.$$

In other words, as n grows large, the probability that $\overline{X}_n$ differs from μ by more than any given positive number E, becomes small. An alternate way of stating this result is as follows :

For any $\varepsilon > 0$, given any positive number δ, we a n choose sufficiently large n such that

$$P\left(|\overline{X}_n-\mu|\geq\varepsilon\right)\leq\delta$$

This result is known as the weak law of large numbers. We now state it as a theorem.

Theorem 2 (Weak law of large nombers) : Suppose X_1, X_2, ..., X_n are i.i.d. random variables with mean μ and finite variance σ^2.

Let

$$\overline{X}_n=\frac{1}{n}\sum_{i=1}^{n}X_i.$$

Then

$$P\left[|\overline{X}_n-\mu|\geq\varepsilon\right]\to 0 \text{ as } n\to\infty.$$

for any E > 0.

The above theorem is true even when the variance is infinite but the mean p is finite However this result does not follow as an application of the Chebyshev's inequality in this general set up. The proof in the general case is beyond the scope of this course.

We make a remark here.

Remark 2 : The above theorem only says that the probability that the value of the difference $|\overline{X}_n - X|$ exceeds any fixed number ε, gets smaller and smaller for successively large values of n. The theorem does not say anything about the limiting case of the actual difference. In fact there is another strong result which talks about the limiting case of the actual values of the differences. This is the reason why Theorem 2 is called 'weak law'. We hove not included the stronger result here since it is beyond the level of this course.

Let us see an example.

Example 4: Suppose a random experiment has two possihle outcomes called success (S) and Failure (F). Let p he the probability of a success. Suppose the experiment is repeated independently n times. Let X_i take the value 1 or 0 according as the outcome in the i-th trial of the experiments is success or a failure. Let us apply Theorem 2 to the set $\{X_i\}_{j=1}^{n}$.

We first note that

$$P[X_i = 1] = p \text{ and } P[X_i = 0] = 1 - p = q,$$

for $1 \le i \le n$. Also you can check that $E(X_i) = p$ ond var $(X_i) = p\,q$ for i = 1, ... n.

Since the mean and the variance are finite, we can apply the weak law of large numbers for the sequence $\{X_1 : 1 \le i \le n\}$. Then we have

$$P\left[\left|\frac{S_n}{n} - p\right| \ge \varepsilon\right] \to 0 \text{ as } b \to \infty$$

Sn for every ε > 0 where $S_n = X_1 + X_2 + ... + X_n$. Now, what is $\frac{S_n}{n}$? S_n is the number of successes observed in n trials and therefore $\frac{S_n}{n}$ is the proportinn of successes in n trials. Then the above result says that as the number of trials increases, the proportion of successes tends stabilize to the probability of a success. Of course, one of the basic assumptions behind this interpretation is that the random experiment can be repeated.

In the next section we shall discuss another limit theorem which gives an approximation to the binomial distrihution.

11.2 Summary

- Suppose X is a random variable with mean μ and finite variance σ^2. Then for every ε > 0.
- The above theorem only says that the probability that the value of the difference $|\overline{X}_n - X|$ exceeds any fixed number ε, gets smaller and smaller for successively large values of n. The theorem does not say anything about the limiting case of the actual difference. In fact there is another strong result which talks about the limiting case of the actual values of the differences. This is the reason why Theorem 2 is called 'weak law'. We hove not included the stronger result here since it is beyond the level of this course.

Notes

- Since the mean and the variance are finite, we can apply the weak law of large numbers for the sequence $\{X_1 : 1 \le i \le n\}$. Then we have

$$P\left[\left|\frac{S_n}{n} - p\right| \ge \varepsilon\right] \to 0 \text{ as } b \to \infty$$

S_n for every $\varepsilon > 0$ where $S_n = X_1 + X_2 + ... + X_n$. Now, what is $\frac{S_n}{n}$? S_n is the number of successes observed in n trials and therefore $\frac{S_n}{n}$ is the proportinn of successes in n trials. Then the above result says that as the number of trials increases, the proportion of successes tends stabilize to the probability of a success. Of course, one of the basic assumptions behind this interpretation is that the random experiment can be repeated.

11.3 Keywords

Chebyshev's inequality is discussed, as an application, weak law of large numbers is derived.

Weak law of large nombers: Suppose $X_1, X_2, ..., X_n$ are i.i.d. random variables with mean m and finite variance σ^2.

11.4 Self Assessment

1. Computation of the probabilities, even when the .... functions are known, is cumbersome at times.
   (a) Chebyshev's inequality (b) limiting distributions
   (c) (absolutely) continuous (d) probability distribution
2. The .... can be used for computation of the probabilities approximately.
   (a) Chebyshev's inequality (b) limiting distributions
   (c) (absolutely) continuous (d) probability distribution
3. .... is discussed, as an application, weak law of large numbers is derived.
   (a) Chebyshev's inequality (b) limiting distributions
   (c) (absolutely) continuous (d) probability distribution
4. Chebyshev's inequality also holds when the distribution of X is neither .... nor discrete.
   (a) Chebyshev's inequality (b) limiting distributions
   (c) (absolutely) continuous (d) probability distribution

11..5 Review Questions

1. Suppose X is $N(\mu, \sigma^2)$. Then $E(X) = \mu$ and $Var(X) = \sigma^2$. Let us compute $P[\ |X - \mu| \ge 3\sigma]$.
2. Suppose X is a random variable such that $P[X = 1] = 1/2 = P[X = -1]$. Let us compute an upper bound for $P[|X - \mu| > 1/2\sigma]$.
3. Suppose a person makes 100 check transactions during a certain period. In balancing his or her check book transactions, suppose he or she rounds off the check entries to the nearest

rupee instead of subtracting the exact amount he or she has used. Let us find an upper bound to the probability that the total error he or she has committed exceeds Rs. 5 after 100 transactions.

Answers: Self Assessment

1. (d) 2. (b) 3. (a) 4. (c)

11.6 Further Readings

Books

Introductory Probability and Statistical Applications by P.L. Meyer

Introduction to Mathematical Statistics by Hogg and Craig

Fundamentals of Mathematical Statistics by S.C. Gupta and V.K. Kapoor

Notes

Unit 12: The Moment Generating Functions

CONTENTS

Objectives

Introduction

12.1 Moment Generating Function of a Random Variable

12.2 Deriving Moments with the mgf

12.3 Characterization of a Distriution via the mgf

12.4 Moment Generating Function - More details

12.4.1 Moment Generating Function of a Linear Transformation

12.4.2 Moment Generating Function of a Sum of Mutually Independent Random Variable

12.5 Summary

12.6 Keywords

12.7 Self Assessment

12.8 Review Questions

12.9 Further Readings

Objectives

After studying this unit, you will be able to:

- Discuss moment generating function of a random variable
- Describe deriving moments with mgf

Introduction

In the previous unit, we learned that the expected value of the sample mean $\overline{X}$ is the population mean m. We also learned that the variance of the sample mean $\overline{X}$ is σ^2/n, that is, the population variance divided by the sample size n. We have not yet determined the probability distribution of the sample mean when, say, the random sample comes from a normal distribution with mean m and variance σ^2. We are going to tackle that in the next lesson! Before we do that, though, we are going to want to put a few more tools into our tollbox. We already have learned a few techniques for finding the probability distribution of a function of random variables, namely the distribution function technique and the change-of-variable technique. In this unit, we'll learn yet another technique called the moment-generating function technique. We'll use the technique in this lesson to learn, among other things, the distribution of sums of chi-square random variables, then, in the next lesson, we'll use the technique to find (finally) the probability distribution of the sample mean when the random sample comes from a normal distribution with mean m and variance σ^2.

Notes

12.1 Moment Generating Function of a Random Variable

Moment Generating Function - Defintion

We start this lecture by giving a definition of moment generating function.

Definition: Let X be a random variable. If the expected value:

$$E[\exp(tX)]$$

exists and is finite for all real numbers t belonging to a closed interval $[-h, h] \subseteq \mathbb{R}$, with $h > 0$, then we say that X possesses a moment generating function and the function $M_X : [-h, h] \to \mathbb{R}$ defined by:

$$Mx(t) = E[\exp(t(X)]$$

is called the moment generating function (or mgf) of X.

Moment Generating Function - Example

The following example shows how the moment generating function of an exponential random variable is calculated.

Example: Let X be an exponential random variable with parameter $\lambda \in \mathbb{R}$...its supposed Rx is the set of positive real numbers:

$$RX = [0, \infty)$$

and its probability density function $f_x(x)$ is:

$$fx(x) = \begin{cases} 1\exp(-lx) & \text{if } x \in Rx \\ 0 & \text{if } x \notin Rx \end{cases}$$

Its moment generating function is computed as follows:

$$\begin{aligned}
E[\exp(tX)] &= \int_{-\infty}^{\infty} \exp(tx)fx(x)dx \\
&= \int_{0}^{\infty} \exp(tx)\lambda\exp(-\lambda x)dx \\
&= \lambda\int_{0}^{\infty} \exp((t-\lambda)x)dx \\
&= \lambda\left[\frac{1}{t-\lambda}\exp((t-\lambda)dx\right]_{0}^{\infty} \\
&= \lambda\left[0-\frac{1}{t-\lambda}\right] \\
&= \frac{\lambda}{t-\lambda}
\end{aligned}$$

Notes

Notes that the above integral is finite for $t \in [-h, h]$ for any $0 < h < \lambda$, so that X possesses a moment generating function.

$$M_X(t) = \frac{\lambda}{\lambda - t}$$

12.2 Deriving Moments with the mgf

The moment generating function takes its name by the fact that it can be used to derive the moments of X, as stated in the following proposition.

Proposition If a random variable X possesses a moment generating function $M_X(t)$, then, for any $n \in \mathbb{N}$, the n-th moment of X (denote it by $\mu_X(n)$) exists and is finite.

Furthermore:

$$\mu_X(n) = E[X^n] = \left.\frac{d^n Mx(t)}{dt^n}\right|_{t=0}$$

where $\left.\frac{d^n Mx(t)}{dt^n}\right|_{t=0}$ is the nth derivative of $M_X(t)$ with respect to t, evaluated at the point t = 0.

Providing the above proposition is quite complicated, because a lot of analytical details must be taken care of (see e.g. Pfeiffer, P.E. (1978) concepts of probability theory, Courier Dover Publications). The intuition, however, is straightforward: since the expected value is a linear operator and differentiation is a linear operation, under appropriate conditions one can differentiate through the expected value, as follows:

$$\frac{d^n M_X(t)}{dt^n} = \frac{d^n}{dt^n} E[\exp(tX)] = E\left[\frac{d^n}{dt^n}\exp(tX)\right] = E[X^n \exp(tX)]$$

which, evaluated at the point t = 0, yields

$$\left.\frac{d^n M_X(t)}{dt^n}\right|_{t=0} = E[X^n \exp(0.X)] = E[X^n] = \mu_X(n)$$

Example: Continuing the example above, the moment generating function of an exponential random variable is

$$MX(t) = \frac{\lambda}{\lambda - t}$$

The expected value of X can be computed by taking the first derivative of the moment generating function.

$$\frac{dM_X(t)}{dt} = \frac{1}{(\lambda - t)^2}$$

and evaluating it at r = 0.

$$E[X] = \left.\frac{dM_X(t)}{dt}\right|_{t=0} = \frac{\lambda}{(\lambda-0)^2} = \frac{1}{\lambda}$$

The second moment of X can be computed by taking the second derivative of the moment generating function

$$\frac{d^2M_X(t)}{dt^2} = \frac{2\lambda}{(\lambda-t)^3}$$

and evaluating it at r = 0

$$E[X_2] = \left.\frac{d^2M_X(t)}{dt^2}\right|_{t=0} = \frac{2\lambda}{(\lambda-t)^3} = \frac{2}{\lambda^2}$$

And so on for the higher moments.

12.3 Characterization of a Distriution via the mgf

The most important property of the moment generating function is the following:

Proposition (Equality of distributions) Let X and Y be two random variables. Denote by $F_X(x)$ and $F_Y(y)$ their distribution functions and by $M_X(t)$ and $M_Y(t)$ their moment generating functions. X and Y have the same distribution (i.e., $F_X(x) = F_Y(x)$ for any x) if and only if they have the same moment generating functions (i.e. $M_X(t) = M_Y(t)$ for any t).

While proving this proposition is beyond the scope of this introductory exposition, it must be stressed that this proposition is extremely important and relevant from a practical viewpoint in many cases where we need to prove that two distributions are equal, it is much easier to prove equality of the moment generating functions than to prove equality of the distribution functions. Also note that equality of the distribution functions can be replaced in the proposition above by equality of the probability mass function (if X and Y are discrete random variables) or by equality of the probability density functions (if X and Y are absolutely continuous random variables).

12.4 Moment Generating Function - More details

12.4.1 Moment Generating Function of a Linear Transformation

Let X be a random variable possessing a moment generating function $M_X(t)$. Define:

$$Y = a + bX$$

where $a, b \in \mathbb{R}$ are two constants and $b \neq 0$. Then, the random variable Y possesses a moment generating function $M_Y(t)$ and

Proof

Using the definition of moment generation function

$$\begin{aligned} MY(t) &= E[\exp(tY)] \\ &= E[\exp(at + btX)] \\ &= E[\exp(at)\exp(btX)] \end{aligned}$$

Notes

$$= \exp(at)E[\exp(btX)]$$

$$= \exp(at)M_X(bt)$$

Obviously, if $M_X(t)$ is defined on a closed interval [–h, h], then $M_Y(t)$ is defined on the interval $\left[-\frac{h}{b}, \frac{h}{b}\right]$.

12.4.2 Moment Generating Function of a Sum of Mutually Independent Random Variable

Let $X_1, ..., X_n$ be n mutually independent random variables. Let Z be their sum

$$Z = \sum_{i=1}^{n} X_i$$

Then, the moment generating function of Z is the product of the moment generating functions of $X_1, ..., X_n$.

$$M_Z(t) = \prod_{i=1}^{n} M_{X_Y}(t)$$

This is easily proved using the definition of moment generating function and the properties of mutually independent variables (mutual independence via expectations):

$$\begin{aligned} MZ(t) &= E[\exp(tZ)] \\ &= E\left[\exp\left(t\sum_{i=1}^{n} X_l\right)\right] \\ &= E\left[\exp\left(\sum_{i=1}^{n} tX_l\right)\right] \\ &= E\left[\prod_{i=1}^{n} \exp(tX_i)\right] \\ &= \prod_{i=1}^{n} E[\exp(tX_i)] \qquad \text{(by mutual independence)} \\ &= \prod_{i=1}^{n} M_{X_i}(t) \qquad \text{(by the definition generation function)} \end{aligned}$$

Example 1: Let X be a discrete random variable having a Bernoulli distribution. Its support R_X is

$$R_X = <0, 1>$$

and its probability mass function $p_X(x)$ is

$$p_X(x) = \begin{cases} p & \text{if } x = 1 \\ 1 - p & \text{if } x = 0 \\ 0 & \text{if } x \notin R_X \end{cases}$$

where $p \in (0, 1)$ is a constant. Derive the moment generating function of X, if it exists.

Example 2: Let X be a random variable with moment generatinf function

$$M_X(t) = \frac{1}{2}(1+\exp(t))$$

Derive the variance of X.

Solution

We can use the following formula for computing the variance

$$Var[X] = E[X^2] - E[X]^2$$

The expected value of X is computed by taking the first derivative of the moment generating function.

$$\frac{dM_X(t)}{dt} = \frac{1}{2}\exp(t)$$

and evaluating it at t = 0

$$E[X] = \left.\frac{dM_X(t)}{dt}\right|_{t=0} = \frac{1}{2}\exp(0) = \frac{1}{2}$$

The second moment of X is computed by taking the second derivative of the moment generating function

$$\frac{d^2M_X(t)}{dt^2} = \frac{1}{2}\exp(t)$$

and evaluating it at t = 0

$$E[X^2] = \left.\frac{d^2M_X(t)}{dt^2}\right|_{t=0} = \frac{1}{2}\exp(0) = \frac{1}{2}$$

$$Var[X] = E[X^2] - E[X]^2$$

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

$$= \frac{1}{2} - \frac{1}{4}$$

$$= \frac{1}{4}$$

Example 3: A random variable X is said to have a Chi-square distribution with n degrees of freedom if its moment generating function is defined for any $t < \frac{1}{2}$ and it is equal to:

$$M_X(t) = (1 - 2t)^{-n/2}$$

Notes

Define

$$Y = X_1 + X_2$$

where X_1 and X_2 are two independent random variables having Chi-square distributions with n_1 and n_2 degrees of freedom respectively. Prove that Y has a Chi-square distribution with $n_1 + n_2$ degrees of freedom.

Solution

The moment generating functions of X_1 and X_2 are

$$M_{X_1}(t) = (1 - 2t)^{-n_1/2}$$

$$M_{X_2}(t) = (1 - 2t)^{-n_2/2}$$

The moment generating function of a sum of independent random variables is just the product of A random variable X is said to have a Chi-square distribution with n degrees of freedom if its moment generating function is defined for any $t < \frac{1}{2}$ and it is equal to:

$$M_X(t) = (1 - 2t)^{-n/2}$$

Define

$$Y = X_1 + X_2$$

where X_1 and X_2 are two independent random variables having Chi-square distributions with n_1 and n_2 degrees of freedom respectively. Prove that Y has a Chi-square distribution with $n_1 + n_2$ degrees of freedom.

Solution

The moment generating functions of X_1 and X_2 are

$$M_{X_1}(t) = (1 - 2t)^{-n_1/2}$$

$$M_{X_2}(t) = (1 - 2t)^{-n_2/2}$$

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

$$M_Y(t) = (1 - 2t)^{-n_1/2}(1 - 2t)^{-n_2/2}$$

$$= (1 - 2t)^{-(n_1 + n_2)/2}$$

Therefore, $M_Y(t)$ is the moment generating function of a Chi-square random variable with $n_1 + n_2$ degrees of freedom. As a consequence, Y has a Chi-square distribution with $n_1 + n_2$ degrees of freedom.

12.5 Summary

- Moment Generating Function - Defintion

  We start this lecture by giving a definition of moment generating function.

  Definition: Let X be a random variable. If the expected value:

$$E[\exp(tX)]$$

exists and is finite for all real numbers t belonging to a closed interval $[-h, h] \subseteq \mathbb{R}$, with $h > 0$, then we say that X possesses a moment generating function and the function $M_X : [-h, h] \to \mathbb{R}$ defined by:

$$Mx(t) = E[exp(t(X)]$$

is called the moment generating function (or mgf) of X.

- Let X be a random variable possessing a moment generating function $M_X(t)$. Define:

$$Y = a + bX$$

where $a, b \in \mathbb{R}$ are two constants and $b \neq 0$.

12.6 Keywords

Moment generating function of a linear transformation: Let X be a random variable possessing a moment generating function $M_X(t)$. Define:

$$Y = a + bX$$

where $a, b \in \mathbb{R}$ are two constants and $b \neq 0$.

Random variable: A random variable X is said to have a Chi-square distribution with n degrees of freedom if its moment generating function is defined for any $t < \frac{1}{2}$ and it is equal to:

$$M_X(t) = (1 - 2t)^{-n/2}$$

12.7 Self Assessment

1. If a random variable X possesses a moment generating function $M_X(t)$, then, for any $n \in \mathbb{N}$, the n-th moment of X (denote it by $\mu_X(n)$) exists and is ....

   (a) random variables (b) same distribution

   (c) $b \neq 0$ (d) finite

2. Let X and Y be two .... Denote by $F_X(x)$ and $F_Y(y)$ their distribution functions and by $M_X(t)$ and $M_Y(t)$ their moment generating functions.

   (a) random variables (b) same distribution

   (c) $b \neq 0$ (d) finite

3. X and Y have the .... (i.e., $F_X(x) = F_Y(x)$ for any x) if and only if they have the same moment generating functions (i.e. $M_X(t) = M_Y(t)$ for any t).

   (a) random variables (b) same distribution

   (c) $b \neq 0$ (d) finite

4. Let X be a random variable possessing a moment generating function $M_X(t)$. Define:

   $$Y = a + bX$$

   where $a, b \in \mathbb{R}$ are two constants and ....

   (a) random variables (b) same distribution

   (c) $b \neq 0$ (d) finite

12.8 Review Questions

1. Let X be a discrete random variable having a Bernoulli distribution. Its support R_X is

$$R_X = <0, 1>$$

and its probability mass function $p_X(x)$ is

$$p_X(x) = \begin{cases} p & \text{if } x = 1 \\ 1-p & \text{if } x = 0 \\ 0 & \text{if } x \notin R_X \end{cases}$$

where $p \in (0, 1)$ is a constant. Derive the moment generating function of X, if it exists.

2. Let X be a discrete random variable having a Bernoulli distribution. Its support R_X is

$$R_X = <0, 1>$$

and its probability mass function $p_X(x)$ is

$$p_X(x) = \begin{cases} p & \text{if } x = 0 \\ 1-p & \text{if } x = 1 \\ 0 & \text{if } x \notin R_X \end{cases}$$

where $p \in (0, 1)$ is a constant. Derive the moment generating function of X, if it exists.

3. Let X be a random variable with moment generatinf function

$$M_X(t) = \frac{1}{3}(1 + \exp(t))$$

Derive the variance of X.

4. Let X be a random variable with moment generatinf function

$$M_X(t) = \frac{4}{3}(1 + \exp(t))$$

Derive the variance of X.

5. A random variable X is said to have a Chi-square distribution with n degrees of freedom if its moment generating function is defined for any $t < \frac{1}{2}$ and it is equal to:

$$M_X(t) = (1 - 3t)^{-n/2}$$

Answers: Self Assessment

1. (d) 2. (a) 3. (b) 4. (c)

12.9 Further Readings

Notes

Books

Introductory Probability and Statistical Applications by P.L. Meyer

Introduction to Mathematical Statistics by Hogg and Craig

Fundamentals of Mathematical Statistics by S.C. Gupta and V.K. Kapoor

Notes

Unit 13: Moment Generating Function - Continue

CONTENTS

Objectives

Introduction

13.1 Joint Moment Generating Function

13.2 Properties of Moment Generating Function

13.3 Summary

13.4 Keywords

13.5 Self Assessment

13.6 Review Questions

13.7 Further Readings

Objectives

After studying this unit, you will be able to:

- Discuss the joint moment generating function
- Describe properties of moment generating function

Introduction

In probability theory and statistics, the **moment-generating function** of any random variable is an alternative definition of its probability distribution. Thus, it provides the basis of an alternative route to analytical results compared with working directly with probability density functions or cumulative distribution functions. There are particularly simple results for the moment-generating functions of distributions defined by the weighted sums of random variables.

In addition to univariate distributions, moment-generating functions can be defined for vector- or matrix-valued random variables, and can even be extended to more general cases.

The moment-generating function does not always exist even for real-valued arguments, unlike the characteristic function. There are relations between the behavior of the moment-generating function of a distribution and properties of the distribution, such as the existence of moments.

13.1 Joint Moment Generating Function

we start this lecture by defining the moment generating function of a random vector.

Definition Let X be a K × 1 random vector. If the expected value

$$E[\exp(t^T X) = E[\exp(t_1 X_1 + t_2 X_2 + \dots t_K X_K)]$$

exists and is finite for all k × 1 real vectors t belonging to a closed rectangle H:

$$H = [-h_1, h_1] \times [-h_2, h_2] \times \dots \times [-h_K, h_K] \subseteq \mathbb{R}^K$$

with hi > 0 for all i = ,...,K, then we say that X possesses a joint moment generating function (or joint mgf) and the function $M_X : H \rightarrow \mathbb{R}$ defined by

$$M_X(t) = E[\exp(t^T X)]$$

is called the joint moment generating function of X.

The following example shows how the joint moment generating function of a standard multivariate normal random vector is calculated:

Example: Let X be a K × 1 standard multivariate normal random vector. Its support R_X is

$$R_X = \mathbb{R}^K$$

and its joint probability density function $f_X(x)$ is

$$f_X(x) = (2\pi)^{-K/2} \exp\left(-\frac{1}{2}x^T x\right)$$

Therefore, the joint moment generating function of X can be derived as follows:

$$M_X(t) = E[\exp(t^T X)]$$

$$= E[\exp(t_1 X_1 + t_2 X_2 + \dots + t_K X_K)]$$

$$= E\left[\prod_{i=1}^{K} \exp(t_i X_i)\right]$$

$$= \prod_{i=1}^{K} E[\exp(t_i X_i)] \qquad \text{(by mutual indepdence of the entries of X)}$$

$$= \prod_{i=1}^{K} M_{X_i}(t_i) \qquad \text{(by the definition of moment generating function)}$$

where we have used the fact that the entries of X are mutually independent (see mutual independence via expectations) and the definition of the moment generating function of a random variable. Since the moment generating function of a standard normal random variable is

$$M_{X_i}(t_i) = \exp\left(\frac{1}{2}t_i^2\right)$$

then the joint moment generating function of X is

$$MX(t) = \prod_{i=1}^{K} M_{X_i}(t_i)$$

$$= \prod_{i=1}^{K} \exp\left(\frac{1}{2}t_i^2\right)$$

$$= \exp\left(\frac{1}{2}\sum_{i=1}^{K} t_i^2\right)$$

$$= \exp\left(\frac{1}{2}t^T t\right)$$

Notes

Notes The moment generating function $M_{X_i}(t_i)$ of a standard normal random variable is defined for any $t_i \in \mathbb{R}$. As a consequence, the joint moment generating function of X is defined for any $t \in \mathbb{R}^K$.

Example 2: Let

$$X = [X_1 \quad X_2]^T$$

be a 2 × 1 random vector with joint moment generating function

$$M_{X_1,X_2}(t_1, t_2) = \frac{1}{3} + \frac{2}{3}\exp(t_1 + 2t_2)$$

Derive the expected value of X_1.

Solution

The moment generating function of X_1 is:

$$\begin{aligned} M_{X_1}(t_1) &= E[\exp(t_1 X_1)] \\ &= E[\exp(t_1 X_1 + 0.X_2)] \\ &= M_{X_1},X_2(t_1,0) \\ &= \frac{1}{3} + \frac{2}{3}\exp(t_1 + 2.0) \\ &= \frac{1}{3} + \frac{2}{3}\exp(t_1) \end{aligned}$$

The expected value of X_1 is obtained by taking the first derivative of its moment generating function:

$$\frac{dM_{X_1}(t_1)}{dt_1} = \frac{2}{3}\exp(t_1)$$

and evaluating it at $t_1 = 0$:

$$E[X_1] = \left.\frac{dM_{X_1}(t_1)}{dt_1}\right|_{t_1=0} = \frac{2}{3}\exp(0) = \frac{2}{3}$$

Example 3: Let

$$X = [X_1 \quad X_2]^T$$

be a 2 × 1 random vector with joint moment generating function

$$M_{X_1,X_2}(t_1, t_2) = \frac{1}{3}[1 + \exp(t_1 + 2t_2) + \exp(2t_1 + t_2)]$$

Derive the covariance between X_1 and X_2.

Solution

Notes

We can use the following covariance formula:

$$\text{Cov}[X_1, X_2] = E[X_1X_2] - E[X_1]E[X_2]$$

The moment generating function of X_1 is:

$$M_{X_1}(t_1) = E[\exp(t_1X_1)]$$

$$= E[\exp(t_1X_1 + 0.X_2)]$$

$$= M_{X_1,X_2}(t_1,0)$$

$$= E[\exp(t_1X_1 + 0.X_2)]$$

$$= M_{X_1,X_2}(t_1, 0)$$

$$= \frac{1}{3}[1+\exp(t_1+2.0)+\exp(2t_1+0)]$$

$$= \frac{1}{3}[1+\exp(t_1)+\exp(2t_1)]$$

The expected value of X_1 is obtained by taking the first derivative of its moment generating function

$$\frac{dM_{X_1}(t_1)}{dt_1} = \frac{1}{3}[\exp(t_1)+2\exp(2t_1)]$$

and evaluating it at $t_1 = 0$

$$E[X_1] = \left.\frac{dM_{X_1}(t_1)}{dt_1}\right|_{t_1=0} = \frac{1}{3}[\exp(0)+2\exp(0)] = 1$$

The moment generating function of X_2 is

$$M_{X_2}(t_2) = E[\exp(t_2X_2)]$$

$$= E[\exp(0.X_1 + t_2X_2)]$$

$$= M_{X_1,X_2}(0,t_2)$$

$$= \frac{1}{3}[1+\exp(0+2t_2)+\exp(2.0+t_2)]$$

$$= \frac{1}{3}[1+\exp(2t_2)+\exp(t_2)]$$

To compute the expected value of X_2 we take the first derivative of its moment generating function

$$\frac{dM_{X2}(t_2)}{dt_2} = \frac{1}{3}[2\exp(t_2)+\exp(t_2)]$$

and evaluating it at $r^2 = 0$

$$E[X_2] = \left.\frac{dM_{X2}(t_2)}{dt_2}\right|_{t_2=0} = \frac{1}{3}[2\exp(0)+\exp(0)] = 1$$

Notes

The second cross-moment of X is computed by taking the second cross-partial derivative of the joint moment generation function

$$\frac{\partial^2 M_{X_1,X_2}(t_1,t_2)}{\partial t_1 \partial t_2} = \frac{\partial}{\delta t_1}\left(\frac{\partial}{\partial t_2}\left(\frac{1}{3}[1+\exp(t_1+2t_2)+\exp(2t_1+t_2)]\right)\right)$$

$$= \frac{\partial}{\delta t_1}\left(\frac{1}{3}[2\exp(t_1+2t_2)+\exp(2t_1+t_2)]\right)$$

$$= \frac{1}{3}[2\exp(t_1+2t_2)+\exp(2t_1+t_2)]$$

and evaluating it at $(t_1, t_2) = (0, 0)$:

$$E[X_1 X_2] = \left.\frac{\partial^2 M_{X_1,X_2}(t_1,t_2)}{\partial t_1 \partial t_2}\right|_{t_1=0,\, t_2=0}$$

$$= \frac{1}{3}[2\exp(0)+2\exp(0)]$$

$$= \frac{4}{3}$$

Therefore

$$Cov[X_1, X_2] = E[X_1X_2] - E[X_1]E[X_2]$$

$$= \frac{4}{3} - 1.1$$

$$= \frac{1}{3}$$

13.2 Properties of Moment Generating Function

(a) The most significant propertyof moment generating function is that "the moment generating function uniquely determines the distribution."

(b) Let a and b be constants, and let MX(t) be the mgf of a random variable X. Then the mgf of the random variable Y = a + bX can be given as follows

$$M_Y(t) = E[e^{tY}] = E[e^{t(a+bX)}] = e^{at}E[e(bt)X] = eatMX(bt)$$

(c) Let X and Y be independent random variables having the respective mgf's $M_X(t)$ and $M_Y(t)$. Recall that $E[g_1(X)g_2(Y)] = E[g_1(X)]E[g_2(Y)]$ for functions g_1 and g_2. We can obtain the mgf Mz(t) of the sum Z = X + Y of random variables as follows.

(d) When t = 0, it clearly follows that M(0) = 1. Now by differentiating M(t)r times, we obtain

$$M^{(r)}(t) = \frac{d^r}{dt^r}E[e^{tX}] = E\left[\frac{d^t}{dt^r}e^{tX}\right] = E[X^r e^{tX}]$$

In particular when t = 0, $M^{(r)}(0)$ generates the r-th moment of X as follows.

$$M^{(r)(0)} = E[X^r],\ r = 1, 2, 3,...$$

Example 4: Find the mgf M(t) for a uniform random variable X on [a, b]. And derive the derivative M′(0) at t = 0 by using the definition of derivative and L'Hospital's rule.

Notes

Distribution	Moment-generating function MX(t)	Characteristic function φ(t)
Bernoulli $P(X = 1) = p$	$1 - p + pe^t$	$1 - p + pe^{it}$
Geometric $(1 - p)^{k-1}p$	$\frac{pe^t}{1-(1-p)e^t}$ for $t < 1 - \ln(1-p)$	$\frac{pe^{it}}{1-(1-p)e^{it}}$
Binomial B(n, p)	$(1 - p + pe^t)^n$	$(1 - p + pe^{it})^n$
Poisson Pois(λ)	$e^{\lambda(e^t-1)}$	$e^{\lambda(e^{it}-1)}$
Uniform U(a, b)	$\frac{e^{tb}-e^{ta}}{t(b-a)}$	$\frac{e^{itb}-e^{ita}}{it(b-a)}$
Normal N(μ, σ2)	$e^{t\mu+\frac{1}{2}\sigma^2t^2}$	$e^{it\mu-\frac{1}{2}\sigma^2t^2}$
Chi-square χ2k	$(1 - 2t)^{-k/2}$	$(1 - 2it)^{-k/2}$
Gamma Γ(k, θ)	$(1 - t\theta)^{-k}$	$(1 - it\theta)^{-k}$
Exponential Exp(λ)	$(1 - t\lambda^1)^{-1}$	$(1 - it\lambda^1)^{-1}$
Multivariate normal N(μ, Σ)	$e^{t^T}\mu+\frac{1}{2}t^T\Sigma t$	$e^{it^T}\mu+\frac{1}{2}t^T\Sigma t$
Degenerate δa	e^{ta}	e^{ita}
Laplace L(μ, b)	$\frac{e^{t\mu}}{1-b^2t^2}$	$\frac{e^{it\mu}}{1+b^2t^2}$
Cauchy Cauchy(μ, θ)	not defined	$e^{it\mu-\theta\|t\|}$
Negative Binomial NB(r, p)	$\frac{(1-p)^r}{(1-pe^t)r}$	$\frac{(1-p)^r}{(1-pe^{it})r}$

13.3 Summary

- Definition Let X be a K × 1 random vector. If the expected value

$$E[\exp(t^TX) = E[\exp(t_1X_1 + t_2X_2 + \ldots t_KX_K)]$$

exists and is finite for all k × 1 real vectors t belonging to a closed rectangle H:

$$H = [-h_1, h_1] \times [-h_2, h_2] \times \ldots \times [-h_K, h_K] \subseteq \mathbb{R}^K$$

with hi > 0 for all i = ,...,K, then we say that X possesses a joint moment generating function (or joint mgf) and the function $M_X : H \rightarrow \mathbb{R}$ defined by

$$M_X(t) = E[\exp(t^TX)]$$

is called the joint moment generating function of X.

- Let a and b be constants, and let MX(t) be the mgf of a random variable X. Then the mgf of the random variable Y = a + bX can be given as follows

$$M_Y(t) = E[e^{tY}] = E[e^{t(a + bX)}] = e^{at}E[e(bt)X] = eatMX(bt)$$

- Let X and Y be independent random variables having the respective mgf's $M_X(t)$ and $M_Y(t)$. Recall that $E[g_1(X)g_2(Y)] = E[g_1(X)]E[g_2(Y)]$ for functions g_1 and g_2. We can obtain the mgf Mz(t) of the sum Z = X + Y of random variables as follows.

Notes

- When t = 0, it clearly follows that M(0) = 1. Now by differentiating M(t)r times, we obtain

  $$M^{(r)}(t) = \frac{d^r}{dt^r}E[e^{tX}] = E\left[\frac{d^t}{dt^r}e^{tX}\right] = E[X^r e^{tX}]$$

  In particular when t = 0, $M^{(r)}(0)$ generates the r-th moment of X as follows.

  $$M^{(r)(0)} = E[X^r],\ r = 1, 2, 3,...$$

13.4 Keywords

Moment-generating function: In probability theory and statistics, the moment-generating function of any random variable is an alternative definition of its probability distribution.

Standard normal random variable: The moment generating function $M_{Xi}(t_i)$ of a standard normal random variable is defined for any $t_i \in \mathbb{R}$. As a consequence, the joint moment generating function of X is defined for any $t \in \mathbb{R}^K$.

13.5 Self Assessment

1. In addition to ...., moment-generating functions can be defined for vector- or matrix-valued random variables, and can even be extended to more general cases.

   (a) finite cross-moments (b) uniquely determines

   (c) univariate distributions, (d) moment-generating

2. The .... function does not always exist even for real-valued arguments, unlike the characteristic function. There are relations between the behavior of the moment-generating function of a distribution and properties of the distribution, such as the existence of moments.

   (a) finite cross-moments (b) uniquely determines

   (c) univariate distributions, (d) moment-generating

3. If a K × 1 random vector X possesses a joint moment generating function $M_X(t)$, then, for any $n \in \mathbb{N}$, X possesses .... of order n.

   (a) finite cross-moments (b) uniquely determines

   (c) univariate distributions, (d) moment-generating

4. The most significant propertyof moment generating function is that "the moment generating function .... the distribution."

   (a) finite cross-moments (b) uniquely determines

   (c) univariate distributions, (d) moment-generating

13.6 Review Questions

1. Let X be a K × 1 standard multivariate normal random vector. Its support R_X is

   $$R_X = \mathbb{R}^K$$

   and its joint probability density function $f_X(x)$ is

   $$f_X(x) = (2\pi)^{-K/2}\exp\left(-\frac{1}{4}x^T x\right)$$

2. Continuing the example above, the joint moment generating function of a 2×1 standard normal random vector X is

$$M_X(t) = \exp\left(\frac{1}{3}t^T t\right) = \exp\left(\frac{1}{3}t_1^2 + \frac{1}{2}t_2^2\right)$$

3. Let X be a 2×1 discrete random vector and denote its components by X_1 and X_2. Let the support of X be

$$RX = \{[1 \quad 1]^T, [2 \quad 2]^T, [0 \quad 0]^T\}$$

and its joint probability mass function be

$$p_X(x) = \begin{cases} \frac{1}{3} & \text{if } x = [1 \quad 1]^T \\ \frac{1}{3} & \text{if } x = [2 \quad 2]^T \\ \frac{1}{3} & \text{if } x = [0 \quad 0]^T \\ 0 & \text{otherwise} \end{cases}$$

Derive the joint moment generating function of X, if it exists.

4. Let

$$X = [X_1 \quad X_2]^T$$

be a 2×1 random vector with joint moment generating function

$$M_{X_1, X_2}(t_1, t_2) = \frac{1}{3} + \frac{1}{3}\exp(t_1 + 2t_2)$$

Derive the expected value of X_1.

Answers: Self Assessment

1. (c) 2. (d) 3. (a) 4. (b)

13.7 Further Readings

Books

Introductory Probability and Statistical Applications by P.L. Meyer

Introduction to Mathematical Statistics by Hogg and Craig

Fundamentals of Mathematical Statistics by S.C. Gupta and V.K. Kapoor

Notes

Unit 14: Theoretical Probability Distributions

CONTENTS

Objectives

Introduction

14.1 Binomial Distribution

14.1.1 Probability Function or Probability Mass Function

14.1.2 Summary Measures of Binomial Distribution

14.1.3 Fitting of Binomial Distribution

14.1.4 Features of Binomial Distribution

14.1.5 Uses of Binomial Distribution

14.2 Hypergeometric Distribution

14.2.1 Binomial Approximation to Hypergeometric Distribution

14.3 Pascal Distribution

14.4 Geometrical Distribution

14.5 Uniform Distribution (Discrete Random Variable)

14.6 Poisson Distribution

14.6.1 Poisson Process

14.6.2 Probability Mass Function

14.6.3 Summary Measures of Poisson Distribution

14.6.4 Poisson Approximation to Binomial

14.6.5 Fitting of a Poisson Distribution

14.6.6 Features of Poisson Distribution

14.6.7 Uses of Poisson Distribution

14.7 Summary

14.8 Keywords

14.9 Self Assessment

14.10 Review Questions

14.11 Further Readings

Objectives

After studying this unit, you will be able to:

- Discuss Binomial Distribution
- Describe Hypergeometric Distribution
- Explain Pascal Distribution
- Discuss Geometrical Distribution
- Describe Uniform Distribution (Discrete Random Variable)
- Explain Poisson Distribution

Introduction

The study of a population can be done either by constructing an observed (or empirical) frequency distribution, often based on a sample from it, or by using a theoretical distribution. We have already studied the construction of an observed frequency distribution and its various summary measures. Now we shall learn a more scientific way to study a population through the use of theoretical probability distribution of a random variable. It may be mentioned that a theoretical probability distribution gives us a law according to which different values of the random variable are distributed with specified probabilities. It is possible to formulate such laws either on the basis of given conditions (a priori considerations) or on the basis of the results (a posteriori inferences) of an experiment.

If a random variable satisfies the conditions of a theoretical probability distribution, then this distribution can be fitted to the observed data.

14.1 Binomial Distribution

Binomial distribution is a theoretical probability distribution which was given by James Bernoulli. This distribution is applicable to situations with the following characteristics:

1. An experiment consists of a finite number of repeated trials.
2. Each trial has only two possible, mutually exclusive, outcomes which are termed as a 'success' or a 'failure'.
3. The probability of a success, denoted by p, is known and remains constant from trial to trial. The probability of a failure, denoted by q, is equal to 1 - p.
4. Different trials are independent, i.e., outcome of any trial or sequence of trials has no effect on the outcome of the subsequent trials.

   The sequence of trials under the above assumptions is also termed as Bernoulli Trials.

14.1.1 Probability Function or Probability Mass Function

Let n be the total number of repeated trials, p be the probability of a success in a trial and q be the probability of its failure so that $q = 1 - p$.

Let r be a random variable which denotes the number of successes in n trials. The possible values of r are 0, 1, 2, ... n. We are interested in finding the probability of r successes out of n trials, i.e., P(r).

Notes

To find this probability, we assume that the first r trials are successes and remaining n - r trials are failures. Since different trials are assumed to be independent, the probability of this sequence is

$$\underbrace{p.p.\;....\;p}_{r\;times}\;\underbrace{q.q.\;....\;q}_{(n-r)\;times}\;\text{ i.e. } p^r q^{n-r}.$$

Since out of n trials any r trials can be success, the number of sequences showing any r trials as success and remaining (n - r) trials as failure is ${}^nC_r$, where the probability of r successes in each trial is $p^r q^{n-r}$. Hence, the required probability is $P(r) = {}^nC_r p^r q^{n-r}$, where r = 0, 1, 2, ... n.

Writing this distribution in a tabular form, we have

r	0	1	2	......	n	*Total*
$P(r)$	${}^nC_0 p^0 q^n$	${}^nC_1 p\; q^{n-1}$	${}^nC_2 p^2 q^{n-2}$	......	${}^nC_n p^n q^0$	1

It should be noted here that the probabilities obtained for various values of r are the terms in the binomial expansion of $(q + p)^n$ and thus, the distribution is termed as Binomial Distribution. $P(r) = {}^nC_r p^r q^{n-r}$ is termed as the probability function or probability mass function (p.m.f.) of the distribution.

14.1.2 Summary Measures of Binomial Distribution

(a) Mean

The mean of a binomial variate r, denoted by μ, is equal to E(r), i.e.,

$$\mu = E(r) = \sum_{r=0}^{n} rP(r) = \sum_{r=1}^{n} r.{}^nC_r p^r q^{n-r} \text{ (note that the term for r = 0 is 0)}$$

$$= \sum_{r=1}^{n} \frac{r.n!}{r!(n-r)!} \cdot p^r q^{n-r} = \sum_{r=1}^{n} \frac{n.(n-1)!}{(r-1)!(n-r)!} \cdot p^r q^{n-r}$$

$$= np\sum_{r=1}^{n} \frac{(n-1)!}{(r-1)!(n-r)!} \cdot p^{r-1} q^{n-r} = np(q+p)^{n-1} = np \qquad [\because q+p=1]$$

(b) Variance

The variance of r, denoted by σ^2, is given by

$$\sigma^2 = E\left[r - E(r)\right]^2 = E\left[r - np\right]^2 = E\left[r^2 - 2npr + n^2p^2\right]$$

$$= E\left(r^2\right) - 2npE(r) + n^2p^2 = E\left(r^2\right) - 2n^2p^2 + n^2p^2$$

$$= E\left(r^2\right) - n^2p^2 \qquad .... (1)$$

Thus, to find σ^2, we first determine $E(r^2)$.

Notes

Now, $E\left(r^2\right)=\sum_{r=1}^{n} r^2.{}^nC_r p^r q^{n-r} = \left[r\left(r-1\right)+r\right]\ {}^nC_r p^r q^{n-r}$

$$= \sum_{r=2}^{n} r\left(r-1\right){}^nC_r p^r q^{n-r} + \sum_{r=1}^{n} r.{}^nC_r p^r q^{n-r} = \sum_{r=2}^{n} \frac{r\left(r-1\right)n!}{r!\left(n-r\right)!}\ .\ p^r q^{n-r} + np$$

$$= \sum_{r=2}^{n} \frac{n!}{\left(r-2\right)!\left(n-r\right)!}\ .\ p^r q^{n-r} + np = \sum_{r=2}^{n} \frac{n\left(n-1\right).\left(n-2\right)!}{\left(r-2\right)!\left(n-r\right)!}\ .\ p^r q^{n-r} + np$$

$$= n\left(n-1\right)p^2 \sum_{r=2}^{n} \frac{\left(n-2\right)!}{\left(r-2\right)!\left(n-r\right)!}\ .\ p^{r-2} q^{n-r} + np$$

$$= n\left(n-1\right)p^2\left(q+p\right)^{n-2} + np = n\left(n-1\right)p^2 + np$$

Substituting this value in equation (1), we get

$$\sigma^2 = n\left(n-1\right)p^2 + np - n^2p^2 = np\left(1-p\right) = npq$$

Or the standard deviation $= \sqrt{npq}$

> *Note* $\sigma^2 = npq = mean \times q$, which shows that $\sigma^2 < mean$, since 0 < q <1.

(c) The values of m_3, m_4, b_1 and b_2

Proceeding as above, we can obtain

$$\mu_3 = E\left(r-np\right)^3 = npq\left(q-p\right)$$

$$\mu_4 = E\left(r-np\right)^4 = 3n^2p^2q^2 + npq\left(1-6pq\right)$$

Also $\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{n^2p^2q^2\left(q-p\right)^2}{n^3p^3q^3} = \frac{\left(q-p\right)^2}{npq}$

The above result shows that the distribution is symmetrical when

p = q = $\frac{1}{2}$, negatively skewed if q < p, and positively skewed if q > p

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{3n^2p^2q^2 + npq\left(1-6pq\right)}{n^2p^2q^2} = 3 + \frac{\left(1-6pq\right)}{npq}$$

The above result shows that the distribution is leptokurtic if 6pq < 1, platykurtic if 6pq > 1 and mesokurtic if 6pq = 1.

(d) Mode

Notes

Mode is that value of the random variable for which probability is maximum.

If r is mode of a binomial distribution, we have

$$P(r - 1) \le P(r) \ge P(r + 1)$$

Consider the inequality P(r) ³ P(r + 1)

or $\quad {}^nC_r p^r q^{n-r} \ge {}^nC_{r+1} p^{r+1} q^{n-r-1}$

or $$\frac{n!}{r!(n-r)!} p^r q^{n-r} \ge \frac{n!}{(r+1)!(n-r-1)!} p^{r+1} q^{n-r-1}$$

or $$\frac{1}{(n-r)}.q \ge \frac{1}{(r+1)}.p \quad or \quad qr + q \ge np - pr$$

Solving the above inequality for r, we get

$$r \ge (n+1)p - 1 \qquad \text{.... (1)}$$

Similarly, on solving the inequality P(r - 1) £ P(r) for r, we can get

$$r \le (n+1)p \qquad \text{.... (2)}$$

Combining inequalities (1) and (2), we get

$$(n+1)p - 1 \le r \le (n+1)p$$

Case I. When (n + 1)p is not an integer

When (n + 1)p is not an integer, then (n + 1)p - 1 is also not an integer. Therefore, mode will be an integer between (n + 1)p - 1 and (n + 1)p or mode will be an integral part of (n + 1)p.

Case II. When (n + 1)p is an integer

When (n + 1)p is an integer, the distribution will be bimodal and the two modal values would be (n + 1)p - 1 and (n + 1)p.

Example 1: An unbiased die is tossed three times. Find the probability of obtaining (a) no six, (b) one six, (c) at least one six, (d) two sixes and (e) three sixes.

Solution.

The three tosses of a die can be taken as three repeated trials which are independent. Let the occurrence of six be termed as a success. Therefore, r will denote the number of six obtained.

Further, n = 3 and $p = \frac{1}{6}$.

(a) Probability of obtaining no six, i.e.,

$$P(r = 0) = {}^3C_0 p^0 q^3 = 1.\left(\frac{1}{6}\right)^0 \left(\frac{5}{6}\right)^3 = \frac{125}{216}$$

Notes

(b) $P(r=1) = {}^3C_1 p^1 q^2 = 3.\left(\frac{1}{6}\right)\left(\frac{5}{6}\right)^2 = \frac{25}{72}$

(c) Probability of getting at least one six = 1 - P(r = 0) $= 1 - \frac{125}{216} = \frac{91}{216}$

(d) $P(r=2) = {}^3C_2 p^2 q^1 = 3.\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right) = \frac{5}{72}$

(e) $P(r=3) = {}^3C_3 p^3 q^0 = 3.\left(\frac{1}{6}\right)^3 = \frac{1}{216}$

Example 2:

Assuming that it is true that 2 in 10 industrial accidents are due to fatigue, find the probability that:

(a) Exactly 2 of 8 industrial accidents will be due to fatigue.

(b) At least 2 of the 8 industrial accidents will be due to fatigue.

Solution.

Eight industrial accidents can be regarded as Bernoulli trials each with probability of success $p = \frac{2}{10} = \frac{1}{5}$. The random variable r denotes the number of accidents due to fatigue.

(a) $P(r=2) = {}^8C_2\left(\frac{1}{5}\right)^2\left(\frac{4}{5}\right)^6 = 0.294$

(b) We have to find P(r ≥ 2). We can write

P(r ≥ 2) = 1 – P(0) – P(1), thus, we first find P(0) and P(1).

We have $P(0) = {}^8C_0\left(\frac{1}{5}\right)^0\left(\frac{4}{5}\right)^8 = 0.168$

and $P(1) = {}^8C_1\left(\frac{1}{5}\right)^1\left(\frac{4}{5}\right)^7 = 0.336$

∴ P(r ≥ 2) = 1 – 0.168 – 0.336 = 0.496

Example 3: The proportion of male and female students in a class is found to be 1 : 2. What is the probability that out of 4 students selected at random with replacement, 2 or more will be females?

Solution.

Notes

Let the selection of a female student be termed as a success. Since the selection of a student is made with replacement, the selection of 4 students can be taken as 4 repeated trials each with probability of success $p = \frac{2}{3}$.

Thus, P(r ≥ 2) = P(r = 2) + P(r = 3) +P(r = 4)

$$= {}^4C_2\left(\frac{2}{3}\right)^2\left(\frac{1}{3}\right)^2 + {}^4C_3\left(\frac{2}{3}\right)^3\left(\frac{1}{3}\right) + {}^4C_4\left(\frac{2}{3}\right)^4 = \frac{8}{9}$$

Note that P(r ≥ 2) can alternatively be found as 1 - P(0) - P(1)

Example 4: The probability of a bomb hitting a target is 1/5. Two bombs are enough to destroy a bridge. If six bombs are aimed at the bridge, find the probability that the bridge is destroyed.

Solution.

Here n = 6 and $p = \frac{1}{5}$

The bridge will be destroyed if at least two bomb hit it. Thus, we have to find P(r ^{3} 2). This is given by

$$P(r \geq 2) = 1 - P(0) - P(1) = 1 - {}^6C_0\left(\frac{4}{5}\right)^6 - {}^6C_1\left(\frac{1}{5}\right)\left(\frac{4}{5}\right)^5 = \frac{1077}{3125}$$

Example 5: An insurance salesman sells policies to 5 men all of identical age and good health. According to the actuarial tables, the probability that a man of this particular age will be alive 30 years hence is 2/3. Find the probability that 30 years hence (i) at least 1 man will be alive, (ii) at least 3 men will be alive.

Solution.

Let the event that a man will be alive 30 years hence be termed as a success. Therefore, n = 5 and $p = \frac{2}{3}$.

(i) P(r ^{3} 1) = 1 - P(r = 0) $= 1 - {}^5C_0\left(\frac{2}{3}\right)^0\left(\frac{1}{3}\right)^5 = \frac{242}{243}$

(ii) P(r ^{3} 3) = P(r = 3) + P(r = 4) +P(r = 5)

$$= {}^5C_3\left(\frac{2}{3}\right)^3\left(\frac{1}{3}\right)^2 + {}^5C_4\left(\frac{2}{3}\right)^4\left(\frac{1}{3}\right) + {}^5C_5\left(\frac{2}{3}\right)^5 = \frac{64}{81}$$

Example 6: Ten percent of items produced on a machine are usually found to be defective. What is the probability that in a random sample of 12 items (i) none, (ii) one, (iii) two, (iv) at the most two, (v) at least two items are found to be defective?

Solution.

Notes

Let the event that an item is found to be defective be termed as a success. Thus, we are given n = 12 and p = 0.1.

(i) $P(r=0) = {}^{12}C_0(0.1)^0(0.9)^{12} = 0.2824$

(ii) $P(r=1) = {}^{12}C_1(0.1)^1(0.9)^{11} = 0.3766$

(iii) $P(r=2) = {}^{12}C_2(0.1)^2(0.9)^{10} = 0.2301$

(iv) $P(r \le 2)$ = P(r = 0) + P(r = 1) +P(r = 2)

= 0.2824 + 0.3766 + 0.2301 = 0.8891

(v) P(r ≥ 2) = 1 - P(0) - P(1) = 1 - 0.2824 - 0.3766 = 0.3410

Example 7: In a large group of students 80% have a recommended statistics book. Three students are selected at random. Find the probability distribution of the number of students having the book. Also compute the mean and variance of the distribution.

Solution.

Let the event that 'a student selected at random has the book' be termed as a success. Since the group of students is large, 3 trials, i.e., the selection of 3 students, can be regarded as independent with probability of a success p = 0.8. Thus, the conditions of the given experiment satisfies the conditions of binomial distribution.

The probability mass function $P(r) = {}^3C_r(0.8)^r(0.2)^{3-r}$,

where r = 0, 1, 2 and 3

The mean is np = 3 × 0.8 = 2.4 and Variance is npq = 2.4 × 0.2 = 0.48

Example 8:

(a) The mean and variance of a discrete random variable X are 6 and 2 respectively. Assuming X to be a binomial variate, find P(5 ≤ X ≤ 7).

(b) In a binomial distribution consisting of 5 independent trials, the probability of 1 and 2 successes are 0.4096 and 0.2048 respectively. Calculate the mean, variance and mode of the distribution.

Solution.

(a) It is given that np = 6 and npq = 2

$$\therefore\ q = \frac{npq}{np} = \frac{2}{6} = \frac{1}{3} \text{ so that } p = 1 - \frac{1}{3} = \frac{2}{3} \text{ and } n = 6 \times \frac{3}{2} = 9$$

Now P(5 ≤ X ≤ 7) = P(X = 5) + P(X = 6) +P(X = 7)

$$= {}^9C_5\left(\frac{2}{3}\right)^5\left(\frac{1}{3}\right)^4 + {}^9C_6\left(\frac{2}{3}\right)^6\left(\frac{1}{3}\right)^3 + {}^9C_7\left(\frac{2}{3}\right)^7\left(\frac{1}{3}\right)^2$$

$$= \frac{2^5}{3^9}\left[{}^9C_5 + {}^9C_6 \times 2 + {}^9C_7 \times 4\right] = \frac{2^5}{3^9} \times 438$$

(b) Let p be the probability of a success. It is given that

Notes

$$^5C_1 p(1-p)^4 = 0.4096 \ \ and \ \ ^5C_2 p^2(1-p)^3 = 0.2048$$

Using these conditions, we can write

$$\frac{5p(1-p)^4}{10p^2(1-p)^3} = \frac{0.4096}{0.2048} = 2 \ \ or \ \ \frac{(1-p)}{p} = 4. \text{ This gives } p = \frac{1}{5}$$

Thus, mean is $np = 5 \times \frac{1}{5} = 1 \ \ and \ \ npq = 1 \times \frac{4}{5} = 0.8$

Since (n +1)p, i.e., $6 \times \frac{1}{5}$ is not an integer, mode is its integral part, i.e., = 1.

Example 9: 5 unbiased coins are tossed simultaneously and the occurrence of a head is termed as a success. Write down various probabilities for the occurrence of 0, 1, 2, 3, 4, 5 successes. Find mean, variance and mode of the distribution.

Solution.

Here n = 5 and $p = q = \frac{1}{2}$.

The probability mass function is $P(r) = {}^5C_r\left(\frac{1}{2}\right)^5$, r = 0, 1, 2, 3, 4, 5.

The probabilities of various values of r are tabulated below :

r	0	1	2	3	4	5	*Total*
$P(r)$	$\frac{1}{32}$	$\frac{5}{32}$	$\frac{10}{32}$	$\frac{10}{32}$	$\frac{5}{32}$	$\frac{1}{32}$	1

Mean = $np = 5 \times \frac{1}{2} = 2.5 \ \ and \ \ variance = 2.5 \times \frac{1}{2} = 1.25$

Since (n+1)p = $6 \times \frac{1}{2} = 3$ is an integer, the distribution is bimodal and the two modes are 2 and 3.

14.1.3 Fitting of Binomial Distribution

The fitting of a distribution to given data implies the determination of expected (or theoretical) frequencies for different values of the random variable on the basis of this data.

The purpose of fitting a distribution is to examine whether the observed frequency distribution can be regarded as a sample from a population with a known probability distribution.

To fit a binomial distribution to the given data, we find its mean. Given the value of n, we can compute the value of p and, using n and p, the probabilities of various values of the random variable. These probabilities are multiplied by total frequency to give the required expected frequencies. In certain cases, the value of p may be determined by the given conditions of the experiment.

Notes

Example 10: The following data give the number of seeds germinating (X) out of 10 on damp filter for 80 sets of seed. Fit a binomial distribution to the data.

X	:	0	1	2	3	4	5	6	7	8	9	10
f	:	6	20	28	12	8	6	0	0	0	0	0

Solution.

Here the random variable X denotes the number of seeds germinating out of a set of 10 seeds. The total number of trials n = 10.

The mean of the given data

$$\overline{X} = \frac{0 \times 6 + 1 \times 20 + 2 \times 28 + 3 \times 12 + 4 \times 8 + 5 \times 6}{80} = \frac{174}{80} = 2.175$$

Since mean of a binomial distribution is np, $\therefore$ np = 2.175. Thus, we get $p = \frac{2.175}{10} = 0.22 \ (approx.)$. Further, q = 1 - 0.22 = 0.78.

Using these values, we can compute $P(X) = {}^{10}C_X (0.22)^X (0.78)^{10-X}$ and then expected frequency [= N × P(X)] for X = 0, 1, 2, ... 10. The calculated probabilities and the respective expected frequencies are shown in the following table :

X	$P(X)$	$N \times P(X)$	*Approximated Frequency*	X	$P(X)$	$N \times P(X)$	*Approximated Frequency*
0	0.0834	6.67	6	6	0.0088	0.71	1
1	0.2351	18.81	19	7	0.0014	0.11	0
2	0.2984	23.87	24	8	0.0001	0.01	0
3	0.2244	17.96	18	9	0.0000	0.00	0
4	0.1108	8.86	9	10	0.0000	0.00	0
5	0.0375	3.00	3	*Total*	1.0000		80

14.1.4 Features of Binomial Distribution

1. It is a discrete probability distribution.
2. It depends upon two parameters n and p. It may be pointed out that a distribution is known if the values of its parameters are known.
3. The total number of possible values of the random variable are n + 1. The successive binomial coefficients are ${}^nC_0$, ${}^nC_1$, ${}^nC_2$, ... ${}^nC_n$. Further, since ${}^nC_r = {}^nC_{n-r}$, these coefficients are symmetric.

   The values of these coefficients, for various values of n, can be obtained directly by using Pascal's triangle.

Pascal's Triangle

Notes

n	*Binomial Coefficients*	*Sum of Coefficients* (2^n)
1	1 1	$2^1 = 2$
2	1 2 1	$2^2 = 4$
3	1 3 3 1	$2^3 = 8$
4	1 4 6 4 1	$2^4 = 16$
5	1 5 10 10 5 1	$2^5 = 32$

We can note that it is very easy to write this triangle. In the first row, both the coefficients will be unity because ${}^1C_0 = {}^1C_1$. To write the second row, we write 1 in the beginning and the end and the value of the middle coefficients is obtained by adding the coefficients of the first row. Other rows of the Pascal's triangle can be written in a similar way.

4. (a) The shape and location of binomial distribution changes as the value of p changes for a given value of n. It can be shown that for a given value of n, if p is increased gradually in the interval (0, 0.5), the distribution changes from a positively skewed to a symmetrical shape. When p = 0.5, the distribution is perfectly symmetrical. Further, for larger values of p the distribution tends to become more and more negatively skewed.

   (b) For a given value of p, which is neither too small nor too large, the distribution becomes more and more symmetrical as n becomes larger and larger.

14.1.5 Uses of Binomial Distribution

Binomial distribution is often used in various decision making situations in business. Acceptance sampling plan, a technique of quality control, is based on this distribution. With the use of sampling plan, it is possible to accept or reject a lot of items either at the stage of its manufacture or at the stage of its purchase.

14.2 Hypergeometric Distribution

The binomial distribution is not applicable when the probability of a success p does not remain constant from trial to trial. In such a situation the probabilities of the various values of r are obtained by the use of Hypergeometric distribution.

Let there be a finite population of size N, where each item can be classified as either a success or a failure. Let there be k successes in the population. If a random sample of size n is taken from this population, then the probability of r successes is given by $P(r) = \frac{\left({}^kC_r\right)\left({}^{N-k}C_{n-r}\right)}{{}^NC_n}$. Here r is a discrete random variable which can take values 0, 1, 2, ... n. Also n ≤ k.

It can be shown that the mean of r is np and its variance is

$$\left(\frac{N-n}{N-1}\right).npq \text{, where } p = \frac{k}{N} \text{ and q = 1 - p.}$$

Example 11: A retailer has 10 identical television sets of a company out which 4 are defective. If 3 televisions are selected at random, construct the probability distribution of the number of defective television sets.

Solution.

Notes

Let the random variable r denote the number of defective televisions. In terms of notations, we can write N = 10, k = 4 and n = 3.

Thus, we can write $P(r) = \dfrac{{}^{4}C_r \times {}^{6}C_{3-r}}{{}^{10}C_3}, \quad r = 0,1,2,3$

The distribution of r is hypergeometric. This distribution can also be written in a tabular form as given below :

r	0	1	2	3	*Total*
$P(r)$	$\frac{5}{30}$	$\frac{15}{30}$	$\frac{9}{30}$	$\frac{1}{30}$	1

14.2.1 Binomial Approximation to Hypergeometric Distribution

In sampling problems, where sample size n (total number of trials) is less than 5% of population size N, i.e., n < 0.05N, the use of binomial distribution will also give satisfactory results. The reason for this is that the smaller the sample size relative to population size, the greater will be the validity of the requirements of independent trials and the constancy of p.

Example 12: There are 200 identical radios out of which 80 are defective. If 5 radios are selected at random, construct the probability distribution of the number of defective radios by using (i) hypergeometric distribution and (ii) binomial distribution.

Solution.

(i) It is given that N = 200, k = 80 and n = 5.

Let r be a hypergeometric random variable which denotes the number of defective radios, then

$$P(r) = \frac{{}^{80}C_r \times {}^{120}C_{5-r}}{{}^{200}C_5}, \quad r = 0,1,2,3,4,5$$

The probabilities for various values of r are given in the following table :

r	0	1	2	3	4	5	*Total*
$P(r)$	0.0752	0.2592	0.3500	0.2313	0.0748	0.0095	1

(ii) To use binomial distribution, we find p = 80/200 = 0.4.

$$P(r) = {}^{5}C_r (0.4)^r (0.6)^{5-r}, r = 0,1,2,3,4,5$$

The probabilities for various values of r are given in the following table :

r	0	1	2	3	4	5	*Total*
$P(r)$	0.0778	0.2592	0.3456	0.2304	0.0768	0.0102	1

We note that these probabilities are in close conformity with the hypergeometric probabilities.

Notes

14.3 Pascal Distribution

In binomial distribution, we derived the probability mass function of the number of successes in n (fixed) Bernoulli trials. We can also derive the probability mass function of the number of Bernoulli trials needed to get r (fixed) successes. This distribution is known as Pascal distribution. Here r and p become parameters while n becomes a random variable.

We may note that r successes can be obtained in r or more trials i.e. possible values of the random variable are r, (r + 1), (r + 2), ... etc. Further, if n trials are required to get r successes, the nth trial must be a success. Thus, we can write the probability mass function of Pascal distribution as follows:

$$P(n) = \begin{pmatrix} \text{Probability of } (r-1) \text{ successes} \\ \text{out of } (n-1) \text{ trials} \end{pmatrix} \times \begin{pmatrix} \text{Probability of a success} \\ \text{in nth trial} \end{pmatrix}$$

$$= {}^{n-1}C_{r-1} p^{r-1} q^{n-r} \times p = {}^{n-1}C_{r-1} p^{r} q^{n-r}, \text{ where n = r, (r + 1), (r + 2), ... etc.}$$

It can be shown that the mean and variance of Pascal distribution are $\frac{r}{p}$ and $\frac{rq}{p^2}$ respectively.

This distribution is also known as Negative Binomial Distribution because various values of P(n) are given by the terms of the binomial expansion of $p^r(1 - q)^{-r}$.

14.4 Geometrical Distribution

When r = 1, the Pascal distribution can be written as

$$P(n) = {}^{n-1}C_0 pq^{n-1} = pq^{n-1}, \quad \text{where } n = 1, 2, 3, ...$$

Here n is a random variable which denotes the number of trials required to get a success. This distribution is known as geometrical distribution. The mean and variance of the distribution are $\frac{1}{p}$ and $\frac{q}{p^2}$ respectively.

14.5 Uniform Distribution (Discrete Random Variable)

A discrete random variable is said to follow a uniform distribution if it takes various discrete values with equal probabilities.

If a random variable X takes values X_1, X_2, ... X_n each with probability $\frac{1}{n}$, the distribution of X is said to be uniform.

14.6 Poisson Distribution

This distribution was derived by a noted mathematician, Simon D. Poisson, in 1837. He derived this distribution as a limiting case of binomial distribution, when the number of trials n tends to become very large and the probability of success in a trial p tends to become very small such that their product np remains a constant. This distribution is used as a model to describe the probability distribution of a random variable defined over a unit of time, length or space. For example, the number of telephone calls received per hour at a telephone exchange, the number of accidents in a city per week, the number of defects per meter of cloth, the number of insurance claims per year, the number breakdowns of machines at a factory per day, the number of arrivals of customers at a shop per hour, the number of typing errors per page etc.

14.6.1 Poisson Process

Let us assume that on an average 3 telephone calls are received per 10 minutes at a telephone exchange desk and we want to find the probability of receiving a telephone call in the next 10 minutes. In an effort to apply binomial distribution, we can divide the interval of 10 minutes into 10 intervals of 1 minute each so that the probability of receiving a telephone call (i.e., a success) in each minute (i.e., trial) becomes 3/10 (note that p = m/n, where m denotes mean). Thus, there are 10 trials which are independent, each with probability of success = 3/10. However, the main difficulty with this formulation is that, strictly speaking, these trials are not Bernoulli trials. One essential requirement of such trials, that each trial must result into one of the two possible outcomes, is violated here. In the above example, a trial, i.e. an interval of one minute, may result into 0, 1, 2, ... successes depending upon whether the exchange desk receives none, one, two, ... telephone calls respectively.

One possible way out is to divide the time interval of 10 minutes into a large number of small intervals so that the probability of receiving two or more telephone calls in an interval becomes almost zero. This is illustrated by the following table which shows that the probabilities of receiving two calls decreases sharply as the number of intervals are increased, keeping the average number of calls, 3 calls in 10 minutes in our example, as constant.

n	$P(\text{one call is received})$	$P(\text{two calls are received})$
10	0.3	0.09
100	0.03	0.0009
1,000	0.003	0.000009
10,000	0.0003	0.00000009

Using symbols, we may note that as n increases then p automatically declines in such a way that the mean m (= np) is always equal to a constant. Such a process is termed as a Poisson Process. The chief characteristics of Poisson process can be summarised as given below:

1. The number of occurrences in an interval is independent of the number of occurrences in another interval.
2. The expected number of occurrences in an interval is constant.
3. It is possible to identify a small interval so that the occurrence of more than one event, in any interval of this size, becomes extremely unlikely.

14.6.2 Probability Mass Function

The probability mass function (p.m.f.) of Poisson distribution can be derived as a limit of p.m.f. of binomial distribution when n $\rightarrow \infty$ such that m (= np) remains constant. Thus, we can write

$$P(r) = \lim_{n \to \infty} {}^nC_r \left(\frac{m}{n}\right)^r \left(1-\frac{m}{n}\right)^{n-r} = \lim_{n \to \infty} \frac{n!}{r!(n-r)!}\left(\frac{m}{n}\right)^r \left(1-\frac{m}{n}\right)^{n-r}$$

$$= \frac{m^r}{r!} . \lim_{n \to \infty} \left[n(n-1)(n-2) \; .... \; (n-r+1) . \frac{1}{n^r} . \left(1-\frac{m}{n}\right)^{n-r} \right]$$

Notes

$$= \frac{m^r}{r!} . \lim_{n \to \infty} \left[\frac{\frac{n}{n}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right) .... \left(1-\frac{(r-1)}{n}\right)\left(1-\frac{m}{n}\right)^n}{\left(1-\frac{m}{n}\right)^r} \right]$$

$$= \frac{m^r}{r!} \lim_{n \to \infty} \left(1-\frac{m}{n}\right)^n$$, since each of the remaining terms will tend to unity as $n \to \infty$.

$$= \frac{m^r . e^{-m}}{r!} \text{, since } \lim_{n \to \infty} \left(1-\frac{m}{n}\right)^n = \lim_{n \to \infty} \left\{\left(1-\frac{m}{n}\right)^{\frac{n}{m}}\right\}^m = e^{-m}.$$

Thus, the probability mass function of Poisson distribution is

$$P(r) = \frac{e^{-m}.m^r}{r!}, \text{ where } r = 0,1,2, ... \infty$$

Here e is a constant with value = 2.71828... . Note that Poisson distribution is a discrete probability distribution with single parameter m.

Total probability $= \sum_{r=0}^{\infty} \frac{e^{-m}.m^r}{r!} = e^{-m}\left(1+\frac{m}{1!}+\frac{m^2}{2!}+\frac{m^3}{3!}+ ....\right)$

$$= e^{-m}.e^{m} = 1.$$

14.6.3 Summary Measures of Poisson Distribution

(a) Mean

The mean of a Poisson variate r is defined as

$$E(r) = \sum_{r=0}^{\infty} r.\frac{e^{-m}.m^r}{r!} = e^{-m}\sum_{r=1}^{\infty}\frac{m^r}{(r-1)!} = e^{-m}\left[m+m^2+\frac{m^3}{2!}+\frac{m^4}{3!}+ ....\right]$$

$$= me^{-m}\left[1+m+\frac{m^2}{2!}+\frac{m^3}{3!}+ ....\right] = me^{-m}e^{m} = m$$

(b) Variance

The variance of a Poisson variate is defined as

$Var(r) = E(r - m)^2 = E(r^2) - m^2$

Now $E(r^2) = \sum_{r=0}^{\infty} r^2 P(r) = \sum_{r=0}^{\infty}\left[r(r-1)+r\right]P(r) = \sum_{r=0}^{\infty}\left[r(r-1)\right]P(r) + \sum_{r=0}^{\infty} r P(r)$

Notes

$$= \sum_{r=2}^{\infty}\left[r(r-1)\right]\frac{e^{-m}.m^r}{r!} + m = e^{-m}\sum_{r=2}^{\infty}\frac{m^r}{(r-2)!} + m$$

$$= m + e^{-m}\left(m^2 + m^3 + \frac{m^4}{2!} + \frac{m^5}{3!} + \ ....\right)$$

$$= m + m^2 e^{-m}\left(1 + m \ + \frac{m^2}{2!} + \frac{m^3}{3!} + \ ....\right) = m + m^2$$

Thus, Var(r) = m + m² - m² = m.

Also standard deviation $\sigma = \sqrt{m}$.

(c) The values of m_3, m_4, b_1 and b_2

It can be shown that m_3 = m and m_4 = m + 3m².

$$\therefore \ \beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{m^2}{m^3} = \frac{1}{m}$$

Since m is a positive quantity, therefore, β_1 is always positive and hence the Poisson distribution is always positively skewed. We note that $\beta_1 \to 0$ as m → ∞, therefore the distribution tends to become more and more symmetrical for large values of m.

Further, $\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{m + 3m^2}{m^2} = 3 + \frac{1}{m} \to 3 \text{ as } m \to \infty$. This result shows that the distribution becomes normal for large values of m.

(d) Mode

As in binomial distribution, a Poisson variate r will be mode if

$$P(r-1) \le P(r) \ge P(r+1)$$

The inequality $P(r-1) \le P(r)$ can be written as

$$\frac{e^{-m}.m^{r-1}}{(r-1)!} \le \frac{e^{-m}.m^r}{r!} \quad \Rightarrow \quad 1 \le \frac{m}{r} \quad \Rightarrow \quad r \le m \quad .... (1)$$

Similarly, the inequality $P(r) \ge P(r+1)$ can be shown to imply that

r ≥ m - 1 .... (2)

Combining (1) and (2), we can write m - 1 £ r £ m.

Case I. When m is not an integer

Notes

The integral part of m will be mode.

Case II. When m is an integer

The distribution is bimodal with values m and m - 1.

Example 13: The average number of customer arrivals per minute at a super bazaar is 2. Find the probability that during one particular minute (i) exactly 3 customers will arrive, (ii) at the most two customers will arrive, (iii) at least one customer will arrive.

Solution.

It is given that m = 2. Let the number of arrivals per minute be denoted by the random variable r. The required probability is given by

(i) $$P(r=3)=\frac{e^{-2}.2^3}{3!}=\frac{0.13534\times 8}{6}=0.18045$$

(ii) $$P(r\le 2)=\sum_{r=0}^{2}\frac{e^{-2}.2^r}{r!}=e^{-2}\left[1+2+\frac{4}{2}\right]=0.13534\times 5=0.6767.$$

(iii) $$P(r\ge 1)=1-P(r=0)=1-\frac{e^{-2}.2^0}{0!}=1-0.13534=0.86464.$$

Example 14: An executive makes, on an average, 5 telephone calls per hour at a cost which may be taken as Rs 2 per call. Determine the probability that in any hour the telephone calls' cost (i) exceeds Rs 6, (ii) remains less than Rs 10.

Solution.

The number of telephone calls per hour is a random variable with mean = 5. The required probability is given by

(i) $$P(r>3)=1-P(r\le 3)=1-\sum_{r=0}^{3}\frac{e^{-5}.5^r}{r!}$$

$$=1-e^{-5}\left[1+5+\frac{25}{2}+\frac{125}{6}\right]=1-0.00678\times\frac{236}{6}=0.7349.$$

(ii) $$P(r\le 4)=\sum_{r=0}^{4}\frac{e^{-5}.5^r}{r!}=e^{-5}\left[1+5+\frac{25}{2}+\frac{125}{6}+\frac{625}{24}\right]=0.00678\times\frac{1569}{24}=0.44324.$$

Example 15: A company makes electric toys. The probability that an electric toy is defective is 0.01. What is the probability that a shipment of 300 toys will contain exactly 5 defectives?

Solution.

Since n is large and p is small, Poisson distribution is applicable. The random variable is the number of defective toys with mean m = np = 300 × 0.01 = 3. The required probability is given by

$$P(r=5)=\frac{e^{-3}.3^5}{5!}=\frac{0.04979\times 243}{120}=0.10082.$$

Example 16: In a town, on an average 10 accidents occur in a span of 50 days. Assuming that the number of accidents per day follow Poisson distribution, find the probability that there will be three or more accidents in a day.

Solution.

The random variable denotes the number accidents per day. Thus, we have $m=\frac{10}{50}=0.2$. The required probability is given by

$$P(r\geq 3)=1-P(r\leq 2)=1-e^{-0.2}\left[1+0.2+\frac{(0.2)^2}{2!}\right]=1-0.8187\times 1.22=0.00119.$$

Example 17: A car hire firm has two cars which it hire out every day. The number of demands for a car on each day is distributed as a Poisson variate with mean 1.5. Calculate the proportion of days on which neither car is used and the proportion of days on which some demand is refused. [$e^{-1.5}$ = 0.2231]

Solution.

When both car are not used, r = 0

$\therefore\ P(r=0)=e^{-1.5}=0.2231$. Hence the proportion of days on which neither car is used is 22.31%.

Further, some demand is refused when more than 2 cars are demanded, i.e., r > 2

$$\therefore\ P(r>2)=1-P(r\leq 2)=1-\sum_{r=0}^{2}\frac{e^{-1.5}(1.5)^r}{r!}=1-0.2231\left[1+1.5+\frac{(1.5)^2}{2!}\right]=0.1913.$$

Hence the proportion of days is 19.13%.

Example 18: A firm produces articles of which 0.1 percent are usually defective. It packs them in cases each containing 500 articles. If a wholesaler purchases 100 such cases, how many cases are expected to be free of defective items and how many are expected to contain one defective item?

Solution.

The Poisson variate is number of defective items with mean

$$m=\frac{1}{1000}\times 500=0.5.$$

Notes

Probability that a case is free of defective items

$P(r=0)=e^{-0.5}=0.6065.$. Hence the number of cases having no defective items = 0.6065 × 100 = 60.65

Similarly, $P(r=1)=e^{-0.5}\times 0.5=0.6065\times 0.5=0.3033.$ Hence the number of cases having one defective item are 30.33.

Example 19: A manager accepts the work submitted by his typist only when there is no mistake in the work. The typist has to type on an average 20 letters per day of about 200 words each. Find the chance of her making a mistake (i) if less than 1% of the letters submitted by her are rejected; (ii) if on 90% of days all the work submitted by her is accepted. [As the probability of making a mistake is small, you may use Poisson distribution. Take e = 2.72].

Solution.

Let p be the probability of making a mistake in typing a word.

(i) Let the random variable r denote the number of mistakes per letter. Since 20 letters are typed, r will follow Poisson distribution with mean = 20 × p.

Since less than 1% of the letters are rejected, it implies that the probability of making at least one mistake is less than 0.01, i.e.,

$P(r \geq 1) \leq 0.01$ or $1 - P(r = 0) \leq 0.01$

$\Rightarrow \quad 1 - e^{-20p} \leq 0.01$ or $e^{-20p} \geq 0.99$

Taking log of both sides

$-20p.\log 2.72 \geq \log 0.99$

$-(20\times 0.4346)p \geq \bar{1}.9956$

$-8.692p \geq -0.0044$ or $p \leq \dfrac{0.0044}{8.692} = 0.00051.$

(ii) In this case r is a Poisson variate which denotes the number of mistakes per day. Since the typist has to type 20 × 200 = 4000 words per day, the mean number of mistakes = 4000p.

It is given that there is no mistake on 90% of the days, i.e.,

$P(r = 0) = 0.90$ or $e^{-4000p} = 0.90$

Taking log of both sides, we have

$-4000p \log 2.72 = \log 0.90$ or $-4000\times 0.4346p=\bar{1}.9542=-0.0458$

$$\therefore \quad p=\frac{0.0458}{4000\times 0.4346}=0.000026.$$

Example 20: A manufacturer of pins knows that on an average 5% of his product is defective. He sells pins in boxes of 100 and guarantees that not more than 4 pins will be defective. What is the probability that the box will meet the guaranteed quality?

Notes

Solution.

The number of defective pins in a box is a Poisson variate with mean equal to 5. A box will meet the guaranteed quality if r £ 4. Thus, the required probability is given by

$$P(r \le 4) = e^{-5}\sum_{r=0}^{4}\frac{5^r}{r!} = e^{-5}\left[1+5+\frac{25}{2}+\frac{125}{6}+\frac{625}{24}\right] = 0.00678\times\frac{1569}{24} = 0.44324.$$

Lot Acceptance using Poisson Distribution

Example 21: Videocon company purchases heaters from Amar Electronics. Recently a shipment of 1000 heaters arrived out of which 60 were tested. The shipment will be accepted if not more than two heaters are defective. What is the probability that the shipment will be accepted? From past experience, it is known that 5% of the heaters made by Amar Electronics are defective.

Solution.

Mean number of defective items in a sample of 60 = 60 × $\frac{5}{100}$ = 3

P(r £ 2) = $\sum_{r=0}^{2}\frac{e^{-3}.3^r}{r!}$

$$= e^{-3}\left[1+3+\frac{3^2}{2!}\right] = e^{-3}.8.5 = 0.4232$$

14.6.4 Poisson Approximation to Binomial

When n, the number of trials become large, the computation of probabilities by using the binomial probability mass function becomes a cumbersome task. Usually, when n ³ 20 and p £ 0.05, Poisson distribution can be used as an approximation to binomial with parameter m = np.

Example 22: Find the probability of 4 successes in 30 trials by using (i) binomial distribution and (ii) Poisson distribution. The probability of success in each trial is given to be 0.02.

Solution.

(i) Here n = 30 and p = 0.02

$$\therefore\ P(r=4) = {}^{30}C_4(0.02)^4(0.98)^{26} = 27405\times 0.00000016\times 0.59 = 0.00259.$$

(ii) Here m = np = 30 × 0.02 = 0.6

$$\therefore\ P(r=4) = \frac{e^{-0.6}(0.6)^4}{4!} = \frac{0.5488\times 0.1296}{24} = 0.00296.$$

Notes

14.6.5 Fitting of a Poisson Distribution

To fit a Poisson distribution to a given frequency distribution, we first compute its mean m. Then the probabilities of various values of the random variable r are computed by using the probability mass function $P(r) = \frac{e^{-m}.m^r}{r!}$. These probabilities are then multiplied by N, the total frequency, to get expected frequencies.

Example 23:

The following mistakes per page were observed in a book :

No. of mistakes per page	:	0	1	2	3
Frequency	:	211	90	19	5

Fit a Poisson distribution to find the theoretical frequencies.

Solution.

The mean of the given frequency distribution is

$$m = \frac{0 \times 211 + 1 \times 90 + 2 \times 19 + 3 \times 5}{211 + 90 + 19 + 5} = \frac{143}{325} = 0.44$$

Calculation of theoretical (or expected) frequencies

We can write $P(r) = \frac{e^{-0.44}(0.44)^r}{r!}$. Substituting r = 0, 1, 2 and 3, we get the probabilities for various values of r, as shown in the following table.

r	$P(r)$	$N \times P(r)$	*Expected Frequencies Approximated to the nearest integer*
0	0.6440	209.30	210
1	0.2834	92.10	92
2	0.0623	20.25	20
3	0.0091	2.96	3
Total			325

14.6.6 Features of Poisson Distribution

(i) It is discrete probability distribution.

(ii) It has only one parameter m.

(iii) The range of the random variable is $0 \le r < \infty$.

(iv) The Poisson distribution is a positively skewed distribution. The skewness decreases as m increases.

14.6.7 Uses of Poisson Distribution

(i) This distribution is applicable to situations where the number of trials is large and the probability of a success in a trial is very small.

(ii) It serves as a reasonably good approximation to binomial distribution when $n \geq 20$ and $p \leq 0.05$.

14.7 Summary

- Binomial distribution is a theoretical probability distribution which was given by James Bernoulli.

  Let n be the total number of repeated trials, p be the probability of a success in a trial and q be the probability of its failure so that q = 1 - p.

- Let r be a random variable which denotes the number of successes in n trials. The possible values of r are 0, 1, 2, ... n. We are interested in finding the probability of r successes out of n trials, i.e., P(r).

- Binomial distribution is often used in various decision making situations in business. Acceptance sampling plan, a technique of quality control, is based on this distribution. With the use of sampling plan, it is possible to accept or reject a lot of items either at the stage of its manufacture or at the stage of its purchase.

- The binomial distribution is not applicable when the probability of a success p does not remain constant from trial to trial. In such a situation the probabilities of the various values of r are obtained by the use of Hypergeometric distribution.

- When n, the number of trials become large, the computation of probabilities by using the binomial probability mass function becomes a cumbersome task. Usually, when n ³ 20 and p £ 0.05, Poisson distribution can be used as an approximation to binomial with parameter m = np.

14.8 Keywords

Binomial distribution is a theoretical probability distribution which was given by James Bernoulli.

Probability distribution: The purpose of fitting a distribution is to examine whether the observed frequency distribution can be regarded as a sample from a population with a known probability distribution.

Geometrical distribution: When r = 1, the Pascal distribution can be written as

$$P(n) = {}^{n-1}C_0 pq^{n-1} = pq^{n-1}, \quad \text{where } n = 1, 2, 3, ...$$

Geometrical distribution: Here n is a random variable which denotes the number of trials required to get a success. This distribution is known as geometrical distribution.

Notes

14.9 Self Assessment

1. .... is a theoretical probability distribution which was given by James Bernoulli.
   (a) expected (b) Binomial distribution
   (c) probability mass function (d) discrete values
2. The fitting of a distribution to given data implies the determination of .... (or theoretical) frequencies for different values of the random variable on the basis of this data.
   (a) expected (b) Binomial distribution
   (c) probability mass function (d) discrete values
3. A discrete random variable is said to follow a uniform distribution if it takes various .... with equal probabilities.
   (a) expected (b) Binomial distribution
   (c) probability mass function (d) discrete values
4. Poisson distribution was derived by a noted mathematician, Simon D. Poisson, in ....
   (a) expected (b) 1837
   (c) probability mass function (d) discrete values
5. The .... (p.m.f.) of Poisson distribution can be derived as a limit of p.m.f. of binomial distribution when $n \to \infty$ such that m (= np) remains constant.
   (a) expected (b) Binomial distribution
   (c) probability mass function (d) discrete values

14.10 Review Questions

1. (a) The probability of a man hitting a target is $\frac{1}{4}$. (i) If he fires 7 times, what is the probability of his hitting the target at least twice? (ii) How many times must he fire so that the probability of his hitting the target at least once is greater than $\frac{2}{3}$?

   (b) How many dice must be thrown so that there is better than even chance of obtaining at least one six?

   Hint : (a)(ii) Probability of hitting the target at least once in n trials is $1-\left(\frac{3}{4}\right)^n$. Find n such that this value is greater than $\frac{2}{3}$. (b) Find n so that $1-\left(\frac{5}{6}\right)^n > \frac{1}{2}$.

2. A machine produces an average of 20% defective bolts. A batch is accepted if a sample of 5 bolts taken from the batch contains no defective and rejected if the sample contains 3 or more defectives. In other cases, a second sample is taken. What is the probability that the second sample is required?

   Hint : A second sample is required if the first sample is neither rejected nor accepted.

Notes

3. A multiple choice test consists of 8 questions with 3 answers to each question (of which only one is correct). A student answers each question by throwing a balanced die and checking the first answer if he gets 1 or 2, the second answer if he gets 3 or 4 and the third answer if he gets 5 or 6. To get a distinction, the student must secure at least 75% correct answers. If there is no negative marking, what is the probability that the student secures a distinction?

   Hint : He should attempt at least 6 questions.

4. What is the most probable number of times an ace will appear if a die is tossed (i) 50 times, (ii) 53 times?

   Hint : Find mode.

5. The number of arrivals of telephone calls at a switch board follows a Poisson process at an average rate of 8 calls per 10 minutes. The operator leaves for a 5 minutes tea break. Find the probability that (a) at the most two calls go unanswered and (b) 3 calls go unanswered, while the operator is away.

   Hint : m = 4.

6. What probability model is appropriate to describe a situation where 100 misprints are distributed randomly throughout the 100 pages of a book? For this model, what is the probability that a page observed at random will contain (i) no misprint, (ii) at the most two misprints, (iii) at least three misprints?

   Hint : The average number of misprint per page is unity.

7. If the probability of getting a defective transistor in a consignment is 0.01, find the mean and standard deviation of the number of defective transistors in a large consignment of 900 transistors. What is the probability that there is at the most one defective transistor in the consignment?

   Hint : The average number of transistors in a consignment is 900×0.01.

8. In a certain factory turning out blades, there is a small chance 1/500 for any one blade to be defective. The blades are supplied in packets of 10. Use Poisson distribution to compute the approximate number of packets containing no defective, one defective, two defective, three defective blades respectively in a consignment of 10,000 packets.

   Hint : The random variable is the number of defective blades in a packet of 10 blades.

Answers: Self Assessment

1. (b) 2. (a) 3. (d) 4. (b) 5. (c)

14.11 Further Readings

Books

Introductory Probability and Statistical Applications by P.L. Meyer

Introduction to Mathematical Statistics by Hogg and Craig

Fundamentals of Mathematical Statistics by S.C. Gupta and V.K. Kapoor

Notes

Unit 15: Exponential Distribution and Normal Distribution

CONTENTS

Objectives

Introduction

15.1 Exponential Distribution

15.1.1 Probability Density Function

15.2 Uniform Distribution (Continuous Variable)

15.3 Normal Distribution

15.3.1 The Conditions of Normality

15.3.2 Probability Density Function

15.3.3 Shape of Normal Probability Curve

15.3.4 Properties of Normal Probability Curve

15.3.5 Probability of Normal Variate in an interval

15.3.6 Normal Approximation to Binomial Distribution

15.9.7 Normal Approximation to Poisson Distribution

15.3.8 Fitting a Normal Curve

15.4 Summary

15.5 Keywords

15.6 Self Assessment

15.7 Review Questions

15.8 Further Readings

Objectives

After studying this unit, you will be able to:

- Discuss the exponential distribution
- Explain uniform distribution
- Describe normal distribution

Introduction

The knowledge of the theoretical probability distribution is of great use in the understanding and analysis of a large number of business and economic situations. For example, with the use of probability distribution, it is possible to test a hypothesis about a population, to take decision in the face of uncertainty, to make forecast, etc.

Notes

Theoretical probability distributions can be divided into two broad categories, viz. discrete and continuous probability distributions, depending upon whether the random variable is discrete or continuous. Although, there are a large number of distributions in each category, we shall discuss only some of them having important business and economic applications.

15.1 Exponential Distribution

The random variable in case of Poisson distribution is of the type ; the number of arrivals of customers per unit of time or the number of defects per unit length of cloth, etc. Alternatively, it is possible to define a random variable, in the context of Poisson Process, as the length of time between the arrivals of two consecutive customers or the length of cloth between two consecutive defects, etc. The probability distribution of such a random variable is termed as Exponential Distribution.

Since the length of time or distance is a continuous random variable, therefore exponential distribution is a continuous probability distribution.

15.1.1 Probability Density Function

Let t be a random variable which denotes the length of time or distance between the occurrence of two consecutive events or the occurrence of the first event and m be the average number of times the event occurs per unit of time or length. Further, let A be the event that the time of occurrence between two consecutive events or the occurrence of the first event is less than or equal to t and f(t) and F(t) denote the probability density function and the distribution (or cumulative density) function of t respectively.

We can write $P(A)+P(\bar{A})=1$ or $F(t)+P(\bar{A})=1$. Note that, by definition, $F(t)=P(A)$. Further, $P(\bar{A})$ is the probability that the length of time between the occurrence of two consecutive events or the occurrence of first event is greater than t. This is also equal to the probability that no event occurs in the time interval t. Since the mean number of occurrence of events in time t is mt, we have , by Poisson distribution,

$$P(\bar{A})=P(r=0)=\frac{e^{-mt}(mt)^0}{0!}=e^{-mt}.$$

Thus, we get $F(t) + e^{-mt} = 1$

or $P(0 \text{ to } t) = F(t) = 1 - e^{-mt}$. .... (1)

To get the probability density function, we differentiate equation (1) with respect to t.

Thus, $f(t) = F'(t) = me^{-mt}$ when $t > 0$

$= 0$ otherwise.

It can be verified that the total probability is equal to unity

$$\text{Total Probability } = \int_0^\infty m.e^{-mt}dt = \left| m.\frac{e^{-mt}}{-m} \right|_0^\infty = \left| -e^{-mt} \right|_0^\infty = 0+1=1.$$

Notes

Mean of t

The mean of t is defined as its expected value, given by

$E(t) = \int_0^\infty t.m.e^{-mt} dt = \frac{1}{m}$, where m denotes the average number of occurrence of events per unit of time or distance.

Example 24: A telephone operator attends on an average 150 telephone calls per hour. Assuming that the distribution of time between consecutive calls follows an exponential distribution, find the probability that (i) the time between two consecutive calls is less than 2 minutes, (ii) the next call will be received only after 3 minutes.

Solution.

Here m = the average number of calls per minute $= \frac{150}{60} = 2.5.$

(i) $P(t \le 2) = \int_0^2 2.5e^{-2.5t} dt = F(2)$

We know that F(t) = 1 - e^{-mt}, \ F(2) = 1 - $e^{-2.5 \times 2}$ = 0.9933

(ii) P(t > 3) = 1 - P(t ≤ 3) = 1 - F(3)

= 1 - [1 - $e^{-2.5 \times 3}$] = 0.0006

Example 25: The average number of accidents in an industry during a year is estimated to be 5. If the distribution of time between two consecutive accidents is known to be exponential, find the probability that there will be no accidents during the next two months.

Solution.

Here m denotes the average number of accidents per month = $\frac{5}{12}$.

$\therefore$ P(t > 2) = 1 - F(2) = $e^{-\frac{5}{12}\times 2} = e^{-0.833} = 0.4347.$

Example 26: The distribution of life, in hours, of a bulb is known to be exponential with mean life of 600 hours. What is the probability that (i) it will not last more than 500 hours, (ii) it will last more than 700 hours?

Solution.

Since the random variable denote hours, therefore $m = \frac{1}{600}$

(i) P(t £ 500) = F(500) $= 1 - e^{-\frac{1}{600}\times 500} = 1 - e^{-0.833} = 0.5653.$

(ii) P(t > 700) = 1 - F(700) $= e^{-\frac{700}{600}} = e^{-1.1667} = 0.3114.$

15.2 Uniform Distribution (Continuous Variable)

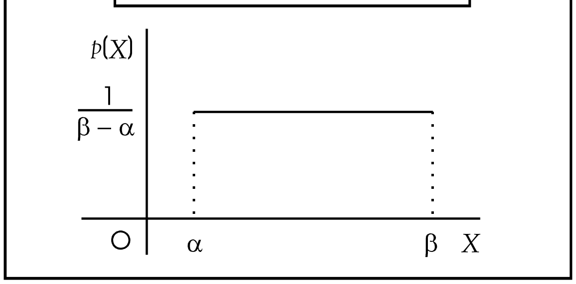
A continuous random variable X is said to be uniformly distributed in a close interval (*a*, *b*) with probability density function p(X) if

$$p(X) = \frac{1}{\beta - \alpha} \qquad \text{for } \alpha \le X \le \beta \quad \text{and}$$

$$= 0 \qquad \text{otherwise.}$$

The uniform distribution is alternatively known as rectangular distribution. The diagram of the probability density function is shown in the figure 15.1.

Figure 15.1



Note that the total area under the curve is unity, i.e. ,

$$\int_{\alpha}^{\beta} \frac{1}{\beta - \alpha} dX = \frac{1}{\beta - \alpha} |X|_{\alpha}^{\beta} = 1$$

Further, $E(X) = \frac{1}{\beta - \alpha} \int_{\alpha}^{\beta} X.dX = \frac{1}{\beta - \alpha} \left| \frac{X^2}{2} \right|_{\alpha}^{\beta} = \frac{\alpha + \beta}{2}$

$$E(X^2) = \frac{1}{\beta - \alpha} \int_{\alpha}^{\beta} X^2.dX = \frac{\beta^3 - \alpha^3}{3(\beta - \alpha)} = \frac{1}{3}(\beta^2 + \alpha\beta + \alpha^2)$$

$$\therefore \qquad Var(X) = \frac{1}{3}(\beta^2 + \alpha\beta + \alpha^2) - \frac{(\alpha + \beta)^2}{4} = \frac{(\beta - \alpha)^2}{12}$$

Example 27: The buses on a certain route run after every 20 minutes. If a person arrives at the bus stop at random, what is the probability that

(a) he has to wait between 5 to 15 minutes,

(b) he gets a bus within 10 minutes,

(c) he has to wait at least 15 minutes.

Solution.

Let the random variable X denote the waiting time, which follows a uniform distribution with p.d.f.

$$f(X) = \frac{1}{20} \qquad \text{for } 0 \le X \le 20$$

Notes

(a) $P(5 \le X \le 15) = \frac{1}{20}\int_{5}^{15} dX = \frac{1}{20}(15-5) = \frac{1}{2}$

(b) $P(0 \le X \le 10) = \frac{1}{20} \times 10 = \frac{1}{2}$

(c) $P(15 \le X \le 20) = \frac{20-15}{20} = \frac{1}{4}.$

15.3 Normal Distribution

The normal probability distribution occupies a place of central importance in Modern Statistical Theory. This distribution was first observed as the normal law of errors by the statisticians of the eighteenth century. They found that each observation X involves an error term which is affected by a large number of small but independent chance factors. This implies that an observed value of X is the sum of its true value and the net effect of a large number of independent errors which may be positive or negative each with equal probability. The observed distribution of such a random variable was found to be in close conformity with a continuous curve, which was termed as the normal curve of errors or simply the normal curve.

Since Gauss used this curve to describe the theory of accidental errors of measurements involved in the calculation of orbits of heavenly bodies, it is also called as Gaussian curve.

15.3.1 The Conditions of Normality

In order that the distribution of a random variable X is normal, the factors affecting its observations must satisfy the following conditions :

(i) A large number of chance factors: The factors, affecting the observations of a random variable, should be numerous and equally probable so that the occurrence or non-occurrence of any one of them is not predictable.

(ii) Condition of homogeneity: The factors must be similar over the relevant population although, their incidence may vary from observation to observation.

(iii) Condition of independence: The factors, affecting observations, must act independently of each other.

(iv) Condition of symmetry: Various factors operate in such a way that the deviations of observations above and below mean are balanced with regard to their magnitude as well as their number.

Random variables observed in many phenomena related to economics, business and other social as well as physical sciences are often found to be distributed normally. For example, observations relating to the life of an electrical component, weight of packages, height of persons, income of the inhabitants of certain area, diameter of wire, etc., are affected by a large number of factors and hence, tend to follow a pattern that is very similar to the normal curve. In addition to this, when the number of observations become large, a number of probability distributions like Binomial, Poisson, etc., can also be approximated by this distribution.

15.3.2 Probability Density Function

If X is a continuous random variable, distributed normally with mean *m* and standard deviation σ, then its p.d.f. is given by

$$p(X) = \frac{1}{\sigma\sqrt{2\pi}}.e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2} \qquad \text{where } -\infty < X < \infty.$$

Here π and e are absolute constants with values 3.14159.... and 2.71828.... respectively.

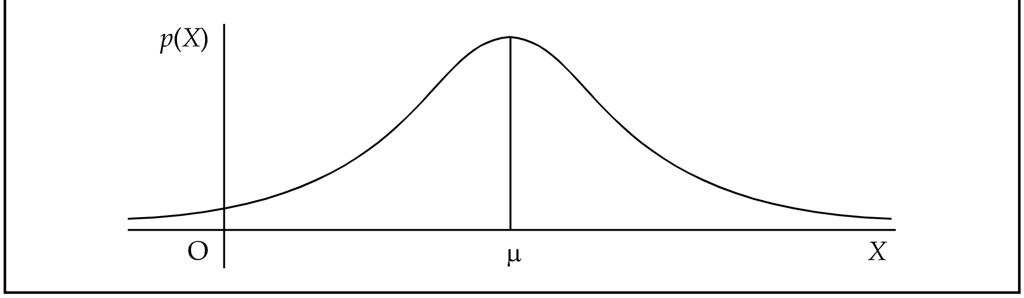
It may be noted here that this distribution is completely known if the values of mean m and standard deviation s are known. Thus, the distribution has two parameters, viz. mean and standard deviation.

15.3.3 Shape of Normal Probability Curve

For given values of the parameters, *m* and *s*, the shape of the curve corresponding to normal probability density function p(X) is as shown in Figure 15.2

It should be noted here that although we seldom encounter variables that have a range from - ∞ to ∞, as shown by the normal curve, nevertheless the curves generated by the relative frequency histograms of various variables closely resembles the shape of normal curve.

Figure 15.2



15.3.4 Properties of Normal Probability Curve

A normal probability curve or normal curve has the following properties:

1. It is a bell shaped symmetrical curve about the ordinate at X = μ . The ordinate is maximum at X = μ .
2. It is unimodal curve and its tails extend infinitely in both directions, i.e., the curve is asymptotic to X axis in both directions.
3. All the three measures of central tendency coincide, i.e.,

   mean = median = mode

4. The total area under the curve gives the total probability of the random variable taking values between - ∞ to ¥. Mathematically, it can be shown that

$$P(-\infty < X < \infty) = \int_{-\infty}^{\infty} p(X)dX = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2} dX = 1.$$

Notes

5. Since median = *m*, the ordinate at X = μ divides the area under the normal curve into two equal parts, i.e.,

$$\int_{-\infty}^{\mu} p(X)dX = \int_{\mu}^{\infty} p(X)dX = 0.5$$

6. The value of p(X) is always non-negative for all values of X, i.e., the whole curve lies above X axis.

7. The points of inflexion (the point at which curvature changes) of the curve are at $X = \mu \pm \sigma$.

8. The quartiles are equidistant from median, i.e., $M_d - Q_1 = Q_3 - M_d$, by virtue of symmetry. Also $Q_1 = \mu - 0.6745\sigma$, $Q_3 = \mu + 0.6745\sigma$, quartile deviation = 0.6745σ and mean deviation = 0.8*s*, approximately.

9. Since the distribution is symmetrical, all odd ordered central moments are zero.

10. The successive even ordered central moments are related according to the following recurrence formula

    $\mu_{2n} = (2n - 1)\sigma^2\mu_{2n-2}$ for n = 1, 2, 3, ...

11. The value of moment coefficient of skewness β_1 is zero.

12. The coefficient of kurtosis $\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{3\sigma^4}{\sigma^4} = 3.$

    Note that the above expression makes use of property 10.

13. Additive or reproductive property

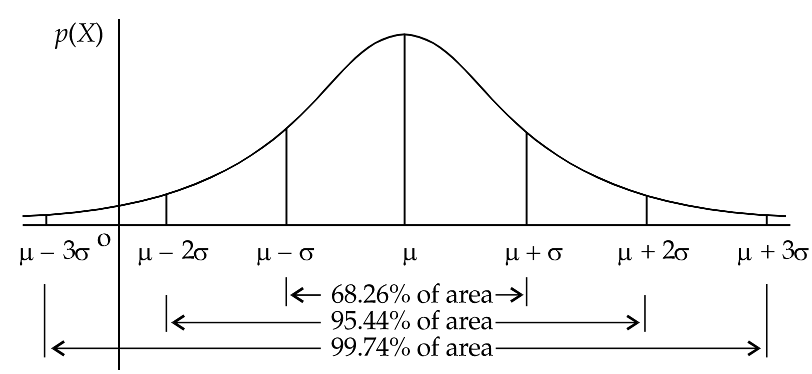
    If $X_1, X_2, ... X_n$ are n independent normal variates with means $\mu_1, \mu_2, ... \mu_n$ and variances $\sigma_1^2, \sigma_2^2, ... \sigma_n^2$, respectively, then their linear combination $a_1X_1 + a_2X_2 + ... + a_nX_n$ is also a normal variate with mean $\sum_{i=1}^{n} a_i\mu_i$ and variance $\sum_{i=1}^{n} a_i^2\sigma_i^2$.

    In particular, if $a_1 = a_2 = ... = a_n = 1$, we have $\sum X_i$ is a normal variate with mean $\sum \mu_i$ and variance $\sum \sigma_i^2$. Thus the sum of independent normal variates is also a normal variate.

14. Area property

    The area under the normal curve is distributed by its standard deviation in the following manner:

Figure 15.3



(i) The area between the ordinates at $\mu - \sigma$ and $\mu + \sigma$ is 0.6826. This implies that for a normal distribution about 68% of the observations will lie between $\mu - \sigma$ and $\mu + \sigma$.

(ii) The area between the ordinates at $\mu - 2\sigma$ and $\mu + 2\sigma$ is 0.9544. This implies that for a normal distribution about 95% of the observations will lie between $\mu - 2\sigma$ and $\mu + 2\sigma$.

(iii) The area between the ordinates at $\mu - 3\sigma$ and $\mu + 3\sigma$ is 0.9974. This implies that for a normal distribution about 99% of the observations will lie between $\mu - 3\sigma$ and $\mu + 3\sigma$. This result shows that, practically, the range of the distribution is 6s although, theoretically, the range is from $-\infty$ to ∞.

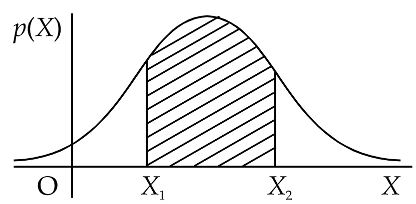
15.3.5 Probability of Normal Variate in an interval

Let X be a normal variate distributed with mean μ and standard deviation σ, also written in abbreviated form as $X \sim N(\mu, \sigma)$ The probability of X lying in the interval (X_1, X_2) is given by

$$P(X_1 \le X \le X_2) = \int_{X_1}^{X_2} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2} dX$$

In terms of figure, this probability is equal to the area under the normal curve between the ordinates at $X = X_1$ and $X = X_2$ respectively.

Figure 15.4



Note It may be recalled that the probability that a continuous random variable takes a particular value is defined to be zero even though the event is not impossible.

It is obvious from the above that, to find $P(X_1 \le X \le X_2)$, we have to evaluate an integral which might be cumbersome and time consuming task. Fortunately, an alternative procedure is available for performing this task. To devise this procedure, we define a new variable $z = \frac{X-\mu}{\sigma}$.

We note that $E(z) = E\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma}\left[E(X) - \mu\right] = 0$

and $Var(z) = Var\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma^2} Var(X-\mu) = \frac{1}{\sigma^2} Var(X) = 1.$

Further, from the reproductive property, it follows that the distribution of z is also normal.

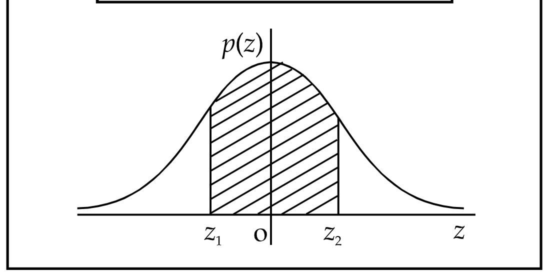
Thus, we conclude that if X is a normal variate with mean *m* and standard deviation *s*, then $z = \frac{X-\mu}{\sigma}$ is a normal variate with mean zero and standard deviation unity. Since the parameters

Notes

of the distribution of z are fixed, it is a known distribution and is termed as standard normal distribution (s.n.d.). Further, z is termed as a standard normal variate (s.n.v.).

It is obvious from the above that the distribution of any normal variateX can always be transformed into the distribution of standard normal variate z. This fact can be utilised to evaluate the integral given above.

Figure 19.5



We can write $P(X_1 \le X \le X_2) = P\left[\left(\frac{X_1 - \mu}{\sigma}\right) \le \left(\frac{X - \mu}{\sigma}\right) \le \left(\frac{X_2 - \mu}{\sigma}\right)\right]$

$= P(z_1 \le z \le z_2)$, where $z_1 = \frac{X_1 - \mu}{\sigma}$ and $z_2 = \frac{X_2 - \mu}{\sigma}$.

In terms of figure, this probability is equal to the area under the standard normal curve between the ordinates at $z = z_1$ and $z = z_2$. Since the distribution of z is fixed, the probabilities of z lying in various intervals are tabulated. These tables can be used to write down the desired probability.

Example 28:

Using the table of areas under the standard normal curve, find the following probabilities :

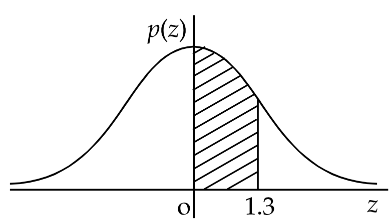
(i) $P(0 \le z \le 1.3)$ (ii) $P(-1 \le z \le 0)$ (iii) $P(-1 \le z \le 2)$

(iv) $P(z \ge 1.54)$ (v) $P(|z| > 2)$ (vi) $P(|z| < 2)$

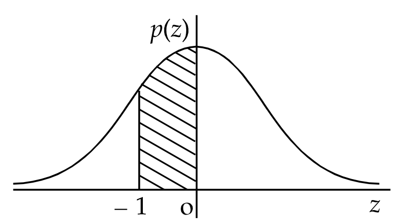
Solution.

The required probability, in each question, is indicated by the shaded are of the corresponding figure.

(i) From the table, we can write $P(0 \le z \le 1.3) = 0.4032$.



(i)



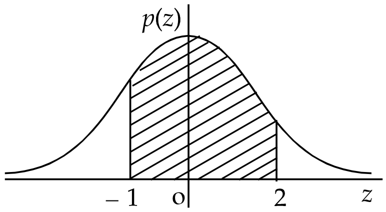
(ii)

(ii) We can write $P(-1 \le z \le 0) = P(0 \le z \le 1)$, because the distribution is symmetrical. From the table, we can write $P(-1 \le z \le 0) = P(0 \le z \le 1) = 0.3413$.

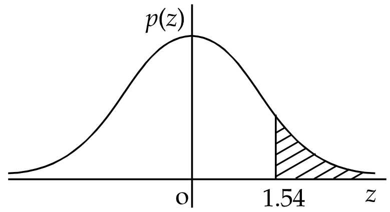
(iii) We can write

$$P(-1 \le z \le 2) = P(-1 \le z \le 0) + P(0 \le z \le 2)$$

$$= P(0 \le z \le 1) + P(0 \le z \le 2) = 0.3413 + 0.4772 = 0.8185.$$



(iii)



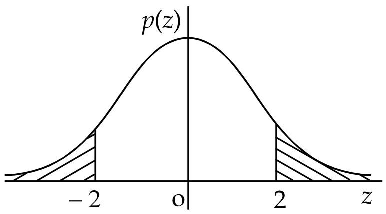
(iv)

(iv) We can write

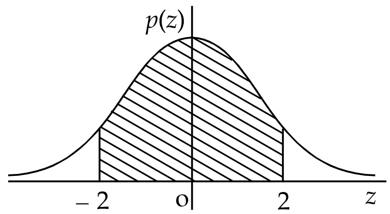
$P(z \geq 1.54) = 0.5000 - P(0 \leq z \leq 1.54) = 0.5000 - 0.4382 = 0.0618.$

(v) $P(|z| > 2) = P(z > 2) + P(z < -2) = 2P(z > 2) = 2[0.5000 - P(0 \leq z \leq 2)]$

$= 1 - 2P(0 \leq z \leq 2) = 1 - 2 \times 0.4772 = 0.0456.$



(v)



(vi)

(vi) $P(|z| < 2) = P(-2 \leq z \leq 0) + P(0 \leq z \leq 2) = 2P(0 \leq z \leq 2) = 2 \times 0.4772 = 0.9544.$

Example 29:

Determine the value or values of z in each of the following situations:

(a) Area between 0 and z is 0.4495.

(b) Area between $-\infty$ to z is 0.1401.

(c) Area between $-\infty$ to z is 0.6103.

(d) Area between - 1.65 and z is 0.0173.

(e) Area between - 0.5 and z is 0.5376.

Solution.

(a) On locating the value of z corresponding to an entry of area 0.4495 in the table of areas under the normal curve, we have z = 1.64. We note that the same situation may correspond to a negative value of z. Thus, z can be 1.64 or - 1.64.

(b) Since the area between $-\infty$ to z < 0.5, z will be negative. Further, the area between z and 0 = 0.5000 – 0.1401 = 0.3599. On locating the value of z corresponding to this entry in the table, we get z = – 1.08.

(c) Since the area between $-\infty$ to z > 0.5000, z will be positive. Further, the area between 0 to z = 0.6103 - 0.5000 = 0.1103. On locating the value of z corresponding to this entry in the table, we get z = 0.28.

(d) Since the area between - 1.65 and z < the area between - 1.65 and 0 (which, from table, is 0.4505), z is negative. Further z can be to the right or to the left of the value -1.65. Thus, when z lies to the right of - 1.65, its value, corresponds to an area (0.4505 - 0.0173) = 0.4332, is given by z = -1.5 (from table). Further, when z lies to the left of - 1.65, its value, corresponds to an area (0.4505 + 0.0173) = 0.4678, is given by z = - 1.85 (from table).

Notes

(e) Since the area between - 0.5 to z > area between - 0.5 to 0 (which, from table, is 0.1915), z is positive. The value of z, located corresponding to an area (0.5376 - 0.1915) = 0.3461, is given by 1.02.

Example 30:

If X is a random variate which is distributed normally with mean 60 and standard deviation 5, find the probabilities of the following events :

(i) $60 \le X \le 70$, (ii) $50 \le X \le 65$, (iii) $X > 45$, (iv) $X \le 50$.

Solution.

It is given that $m = 60$ and $s = 5$

(i) Given $X_1 = 60$ and $X_2 = 70$, we can write

$$z_1 = \frac{X_1 - \mu}{\sigma} = \frac{60-60}{5} = 0 \text{ and } z_2 = \frac{X_2 - \mu}{\sigma} = \frac{70-60}{5} = 2.$$

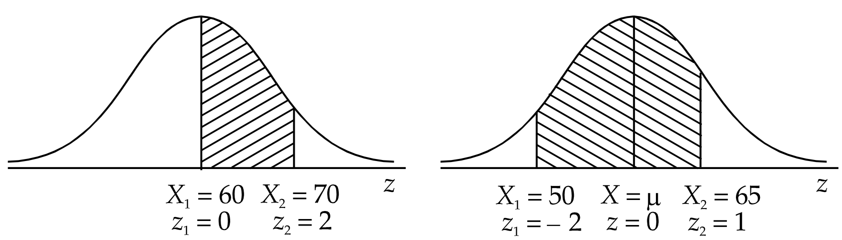
∴ $P(60 \le X \le 70) = P(0 \le z \le 2) = 0.4772$ (from table).

(ii) Here $X_1 = 50$ and $X_2 = 65$, therefore, we can write

$$z_1 = \frac{50-60}{5} = -2 \text{ and } z_2 = \frac{65-60}{5} = 1.$$

Hence $P(50 \le X \le 65) = P(-2 \le z \le 1) = P(0 \le z \le 2) + P(0 \le z \le 1)$

$= 0.4772 + 0.3413 = 0.8185$



(i) (ii)

(iii) $P(X > 45) = P\left(z \ge \frac{45-60}{5}\right) = P(z \ge -3)$

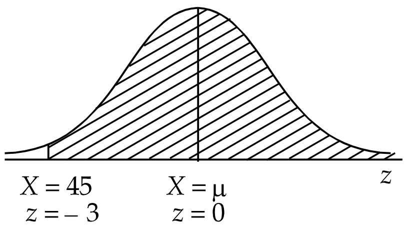
$$= P(-3 \le z \le 0) + P(0 \le z \le \infty) = P(0 \le z \le 3) + P(0 \le z \le \infty)$$

$= 0.4987 + 0.5000 = 0.9987$

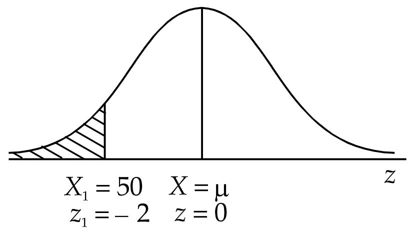
(iv) $P(X \le 50) = P\left(z \le \frac{50-60}{5}\right) = P(z \le -2)$

$$= 0.5000 - P(-2 \le z \le 0) = 0.5000 - P(0 \le z \le 2)$$

$= 0.5000 - 0.4772 = 0.0228$



(iii)



(iv)

Example 31:

The average monthly sales of 5,000 firms are normally distributed with mean Rs 36,000 and standard deviation Rs 10,000. Find:

(i) The number of firms with sales of over Rs 40,000.

(ii) The percentage of firms with sales between Rs 38,500 and Rs 41,000.

(iii) The number of firms with sales between Rs 30,000 and Rs 40,000.

Solution.

Let X be the normal variate which represents the monthly sales of a firm. Thus X ~ N(36,000, 10,000).

(i) $P(X > 40000) = P\left(z > \frac{40000 - 36000}{10000}\right) = P(z > 0.4)$

$$= 0.5000 - P(0 \le z \le 0.4) = 0.5000 - 0.1554 = 0.3446.$$

Thus, the number of firms having sales over Rs 40,000

$$= 0.3446 \times 5000 = 1723$$

(ii) $P(38500 \le X \le 41000) = P\left(\frac{38500 - 36000}{10000} \le z \le \frac{41000 - 36000}{10000}\right)$

$$= P(0.25 \le z \le 0.5) = P(0 \le z \le 0.5) - P(0 \le z \le 0.25)$$

$$= 0.1915 - 0.0987 = 0.0987.$$

Thus, the required percentage of firms =0.0987 × 100 = 9.87%.

(iii) $P(30000 \le X \le 40000) = P\left(\frac{30000 - 36000}{10000} \le z \le \frac{40000 - 36000}{10000}\right)$

$$= P(-0.6 \le z \le 0.4) = P(0 \le z \le 0.6) + P(0 \le z \le 0.4)$$

$$= 0.2258 + 0.1554 = 0.3812.$$

Thus, the required number of firms = 0.3812 × 5000 = 1906

Example 32: In a large institution, 2.28% of employees have income below Rs 4,500 and 15.87% of employees have income above Rs. 7,500 per month. Assuming the distribution of income to be normal, find its mean and standard deviation.

Notes

Solution.

Let the mean and standard deviation of the given distribution be *m* and *s* respectively.

It is given that $P(X < 4500) = 0.0228$ or $P\left(z < \frac{4500 - \mu}{\sigma}\right) = 0.0228$

On locating the value of z corresponding to an area 0.4772 (0.5000 - 0.0228), we can write

$\frac{4500 - \mu}{\sigma} = -2$ or $4500 - \mu = -2\sigma$ .... (1)

Similarly, it is also given that

$$P(X > 7500) = 0.1587 \text{ or } P\left(z > \frac{7500 - \mu}{\sigma}\right) = 0.1587$$

Locating the value of z corresponding to an area 0.3413 (0.5000 - 0.1587), we can write

$\frac{7500 - \mu}{\sigma} = 1$ or $7500 - \mu = \sigma$ .... (2)

Solving (1) and (2) simultaneously, we get

m = Rs 6,500 and *s* = Rs 1,000.

Example 33: Marks in an examination are approximately normally distributed with mean 75 and standard deviation 5. If the top 5% of the students get grade A and the bottom 25% get grade F, what mark is the lowest A and what mark is the highest F?

Solution.

Let A be the lowest mark in grade A and F be the highest mark in grade F. From the given information, we can write

$$P(X \geq A) = 0.05 \text{ or } P\left(z \geq \frac{A - 75}{5}\right) = 0.05$$

On locating the value of z corresponding to an area 0.4500 (0.5000 - 0.0500), we can write

$\frac{A - 75}{5} = 1.645 \Rightarrow A = 83.225$

Further, it is given that

$$P(X \leq F) = 0.25 \text{ or } P\left(z \leq \frac{F - 75}{5}\right) = 0.25$$

On locating the value of z corresponding to an area 0.2500 (0.5000 - 0.2500), we can write

$\frac{F - 75}{5} = -0.675 \Rightarrow F = 71.625$

Example 34: The mean inside diameter of a sample of 200 washers produced by a machine is 5.02 mm and the standard deviation is 0.05 mm. The purpose for which these washers are intended allows a maximum tolerance in the diameter of 4.96 to 5.08 mm, otherwise the washers are considered as defective. Determine the percentage of defective washers produced by the machine on the assumption that diameters are normally distributed.

Solution.

Let X denote the diameter of the washer. Thus, X ~ N (5.02, 0.05).

The probability that a washer is defective = 1 - P(4.96 £ X £ 5.08)

$$= 1 - P\left[\left(\frac{4.96 - 5.02}{0.05}\right) \le z \le \left(\frac{5.08 - 5.02}{0.05}\right)\right]$$

$$= 1 - P(-1.2 \le z \le 1.2) = 1 - 2P(0 \le z \le 1.2) = 1 - 2 \times 0.3849 = 0.2302$$

Thus, the percentage of defective washers = 23.02.

Example 35: The average number of units produced by a manufacturing concern per day is 355 with a standard deviation of 50. It makes a profit of Rs 1.50 per unit. Determine the percentage of days when its total profit per day is (i) between Rs 457.50 and Rs 645.00, (ii) greater than Rs 682.50 (assume the distribution to be normal). The area between z = 0 to z = 1 is 0.34134, the area between z = 0 to z = 1.5 is 0.43319 and the area between z = 0 to z = 2 is 0.47725, where z is a standard normal variate.

Solution.

Let X denote the profit per day. The mean of X is 355×1.50 = Rs 532.50 and its S.D. is 50×1.50 = Rs 75. Thus, X ~ N (532.50, 75).

(i) The probability of profit per day lying between Rs 457.50 and Rs 645.00

$$P(457.50 \le X \le 645.00) = P\left(\frac{457.50 - 532.50}{75} \le z \le \frac{645.00 - 532.50}{75}\right)$$

$$= P(-1 \le z \le 1.5) = P(0 \le z \le 1) + P(0 \le z \le 1.5) = 0.34134 + 0.43319 = 0.77453$$

Thus, the percentage of days = 77.453

(ii) $$P(X \ge 682.50) = P\left(z \ge \frac{682.50 - 532.50}{75}\right) = P(z \ge 2)$$

$$= 0.5000 - P(0 \le z \le 2) = 0.5000 - 0.47725 = 0.02275$$

Thus, the percentage of days = 2.275

Example 36:

The distribution of 1,000 examinees according to marks percentage is given below :

% Marks	less than 40	40 - 75	75 or more	Total
No. of examinees	430	420	150	1000

Assuming the marks percentage to follow a normal distribution, calculate the mean and standard deviation of marks. If not more than 300 examinees are to fail, what should be the passing marks?

Notes

Solution.

Let X denote the percentage of marks and its mean and S.D. be μ and σ respectively. From the given table, we can write

$P(X < 40) = 0.43$ and $P(X \geq 75) = 0.15$, which can also be written as

$$P\left(z < \frac{40-\mu}{\sigma}\right) = 0.43 \text{ and } P\left(z \geq \frac{75-\mu}{\sigma}\right) = 0.15$$

The above equations respectively imply that

$$\frac{40-\mu}{\sigma} = -0.175 \text{ or } 40-\mu = -0.175\sigma \qquad \text{.... (1)}$$

and $$\frac{75-\mu}{\sigma} = 1.04 \text{ or } 75-\mu = 1.04\sigma \qquad \text{.... (2)}$$

Solving the above equations simultaneously, we get μ = 45.04 and σ = 28.81.

Let X_1 be the percentage of marks required to pass the examination.

Then we have $P(X < X_1) = 0.3$ or $P\left(z < \frac{X_1 - 45.04}{28.81}\right) = 0.3$

$$\therefore \quad \frac{X_1 - 45.04}{28.81} = -0.525 \Rightarrow X_1 = 29.91 \text{ or } 30\% \text{ (approx.)}$$

Example 37: In a certain book, the frequency distribution of the number of words per page may be taken as approximately normal with mean 800 and standard deviation 50. If three pages are chosen at random, what is the probability that none of them has between 830 and 845 words each?

Solution.

Let X be a normal variate which denotes the number of words per page. It is given that $X \sim N(800, 50)$.

The probability that a page, select at random, does not have number of words between 830 and 845, is given by

$$1 - P(830 < X < 845) = 1 - P\left(\frac{830-800}{50} < z < \frac{845-800}{50}\right)$$

$$= 1 - P(0.6 < z < 0.9) = 1 - P(0 < z < 0.9) + P(0 < z < 0.6)$$

$$= 1 - 0.3159 + 0.2257 = 0.9098 » 0.91$$

Thus, the probability that none of the three pages, selected at random, have number of words lying between 830 and 845 = $(0.91)^3 = 0.7536$.

Example 38: At a petrol station, the mean quantity of petrol sold to a vehicle is 20 litres per day with a standard deviation of 10 litres. If on a particular day, 100 vehicles took 25 or more litres of petrol, estimate the total number of vehicles who took petrol from the station on that day. Assume that the quantity of petrol taken from the station by a vehicle is a normal variate.

Notes

Solution.

Let X denote the quantity of petrol taken by a vehicle. It is given that X ~ N(20, 10).

$$\therefore \quad P(X \geq 25) = P\left(z \geq \frac{25-20}{10}\right) = P(z \geq 0.5)$$

$$= 0.5000 - P(0 \leq z \leq 0.5) = 0.5000 - 0.1915 = 0.3085$$

Let N be the total number of vehicles taking petrol on that day.

$\therefore$ 0.3085×× N = 100 or N = 100/0.3085 = 324 (approx.)

15.3.6 Normal Approximation to Binomial Distribution

Normal distribution can be used as an approximation to binomial distribution when n is large and neither p nor q is very small. If X denotes the number of successes with probability p of a success in each of the n trials, then X will be distributed approximately normally with mean np and standard deviation $\sqrt{npq}$.

$$\text{Further, } z = \frac{X - np}{\sqrt{npq}} \sim N(0,1).$$

It may be noted here that as X varies from 0 to n, the standard normal variate z would vary from - ∞ to ∞ because

$$\text{when } X = 0, \quad \lim_{n \to \infty} \left(\frac{-np}{\sqrt{npq}}\right) = \lim_{n \to \infty} \left(-\sqrt{\frac{np}{q}}\right) = -\infty$$

$$\text{and when } X = n, \quad \lim_{n \to \infty} \left(\frac{n - np}{\sqrt{npq}}\right) = \lim_{n \to \infty} \left(\frac{nq}{\sqrt{npq}}\right) = \lim_{n \to \infty} \left(\sqrt{\frac{nq}{p}}\right) = \infty$$

Correction for Continuity

Since the number of successes is a discrete variable, to use normal approximation, we have make corrections for continuity. For example,

$P(X_1 \leq X \leq X_2)$ is to be corrected as $P\left(X_1 - \frac{1}{2} \leq X \leq X_2 + \frac{1}{2}\right)$, while using normal approximation to binomial since the gap between successive values of a binomial variate is unity. Similarly, $P(X_1 < X < X_2)$ is to be corrected as $P\left(X_1 + \frac{1}{2} \leq X \leq X_2 - \frac{1}{2}\right)$, since $X_1 < X$ does not include X_1 and $X < X_2$ does not include X_2.

Note The normal approximation to binomial probability mass function is good when $n \geq 50$ and neither p nor q is less than 0.1.

Notes

Example 39: An unbiased die is tossed 600 times. Use normal approximation to binomial to find the probability obtaining

(i) more than 125 aces,

(ii) number of aces between 80 and 110,

(iii) exactly 150 aces.

Solution.

Let X denote the number of successes, i.e., the number of aces.

$$\therefore \quad \mu = np = 600 \times \frac{1}{6} = 100 \text{ and } \sigma = \sqrt{npq} = \sqrt{600 \times \frac{1}{6} \times \frac{5}{6}} = 9.1$$

(i) To make correction for continuity, we can write

$$P(X > 125) = P(X > 125 + 0.5)$$

Thus, $P(X \geq 125.5) = P\left(z \geq \frac{125.5 - 100}{9.1}\right) = P(z \geq 2.80)$

$$= 0.5000 - P(0 \leq z \leq 2.80) = 0.5000 - 0.4974 = 0.0026.$$

(ii) In a similar way, the probability of the number of aces between 80 and 110 is given by

$$P(79.5 \leq X \leq 110.5) = P\left(\frac{79.5 - 100}{9.1} \leq z \leq \frac{110.5 - 100}{9.1}\right)$$

$$= P(-2.25 \leq z \leq 1.15) = P(0 \leq z \leq 2.25) + P(0 \leq z \leq 1.15)$$

$$= 0.4878 + 0.3749 = 0.8627$$

(iii) $P(X = 120) = P(119.5 \leq X \leq 120.5) = P\left(\frac{19.5}{9.1} \leq z \leq \frac{20.5}{9.1}\right)$

$$= P(2.14 \leq z \leq 2.25) = P(0 \leq z \leq 2.25) - P(0 \leq z \leq 2.14)$$

$$= 0.4878 - 0.4838 = 0.0040$$

15.9.7 Normal Approximation to Poisson Distribution

Normal distribution can also be used to approximate a Poisson distribution when its parameter $m \geq 10$. If X is a Poisson variate with mean m, then, for m ³ 10, the distribution of X can be taken as approximately normal with mean m and standard deviation $\sqrt{m}$ so that $z = \frac{X - m}{\sqrt{m}}$ is a standard normal variate.

Example 40: A random variable X follows Poisson distribution with parameter 25. Use normal approximation to Poisson distribution to find the probability that X is greater than or equal to 30.

Notes

Solution.

P(X ≥ 30) = P(X ≥ 29.5) (after making correction for continuity).

$$= P\left(z \geq \frac{29.5 - 25}{5}\right) = P(z \geq 0.9)$$

= 0.5000 – P(0 ≤ z ≤ 0.9) = 0.5000 - 0.3159 = 0.1841

15.3.8 Fitting a Normal Curve

A normal curve is fitted to the observed data with the following objectives :

1. To provide a visual device to judge whether it is a good fit or not.
2. Use to estimate the characteristics of the population.

   The fitting of a normal curve can be done by

   (a) The Method of Ordinates or

   (b) The Method of Areas.

(a) Method of Ordinates

In this method, the ordinate f(X) of the normal curve, for various values of the random variate X are obtained by using the table of ordinates for a standard normal variate.

We can write $f(X) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2} = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}z^2} = \frac{1}{\sigma}\phi(z)$

where $z = \frac{X-\mu}{\sigma}$ and $\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$.

The expected frequency corresponding to a particular value of X is given by $y = N.f(X) = \frac{N}{\sigma}\phi(z)$ and therefore, the expected frequency of a class = y ´ h, where h is the class interval.

Example 41:

Fit a normal curve to the following data :

Class Intervals	:	10-20	20-30	30-40	40-50	50-60	60-70	70-80	*Total*
Frequency	:	2	11	24	33	20	8	2	100

Notes

Solution.

First we compute mean and standard deviation of the given data.

Class Intervals	Mid - values (X)	Frequency (f)	$d = \frac{X - 45}{10}$	fd	fd^2
10 - 20	15	2	- 3	- 6	18
20 - 30	25	11	- 2	- 22	44
30 - 40	35	24	- 1	- 24	24
40 - 50	45	33	0	0	0
50 - 60	55	20	1	20	20
60 - 70	65	8	2	16	32
70 - 80	75	2	3	6	18
Total		100		- 10	156

Note: If the class intervals are not continuous, they should first be made so.

$$\therefore \quad \mu = 45 - 10 \times \frac{10}{100} = 44$$

and $\sigma = 10\sqrt{\frac{156}{100} - \left(\frac{10}{100}\right)^2} = 10\sqrt{1.55} = 12.4$

Table for the fitting of Normal Curve

Class Intervals	Mid - values (X)	$z = \frac{X - m}{s}$	$f(z)$ (from table)	$y = \frac{N}{s} f(z)$	f_e *
10 - 20	15	- 2.34	0.0258	0.2081	2
20 - 30	25	- 1.53	0.1238	0.9984	10
30 - 40	35	- 0.73	0.3056	2.4645	25
40 - 50	45	0.08	0.3977	3.2073	32
50 - 60	55	0.89	0.2685	2.1653	22
60 - 70	65	1.69	0.0957	0.7718	8
70 - 80	75	2.50	0.0175	0.1411	1

(b) Method of Areas

Under this method, the probabilities or the areas of the random variable lying in various intervals are determined. These probabilities are then multiplied by N to get the expected frequencies. This procedure is explained below for the data of the above example.

Class Intervals	Lower Limit (X)	$z = \frac{X - 44}{12.4}$	Area from 0 to z	Area under the class	f_e *
10 - 20	10	- 2.74	0.4969	0.0231	2
20 - 30	20	- 1.94	0.4738	0.1030	10
30 - 40	30	- 1.13	0.3708	0.2453	25
40 - 50	40	- 0.32	0.1255	0.3099	31
50 - 60	50	0.48	0.1844	0.2171	22
60 - 70	60	1.29	0.4015	0.0806	8
70 - 80	70	2.10	0.4821	0.0160	2
	80	2.90	0.4981		

*Expected frequency approximated to the nearest integer.

15.4 Summary

Notes

- The random variable in case of Poisson distribution is of the type ; the number of arrivals of customers per unit of time or the number of defects per unit length of cloth, etc. Alternatively, it is possible to define a random variable, in the context of Poisson Process, as the length of time between the arrivals of two consecutive customers or the length of cloth between two consecutive defects, etc. The probability distribution of such a random variable is termed as Exponential Distribution.

  Let t be a random variable which denotes the length of time or distance between the occurrence of two consecutive events or the occurrence of the first event and m be the average number of times the event occurs per unit of time or length. Further, let A be the event that the time of occurrence between two consecutive events or the occurrence of the first event is less than or equal to t and f(t) and F(t) denote the probability density function and the distribution (or cumulative density) function of t respectively.

- We can write $P(A)+P(\bar{A})=1$ or $F(t)+P(\bar{A})=1$. Note that, by definition, $F(t)=P(A)$.

  Further, $P(\bar{A})$ is the probability that the length of time between the occurrence of two consecutive events or the occurrence of first event is greater than t. This is also equal to the probability that no event occurs in the time interval t. Since the mean number of occurrence of events in time t is mt, we have , by Poisson distribution.

- A large number of chance factors: The factors, affecting the observations of a random variable, should be numerous and equally probable so that the occurrence or non-occurrence of any one of them is not predictable.
- Condition of homogeneity: The factors must be similar over the relevant population although, their incidence may vary from observation to observation.
- Condition of independence: The factors, affecting observations, must act independently of each other.
- Condition of symmetry: Various factors operate in such a way that the deviations of observations above and below mean are balanced with regard to their magnitude as well as their number.
- Normal distribution can also be used to approximate a Poisson distribution when its parameter m ≥ 10. If X is a Poisson variate with mean m, then, for m ³ 10, the distribution of X can be taken as approximately normal with mean m and standard deviation $\sqrt{m}$ so that $z=\frac{X-m}{\sqrt{m}}$ is a standard normal variate.

15.5 Keywords

Continuous random variable: A continuous random variable X is said to be uniformly distributed in a close interval (*a*, *b*) with probability density function p(X) if

$$p(X)=\frac{1}{\beta-\alpha} \quad \text{for } \alpha \le X \le \beta \quad \text{and}$$

$$= 0 \quad \text{otherwise.}$$

Notes

Normal probability distribution: The normal probability distribution occupies a place of central importance in Modern Statistical Theory.

Binomial distribution: Normal distribution can be used as an approximation to binomial distribution when n is large and neither p nor q is very small.

15.6 Self Assessment

1. .... distributions can be divided into two broad categories, viz. discrete and continuous probability distributions, depending upon whether the random variable is discrete or continuous.

   (a) Theoretical probability (b) uniform distribution

   (c) continuous random variable (d) normal probability distribution

2. A .... X is said to be uniformly distributed in a close interval (*a*, *b*) with probability density function p(X) if

   $$p(X) = \frac{1}{\beta - \alpha} \quad \text{for } \alpha \le X \le \beta \quad \text{and}$$

   $$= 0 \quad \text{otherwise.}$$

   (a) Theoretical probability (b) uniform distribution

   (c) continuous random variable (d) normal probability distribution

3. The .... is alternatively known as rectangular distribution.

   (a) Theoretical probability (b) uniform distribution

   (c) continuous random variable (d) normal probability distribution

4. The .... occupies a place of central importance in Modern Statistical Theory.

   (a) Theoretical probability (b) uniform distribution

   (c) continuous random variable (d) normal probability distribution

5. Normal distribution can be used as an approximation to .... when n is large and neither p nor q is very small.

   (a) Theoretical probability (b) uniform distribution

   (c) continuous random variable (d) normal probability distribution

15.7 Review Questions

1. In a metropolitan city, there are on the average 10 fatal road accidents in a month (30 days). What is the probability that (i) there will be no fatal accident tomorrow, (ii) next fatal accident will occur within a week?

   Hint : Take m = 1/3 and apply exponential distribution.

2. A counter at a super bazaar can entertain on the average 20 customers per hour. What is the probability that the time taken to serve a particular customer will be (i) less than 5 minutes, (ii) greater than 8 minutes?

   Hint : Use exponential distribution.

Notes

3. The marks obtained in a certain examination follow normal distribution with mean 45 and standard deviation 10. If 1,000 students appeared at the examination, calculate the number of students scoring (i) less than 40 marks, (ii) more than 60 marks and (iii) between 40 and 50 marks.

   Hint : See example 30.

4. The ages of workers in a large plant, with a mean of 50 years and standard deviation of 5 years, are assumed to be normally distributed. If 20% of the workers are below a certain age, find that age.

   Hint : Given $P(X < X_1) = 0.20$, find X_1.

5. The mean and standard deviation of certain measurements computed from a large sample are 10 and 3 respectively. Use normal distribution approximation to answer the following:

   (i) About what percentage of the measurements lie between 7 and 13 inclusive?

   (ii) About what percentage of the measurements are greater than 16?

   Hint : Apply correction for continuity.

6. There are 600 business students in the post graduate department of a university and the probability for any student to need a copy of a particular text book from the university library on any day is 0.05. How many copies of the book should be kept in the library so that the probability that none of the students, needing a copy, has to come back disappointed is greater than 0.90? (Use normal approximation to binomial.)

   Hint : If X_1 is the required number of copies, $P(X \leq X_1) \geq 0.90$.

7. The grades on a short quiz in biology were 0, 1, 2, 3, ... 10 points, depending upon the number of correct answers out of 10 questions. The mean grade was 6.7 with standard deviation of 1.2. Assuming the grades to be normally distributed, determine (i) the percentage of students scoring 6 points, (ii) the maximum grade of the lowest 10% of the class.

   Hint : Apply normal approximation to binomial.

8. The following rules are followed in a certain examination. "A candidate is awarded a first division if his aggregate marks are 60% or above, a second division if his aggregate marks are 45% or above but less than 60% and a third division if the aggregate marks are 30% or above but less than 45%. A candidate is declared failed if his aggregate marks are below 30% and awarded a distinction if his aggregate marks are 80% or above. "

   At such an examination, it is found that 10% of the candidates have failed, 5% have obtained distinction. Calculate the percentage of students who were placed in the second division. Assume that the distribution of marks is normal. The areas under the standard normal curve from 0 to z are:

z	:	1.28	1.64	0.41	0.47
Area	:	0.4000	0.4500	0.1591	0.1808

   Hint : First find parameters of the distribution on the basis of the given information.

9. For a certain normal distribution, the first moment about 10 is 40 and the fourth moment about 50 is 48. What is the mean and standard deviation of the distribution?

   Hint : Use the condition $b_2 = 3$, for a normal distribution.

10. In a test of clerical ability, a recruiting agency found that the mean and standard deviation of scores for a group of fresh candidates were 55 and 10 respectively. For another experienced group, the mean and standard deviation of scores were found to be 62 and 8 respectively.

Notes

Assuming a cut-off scores of 70, (i) what percentage of the experienced group is likely to be rejected, (ii) what percentage of the fresh group is likely to be selected, (iii) what will be the likely percentage of fresh candidates in the selected group? Assume that the scores are normally distributed.

Hint : See example 33.

11. 1,000 light bulbs with mean life of 120 days are installed in a new factory. Their length of life is normally distributed with standard deviation of 20 days. (i) How many bulbs will expire in less than 90 days? (ii) If it is decided to replace all the bulbs together, what interval should be allowed between replacements if not more than 10 percent bulbs should expire before replacement?

    Hint : (ii) $P(X \leq X_1) = 0.9$.

Answers: Self Assessment

1. (a) 2. (c) 3. (b) 4. (d) 5. (a)

15.8 Further Readings

Books

Introductory Probability and Statistical Applications by P.L. Meyer

Introduction to Mathematical Statistics by Hogg and Craig

Fundamentals of Mathematical Statistics by S.C. Gupta and V.K. Kapoor

Unit 16: Reliability Theory

CONTENTS

Objectives

Introduction

16.1 Classical Theory of Parallel Tests

16.2 Domain Sampling Theory-1

16.2.1 Domain Sampling Theory-2

16.3 Summary

16.4 Keywords

16.5 Self Assessment

16.6 Review Questions

16.7 Further Readings

Objectives

After studying this unit, you will be able to:

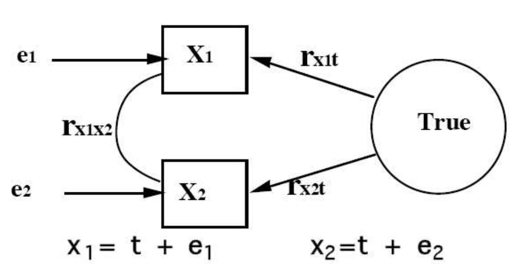
- Discuss classical theory of parallel tests
- Describe domain sampling theory

Introduction

In last unit you have studied about realiability system. Ths unit will explain you theory related to realiability.

16.1 Classical Theory of Parallel Tests

Consider two tests (X_1 and X_2) which both measure the same construct (T). Assume that for every individual that $t_1 = t_2 = t$.



Notes

Even if $e_1 \neq e_2$, we can assume that $V_{e_1} = V_{e_2}$ Then: $V_{x_1} = V_t + V_{e_1} = V_t + V_{e_2} = V_{x_2} = V_x$

$$C_{x_1x_2} = \frac{\Sigma(x_1 * x_2)}{N} = \frac{\Sigma[(t+e_1)(t+e_2)]}{N} \Rightarrow$$

$$C_{x_1x_2} = V_t + Ct_{e_1} + Ct_{e_2} + C_{e_1e_2} = V_t$$

$$r_{x_1x_2} = \frac{C_{x_1x_2}}{\sqrt{V_{x_1} * V_{x_2}}} = \frac{C_{x_1x_2}}{V_x} = \frac{V_t}{V_x}$$

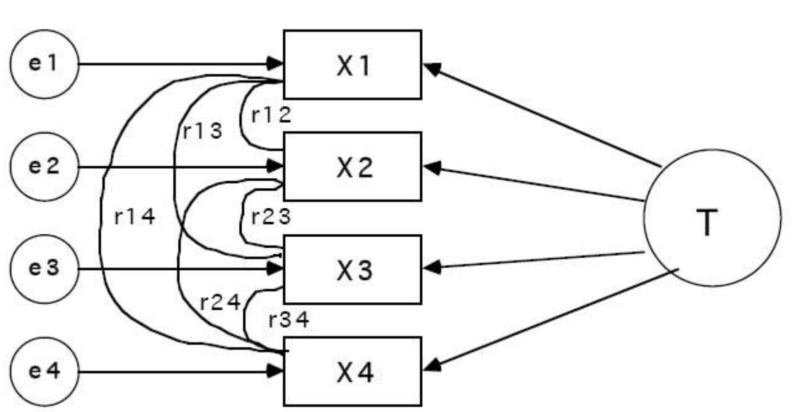
$$r_{x_1x_2} = \frac{V_t}{V_x} = rxt^2$$

The reliability is the correlation between two parallel tests and is equal to the squared correlation of the test with the construct.

$r_{xx} = \frac{V_t}{V_x}$ = percent of test variance which is construct variance. rxt = $\sqrt{rxx}$ $\Rightarrow$ the validity of a test is bounded by the square root of the reliability.

How do we tell if one of the two "parallel" tests is not as good as the other? That is, what if the two tests are not parallel?

Congeneric Measurement Theory



This matrix will have the following covariances:

	x_1	x_2	x_3	x_4
x_1	Vx_1			
x_2	Cx_1x_2	Vx_2		
x_3	Cx_1x_3	Cx_2x_3	Vx_3	
x_4	Cx_1x_4	Cx_2x_4	Cx_3x_4	Vx_4

These covariances reflect the following parameters:

Notes

	x_1	x_2	x_3	x_4
x_1	$Vt+Ve_1$			
x_2	Cx_1tCx_2tVt	$Vt+Ve_2$		
x_3	Cx_1tCx_3tVt	Cx_2tCx_3tVt	$Vt+Ve_3$	
x_4	$Cx_1tCx_4tV_t$	Cx_2tCx_4tVt	Cx_3tCx_4tVt	$Vt+Ve_4$

We need to estimate the following parameters:

Vt, Ve_1, Ve_2, Ve_3, Ve_4, Cx_1t, Cx_2t, Cx_3t, Cx_4t

Parallel tests assume $Ve_1 = Ve_2 = Ve_3 = Ve_4$, and $Cx_1t = Cx_2t$

$= Cx_3t = Cx_4t$ and only need two tests.

Tau equivalent tests assume: $Cx_1t = Cx_2t = Cx_3t = Cx_4t$ and need at least three tests to estimate parameters.

Congeneric tests allow all parameters to vary but require at least four tests to estimate parameters.

16.2 Domain Sampling Theory-1

Consider a domain (D) of k items relevant to a construct. (E.g., English vocabulary items, expresions of impulsivity). Let D_i represent the number of items in D which the i^{th} subject can pass (or endorse in the keyed direction) given all D items. Call this the domain score for subject i.

What is the correlation of scores on an itemj with domain scores?

$$C_jd = Vj + \sum_{l=1}^{k} Cl_j = V_j + (k-1) * (\text{average covariance of j})$$

Domain variance = $\sum_{l=1}^{k} Vl + \sum_{j\neq 1}^{k} Cl_j = \Sigma(\text{variance}) + \Sigma(\text{covariances})$

Vd = k* (average variance) + k*(k – 1) * (average covariance)

Let Va = average variance and Ca = average covariance then Vd= k(Va + (k-1)Ca).

Assume that Vj = Va and that Cj l = Ca.

$$rjd = \frac{Cjd}{\sqrt{Vj * Vd}} = \frac{Va + (k-1) * Ca}{\sqrt{Va * k(Va + (K-1)Ca)}}$$

$$rjd^2 = \frac{(Va + (k-1) * Ca) * (Va + (k-1) * Ca)}{Va * k * (Va + (k-1)Ca)}$$

Now, find the limit of rj d2 as k becomes large:

$\lim_{k\to\infty} rjd^2 = \frac{C_a}{V_a}$ =average covariance/average variance i.e., the amount of domain variance in an item (the squared correlation of the item with the domain) is the averge intercorrelation in the domain.

Notes

16.2.1 Domain Sampling Theory-2

What is the correlation of a test of n items with the domain score?

Domain variance = $\sum_{I=1}^{k} V_I + \sum_{j \neq 1}^{k} Clj$ = Σ (variances) + Σ (covariances)

Let Va = average variance and Ca = average covariance then Vd= k(Va + (k-1)Ca), Cnd = n*Va +n*(k-1)Ca

Vn = variance of an n-item test = ΣVj + ΣCjl = Vn = n*Va + n*(n-1)*Ca

Vn = n*Va+n*(n-1)*Ca

$$rnd = \frac{Cnd}{\sqrt{Vn * Vd}} \Rightarrow rnd^2 = \frac{Cnd^2}{Vn * Vd}$$

$$rnd^2 = \frac{\{n * Va + n * (k-1)Ca\} * \{n * Va + n * (k-1)Ca\}}{\{n * Va + n * (n-1) * Ca\} * \{k(Va + (k-1)Ca)\}}$$

$$rnd^2 = \frac{\{Va + (k-1)Ca\} * \{n * Va + n * (k-1)Ca\}}{\{Va + (n-1) * Ca\} * \{k(Va + (k-1)Ca)\}}$$

$$rnd^2 = \frac{\{n * Va + n * (k-1)Ca\}}{\{Va + (k-1) * Ca)\} * \{k\}}$$

lim as k ⇒ ∞ of rnd² = $\frac{n * Ca}{Va + (n-1)Ca}$

i.e., the amount of domain variance in a n-item test (the squared correlation of the test with the domain) is a function of the number of items and the average covariance within the test.

Coefficient Alpha - 1

Consider a test made up of k items with an average intercorrelation r.

1) What is the correlation of this test with another test sampled from the same domain of items?

2) What is the correlation of this test with the domain?

	Test 1	Test 2
Test 1	V1	C1 2
Test 2	C1 2	V2

Let r1 be the average correlation of items within test 1

Let r2 be the average correlation of items within test 2

Notes

Let r12 be the average intercorrelation of items between the two tests.

$$rx1x2 = \frac{C12}{\sqrt{V1*V2}}$$

	Test 1	Test 2
Test 1	V1 = k * [1+(k-1) *r1]	C12= k*k* r12
Test 2	C12= k*k* r12	V2 = k * [1+(k-1) *r2]

$$rx1x2 = \frac{k * k * r12}{\sqrt{k*[1+(k-1)*r1]*k*[1+(k-1)*r2]}}$$

But, since the two tests are composed of randomly equivalent items, r1 = r2 = r and

$$rx1x2 = \frac{k*r}{1+(k-1)r} = alpha = a$$

Note That is the same as the squared correlation of a test with the domain. Alpha is the correlation of a test with a test just like it, and is the percentage of test variance which is domain variance.

Internal Consistency and Coefficient alpha - 2

Consider a test made up of k items with average variance v_i . What is the correlation of this test with another test sampled from the same domain of items?

	Test 1	Test 2
Test 1	V1	C12
Test 2	C12	V2

What is the correlation of this test with the domain?

Let V_t be the total test variance for Test 1= V_1=V_2

Let v_i be the average variance of an item within the test.

$$rx1x2 = \frac{C12}{\sqrt{V1*V2}}$$

We need to estimate the covariance with the other test:

	Test 1	Test 2
Test 1	V1 = k * [vi +(k-1) *c1]	C12= k*k* r12
Test 2	C12= k2 c12,	V2 = k * [vi +(k-1) *c2]

C12 = k2 c12, but what is the average c12?

Notes

$Vt = V1 = !V2 \Rightarrow c_1 = c_2 = c_{12} \Rightarrow$

$$c1 = \frac{Vt - \Sigma v_i}{k^*(k\text{-}1)} = \text{average covariance}$$

$$C12 = k^2\, c12 \implies C12 = k^2 * \frac{V_t - \Sigma v_i}{k^*(k\text{-}1)}$$

$$rx1x2 = \frac{k^2 * \frac{V_t - \Sigma v_i}{k^*(k\text{-}1)}}{V_t} = \frac{V_t - \Sigma v_i}{V_t} * \frac{k}{k-1}$$

This allows us to find coefficient alpha without finding the average interitem correlation.

The effect of test length of internal consistency reliability.

Number of items	Average r	Average r
	0.2	0.1
1	0.20	0.10
2	0.33	0.18
4	0.50	0.31
8	0.67	0.47
16	0.80	0.64
32	0.89	0.78
64	0.94	0.88
128	0.97	0.93

Estimates of reliability reflect both the length of the test as well as the average inter-item correlation. To report the internal consistency of a domain (rather than a specific test with a specific length, it is possible to report the "alpha1" for the test.

$$\text{Average interitem r} = \text{alpha1} = \frac{\text{alpha}}{\text{alpha} + k^*(1\text{-alpha})}$$

This allows us to find the average internal consistency of a scale independent of test length.

because $\alpha = \frac{V_t - \Sigma v_i}{V_t} * \frac{k}{k-1}$ is easy to estimate from the basic test statistics and is an estimate of the amount of test variance that is construct related, it should be reported whenever a particular inventory is used.

Coefficients Alpha, Beta and Omega - 1

Components of variance associated with a test score include general test variance, group variance, specific item variance, and error variance.

General	Group	Specific	Error
Reliable Variance			
Common Shared Variance			

Notes

Coefficient alpha is the average of all possible splits, and over estimates the general and underestimates the total common variance. It is a lower bound estimate of reliable variance.

Now, consider a test with general and group variance. Each Subtest has general variance but also has Group, Specific, and Error. The subtests share only general variance. How do we estimate the amount of General variance? What would be to correlation of this test with another test with the same general structure, but with different group structures?

Find the two most unrelated subtests within each test.

	Subtest A-1	Subtest A-2	Subtest B-3	Subtest B-4
Subtest A-1	g+G1+S+E	g	g	g
Subtest A-2	g	g+G2+S+E	g	g
Subtest B-3	g	g	g+G3+S+E	
Subtest B-4	g	g		g+G4+S+E

$$r_{ab} = \frac{Cab}{\sqrt{Va*Vb}} = \frac{4g}{\sqrt{2*(g+G_i+S+E+g)*2*(g+G_i+S+E+g)}}$$

$$\frac{2g}{g+G_1+S+E+g} = \frac{2r_{a1a2}}{1+r_{a1a2}} = \text{"Coefficient Beta"}$$

Coefficient beta is the worst split half reliability and is thus an estimate of the general saturation of the test.

Coefficients Alpha, Beta and Omega - 2

Consider a test with two subtests which are maximally different (the worst split half). What is the predicted correlation with another test formed in the same way?

	Subtest A-1	Subtest A-2	Subtest B-3	Subtest B-4
Subtest A-1	g+G1+S+E	g	g	g
Subtest A-2	g	g+G2+S+E	g	g
Subtest B-3	g	g	g+G3+S+E	
Subtest B-4	g	g		g+G4+S+E

		Test Size = 10 items		Test Size = 20 items	
General Factor	Group Factor	Alpha	Beta	Alpha	Beta
0.25	0.00	0.77	0.77	0.87	0.87
0.20	0.05	0.75	0.71	0.86	0.83
0.15	0.10	0.73	0.64	0.84	0.78
0.10	0.15	0.70	0.53	0.82	0.69
0.05	0.20	0.67	0.34	0.80	0.51
0.00	0.25	0.63	0.00	0.77	0.00

Notes

Note Although alpha is relatively insensitive to the relative contributions of group and general factor, beta is very sensitive. Alpha, however, can be found from item and test statistics, beta needs to be estimated by finding the worst split half. Such an estimate is computationally much more difficult.

Omega, a more general estimate, based upon the factor structure of the test, allows for bette estimate of the first factor saturation.

Generalizabilty Theory Reliability across facets:

The consistency of Individual Differences across facets may be assessed by analysing variance components associated with each facet. i.e., what amount of varianceis associated with a particular facet across which one wants to generalize?

Facets of reliability

Across Items	Domain Sampling
	Internal Consistency
Across Time	Temporal Stability
Across Forms	Alternate Form Reliability
Across Raters	Inter-rater agreement
Across Situations	Situational Stability
Across "Tests" (facets unspecified)	Parallel Test reliability

Generalizability theory is a decomposition of variance components to estimate sources of variance across which one wants to generalize.

All of these conventional approaches are concerned with generalizing about individual differences (in response to an item, time, form, rater, or situation) between people. Thus, the emphasis is upon consistency of rank orders. Classical reliability is a function of large between subject variability and small within subject variability. It is unable to estimate the within subject precision.

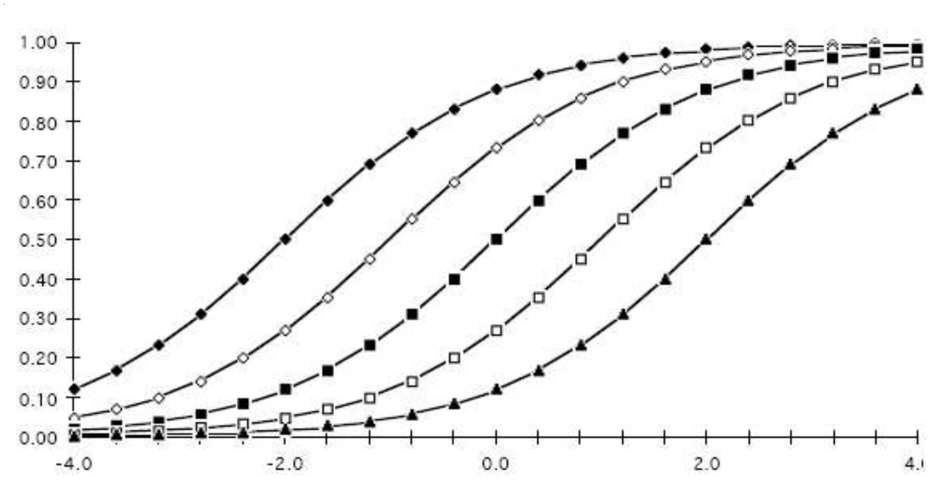
An alternative method (Latent Response Theory or Item Response Theory) is to determine the precision of the estimate of a particular person's position on a latent variable.

Item Response Theory - 1

A model for item response as a function of increasing level of subject ability and increasing levels of item difficulty. This model estimates the probability of making a particular response (generally, correct or incorrect) as a joint function of the subject's value on a latent attribute dimension, and the difficulty (item endorsement rate) of a particular item.

Model 1: the Rasch model: Probability of endorsing an item given ability (ø) and difficulty (diff):

$$P(y \mid \text{ø}, diff) = \frac{1}{1+e(diff-\text{ø})}$$



Attribute Value —>

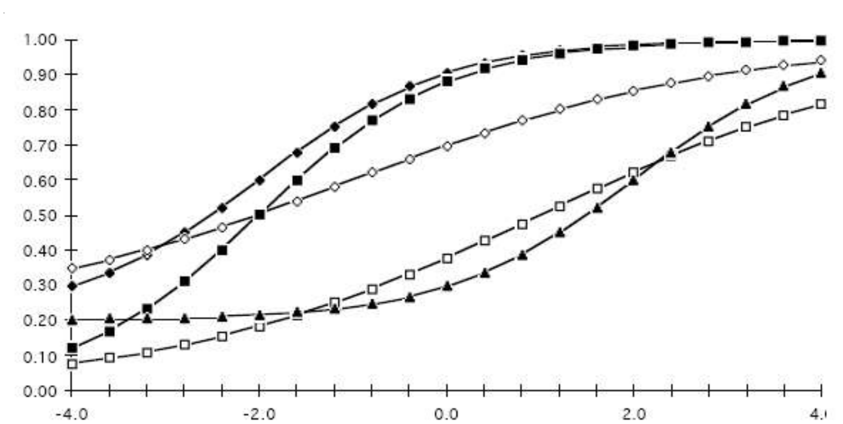
This procedure is (theoretically) not concerned with rank orders of respondents, but rather with the error of estimate for a particular respondent. This technique allows for computerized adaptive testing.

Item Response Theory - 2

A model for item response as a function of increasing level of subject ability and increasing levels of item difficulty.

Model 2: the 3 parameter model: Probability of endorsing an item given ability (ø) ,difficulty (diff), guessing (guessing), and item discrimination sensitivity:

$$P(y \mid ø,diff,guess,sensitivity) = guessing + \frac{guessing}{1 + esensitivity*(diff-ø)}$$



Attribute value —>

Note that with this model, even though the probability of item endorsement for a particular item may be a monotonic function of attribute value, item endorsement probabilities for different items may be a non-monotonic function of the attribute.

Notes

16.3 Summary

- The reliability is the correlation between two parallel tests and is equal to the squared correlation of the test with the construct.

  $r_{xx} = \frac{V_t}{V_x}$ = percent of test variance which is construct variance. rxt = $\sqrt{rxx}$ $\Rightarrow$ the validity of a test is bounded by the square root of the reliability.

  How do we tell if one of the two "parallel" tests is not as good as the other? That is, what if the two tests are not parallel?

- Although alpha is relatively insensitive to the relative contributions of group and general factor, beta is very sensitive. Alpha, however, can be found from item and test statistics, beta needs to be estimated by finding the worst split half. Such an estimate is computationally much more difficult.

- All of these conventional approaches are concerned with generalizing about individual differences (in response to an item, time, form, rater, or situation) between people. Thus, the emphasis is upon consistency of rank orders. Classical reliability is a function of large between subject variability and small within subject variability. It is unable to estimate the within subject precision.

  An alternative method (Latent Response Theory or Item Response Theory) is to determine the precision of the estimate of a particular person's position on a latent variable.

16.4 Keywords

Reliability: The reliability is the correlation between two parallel tests and is equal to the squared correlation of the test with the construct.

Congeneric tests allow all parameters to vary but require at least four tests to estimate parameters.

Domain variance: Domain variance = $\sum_{I=1}^{k} V_I + \sum_{j \neq 1}^{k} Clj$ = Σ (variances) + Σ (covariances)

Components of variance: Components of variance associated with a test score include general test variance, group variance, specific item variance, and error variance.

Coefficient alpha is the average of all possible splits, and over estimates the general and underestimates the total common variance. It is a lower bound estimate of reliable variance.

16.5 Self Assessment

1. The .... is the correlation between two parallel tests and is equal to the squared correlation of the test with the construct.
2. .... tests allow all parameters to vary but require at least four tests to estimate parameters.
3. .... associated with a test score include general test variance, group variance, specific item variance, and error variance.
4. .... is the average of all possible splits, and over estimates the general and underestimates the total common variance. It is a lower bound estimate of reliable variance.

5. ...., a more general estimate, based upon the factor structure of the test, allows for bette estimate of the first factor saturation.
6. The consistency of .... across facets may be assessed by analysing variance components associated with each facet.
7. .... is a decomposition of variance components to estimate sources of variance across which one wants to generalize.

16.6 Review Questions

Answers: Self Assessment

1. reliability	2. Congeneric	3. Components of variance	4. Coefficient alpha
5. Omega,	6. Individual Differences	7. Generalizability theory	

16.7 Further Readings

Books

Sheldon M. Ross, Introduction to Probability Models, Ninth Edition, Elsevier Inc., 2007.

Jan Pukite, Paul Pukite, Modeling for Reliability Analysis, IEEE Press on Engineering of Complex Computing Systems, 1998.

Notes

Unit 17: System Reliability

CONTENTS

Objectives

Introduction

17.1 State Vectors

17.1.1 Structure Functions

17.1.2 The Series Structure

17.1.3 The Parallel Structure

17.1.4 The k-out-of-n Structure

17.2 Order and Monotonicity

17.2.1 Minimal Path Sets

17.3 The Bridge Structure

17.3.1 Minimal Cut Sets

17.4 Bounds on Reliability

17.5 System Life in Systems Without Repair

17.6 Distribution Functions for Modeling Component Lifetimes

17.7 The Bathtub Curve and Failure Rate

17.8 Expected System Life

17.9 Systems with Repair

17.9.1 System with Standby Components and Repair

17.9.2 System with Interrelated Repair

17.10 Summary

17.11 Keywords

17.12 Self Assessment

17.13 Review Questions

17.14 Further Readings

Objectives

After studying this unit, you will be able to:

- Discuss state vectors
- Describe order and monotonicity
- Explain bridge structure

Notes

Introduction

The ability of a system or component to perform its required functions under stated conditions for a specified period of time

System reliability is a function of:

- the reliability of the components
- the interdependence of the components
- the topology of the components

17.1 State Vectors

Consider a system comprised on *n* components, where each component is either functioning or has failed. Define

$$x_i = \begin{cases} 1, & \text{if the } i^{th} \text{ component is functioning} \\ 0, & \text{if the } i^{th} \text{ component has failed} \end{cases}$$

The vector $\boldsymbol{x} = \{x_1, \ldots, x_n\}$ is called the *state vector*.

17.1.1 Structure Functions

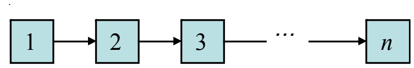
Assume that whether the system as a whole is functioning is completely determined by the state vector $\boldsymbol{x}$. Define

$$\phi(\boldsymbol{x}) = \begin{cases} 1, & \text{if the system is functioning when the state vector is } \boldsymbol{x} \\ 0, & \text{if the system has failed when the state vector is } \boldsymbol{x} \end{cases}$$

The function $\phi(\boldsymbol{x})$ is called the *structure function* of the system.

17.1.2 The Series Structure

A series system functions if and only if all of its *n* components are functioning:

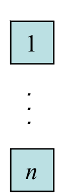


Its structure function is given by

$$\phi(\boldsymbol{x}) = \prod_{i=1}^{n} x_i.$$

17.1.3 The Parallel Structure

A parallel system functions if and only if at least one of its *n* components are functioning:



Notes

Its structure function is given by

$$\phi(\boldsymbol{x}) = \max_{i=1,\ldots,n} x_i .$$

17.1.4 The *k*-out-of-*n* Structure

A *k*-out-of-*n* system functions if and only if at least *k* of its *n* components are functioning:

Its structure function is given by

$$\phi(\boldsymbol{x}) = \begin{cases} 1, & \text{if } \sum_{i=1}^{n} x_i \geq k \\ 0, & \text{if } \sum_{i=1}^{n} x_i < k. \end{cases}$$

17.2 Order and Monotonicity

A partial order is defined on the set of state vectors as follows. Let x and y be two state vectors. We define

$$\boldsymbol{x} \leq \boldsymbol{y} \text{ if } x_i \leq y_i,\ i = 1, \ldots, n.$$

Furthermore,

$$\boldsymbol{x} < \boldsymbol{y} \text{ if } \boldsymbol{x} \leq \boldsymbol{y} \text{ and } x_i < y_i \text{ for some } i.$$

We assume that if $\boldsymbol{x} \leq \boldsymbol{y}$ then $\phi(\boldsymbol{x}) \leq \phi(\boldsymbol{y})$. In this case we say that the system is *monotone*.

17.2.1 Minimal Path Sets

- A state vector $\boldsymbol{x}$ is call a *path vector* if $\phi(\boldsymbol{x}) = 1$.
- If $\phi(\boldsymbol{y}) = 0$ for all $\boldsymbol{y} < \boldsymbol{x}$, then $\boldsymbol{x}$ is a *minimal path vector*.
- If $\boldsymbol{x}$ is a minimal path vector, then the set $A = \{i : x_i = 1\}$ is a *minimal path set*.

Examples:

1. ***The Series System:*** There is only one minimal path set, namely the entire system.
2. ***The Parallel System:*** There are n minimal path sets, namely the sets consisting of one component.
3. ***The k-out-of-n System:*** There are $\binom{n}{k}$ minimal path sets, namely all of the sets consisting of exactly k components.

Let $A_1, \ldots, A_s$ be the minimal path sets of a system. A system will function if and only if all the components of at least one minimal path set are functioning, so that

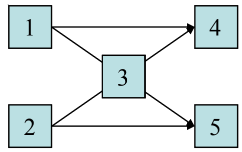
$$\phi(\boldsymbol{x}) = \max_{j} \prod_{i \in A_j} x_i .$$

This expresses the system as a parallel arrangement of series systems.

Notes

17.3 The Bridge Structure

The system whose structure is shown below is called the bridge system. Its minimal path sets are: {1, 4}, {1, 3, 5}, {2, 5}, {2, 3, 4}.



For example, the system will work if only 1 and 4 are working, but will not work if only 1 is working.

Its structure function is given by

$$\phi(\boldsymbol{x}) = \max\{x_1x_4, x_1x_3x_5, x_2x_5, x_2x_3x_4\}$$
$$= 1 - (1 - x_1x_4)(1 - x_1x_3x_5)(1 - x_2x_5)(1 - x_2x_3x_4).$$

17.3.1 Minimal Cut Sets

- A state vector $\boldsymbol{x}$ is call a *cut vector* if $\phi(\boldsymbol{x}) = 0$.
- If $\phi(\boldsymbol{y}) = 1$ for all $\boldsymbol{y} > \boldsymbol{x}$, then $\boldsymbol{x}$ is a *minimal cut vector*.
- If $\boldsymbol{x}$ is a minimal cut vector, then the set $C = \{i: x_i = 0\}$ is a *minimal cut set*.

Examples:

1. ***The Series System:*** There are *n* minimal cut sets, namely, the sets consisting of all but one component.
2. ***The Parallel System:*** There is one minimal cut set, namely, the empty set.
3. ***The k-out-of-n System:*** There are $\binom{n}{n-k+1}$ minimal cut sets, namely all of the sets consisting of exactly $n - k + 1$ components.

Let $C_1, \ldots, C_k$ be the minimal cut sets of a system. A system will not function if and only if all the components of at least one minimal cut set are not functioning, so that

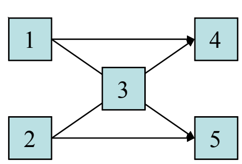
$$\phi(\boldsymbol{x}) = \prod_{j=1}^{k} \max_{i \in C_j} x_i.$$

This expresses the system as a series arrangement of parallel systems.

The Bridge Structure

The system whose structure is shown below is called the bridge system. Its minimal cut sets are:

{1, 2}, {1, 3, 5}, {4, 5}, {2, 3, 4}.



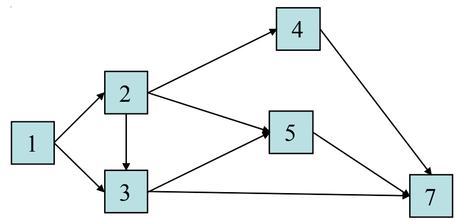
Notes

For example, the system will work if 1 and 2 are not working, but it can work if either 1 or 2 are working.

Its structure function is given by

$$\phi(\boldsymbol{x}) = \max\{x_1, x_2\}\max\{x_1, x_3, x_5\}\max\{x_4, x_5\}\max\{x_2, x_3, x_4\}.$$

Example:



What are the minimal path sets for this system?

What are the minimal cut sets?

System Reliability

Component Reliability:

$$p_i = P\{x_i = 1\}.$$

System Reliability:

$$r = P\{\phi(x) = 1\} = E[\phi(x)].$$

When the components are independent, then r can be expressed as a function of the component reliabilities:

$$r = r(p), \text{ where } p = (p_1, \ldots, p_n).$$

The function $r(\mathbf{p})$ is called the *reliability function*.

Example:

1. The Series System

$$\begin{aligned} r(\mathbf{p}) &= P\{\phi(\boldsymbol{x}) = 1\} \\ &= P\{x_i = 1 \text{ for all } i = 1,\ldots n\} \\ &= \prod_{i=1}^{n} p_i. \end{aligned}$$

2. The Parallel System

$$\begin{aligned} r(\mathbf{p}) &= P\{\phi(\boldsymbol{x}) = 1\} \\ &= P\{x_i = 1 \text{ for some } i = 1,\ldots n\} \\ &= 1 - \prod_{i=1}^{n} (1 - p_i). \end{aligned}$$

3. The k-out-of-n System. If $p_i = p$ for all $i = 1, \ldots, n$, then

$$r(\mathbf{p}) = P\{\phi(\boldsymbol{x}) = 1\}$$

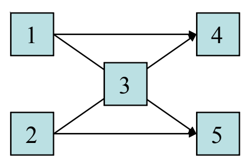
$$= P\{\sum_{i=1}^{n} x_i \geq k\}$$

$$= \sum_{i=k}^{n} \binom{n}{i} p^i (1-p)^{n-i}.$$

The Bridge Structure

Assume that all components have the same reliability p.

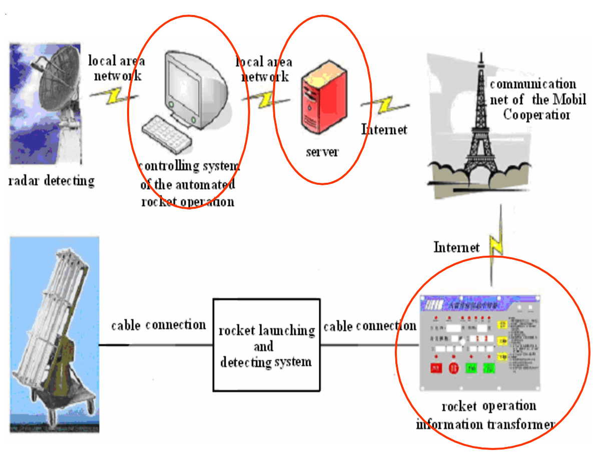
p	r(p)
0.8	0.91136
0.9	0.97848
0.95	0.99478
0.99	0.99980



System Reliability

Theorem 1. If r(p) is the reliability function of a system of independent components, then r(p) is an increasing function of p.

Example: Communications System



Suppose we are concerned about the reliability of the controller, server and transformer. Assume that these components are independent, and that

$p_{\text{controller}} = .95$

$p_{\text{server}} = .96$

$p_{\text{transformer}} = .99$

Notes

Since these three components connect in series, the system A consisting of these components has reliability

r_{system_A}

$= p_{controller} \cdot p_{server} \cdot p_{transformer}$

= .90

Suppose that we want to increase the reliability of system A. What are our options?

Suppose that we have two controllers, two servers, and two transformers.

Theorem 2. For any reliability function r and vectors,

$$p_1, p_2, r[1 - (1 - p_1)(1 - p_2)] \geq 1 - [1 - r(p_1)][1 - r(p_2)].$$

Note $1(1 - p_1)(1 - p_2) = (1 - (1 - p_{11})(1 - p_{21}), \ldots, 1 - (1 - p_{1n})(1 - p_{2n}))$

17.4 Bounds on Reliability

Let $A_1, \ldots, A_s$ be the minimal path sets of a system. Since the system will function if and only if all the components of at least one minimal path set are functioning, then

$$\begin{aligned} r(\mathbf{p}) &= P\left(\bigcup_{j=1}^{s}\{\text{all components } i \in A_j \text{ function}\}\right) \\ &\leq \sum_{j=1}^{s} P\{\text{all components } i \in A_j \text{ function}\} \\ &= \sum_{j=1}^{s}\prod_{i \in A_j} p_i \,. \end{aligned}$$

This bound works well only if p_i is small (< 0.2) for each component.

Similarly, let $C_1, \ldots, C_k$ be the minimal cut sets of a system. Since the system will not function if and only if all the components of at least one minimal cut set are not functioning, then

$$\begin{aligned} r(\mathbf{p}) &= 1 - P\left(\bigcup_{j=1}^{k}\{\text{all components } i \in C_j \text{ are not functioning}\}\right) \\ &\geq 1 - \sum_{j=1}^{k} P\{\text{all components } i \in C_j \text{ are not functioning}\} \\ &= 1 - \sum_{j=1}^{k}\left(\prod_{i \in C_j}(1 - p_i)\right). \end{aligned}$$

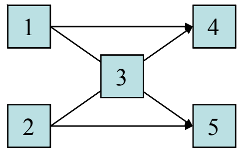
This bound works well only if p_i is large (> 0.8) for each component.

Example: The Bridge Structure

The minimal cut sets are:

{1, 2}, {1, 3, 5}, {4, 5}, {2, 3, 4}.

Notes



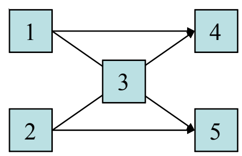
If each component has reliability *p*, then

$$r(\mathbf{p}) = 1 - P\left(\bigcup_{j=1}^{4}\{\text{all components } i \in C_j \text{ are not functioning}\}\right)$$

$$\geq 1 - \sum_{j=1}^{4} P\{\text{all components } i \in C_j \text{ are not functioning}\}$$

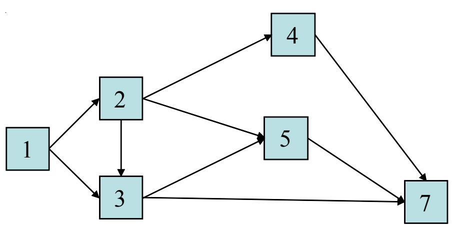
$$= 1 - 2(1-p)^2 - 2(1-p)^3$$

p	r(p)	lower bound
0.8	0.91136	0.90400
0.9	0.97848	0.97800
0.95	0.99478	0.99475
0.99	0.99980	0.99980

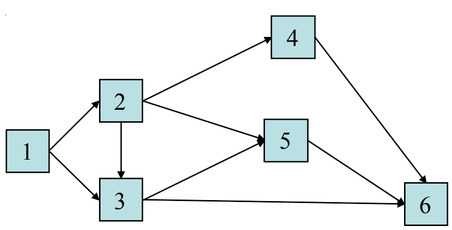


Example:

1. Calculate a lower bound on the reliability of this system. Assume that all components have the same reliability *p*.



2. The minimal cut sets are {1}, {2, 3}, {3, 4, 5}, and {7}. Hence $r \geq 1 - 2(1-p) - (1-p)^2 - (1-p)^3$.



17.5 System Life in Systems Without Repair

Suppose that the i^{th} component in an *n*-component system functions for a random lifetime having distribution function F_i and then fails.

Notes

Let $P_i(t)$ be the probability that component i is functioning at time t. Then

$$\begin{aligned} P_i(t) &= P\{\text{component i is functioning at time t}\} \\ &= P\{\text{lifetime of i} > t\} \\ &= 1 - F_i(t) \\ &\equiv \overline{F}_i(t). \end{aligned}$$

Now let F be the distribution function for the lifetime of the system. How does F relate to the F_i?

Let r(p) be the reliability function for the system, then

$$\begin{aligned} \overline{F}(t) &\equiv 1 - F(t) \\ &= P\{\text{lifetime of system} > t\} \\ &= P\{\text{system is functioning at time t}\} \\ &= r(P_1(t), \ldots P_n(t)) \\ &= r(\overline{F}_1(t), \ldots, \overline{F}_n(t)). \end{aligned}$$

Example 1: The Series System

$$r(\mathbf{p}) = \prod_{i=1}^{n} p_i,$$

so that

$$\overline{F}(t) = \prod_{i=1}^{n} \overline{F}_i(t).$$

Example 2: The Parallel System

$$r(\mathbf{p}) = 1 - \prod_{i=1}^{n} (1 - p_i),$$

so that

$$\begin{aligned} \overline{F}(t) &= 1 - \prod_{i=1}^{n} (1 - \overline{F}_i(t) \\ &= 1 - \prod_{i=1}^{n} F_i. \end{aligned}$$

Failure Rate

- For a continuous distribution F with density f, the *failure (or hazard) rate function of F*, $l(t)$, is given by

$$\lambda(t) = \frac{f(t)}{\overline{F}(t)}.$$

- If the lifetime of a component has distribution function F, then $l(t)$ is the conditional probability that the component of age t will fail.

Notes

- *F* is an *increasing failure rate* (*IFR*) distribution if *l*(*t*) is an increasing function of *t*.
  This is analogous to "wearing out".
- *F* is a *decreasing failure rate* (*DFR*) distribution if *l*(*t*) is a decreasing function of *t*.
  This is analogous to "burning in".

17.6 Distribution Functions for Modeling Component Lifetimes

- Exponential Distribution
- Weibull Distribution
- Gamma Distribution
- Log-Normal Distribution

The exponential distribution with parameters *l* > 0 has distribution function

$$G(t) = 1 - e^{-(\lambda t)}, \quad t \geq 0.$$

Its failure rate function is given by

$$\lambda(t) = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda.$$

It is considered both IFR and DFR.

The Weibull distribution with parameters λ > 0, α > 0 has distribution function

$$G(t) = 1 - e^{-(\lambda t)^{\alpha}}, \quad t \geq 0.$$

Its failure rate function is given by

$$\lambda(t) = \alpha\lambda(\lambda t)^{\alpha-1}$$

It is IFR if a ≥ 1 and DFR if 0 < a ≤ 1.

The gamma distribution with parameters λ > 0, α > 0 has density function

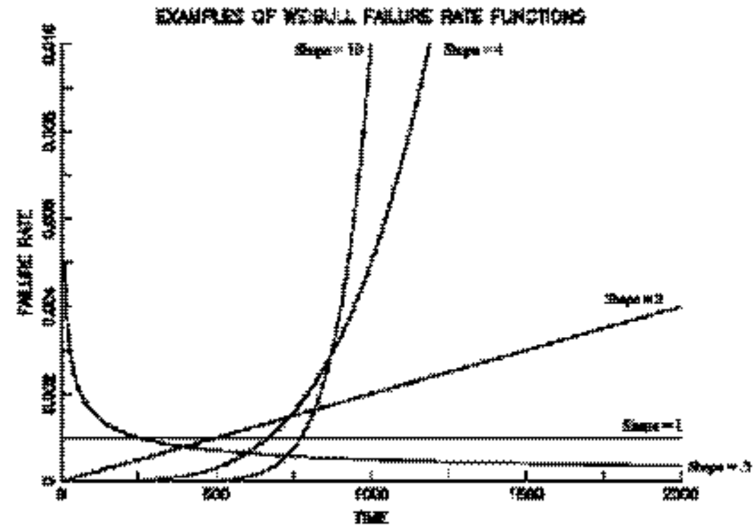
$$g(t) = \frac{\lambda e^{-\lambda t}(\lambda t)^{\alpha-1}}{\Gamma(\alpha)} = \frac{\lambda e^{-\lambda t}(\lambda t)^{\alpha-1}}{\int_0^{\infty} e^{-x}x^{\alpha-1}dx}, \quad t \geq 0.$$

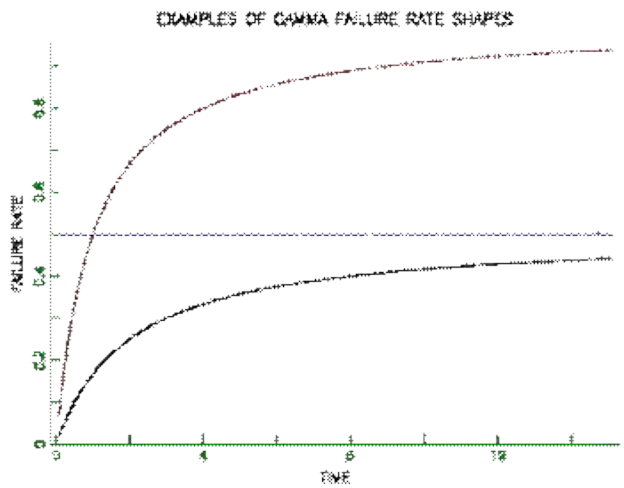
Its failure rate function is given by

$$\frac{1}{\lambda(t)} = \int_0^{\infty} e^{-\lambda x}\left(1 + \frac{x}{t}\right)^{\alpha-1} dx.$$

It is IFR if a ≥ 1 and DFR if 0 < a ≤ 1.

Notes





The log-normal distribution with parameters μ and σ > 0 has density function

$$g(t) = \frac{1}{\sqrt{2\pi}\,\sigma t} e^{-\frac{(\ln(t)-\mu)^2}{2\sigma^2}}, \quad t > 0.$$

Its failure rate function is given by

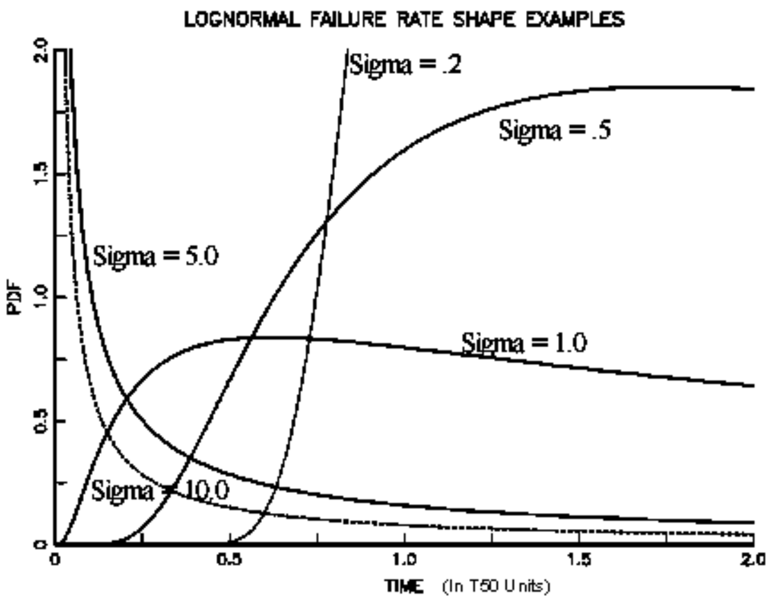
$$\lambda(t) = \frac{1}{\sigma t} Z\left(\frac{\log t - \mu}{\sigma}\right),$$

where *Z* is the standard normal hazard function, which is IFR.

The behavior of the failure rate function for the log-normal distribution depends on *s* :

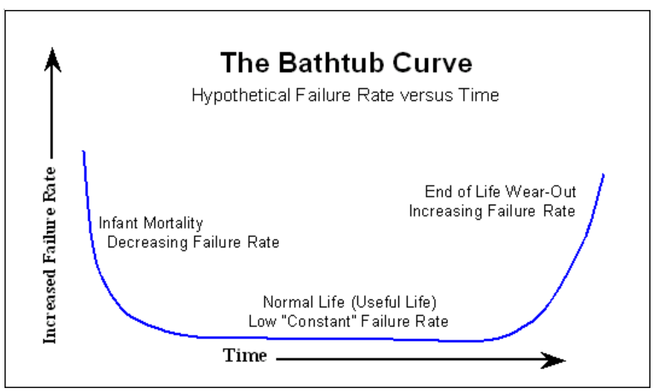
"For s ≈ 1.0, λ(t) is roughly constant. For σ ≤ 0.4, λ(t) increases…. For σ ≥ 1.5, l(t) decreases. This flexibility makes the lognormal distribution popular and suitable for many products."

Notes



from William Grant Ireson, Clyde F. Coombs, Richard Y., *Handbook of Reliability Engineering and Management*.

17.7 The Bathtub Curve and Failure Rate



Theorem 3. Consider a monotone system in which each component has the same IFR lifetime distribution. Define

$$r(p) = r(p, \ldots, p).$$

Then the distribution of system lifetime is IFR if

$$p\, r = (p)/r(p)$$

is a decreasing function of p.

Example 1: A k-out-of-n system with identical components is IFR if the individual components are IFR.

Notes

Example 2: A parallel system with two independent components with different exponential lifetime distributions is not IFR; in fact, $\lambda(t)$ is initially strictly increasing, and then strictly decreasing.

17.8 Expected System Life

Since system lifetime is non-negative, then

$$E[\text{system life}] = \int_0^\infty P\{\text{system life} > t\}dt$$
$$= \int_0^\infty r(\overline{\mathbf{F}}(t))dt,$$

where

$$\overline{\mathbf{F}}(t) = (\overline{F}_1(t), \ldots, \overline{F}_n(t)).$$

Example 3: *k*-out-of-*n* System. If each component has the same distribution function *G*, then

$$E[\text{system life}] = \int_0^\infty \sum_{i=k}^{n} \binom{n}{i} \left[\overline{G}(t)\right]^i \left[G(t)\right]^{n-i} dt.$$

Example 4: *k*-out-of-*n* System. Assume the expected component lifetime has mean θ. Uniformly distributed component lifetimes:

$$E[\text{system life}] = \int_0^{2\theta} \sum_{i=k}^{n} \binom{n}{i} \left[1 - \frac{t}{2\theta}\right]^i \left[\frac{t}{2\theta}\right]^{n-i} dt$$

$$= \frac{2\theta(n-k+1)}{n+1}.$$

Example 5: *k*-out-of-*n* System (uniformly distributed component lifetimes).

Parallel system (1-out-of-*n*):

$$E[\text{system life}] = \frac{2\theta n}{n+1}.$$

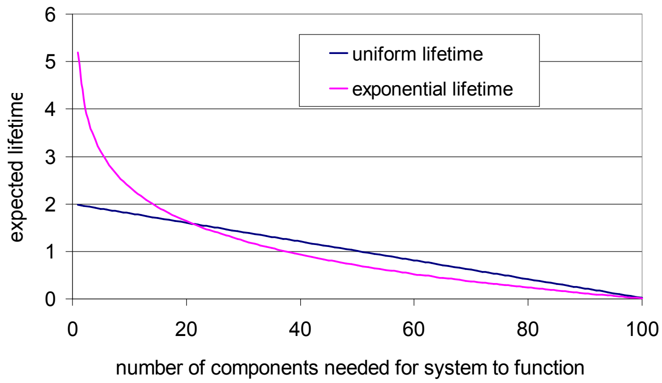
Serial system (*n*-out-of-*n*):

$$E[\text{system life}] = \frac{2\theta}{n+1}.$$

Example 6: *k*-out-of-*n* System (n = 100, θ = 1).

	Uniform	Exponential
k = 10	1.80	2.36
k = 50	1.01	0.71

Example 7: *k*-out-of-*n* System (n = 100, θ = 1):

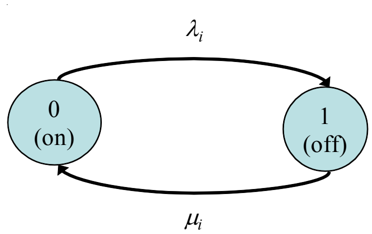


17.9 Systems with Repair

Consider a n-component system with reliability function r(p). Suppose that:

- each component i functions for an exponentially distributed time with rate λ_i and then fails;
- once failed, component i takes an exponential time with rate μ_i to be repaired;
- all components are functioning at time 0;
- all components act independently.

The state of component *i* (on or off) can be modeled as a two-state Markov process:



Let $A_i(t)$ be the availability of component *i* at time *t*, i.e., the probability that component *i* is functioning at time *t*. $A_i(t)$ is given by (see Ross example 6.11):

$$A_i(t) = P_{oo}(t) = \frac{\mu_i}{\mu_i + \lambda_i} + \frac{\lambda_i}{\mu_i + \lambda_i} e^{-(\lambda_i + \mu_i)t}.$$

Notes

The availability of system at time *t*, *A*(*t*), is given by

$$A(t) = r(A_i(t), \ldots, A_n(t))$$

$$= r\left(\frac{\mu}{\mu+\lambda} + \frac{\lambda}{\mu+\lambda} e^{-(\lambda+\mu)t}\right).$$

The limiting availability *A* is given by

$$A = \lim_{t\to\infty} A(t) = r\left(\frac{\mu}{\mu+\lambda}\right).$$

Example 1: The Series System

The availability of system at time *t*, *A*(*t*), is given by

$$A(t) = \prod_{i=1}^{n}\left[\frac{\mu_i}{\mu_i+\lambda_i} + \frac{\lambda_i}{\mu_i+\lambda_i} e^{-(\lambda_i+\mu_i)t}\right]$$

and

$$A = \prod_{i=1}^{n} \frac{\mu_i}{\mu_i+\lambda_i}.$$

Example 2: The Parallel System

The availability of system at time *t*, *A*(*t*), is given by

$$A(t) = 1 - \prod_{i=1}^{n}\left[\frac{\lambda_i}{\mu_i+\lambda_i}(1 - e^{-(\lambda_i+\mu_i)t})\right]$$

and

$$A = 1 - \prod_{i=1}^{n} \frac{\lambda_i}{\mu_i+\lambda_i}.$$

The average uptime *U* and downtime *D* are given respectively by

$$U = \frac{r\left(\frac{\mu}{\mu+\lambda}\right)}{\sum_{i=1}^{n} \frac{\lambda_i\mu_i}{\lambda_i+\mu_i}\left[r\left(1_i, \frac{\mu}{\mu+\lambda}\right) - r\left(0_i, \frac{\mu}{\mu+\lambda}\right)\right]}$$

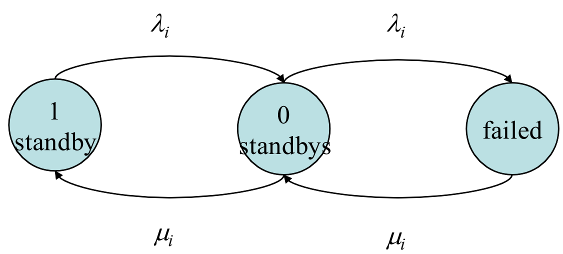
$$D = \frac{\left[1 - r\left(\frac{\mu}{\mu+\lambda}\right)\right]U}{r\left(\frac{\mu}{\mu+\lambda}\right)}.$$

17.9.1 System with Standby Components and Repair

Consider a *n*-component system with reliability function $r(\mathbf{p})$. Suppose that

- each component *i* functions for an exponentially distributed time with rate λ_i and then fails;
- once failed, component *i* takes an exponential time to be repaired;
- component *i* has a standby component that begins functioning if the primary component fails;
- if the standby component fails, it is also repaired;
- the repair rate is μ_i regardless of the number of failed type *i* components; the repair rate of type *i* components is independent of the number of other failed components;
- all components act independently.

The state of component *i* can be modeled as a three-state Markov process:



Note This is the same model we used for an M/M/1/2 queueing system. In equilibrium

$$P_{1\ \text{standby}} = \frac{(1-\lambda_i/\mu_i)}{1-(\lambda_i/\mu_i)^3}$$

$$P_{0\ \text{standbys}} = \frac{(\lambda_i/\mu_i)(1-\lambda_i/\mu_i)}{1-(\lambda_i/\mu_i)^3}$$

$$P_{\text{failed}} = \frac{(\lambda_i/\mu_i)^2(1-\lambda_i/\mu_i)}{1-(\lambda_i/\mu_i)^3}.$$

The equilibrium availability *A* of the system is given by

$$A = r\left(1 - \frac{(\lambda/\mu)^2(1-\lambda/\mu)}{1-(\lambda/\mu)^3}\right).$$

17.9.2 System with Interrelated Repair

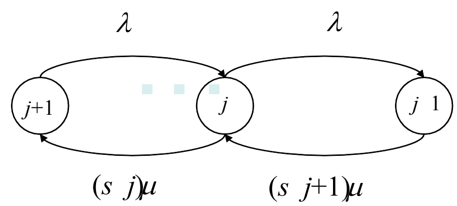
Consider an *s*-component parallel system with one repairman.

- All components have the same exponential lifetime and repair distributions.
- The repair rate is independent of the number of failed components.

Notes

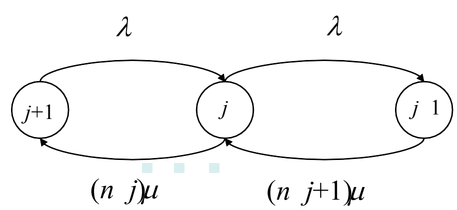
Let

- the state of the system be the number of failed components
- μ = the failure rate of each component
- λ = the repair rate



This looks like an M/M/s/s queue (Erlang Loss system).

$$A = 1 - P_s = 1 - \left(\sum_{j=0}^{s} \frac{1}{j!} \left(\frac{\lambda}{\mu} \right)^j \right)^{-1}.$$



17.10 Summary

- The ability of a system or component to perform its required functions under stated conditions for a specified period of time

  System reliability is a function of:

  - the reliability of the components
  - the interdependence of the components
  - the topology of the components

- Consider a system comprised on *n* components, where each component is either functioning or has failed. Define

$$x_i = \begin{cases} 1, & \text{if the i}^{\text{th}} \text{ component is functioning} \\ 0, & \text{if the i}^{\text{th}} \text{ component has failed} \end{cases}$$

  The vector $\boldsymbol{x} = \{x_1, \ldots, x_n\}$ is called the *state vector*.

- Assume that whether the system as a whole is functioning is completely determined by the state vector $\boldsymbol{x}$. Define

$$\phi(\boldsymbol{x}) = \begin{cases} 1, & \text{if the system is functioning when the state vector is } \boldsymbol{x} \\ 0, & \text{if the system has failed when the state vector is } \boldsymbol{x} \end{cases}$$

  The function f($\boldsymbol{x}$) is called the *structure function* of the system.

Notes

- A state vector $\boldsymbol{x}$ is call a *path vector* if $\phi(\boldsymbol{x}) = 1$.
- If $\phi(\boldsymbol{y}) = 0$ for all $\boldsymbol{y} < \boldsymbol{x}$, then $\boldsymbol{x}$ is a *minimal path vector*.
- If $\boldsymbol{x}$ is a minimal path vector, then the set $A = \{i : x_i = 1\}$ is a *minimal path set*.
- For a continuous distribution F with density f, the *failure (or hazard) rate function of F*, $l(t)$, is given by

$$\lambda(t) = \frac{f(t)}{\overline{F}(t)}.$$

- If the lifetime of a component has distribution function F, then $l(t)$ is the conditional probability that the component of age t will fail.
- The exponential distribution with parameters $l > 0$ has distribution function

$$G(t) = 1 - e^{-(\lambda t)}, \quad t \geq 0.$$

Its failure rate function is given by

$$\lambda(t) = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda.$$

It is considered both IFR and DFR.

The Weibull distribution with parameters $\lambda > 0$, $\alpha > 0$ has distribution function

$$G(t) = 1 - e^{-(\lambda t)^{\alpha}}, \quad t \geq 0.$$

Its failure rate function is given by

$$\lambda(t) = \alpha\lambda(\lambda t)^{\alpha-1}$$

It is IFR if $a \geq 1$ and DFR if $0 < a \leq 1$.

The gamma distribution with parameters $\lambda > 0$, $\alpha > 0$ has density function

$$g(t) = \frac{\lambda e^{-\lambda t}(\lambda t)^{\alpha-1}}{\Gamma(\alpha)} = \frac{\lambda e^{-\lambda t}(\lambda t)^{\alpha-1}}{\int_0^{\infty} e^{-x}x^{\alpha-1}dx}, \quad t \geq 0.$$

Its failure rate function is given by

$$\frac{1}{\lambda(t)} = \int_0^{\infty} e^{-\lambda x}\left(1 + \frac{x}{t}\right)^{\alpha-1} dx.$$

It is IFR if $a \geq 1$ and DFR if $0 < a \leq 1$.

17.11 Keywords

State vector: Consider a system comprised on n components, where each component is either functioning or has failed. Define

$$x_i = \begin{cases} 1, & \text{if the } i^{th} \text{ component is functioning} \\ 0, & \text{if the } i^{th} \text{ component has failed} \end{cases}$$

The vector $\boldsymbol{x} = \{x_1, \ldots, x_n\}$ is called the state vector.

Notes

Structure function: Assume that whether the system as a whole is functioning is completely determined by the state vector $\boldsymbol{x}$. Define

$$\phi(\boldsymbol{x}) = \begin{cases} 1, & \text{if the system is functioning when the state vector is } \boldsymbol{x} \\ 0, & \text{if the system has failed when the state vector is } \boldsymbol{x} \end{cases}$$

The function f(x) is called the structure function of the system.

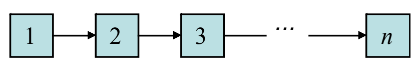
The Series System: There is only one minimal path set, namely the entire system.

The Parallel System: There are n minimal path sets, namely the sets consisting of one component.

The k-out-of-n System: There are $\binom{n}{k}$ minimal path sets, namely all of the sets consisting of exactly k components.

17.12 Self Assessment

1. A series system functions if and only if all of its n components are functioning:



Its structure function is given by ....

(a) k-out-of-n system (b) stated conditions

(c) $\phi(\boldsymbol{x}) = \prod_{i=1}^{n} x_i$ (d) partial order

2. A .... functions if and only if at least k of its n components are functioning: Its structure function is given by

$$\phi(\boldsymbol{x}) = \begin{cases} 1, & \text{if } \sum_{i=1}^{n} x_i \geq k \\ 0, & \text{if } \sum_{i=1}^{n} x_i < k. \end{cases}$$

(a) k-out-of-n system (b) stated conditions

(c) $\phi(\boldsymbol{x}) = \prod_{i=1}^{n} x_i$ (d) partial order

3. A .... is defined on the set of state vectors as follows. Let x and y be two state vectors. We define

$$\boldsymbol{x} \leq \boldsymbol{y} \text{ if } x_i \leq y_i,\ i = 1, \ldots, n.$$

(a) k-out-of-n system (b) stated conditions

(c) $\phi(\boldsymbol{x}) = \prod_{i=1}^{n} x_i$ (d) partial order

4. The ability of a system or component to perform its required functions under .... for a specified period of time.

   (a) k-out-of-n system (b) stated conditions

   (c) $\phi(\boldsymbol{x}) = \prod_{i=1}^{n} x_i$ (d) partial order

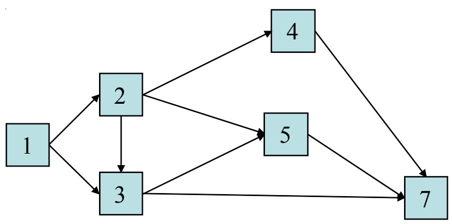
17.13 Review Questions

1. The Bridge Structure

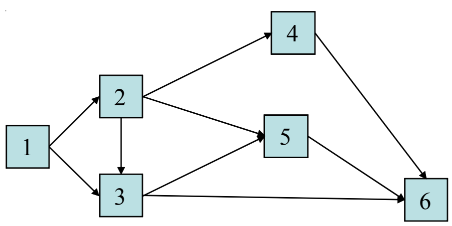
   The minimal cut sets are:

   {2}, {1, 4, 5}, {4, 6}, {2, 4, 5}.

2. Calculate a lower bound on the reliability of this system. Assume that all components have the same reliability p.



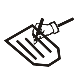
3. The minimal cut sets are {1}, {2, 3}, {3, 4, 5}, and {7}. Hence $r \geq 1 - 2(1 - p) - (1 - p)^2 - (1 - p)^3$.



Answers: Self Assessment

1. (a) 2. (a) 3. (d) 4. (d) 5. (b)

17.14 Further Readings



Books

Sheldon M. Ross, Introduction to Probability Models, Ninth Edition, Elsevier Inc., 2007.

Jan Pukite, Paul Pukite, Modeling for Reliability Analysis, IEEE Press on Engineering of Complex Computing Systems, 1998.

Notes

Unit 18: The Weak Law

CONTENTS

Objectives

Introduction

18.1 Summary

18.2 Keywords

18.3 Self Assessment

18.4 Review Questions

18.5 Further Readings

Objectives

After studying this unit, you will be able to:

- Discuss the weak laws
- Describe some examples related to weak law

Introduction

James Bernoulli proved the weak law of large numbers (WLLN)around 1700 which was published posthumously in 1713 in his treatise Ars Conjectandi. Poisson generalized Bernoulli's theorem around 1800, and in 1866 Tchebychev discovered the method bearinghis name. Later on one of his students, Markov observed that Tchebychev's reasoning can be used to extend Bernoulli's theoremto dependent random variables as well.

In 1909 the French mathematician Emile Borel proved adeeper theorem known as the strong law of large numbers that furthergeneralizes Bernoulli's theorem. In 1926 Kolmogorov derived conditions that were necessary and sufficient for a set of mutually independent random variables to obey the law of large numbers.

18.1 Weak Law of Number

Let X_i be independent, identically distributed Bernoulli randomVariables such that

$$P(X_i) = p, \qquad P(X_i = 0) = 1 - p = q,$$

and let $k = X_1 + X_2 + ... + X_n$ represent the number of "successes"in n trials. Then the weak law due to Bernoulli states that [see Theorem 3-1, page 58, Text]

$$P\left\{\left|\frac{k}{h} - p\right| > \varepsilon\right\} \le \frac{pq}{n\varepsilon^2} \qquad ...(18.1)$$

i.e., the ratio "total number of successes to the total numberof trials" tends to p in probability as nincreases.

Notes

A stronger version of this result due to Borel and Cantellistates that the above ratio k/n tends to p not only in probability, but with probability 1. This is the strong law of large numbers (SLLN).

What is the difference between the weak law and the strong law? The strong law of large numbers states that if $\{\varepsilon_n\}$ is a sequence of positive numbers converging to zero, then

$$\sum_{n=1}^{\infty} P\left\{\left|\frac{k}{h} - p\right| \geq \varepsilon_n\right\} < \infty \qquad ...(18.2)$$

From Borel-Cantelli lemma [see (2-69) Text], when (13-2) is satisfied the events $An \stackrel{\Delta}{=} \left\{\left|\frac{k}{h} - p\right| \geq \varepsilon_n\right\}$ can occur only for a finitenumber of indices n in an infinite sequence, or equivalently, the events $\left\{\left|\frac{k}{h} - p\right| \geq \varepsilon_n\right\}$ occur infinitely often, i.e., the event k/nconverges to palmost-surely.

Proof: To prove (18.2), we proceed as follows. Since

$$\left|\frac{k}{h} - p\right| \geq \varepsilon \Rightarrow |k - np|^4 \geq \varepsilon^4 n^4$$

we have

$$\sum_{k=0}^{n} (k - np)^4 p_n(k) \geq \varepsilon^4 n^4 = \varepsilon^4 n^4 \left(P\left\{\left|\frac{k}{n} - p\right| \geq \varepsilon\right\} + P\left\{\left|\frac{k}{n} - p\right| < \varepsilon\right\}\right)$$

and hence

$$P\left\{\left|\frac{k}{n} - p\right| \geq \varepsilon\right\} \leq \frac{\sum_{k=0}^{n} (k - np)^4 p_n(k)}{\varepsilon^4 n^4} \qquad ...(13.3)$$

where

$$p_n(k) = P\left\{\sum_{i=1}^{n} X_i = k\right\} = \binom{n}{k} p^k q^{n-k}$$

By direct computation

$$\sum_{k=0}^{n} (k - np)4pn(k) = E\left\{\left(\sum_{i=1}^{n} X_i - np\right)^4\right\} = E\left\{\left(\sum_{i=1}^{n} X_i - p\right)^4\right\}$$

$$= E\left\{\left(\sum_{i=1}^{n} Y_i\right)^4\right\} = \sum_{i=1}^{n}\sum_{k=1}^{n}\sum_{j=1}^{n}\sum_{l=1}^{n} E(Y_i Y_k Y_j Y_l)$$

$$= \sum_{i=1}^{n} E(Y_i^4) + 4n(n-1)\sum_{i=1}^{n}\sum_{j=1}^{n} E(Y_i^3)E(X_j) + 3n(n-1)\sum_{i=1}^{n}\sum_{j=1}^{n} E(Y_i^2)E(Y_j^2)$$

$$= n(p^3 + q^3)pq + 3n(n-1)(pq)^2 \leq [n + 3n(n-1)]pq$$

$$= 3n^2 pq, \qquad ...(18.4)$$

Notes

since

$$p^3 + q^3 = (p + q)^3 - 3p^2q - 3pq^2 < 1,\ pq \le 1/2 < 1$$

Substituting (18.4) also (18.3) we obtain

$$P\left\{\left|\frac{k}{n} - p\right| \ge \varepsilon\right\} \le \frac{3pq}{n^2\varepsilon^4}$$

Let $\varepsilon = \frac{1}{n^{1/8}}$ so that the above integral reads and hence

$$\sum_{n=1}^{\infty} P\left\{\left|\frac{k}{n} - p\right| \ge \frac{1}{n^{1/8}}\right\} \le 3pq \sum_{n=1}^{\infty} \frac{1}{n^{3/2}} \le 3pq(1 + \int_1^{\infty} x^{-3/2} dx)$$

$$= 3pq(1 + 2) = 9pq < \infty, \qquad ...(18.5)$$

thus proving the strong law by exhibiting a sequence of positive numbers $\varepsilon_n = 1/n^{1/8}$ that converges to zero and satisfies (13-2).

We return back to the same question: "What is the difference between the weak law and the strong law?.

"The weak law states that for every n that is large enough, the ratio $\left(\sum_{i=1}^{n} X_i\right)/n = k/n$ is likely to be near p with certain probability that tends to 1 as nincreases. However, it does not say that k/n is boundto stay near p if the number of trials is increased. Suppose (18.1) is satisfied for a given ε in a certain number of trials n_0. If additional trials are conducted beyond n_0, the weak law does not guarantee that the new k/n is bound to stay near p for such trials. In fact there can be events for which $k/n > p + e$, for $n > n_0$ in some regular manner. The probability for such an event is the sum of a large number of very small probabilities, and the weak law is unable to say anything specific about the convergence of that sum.

However, the strong law states (through (18.2)) that not onlyall such sums converge, but the total number of all such events where $k/n > p + \varepsilon$ is in fact finite! This implies that the probability $\left\{\left|\frac{k}{n} - p\right| > \varepsilon\right\}$ of the events as n increases becomes and remains small, since with probability 1 only finitely many violations to the above inequality takes place as $n \to \infty$.

Interestingly, if it possible to arrive at the same conclusion using a powerful bound known as Bernstein's inequality that is based on the WLLN.

Bernstein's inequality : Note that

$$\left|\frac{k}{n} - p\right| > \varepsilon \Rightarrow k > n(p + \varepsilon)$$

and for any $\lambda > 0$, this gives $e^{\lambda(k - n(p+\varepsilon))} > 1$.

Thus

$$P\left\{\frac{k}{n} - p > \varepsilon\right\} = \sum_{k=[n(p+\varepsilon)]}^{n} \binom{n}{k} p^k q^{n-k}$$

$$\le \sum_{k=[n(p+\varepsilon)]}^{n} e^{\lambda(k-n(p+\varepsilon))} \binom{n}{k} p^k q^{n-k}$$

$$\le \sum_{k=0}^{n} e^{\lambda(k-n(p+\varepsilon))}\binom{n}{k}p^k q^{n-k}$$

$$P\left\{\frac{k}{n}-p>\varepsilon\right\}=e^{-\lambda n\varepsilon}\sum_{k=0}^{n}\binom{n}{k}(pe^{\lambda q})^k(qe^{-\lambda p})^{n-k}$$

$$= e^{-\lambda n\varepsilon}\left(pe^{\lambda q}+qe^{-\lambda p}\right)^n \qquad ...(18.6)$$

Since $e^x \le x + e^{x^2}$ for any real x,

$$pe^{\lambda q} + qe^{-\lambda p} \le p(\lambda q + e^{\lambda^2 q^2}) + q(-\lambda p + e^{\lambda^2 p^2})$$

$$= pe^{\lambda^2 q^2} + qe^{\lambda^2 p^2} \le e^{\lambda^2}. \qquad ...(18.7)$$

Substituting (18.7) into (18.6), we get

$$P\left\{\frac{k}{n}-p>\varepsilon\right\}\le e^{\lambda^2 n-\lambda n\varepsilon}.$$

But $\lambda^2 n - \lambda n\varepsilon$ is minimum for $\lambda = \varepsilon/2$ and hence

$$P\left\{\frac{k}{n}-p>\varepsilon\right\}\le e^{-n\varepsilon^2/4},\ \varepsilon>0. \qquad ...(18.8)$$

Similarly

$$P\left\{\frac{k}{n}-p<-\varepsilon\right\}£\ e^{-n\varepsilon 2/4}$$

and hence we obtain Bernstein's inequality

$$P\left\{\left|\frac{k}{n}-p\right|>\varepsilon\right\}£\ 2e^{-n\varepsilon^2/4}. \qquad ...(18.9)$$

Bernstein's inequality is more powerful than Tchebyshev's inequalityas it states that the chances for the relative frequency k /n exceeding its probability p tends to zero exponentially fast as $n \to \infty$.

Chebyshev's inequality gives the probability of k /nto lie between and for a specific n. We can use Bernstein's inequality to estimate the probability for k /nto lie between and for all large n

Towards this, let

$$y_n=\left\{p-\varepsilon\le\frac{k}{n}<p+\varepsilon\right\}$$

so that

$$P(y_n^c)=P\left\{\left|\frac{n}{k}-p\right|>\varepsilon\right\}\le 2e^{-n\varepsilon 2/4}$$

Notes

To compute the probability of the event $\bigcap_{n=m}^{\infty} y_n$, note that its complement is given by

$$\left(\bigcap_{n=m}^{\infty} y_n\right)^c = \bigcup_{n=m}^{\infty} y_n^c$$

and using Eq. (2-68) Text,

$$P\left(\bigcup_{n=m}^{\infty} y_n^c\right) \le \sum_{n=m}^{\infty} P(y_n^c) \le \sum_{n=m}^{\infty} 2e^{-n\varepsilon^2/4} = \frac{2e^{-m\varepsilon^2/4}}{1-e^{-\varepsilon^2/4}}.$$

This gives

$$P\left(\bigcup_{n=m}^{\infty} y_n\right) = \left\{1 - P\left(\bigcup_{n=m}^{\infty} \overline{y}_n\right)\right\} \ge 1 - \frac{2e^{-m\varepsilon^2/4}}{1-e^{-\varepsilon^2/4}} \to 1 \text{ as } m \to \infty$$

or,

$$P\left\{p - \varepsilon \le \frac{k}{n} \le p + \varepsilon, \text{ for all } n \ge m\right\} \to 1 \text{ as } m \to \infty.$$

Thus k/n is bound to stay near p for all large enough n, in probability,a conclusion already reached by the SLLN.

Discussion: Let Thus if we toss a fair coin 1,000 times, from the weak law

$$P\left\{\left|\frac{k}{n} - \frac{1}{2}\right| \ge 0.01\right\} \le \frac{1}{40}.$$

Thus on the average 39 out of 40 such events each with 1000 or more trials will satisfy the inequality $\left\{\left|\frac{k}{n} - \frac{1}{2}\right| \le 0.1\right\}$ or, it is quite possible that one out of 40 such events may not satisfy it.

As a result if we continue the coin tossing experiment for an additional 1000 moretrials, with k representing the total number of successes up to the current trial n, for n = 1000 $\to$ 2000, it is quite possible that for few such n the above inequality may be violated. This is still consistent with the weak law, but "not so often" says the strong law. According to the strong law such violations can occur only a finite number of times each with a finite probability in an infinite sequence of trials, and hence almost always the above inequality will be satisfied, i.e., the sample space of k/n coincides with that of p as $n \to \infty$.

Next we look at an experiment to confirm the strong law:

Example: 2n red cards and 2n black cards (all distinct) are shuffled together to form a single deck, and then split into half. What is the probability that each half will contain n red and n black cards?

Notes

Solution: From a deck of 4n cards, 2n cards can be chosen $\binom{4n}{2n}$ in different ways. To determine the number of favorable draws of n red and n black cards in each half, consider the unique draw consisting of 2n red cards and 2n black cards in each half. Among those 2n red cards, n of them can be chosen in $\binom{2n}{n}$ different ways; similarly for each such draw there are $\binom{2n}{n}$ ways of choosing n black cards. Thus the total number of favorable draws containing n red and n black cards in each half are $\binom{2n}{n}\binom{2n}{n}$ among a total of $\binom{4n}{2n}$ draws. This gives the desired probability p_n to be

$$p_n \simeq \frac{\binom{2n}{n}\binom{2n}{n}}{\binom{4n}{2n}} = \frac{(2n!)^4}{(4n)!(n!)^4}.$$

For large n, using Stingling's formula we get

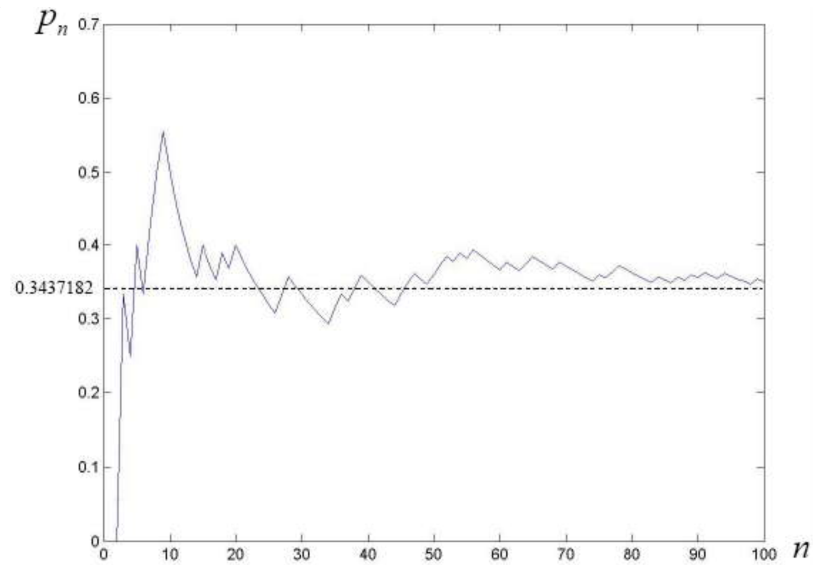
Table 18.1

Expt	Number of successes	Expt	Number of successes	Expt	Number of successes	Expt	Number of successes	Expt	Number of successes
1	0	21	8	41	14	61	23	81	29
2	0	22	8	42	14	62	23	82	29
3	1	23	8	43	14	63	23	83	30
4	1	24	8	44	14	64	24	84	30
5	2	25	8	45	15	65	25	85	30
6	2	26	8	46	16	66	25	86	31
7	3	27	9	47	17	67	25	87	31
8	4	28	10	48	17	68	25	88	32
9	5	29	10	49	17	69	26	89	32
10	5	30	10	50	18	70	26	90	32
11	5	31	10	51	19	71	26	91	33
12	5	32	10	52	20	72	26	92	33
13	5	33	10	53	20	73	26	93	33
14	5	34	10	54	21	74	26	94	34
15	6	35	11	55	21	75	27	95	34
16	6	36	12	56	22	76	27	96	34
17	6	37	12	57	22	77	28	97	34
18	7	38	13	58	22	78	29	98	34
19	7	39	14	59	22	79	29	99	34
20	8	40	14	60	22	80	29	100	35

The figure below shows results of an experiment of 100 trials.

Notes

Figure 18.1



18.1 Summary

Let X_i be independent, identically distributed Bernoulli randomVariables such that

$$P(X_i) = p, \qquad P(X_i = 0) = 1 - p = q,$$

and let $k = X_1 + X_2 + ... + X_n$ represent the number of "successes"in n trials. Then the weak law due to Bernoulli states that [see Theorem 3-1, page 58, Text]

$$P\left\{\left|\frac{k}{h} - p\right| > \varepsilon\right\} \leq \frac{pq}{n\varepsilon^2} \qquad ...(18.1)$$

i.e., the ratio "total number of successes to the total numberof trials" tends to p in probability as nincreases.

A stronger version of this result due to Borel and Cantellistates that the above ratio k/n tends to p not only in probability, but with probability 1. This is the strong law of large numbers (SLLN).

18.2 Keywords

Strong law of large numbers: A stronger version of this result due to Borel and Cantellistates that the above ratio k/n tends to p not only in probability, but with probability 1. This is the strong law of large numbers (SLLN).

Bernstein's inequality is more powerful than Tchebyshev's inequalityas it states that the chances for the relative frequency k /n exceeding its probability p tends to zero exponentially fast as $n \rightarrow \infty$.

18.3 Self Assessment

Notes

1. .... generalized Bernoulli's theorem around 1800, and in 1866 Tchebychev discovered the method bearinghis name.
2. In .... the French mathematician Emile Borel proved adeeper theorem known as the strong law of large numbers that further generalizes Bernoulli's theorem.
3. In .... Kolmogorov derived conditions that were necessary and sufficient for a set of mutually independent random variables to obey the law of large numbers.
4. A .... of this result due to Borel and Cantellistates that the above ratio k/n tends to p not only in probability, but with probability 1. This is the strong law of large numbers (SLLN).
5. The strong law of large numbers states that if $\{e_n\}$ is a sequence of .... to zero, then

$$\sum_{n=1}^{\infty} P\left\{\left|\frac{k}{h}-p\right| \geq \varepsilon_n\right\} < \infty$$

18.4 Review Questions

1. 2n red cards and 2n black cards (all distinct) are shuffled together to form a single deck, and then split into half. What is the probability that each half will contain n red and n black cards?
2. 3n red cards and n black cards (all distinct) are shuffled together to form a single deck, and then split into half. What is the probability that each half will contain n red and n black cards?
3. 4n red cards and 4n black cards (all distinct) are shuffled together to form a single deck, and then split into half. What is the probability that each half will contain n red and n black cards?
4. n red cards and 2n black cards (all distinct) are shuffled together to form a single deck, and then split into half. What is the probability that each half will contain n red and n black cards?

Answers: Self Assessment

1. Poisson 2. 1909 3. 1926 4. stronger version

5. positive numbers converging

18.5 Further Readings

Books

Sheldon M. Ross, Introduction to Probability Models, Ninth Edition, Elsevier Inc., 2007.

Jan Pukite, Paul Pukite, Modeling for Reliability Analysis, IEEE Press on Engineering of Complex Computing Systems, 1998.

Notes

Unit 19: The Laws of Large Numbers Compared

CONTENTS

Objectives

Introduction

19.1 Strong Law of Large Numbers

19.2 Summary

19.3 Keywords

19.4 Self Assessment

19.5 Review Questions

19.6 Further Readings

Objectives

After studying this unit, you will be able to:

- Discuss the strong law of large number
- Discuss examples related to large number

Introduction

Probability Theory includes various theorems known as Laws of Large Numbers; for instance, see [Fel68, Hea71, Ros89]. Usually two major categories are distinguished: Weak Laws versus Strong Laws. Within these categories there are numerous subtle variants of differing generally. Also the Central Limit Theorems are often brought up in this context.

Many introductory probability texts treat this topic superficially, and more than once their vague formulations are misleading or plainly wrong. In this note, we consider a special case to clarify the relationship between the Weak and Strong Laws. The reason for doing so is that I have not been able to find a concise formal exposition all in one place. The material presented here is certainly not new and was gleaned from many sources.

In the following sections, X1, X2, ... is a sequence of independent and indentically distributed random variabels with finite expectation μ. We define the associated sequence $\overline{X}_i$ of partial sample means by

$$\overline{X}_i = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

The Laws of Large Numbers make statements about the convergence of $\overline{X}_n$ to m. Both laws relate bounds on sample size, accuracy of approximation, and degree of confidence. The Weak Laws deal with limits of probabilities involving $\overline{X}_n$. The Strong Laws deal with probabilities involving limits of $\overline{X}_n$. Especially the mathematical underpinning of the Strong Laws requires a caretful approach ([Hea71, Ch. 5] is an accesible presentation).

19.1 Strong Law of Large Numbers

Notes

We are now ready to give Etemadi's proof of

(7.1) Strong law of large numbers. Let X1, X2, ... be pairwise independent identically distributed random variables with $E\,|\,X_i\,| < \infty$. Let $EX_i = \mu$ and $S_n = X_1 + ... + X_n$. Then $S_n/n \to \mu$ a.s. as $n \to \infty$.

Proof : As in the proof of weak law of large numbers, we begin by truncating.

(a) **Lemma.** Let $Y_k = K_k 1_{(|X_k| \le k)}$ and $T_n = Y_1 + ... + Y_n$. It is sufficient to prove that $T_n/n \to \mu$ a.s.

Proof $\sum_{k=1}^{\infty} P(|X_k| > k) \le \int_0^{\infty} P(|X_1| > t)dt = E|X_1| < \infty$ so $P(X_k \ne Yk \text{ i.o.}) = 0$. This shows that $|Sn(w) - Tn(w) < \infty$ a.s. for all n, from which the desired result follows.

The second step is not so intuitive but it is an important part of this proof and the one given in Section 1.8.

(b) **Lemma.** $\sum_{k=1}^{\infty} \text{var}(Y_k)/k^2 \le 4E|X_1| < \infty$.

Proof To bound the sum, we observe

$$\text{var}(Y_k) \le E(Y_k^2) = \int_0^{\infty} 2yP(|Y_k| > y)dy \le \int_0^{k} 2yP(|X_1| > y)dy$$

so using Fubini's theorem (since everything is ≤ 0 and the sum is just an integral with respect to counting measure on {1, 2, ...})

$$\sum_{k=1}^{\infty} E(Y_k^2)/k^2 \le \sum_{k=1}^{\infty} k^{-2} \int_0^{\infty} 1_{(y<k)} 2yP(|X_1| > y)dy$$

$$= \int_0^{\infty} \left\{ \sum_{k=1}^{\infty} k^{-2} 1_{(y<k)} \right\} 2yP(|X_1| > y)dy$$

Since $E|X_1| = \int_0^{\infty} P(|X_1| > y)dy$, we can complete the proof by showing

(c) Lemma. If $y \ge 0$ then $2y \sum_{k>y} k^{-2} \ge 4$.

Proof We being with the observation that if $m \ge 2$ then

$$\sum_{k \ge m} k^{-2} \le \int_{m-1}^{\infty} x^{-2} dx = (m-1)^{-1}$$

When $y \ge 1$ the sum starts with $k = [y] + 1 \ge 2$ so

$$2y \sum_{k \ge m} k^{-2} \le 2y/[y] \le 4$$

since $y/[y] \le 2$ for $y \ge 1$ (the worst case being y close to 2). To cover $0 \le y < 1$ we note that in this case

$$2y \sum_{k \ge y} k^{-2} \le 2y \left(1 + \sum_{k=2}^{\infty} k^{-2} \right) \le 4$$

The first two steps, (a) and (b) above, are standard. Etemadi's inspiration was that since $X_n^+, n \ge 1$, and $X_n^-, n \ge 1$, satisfy the assumptions of the theorem of $X_n = X_n^+ - X_n^-$, we can without loss of generality suppose $X_n \ge 0$, As in proof of (6.8) we will prove the result first for a subsequence

Notes

and then use monotonicity to control the values in between. This time however, we let $\alpha > 1$, and $k(n) = [a^n]$. Chebyshev's inequality implies that if $\in > 0$

$$\sum_{n=1}^{\infty} P(|T_{k(n)} - ET_{k(n)}| > \in k(n)) \le \in -2 \sum_{n=1}^{\infty} var(T_{k(n)})/k(n)^2$$

$$= \in^{-2} \sum_{n=1}^{\infty} k(n)^{-2} \sum_{m=1}^{k(n)} var(Y_m)$$

$$= \in^{-2} \sum_{m=1}^{\infty} var(Y_m) \sum_{n:k(n)\ge m} k(n)^{-2}$$

where we have used Fubini's theorem to interchange the two summations (everything is ≥ 0). Now $k(n) = [\alpha^n]$ and $[\alpha^n] \ge \alpha^n/2$ for $n \ge 1$, so summing the geometric series and noting that the first term is $\le m^{-2}$

$$\sum_{n:\alpha^n \ge m} [\alpha^n]^{-2} \ge 4 \sum_{n:\alpha^n \ge m} \alpha^{-2n} \le 4(1-\alpha^{-2})^{-1} m^{-2}$$

Combining our computations shows

$$\sum_{n=1}^{\infty} P(|T_{k(n)} - ET_{k(n)}| > \in k(n)) \le 4(1-\alpha^{-2})^{-1} \in^{-2} \sum_{m=1}^{\infty} E(Y_m^2) m^{-2} < \infty$$

by (b). Since $\in$ is arbitrary $(T_{k(n)} - ET_{k(n)})/k(n) \to 0$. The dominated convergence theorem implies $EY_k \to EX_1$ as $k \to \infty$, so $ET_{k(n)}/k(n) \to Ex_1$ and we have shown $T_{k(n)}/k(n) \to EX_1$ a.s. To handle the intermediate values, we observe that if $k(n) \le m < k(n + 1)$

$$\frac{T_{k(n)}}{k(n+1)} \le \frac{T_m}{m} \le \frac{T_{k(n+1)}}{k(n)}$$

(here we use $Y_i \ge 0$), so recalling $k(n) = [\alpha^n]$ we have $k(n + 1)/k(n) \to \alpha$ and

$$\frac{1}{\alpha} EX_1 \le \lim_{n\to\infty} \inf Tm/m \le \lim_{m\to\infty} \sup T_m/m \le \alpha EX_1$$

Since $\alpha > 1$ is arbitrary the proof is complete.

The next result shows that the strong law holds whenever EX_i exists.

(7.2) Theorem. Let $X_1, X_2, \ldots$ be i.i.d. with $EX_i^+ = \infty$ and $EX_i^- < \infty$. If $S_n = X_1 + \ldots + X_n$ then $Sn/n \to \infty$ a.s.

Proof Let $M > 0$ and $X_i^M = Xi \wedge M$. The X_i^M are i.i.d with $E|X_i^M| < \infty$ so if $S_i^M = X_i^M + \ldots + X_n^M$ then (7.1) implies $S_n^M/n \to EX_i^M$. Since $X_i \ge X_i^M$ it follows that

$$\lim_{n\to\infty} \inf S_n/n \ge \lim_{n\to\infty} S_n^M/n = EX_i^M$$

The monotone convergence theorem implies $E(X_i^M)+ \uparrow EX_i^+ = \infty$ as $M\uparrow\infty$, so $EX_i^+ = E(X_i^M)^+ - E(X_i^M)^- \uparrow \infty$ and we have $\lim \inf_{n\to\infty} S_n/n \ge \infty$ which imlies the desired result.

The rest of this section is devoted to applications of the strong law of large numbers.

Notes

Example: Renewal theory. Let $X_1, X_2, \ldots$ be i.i.d. with $0 < X_i < \infty$. Let $T_n = X_1 + \ldots + X_n$ and think of T_n as the time of nth occurence of some event. For a concrete situation consider a diligent janitor who replaces a light bulb the instant it burns out. Suppose the first bulb is put in at time 0 and let X_i be the lifetime of the ith lightbulb. In this interpretation T_n is the time the nth light bulb burns out and $N_t = \sup\{n : T_n \le t\}$ is the number of light bulbs that have burns out by time t.

Theorem. If $EX_1 = \mu \le \infty$ then as $t \to \infty$, $N_t/t \to 1/\mu$ a.s. $(1/\infty = 0)$

19.2 Summary

- Many introductory probability texts treat this topic superficially, and more than once their vague formulations are misleading or plainly wrong. In this note, we consider a special case to clarify the relationship between the Weak and Strong Laws. The reason for doing so is that I have not been able to find a concise formal exposition all in one place. The material presented here is certainly not new and was gleaned from many sources.

  In the following sections, X1, X2, ... is a sequence of independent and indentically distributed random variabels with finite expectation m. We define the associated sequence $\overline{X}_i$ of partial sample means by

$$\overline{X}_i = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

- Lemma. Let $Y_k = K_k 1_{(|X_k| \le k)}$ and $T_n = Y_1 + \ldots + Y_n$. It is sufficient to prove that $T_n/n \to \mu$ a.s.

- Lemma. $\sum_{k=1}^{\infty} \mathrm{var}(Y_k)/k^2 \le 4E|X_1| < \infty$.

- Lemma. If $y \ge 0$ then $2y\sum_{k>y} k^{-2} \ge 4$.

- Implies $S_n^M/n \to EX_i^M$. Since $X_i \ge X_i^M$ it follows that

$$\lim_{n\to\infty} \inf S_n/n \ge \lim_{n\to\infty} S_n^M/n = EX_i^M$$

  The monotone convergence theorem implies $E(X_i^M)+ \uparrow EX_i^+ = \infty$ as $M \uparrow \infty$, so $EX_i^+ = E(X_i^M)^+ - E(X_i^M)^- \uparrow \infty$ and we have $\lim\inf_{n\to\infty} S_n/n \ge \infty$ which imlies the desired result.

19.3 Keywords

Probability Theory includes various theorems known as Laws of Large Numbers.

Strong law of large numbers. Let X1, X2, ... be pairwise independent identically distributed random variables with $E|X_i| < \infty$. Let $EX_i = \mu$ and $S_n = X_1 + \ldots + X_n$. Then $S_n/n \to \mu$ a.s. as $n \to \infty$.

19.4 Self Assessment

1. .... includes various theorems known as Laws of Large Numbers.
2. The Laws of Large Numbers make statements about the convergence of .... to m.
3. Lemma. $\sum_{k=1}^{\infty} \mathrm{var}(Y_k)/k^2 \le$ ....
4. Lemma. If $y \ge 0$ then ....

Notes

5. The .... $E(X_i^M)+\uparrow EX_i^+ = \infty$ as $M\uparrow\infty$, so $EX_i^+ = E(X_i^M)^+ - E(X_i^M)^- \uparrow \infty$ and we have lim $\inf_{n\to\infty} S_n/n \geq \infty$ which imlies the desired result.

6. If $EX_1 = \mu \leq \infty$ then as $t \to \infty$, $N_t/t \to 1/\mu$ a.s. ....

19.5 Review Questions

1. Discuss the strong law of large number.

2. Discuss examples related to large number.

Answers: Self Assessment

1. Probability Theory 2. $\overline{X}_n$ 3. $4E|X_1| < \infty$ 4. $2y\sum_{k>y} k^{-2} \geq 4$

5. monotone convergence theorem implies 6. $(1/\infty = 0)$

19.6 Further Readings

Books

Sheldon M. Ross, Introduction to Probability Models, Ninth Edition, Elsevier Inc., 2007.

Jan Pukite, Paul Pukite, Modeling for Reliability Analysis, IEEE Press on Engineering of Complex Computing Systems, 1998.

Unit 20: Control Limit Theorem

Notes

CONTENTS

Objectives

Introduction

20.1 Central Limit Theorem

20.2 Summary

20.3 Keywords

20.4 Self Assessment

20.5 Review Questions

20.6 Further Readings

Objectives

After studying this unit, you will be able to:

- Define the central limit theorem
- Describe control limit theorem

Introduction

In Binomial distribution with parameters n and p is shown to be approximable by a Poisson distribution whenever n is large and p is such that np is a constant A z 0. An important limit theorem, known as the central limit theorem, is studied in Section 14.4. Central limit theorem essentially states that whatever the original distribution is (as long as it has finite variance), the sample mean computed from the observations following that distribution has an approximate normal distribution as long as the sample size (number of observations) is large. An important special case of this result is that binomial distribution can be approximated by an appropriate normal distribution for large samples.

20.1 Central Limit Theorem

The Central Limit Theorem (CLT) is one of the most important and useful results in probability theory. We have already seen that the sum of a finite number of independent normal random variables is normally distributed. However the sum of a finite number of independent non-normal random variables need not be normally distributed. Even then, according to the central limit theorem, the sum of a large number of independent random variables has a distribution that is approximately normal under general conditions. The CLT provides a simple method of computing the probabilities for the sum of independent random variables approximately. This theorem also suggests the reasoning behind why most of the data observed in practice leads to bell-shaped curves.

Notes

Let us now state the main theorem.

Theorem 3 (Central Limit Theorem) : Let X_1, X_2, ... be an infinite sequence of independent and identically distributed random variables with mean p and finite variance σ^2. Then, for any real x,

$$P\left[\frac{X_1 + ... + X_n - n\mu}{\sigma\sqrt{n}}\right] \to \phi(x) \text{ as } n \to \infty \qquad ...(5)$$

where $\phi(x)$ is the standard normal distribution function.

We have omitted the proof because proof of this result involves complex analysis and other concepts which are beyond the scope of this course. Let us try to understand the above statement more clearly. Let $S_n = X_1 + X_2 + ... + X_n$. Then we know that $P\left[\frac{S_n - n\mu}{\sigma\sqrt{n}} \le x\right]$ represents the distribution of the random variable $\frac{S_n - n\mu}{\sigma\sqrt{n}}$. Then the theorem says that the distribution of $\frac{S_n - n\mu}{\sigma\sqrt{n}}$ is approximately a standard normal distribution for sufficiently large n. Therefore the distribution of S_n will be approximately normal with mean $n\pi$ and variance $n\sigma^2$. In other words the theorem asserts that if $X_1 + X_2 ... + X_n$ are i.i.d.r. v' s of any kind (discrete or continuous) with finite variances, the $\Sigma\, S_n = X_1 + X_2 + ... + X_n$ will approximately be a normal distribution for sufficiently large n. The importance of the theorem lies in this fact. This theorem has got many applications. An important application is to a sequence of Bernoulli random variables.

Normal Approximation to the Binomial Distribution

Let X_i, $i \ge 1$ be a sequence of i.i.d. random variables such that

$$P[X_i = 1] = p, P[X_i = 0] = 1 - p$$

where 0 < p < l.

Observe that $S_n = X_1 + ... + X_n$ has the binomial distribution with parameters n qnd p. You can check that $E(X_i) = p$ and Var $(X_i) = p$ (1-p) for any i which is finite and positive. An application of the central limit theorem gives the following result:

For every real x,

$$P\left[\frac{Sn - np}{\sqrt{n}\sqrt{[p(1-p)}} \le x\right] \to f(x) \qquad \text{as } n \to \infty.$$

In other words, for large n

$$P[\left[S_n \le np + x\sqrt{np(1-p)}\right] \simeq \phi(x) \qquad ...(6)$$

where $\simeq$ denotes that the quantities on both sides are approximately equal to each other.

An alternate way of interrupting the above approximation that a binomial distribution tends to be close to a normal distribution for large n. Let us explain this in mote detail.

Suppose S_n has binomial distribution with parameters n and p. Then, for $1 \le r \le n$,

$$P[S_n \le r] = P\left[\frac{S_n - np}{\sqrt{np(1-p)}} \le \frac{r - np}{\sqrt{np(1-p)}}\right]$$

$$\simeq \phi\left[\frac{r - np}{\sqrt{np(1-p)}}\right]$$

for large n by (2). In general, it is computationally difficult to calculate the exact probability

$$P[S_n \le r] = \sum_{j=0}^{r} \begin{bmatrix} n \\ j \end{bmatrix} p^n (1-p)^{n-j}$$

when n is large. A close approximation to this probability can be obtained by computing

$$\phi\left(\frac{r - np}{\sqrt{np(1-p)}}\right)$$

where $ is the standard normal distribution function. It has been found from empirical studies that this approximation is good when $n \ge 30$ and a better approximation is obtained by applying a slight correction, namely,

$$\phi\left[\frac{r + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right]$$

Let us illustrate these results by an example.

Example 6: The ideal size of a first year class in a college is 150. It is known from an earlier data that on the average only 30% of those accepted for admission will actually attend. Suppose the college admits 450 students. What is the probability that more than 150 first year students attend the college?

Let us denote by S_n the number of sludents that attend the college when n are admitted. Assuming that all the students take independent decision of either attending or not attending the college, we can suppose that ***S,*** has the binomial distribution with parameters n and p = 0.3. Here n = 450 and we are interested in finding the

$$P[S_n \ge 150].$$

Note that $E(S_n) = np = (450)\,(0.3) = 135$ and

$Var(S_n) = np(1 - p) = (135)(.7)$

Further more

$$P[S_n \ge 150] = 1 - P[S_n < 150]$$

$$\simeq 1 - P[S_n \le 149]$$

and

$$P[S_n \le 149] = \phi\left[\frac{149 + \frac{1}{2} - 135}{\sqrt{(135)(.7)}}\right]$$

$$= \phi(1.59)$$

Hence

$$P[S_n \ge 150] = 1 - \phi(1.59)$$

$$= .0559$$

This shows that the probability that more than 150 first year students attend is less than 6%. Let us now consider a different type of application of the central limit theorem.

Example 7: Suppose X_1, X_2 .... is a sequence of i.i.d. random variables each N(0, l). Then $X_1^2, X_2^2,$ .... is a sequence of i.i.d. random variables each with X_1^2 -distribution.

Note that E(X_1^2) = 1 and Var (X_1^2) = 2 for any i. Hence by central limit theorem we get

$$P\left[\frac{X_1^2 + ... + X_n^2 - n}{\sqrt{2n}} \le x\right] \to \phi(x) \text{ as } n \to \infty.$$

But S_n = X_1^2 + ... + X_1^2 has X_n^2 distribution. What we have shown just now is that if Sn has X_n^2 distribution, then $\frac{S_n - n}{\sqrt{2n}}$ has an approximate standard normal distribution for large n. In other words, for every real x,

$$P\left[\frac{S_n - n}{\sqrt{2n}} \le x\right] \simeq \phi(x)$$

for large n whenever S, has *Xz* -distribution.

We make a remark now.

Remark 3 : The central limit theorem is central to the distribution theory needed for statistical inferential techniques to he developed in Block 4. You must have noted that the distribution of individual Xi in CLT could be discrete or continuous. The only condition that is imposed is that its variance has to be finite. In general, it is not easy to specify the size of n for a good approximation as it depends on the underlying distribution of {X_i}. However, it is found in practice that, in most cases, a good approximation is obtained whenever n is greater than or equal to 30.

We will stop our discussion on limit theorem now, though we shall refer to them off and on in the next block. Let us now do quick review of what we have covered in this unit.

20.2 Summary

- Obtain Poisson approximation to binomial;
- Discussed the central limit theorem and obtained normal approximation to binomial as an application.

As usual we suggest that you go back to the beginning of the unit and see if you have achieved the objectives. We have given our solutions to the exercises in the unit in the last section. Please go through them too. With this we have come to the end of this block.

- The Central Limit Theorem (CLT) is one of the most important and useful results in probability theory. We have already seen that the sum of a finite number of independent normal random variables is normally distributed. However the sum of a finite number of independent non-normal random variables need not be normally distributed. Even then, according to the central limit theorem, the sum of a large number of independent random variables has a distribution that is approximately normal under general conditions. The CLT provides a simple method of computing the probabilities for the sum of independent random variables approximately. This theorem also suggests the reasoning behind why most of the data observed in practice leads to bell-shaped curves.

20.3 Keywords

Binomial distribution with parameters n and p is shown to be approximable by a Poisson distribution whenever n is large and p is such that np is a constant A z 0.

Central Limit Theorem (CLT): The Central Limit Theorem (CLT) is one of the most important and useful results in probability theory.

20.4 Self Assessment

1. .... with parameters n and p is shown to be approximable by a Poisson distribution whenever n is large and p is such that np is a constant A z 0.
2. An important special case of this result is that binomial distribution can be approximated by an appropriate .... for large samples.
3. The .... is one of the most important and useful results in probability theory.
4. The CLT provides a simple method of computing the probabilities for the sum of .... approximately.

20.5 Review Questions

1. If X is binomial with n = 100 and p = 1/2, find an approximation for P[X = 50].
2. Suppose X is binomial with parameters n and p = 0.55. Determine the smallest n for which

$$P\left[\frac{X}{n} > \frac{1}{2}\right] \geq 0.95$$

   approximately.
3. If 10 fair dice are rolled, find the approximate probability that the sum of the numbers observed is between 30 and 40.
4. Suppose X is binomial with n = 100 and p = 0.1. Find the approximate value of P(12 ≤ X ≤ 14) using
   (a) the normal approximation
   (b) the poisson approximation, and
   (c) the binomial distribution.

Notes

Answers: Self Assessment

1. Binomial distribution
2. normal distribution
3. Central Limit Theorem (CLT)
4. independent random variables

20.6 Further Readings

Books

Sheldon M. Ross, Introduction to Probability Models, Ninth Edition, Elsevier Inc., 2007.

Jan Pukite, Paul Pukite, Modeling for Reliability Analysis, IEEE Press on Engineering of Complex Computing Systems, 1998.

Unit 21: Confidence Intervals

CONTENTS

Objectives

Introduction

21.1 Some Common Tests of Hypothesis for Normal Populations

21.2 Confidence Intervals

21.3 Summary

21.4 Keywords

21.5 Self Assessment

21.6 Review Questions

21.7 Further Readings

Objectives

After studying this unit, you will be able to:

- Discuss statistic for various testing of hypotheses problems as well as to derive power functions
- Explain confidence intervals for parameters of various distributions
- Describe large sample tests.

Introduction

You have been introduced to the problem of testing of hypothesis and also to some basic concepts of the theory of testing of hypothesis. There you have studied two important procedures fortesting statistical hypotheses,viz. using Neyman-Pearson Lemma and the likelihood ratio test. In this unit, you will be exposed to the problem of testing statistical hypotheses involving the parameters of some important distributions through some selected examples. In this unit, you will also be exposed to the problem of constructing confidence intervals for parameters of some important distributions through some selected examples. You will also learn the use of chi-square test for goodness of fit.

21.1 Some Common Tests of Hypothesis for Normal Populations

We have already described with examples two procedures for testing statistical hypotheses. In this section we will employ Neyman-Pearson Lamma and likelihood ratio test for testing of hypothesis related to a normal population.

Example 1: Let $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ be independent random samples from $N(\mu_1, \sigma^2)$ and $N(\mu_2, \sigma^2)$, respectively. It is desired to obtain a test statistic for testing $H_0 : \mu_1 = \mu_2$ against $\mu_1 : \mu_1 \neq \mu_2$ when σ^2 (> 0) is unknown.

Notes

In order to obtain the test statistic, we use the likelihood ratio test. We have

$$\Omega = \{ (\mu_1, \mu_2, \sigma^2) : -\infty < \mu_1, \mu_2 < \infty, \sigma^2 > 0 \}$$

$$\Omega_0 = \{ \mu_1 = \mu_2 = \mu \text{ (say)}, \sigma^2) : -\infty < \mu < \infty, \sigma^2 > 0\}$$

We shall write $\theta = (\mu_1, \mu_2, \sigma^2)$

We have

$$\{\sup_{\theta \in Q_0} L(\theta | X, Y)$$

$$= \text{Sup} \frac{1}{(2\pi)^{\frac{m+n}{2}} (\sigma^2)^{\frac{m+n}{2}}} \exp\left\{ -\frac{1}{2\sigma^2} \left[\sum_1^m (X_i - \mu_1)^2 + \sum_1^n (Y_i - \mu_2)^2 \right] \right\}$$

Under H_0, $\mu_1 = \mu_2 = \mu$ and the maximum likelihood estimate of μ is

$$\hat{\mu} = \frac{m\overline{X} + n\overline{Y}}{m+n} \text{ and of } \sigma^2 \text{ is}$$

$$\hat{\sigma}^2 = \frac{1}{m+n} \left[\sum_1^m (X_i - \mu_1)^2 + \sum_1^n (X_1 - \mu_2)^2 + \frac{mn}{(m+n)} (\overline{X} - \overline{Y})^2 \right]$$

= u (say)

$$\text{Thus } \sup_{\theta \in \Omega_0} L(\theta | X, Y) = \frac{1}{(2pu')^{\frac{m+n}{2}}} \exp\left[-\frac{1}{2u'} (m+n) u' \right]$$

$$= \left(\frac{1}{2\pi u'} \right)^{\frac{m+n}{2}} \exp\left(-\frac{(m+n)}{2} \right)$$

Under H_1, the maximum likelihood estimates of μ_1, μ_2 and σ^2 are respectively

$$\hat{\mu}_1 = \overline{X}, \hat{\mu}_2 = \overline{Y}, \hat{\sigma}^2 = \frac{\sum_1^m (X_i - \overline{X})^2 + \sum_1^n (Y - \overline{Y})^2}{m+n} = u(\text{say})$$

and

$$\{\sup_{\theta \in Q_0} L(\theta | X, Y)$$

$$= \left(\frac{1}{2\pi u} \right)^{\frac{m+n}{2}} \exp\left(-\frac{m+n}{2} \right)$$

The likelihood ratio test is thus

$$\lambda(X, Y) = \frac{\{\sup_{\theta \in \Omega_0} L(\underline{\theta} | X, Y)}{\{\sup_{\theta \in \Omega} L(\underline{\theta} | X, Y)}$$

Notes

$$= \left(\frac{u}{u'}\right)^{\left(\frac{m+n}{2}\right)}$$

$$= \left[\frac{\sum_1^m (Xi-\overline{X})^2 + \sum_1^n (Yi-\overline{Y})^2}{\sum_1^m (Xi-\overline{X})^2 + \sum_1^n (Yi-\overline{Y})^2 + \frac{mn}{(m+n)}(Xi-\overline{X})^2}\right]^{\frac{m+n}{2}}$$

Now under null hypothesis, $\mu_1 = \mu_2 = \mu$, and $t = \frac{\overline{X}-\overline{Y}}{S\sqrt{\left(\frac{1}{n}+\frac{1}{m}\right)}}$ follows a Student's t distributions

with m + n – 2 degrees of freedom, where $S^2 = \frac{u(m+n)}{m+n-2}$

Thus

$$t^2 = \frac{(m+n-2)mn(\overline{X}-\overline{Y})^2}{(m+n)\left\{\sum_1^m (X_i-\overline{X})^2 + \sum_1^n (Y-\overline{Y})^2\right\}}$$

and

$$\lambda(X, Y) = \left[\frac{1}{1+\frac{t^2}{m+n-2}}\right]^{\frac{m+n}{2}} < c$$

The likelihood ratio critical region is given by

$$\lambda(X, Y) = \left[\frac{1}{1+\frac{t^2}{m+n-2}}\right]^{\frac{m+n}{2}} < c$$

where c is to be determined so that

$$\underset{\theta\in\Omega_0}{\text{Sup}}\, P_\theta\left[\lambda(X,Y) < c\right] = \alpha$$

Since $\lambda(X, Y)$ is a decreasing function of $t^2/(m + n - 2)$ we reject H_0

if

$$\frac{t^2}{(m+n-2)} > c^{2/(m+n)}$$

Notes

or

$$|t| > c_1$$

where c_1 is so chosen that

Let $c_1 = t_{m+n-2, \alpha/2}$ in accordance with the diskbution oft under Ho. Thus, the two sided test obtained is

$$\left| \frac{(\overline{X} - \overline{Y}}{S} \sqrt{\frac{mn}{(m+n)}} \right| > t_{m+n-2,\alpha/2}$$

Example 2: Let $X_1, \ldots, X_n$, be a random sample from N (μ, σ^2), p is known and $\sigma^2 > 0$, is unknown. We wish to obtain a test statistic for testing $H_0 : \sigma^2 - \sigma_0^2$ against an alternative $H_1 : \sigma^2 = \sigma_1^2 \ (> \sigma_0^2)$.

We have

$$P_{\theta_1}(\underline{X}) = \frac{1}{(2\pi\sigma_1)^{n/2}} \exp\left[-\frac{1}{2\sigma_1^2}\sum_1^n (Xi - \mu)^2\right]$$

$$P_{\theta_0}(\underline{X}) = \frac{1}{(2\pi\sigma_0)^{n/2}} \exp\left[-\frac{1}{2\sigma_0^2}\sum_1^n (Xi - \mu)^2\right]$$

Using Neyman-Pearson Lemma, the test statistic is

$$T(X) = \frac{P_{\theta_1}(\underline{X})}{P_{\theta_0}(\underline{X})} \geq k$$

$$\Rightarrow \left(\frac{\sigma_0}{\sigma_1}\right)^{n/2} \exp\left\{1/2\left(\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2}\right)\sum_1^n (X_i - \mu)^2\right.$$

$$\Rightarrow (\sigma_1^2 - \sigma_0^2)\sum_1^n (X_i - \mu^2) \geq k \text{, taking logarithms}$$

$$\Rightarrow \sum_1^n (X_i - \mu^2) \geq k_1 \text{, since } \sigma_1^2 > \sigma_0^2, \text{under } H_1$$

Here k_1 is so determined that

$$P_{\theta_0}(T,(\underline{X}) \geq k) = \alpha$$

$$\Rightarrow P_{\theta_0}\left[\sum_1^n (X_i - \mu)^2 \geq k_1\right] = \alpha$$

$$\Rightarrow P_{\theta_0}\left[\sum_1^n (X_i - \mu)^2 / \sigma_0^2 \geq k_1 / \sigma_0^2\right]$$

Notes

Under the null hypothesis, since σ² = $\sigma_0^2, \sum_1^n (X_i - \mu)^2 / \sigma_0^2\}$ has a χ_n^2 distribution (chi-square distribution with n degrees of freedom). Let $\chi_{n,\alpha}^2$ be the upper -α probability point of χ_n^2. The test statistic is thus

$$\sum_1^n (X_i - \mu)^2 >= k1 \text{ and hence}$$

$$C0 = \left\{ X \mid \sum_1^n (X_i - \mu)^2 / \sigma_0^2 > c_{n,\alpha}^2 \right\}$$

On the other hand, if the alternative hypothesis is $H_1 : \sigma^2 = \sigma_1^2 \; (\sigma_1^2 < \sigma_0^2)$, then the test statistic is and hence

$$\sum_1^n (X_i - \mu)^2 < k_2$$

where $\chi_{n,1-\alpha}^2$ is the lower α -probability point of the χ^2 distribution with n degrees of freedom.

Example 3: Let $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ be independent random samples from $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$. We wish to obtain a test statistic for testing $H_0 : \sigma_1^2 = \sigma_2^2$ against $H_1 : \sigma_1^2 \neq \sigma_2^2$.

Here Ω = $\{(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2) : -\infty < \mu_i < \infty, \sigma_1^2 > 0, i = 1,2\}$

and Ω_0 = $\{(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2) : -\infty < \mu_i < \infty, i = 1,2, \sigma_1^2 = \sigma_2^2 = \sigma^2 > 0\}$

We shall use $\underline{\theta} = (\mu_1, \mu_2, \sigma_1^2, \sigma_2^2)$.

Also L (8 | X, Y)

$$= \left(\frac{1}{2\pi}\right)^{\frac{m+n}{2}} \left(\frac{1}{\sigma_1^2}\right)^{m/2} \left(\frac{1}{\sigma_1^2}\right)^{n/2} \exp\left\{ -\frac{1}{2\sigma_1^2} \sum_1^n (X_i - \mu_1)^2 - \frac{1}{2\sigma_2^2} \sum_1^n (Y_i - \mu_2)^2 \right\}$$

The maximum likelihood estimates of μ_1, μ_2, σ_1^2, σ_2^2 are respectively

$$\hat{\mu}_1 = \frac{1}{m} \sum_1^m X_i = \overline{X}, \hat{\mu}_2 = \frac{1}{n} \sum_1^n Y_i = \overline{Y}$$

$$\hat{\sigma}_1^2 = \frac{1}{m} \sum_1^m (X_i - \overline{X})^2, \hat{\sigma}_2^2 = \frac{1}{n} \sum_1^n (Y_i - \overline{Y})^2$$

Further, if $\sigma_1^2 = \sigma_2^2 = \sigma^2$, the maximum likelihood estimate of σ^2 is

$$\hat{\sigma}^2 = \frac{1}{(m+n)} \left[\sum_1^m (X_i - \overline{X})^2 + \sum_1^n (Y_i - \overline{Y})^2 \right]$$

Notes

Thus

$$\underset{\theta \in Q_0}{\text{Sup}}\, L(\theta | X, Y)$$

$$= \frac{\exp\{-(m+n)/2\}}{[2\pi/(m+n)]^{\frac{m+n}{2}} \left\{\sum_1^m (X_i - \overline{X})^2 + \sum_1^n (Y_i - \overline{Y})^2\right\}^{\frac{m+n}{2}}}$$

and

$$\underset{\theta \in Q_0}{\text{Sup}}\, L(\theta | X, Y)$$

$$= \frac{\exp\{-(m+n)/2\}}{(2\pi/m)^{m/2}(2\pi/n)^{n/2} \left\{\sum_1^m (X_i - \overline{X})^2\right\}^{\frac{m}{2}} \left\{\sum_1^n (Y_i - \overline{Y})^2\right\}^{\frac{n}{2}}}$$

$$= \frac{\exp\{-(m+n)/2\}}{(2\pi/m)^{m/2}(2\pi/n)^{n/2} \left\{\sum_1^m (X_i - \overline{X})^2\right\}^{\frac{m}{2}} \left\{\sum_1^n (Y_i - \overline{Y})^2\right\}^{\frac{n}{2}}}$$

The likelihood ratio test is thus

$$\lambda\,(X, Y) = \frac{\underset{\theta \in \Omega_0}{\text{Sup}}\, L(\theta | X, Y)}{\underset{\theta \in \Omega_0}{\text{Sup}}\, L(\theta | X, Y)}$$

$$= \left(\frac{m}{m+n}\right)^{m/2} \left(\frac{n}{m+n}\right)^{n/2} \frac{\left\{\sum_1^m (X_i - \overline{X})^2\right\}^{\frac{m}{2}} \left\{\sum_1^n (Y_i - \overline{Y})^2\right\}^{\frac{n}{2}}}{\left\{\sum_1^m (X_i - \overline{X})^2 + \sum_1^n (Y_i - \overline{Y})^2\right\}^{\frac{m+n}{2}}}$$

Now

$$\frac{\left\{\sum_1^m (X_i - \overline{X})^2\right\}^{\frac{m}{2}} \left\{\sum_1^n (Y_i - \overline{Y})^2\right\}^{\frac{n}{2}}}{\left\{\sum_1^m (X_i - \overline{X})^2 + \sum_1^n (Y_i - \overline{Y})^2\right\}^{\frac{m+n}{2}}}$$

We have

$$\lambda\,(X, Y) = \frac{\left(\frac{m}{m+n}\right)^{m/2} \left(\frac{n}{m+n}\right)^{m/2}}{\left[1 + \frac{(m-1)}{(m-1)} f\right]^{n/2} \left[1 + \frac{(n-1)}{(m-1)}(1/f)\right]^{m/2}}$$

The likelihood ratio test criterion rejects Ho if λ (X, Y) < c

It is easy to see that λ (X, Y) is a monotonic function of f and h (X, Y) < c is equivalent to f < c_1 or f > c. Under H_0,

$$f = \frac{\sum_{1}^{m}(X_i - \overline{X})^2/(m-1)}{\sum_{1}^{n}(Y_i - \overline{Y})^2/(n-1)}$$

has Snedecar's F (m – 1, n – 1) distribution, so that c_1, c_2 can be selected, such that

$$\underset{\theta \in \Omega_0}{\text{Sup}}\ P_{\underset{\sim}{\theta}}\,[\lambda(X, Y) < c] = \alpha$$

or

$$P(F \leq c_1) = P(F \geq c_2) = \alpha/2$$

Thus c_2 = F(m – l, n – 1, α/2) is the upper α/2 probability point of F (m – 1, n – 1) distribution and c_1 = F (m – 1, n – 1, l – α/2) is the lower α/2 probability point of F (m - 1, n - 1).

21.2 Confidence Intervals

In you have been briefly exposed to some notions of interval estimation of a parameter. In this section we discuss in detail the problem of obtaining interval estimates of parameters and describe, through examples, some methods of constructing interval extimates of parameters. We may remind you again that an interval estimate is also called a confidence interval or a confidence set. We first illustrate through small examples the need for constructing confidence intervals. Suppose X denotes the tensile strength of a copper wire. A potential user may desire to know the lower bound for the mean of X, so that he can use the wire if the average tensile strength is not less than say go. Similarly, if the random 'variable X measures the toxicity of a drug, a doctor may wish to have a knowledge t of the upper bound for the hean of X in order to prescribe this dmg. If the random variable X measures the waiting times at the emergency room of a large city hospital, one may be interested in the mean waiting time at this emergency room. In this case we wish to obtain both the lower and upper bounds for the waiting time.

In this unit we are concerned with the problem of determining confidence intervals for a parameter. A formal definition of a confidence interval has been given in Section 15.6. However, for the sake of completeness we define some terms here.

Let $X_1, X_2, \ldots, X_n$ be a random sample from a population with density (or, mass) function f (x, θ), $\theta \in \Omega \subseteq R^1$. The object is to find statistics r_L ($X_1, \ldots, X_n$) and r_U ($X_1, \ldots, X_n$) such that

P_θ { (r_L (X_1,..., X_n) $\leq \theta \leq r_U$(X_1 ,..., X_n)] $\geq 1 - \alpha$ for all $\theta \in \Omega \subseteq R^1$. The interval $(r_L(\underset{\sim}{X}), r_U(\underset{\sim}{X}))$ is called a confidence interval and the quantity

$$\inf P_\theta [r_L (X_1,\ldots,X_n) \leq \theta \leq r_U (X_1,\ldots,X_n)]$$

will be referred to as the confidence co-efficient associated with the random interval.

We now give some examples of construction of confidence intervals.

Example 4: Let $X_1, X_2, \ldots, X_n$ be a random sample from a normal population, N (μ, σ^2). We wish to obtain a (1 – α) level confidence interval for μ.

Notes

Let $\overline{X} = n^{-1}\sum_{1}^{n} X_i$. Consider the interval $(\overline{X} - a, \overline{X} + b)$. In order for this to be a (1 – α)-level confidence interval, we must have

$$P\{\overline{X} - a < \mu < \overline{X} + b\} \geq 1 - \alpha$$

Thus

$$P\left\{-\frac{b}{\sigma}\sqrt{n} < \frac{(\overline{X} - \mu)}{\sigma}\sqrt{n} < \frac{a}{b}\sqrt{n}\right\} \geq 1 - \alpha$$

Since, $\frac{(\overline{X} - \mu)}{\sigma}\sqrt{n} / \sim N(0,1)$ we can choose a and b to satisfy

$$P\left\{-\frac{b}{\sigma}\sqrt{n} < \frac{(\overline{X} - \mu)\sqrt{n}}{\sigma} < \frac{a}{\sigma}\sqrt{n}\right\} = 1 - a$$

provided that a is known. There are infinitely many such pairs of values (a, b). In Inference particular, an intuitively reasonable choice is a = b = c , say

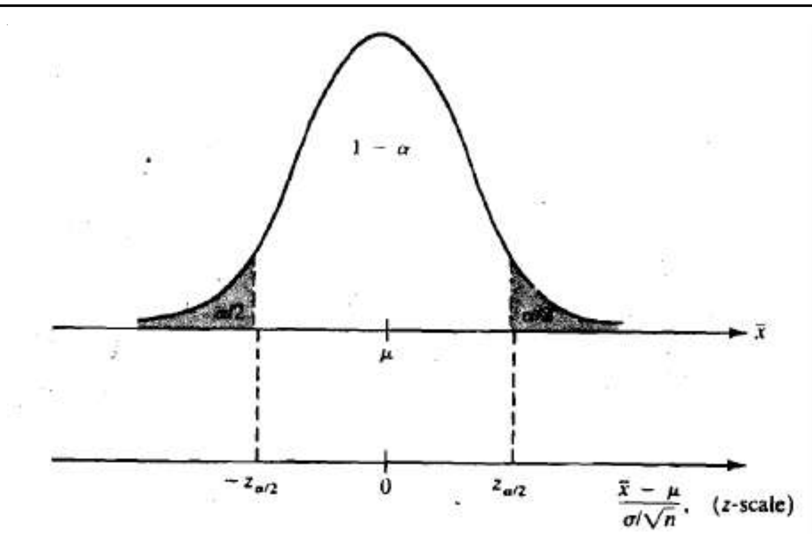
In that case

$\frac{c\sqrt{n}}{\sigma} = Z_{\alpha/2}$ where $Z_{\alpha/2}$ is the α/2 percent point of the standard normal distribution, and the confidence interval is

$$(\overline{X} - (\sigma/\sqrt{n})Z_{\alpha/2}, \overline{X} + (\sigma/\sqrt{n})Z_{\alpha/2})$$

The length of the interval is $(2\sigma/\sqrt{n})\ Z_{\alpha/2}$ Given a and a one can choose n to get a confidence interval of desired length.

Figure 21.1: Probability density curve of normal distribution with mean m and variance σ²/n. Shows area σ/2 in each of two talls



If σ^2 is unknown, we have from

$$P\{-b < \overline{X} - \mu < a\} \geq 1 - \alpha$$

that

$$P\left\{-\frac{b}{S}\sqrt{n} < \frac{(\overline{X}-\mu)}{S} < \frac{a}{S}\sqrt{n}\right\} \geq 1-\alpha$$

It is known that $\frac{\overline{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$. We can choose pairs of values (a, b) using a students t-distribution with (n – 1) degrees of freedom such

$$P\left\{-\frac{b\sqrt{n}}{S} < \frac{\overline{X}-\mu}{Sb/\sqrt{n}} < \frac{a\sqrt{n}}{S}\right\} = 1-\alpha$$

In particular, an intuitively reasonable choice is a = b = c say. Then

$$\frac{c\sqrt{n}}{S} = t_{n-1,\alpha/2}$$

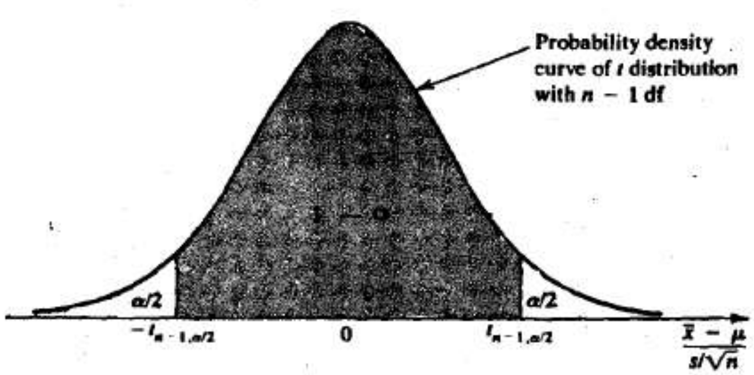
and $(\overline{X}-(S/\sqrt{n})t_{n-1,\alpha/2}, \overline{X}+(S/\sqrt{n})t_{n-1,\alpha/2})$ is 1 – α level confidence interval for μ. The length of the interval is $(2S/\sqrt{n})t_{n-1,\alpha/2}$, which is no longer constant.

Therefore, in this case one cannot choose n to get a fixed length confidence interval of level 1 – α. The expected length is, however,

$$\frac{2}{\sqrt{n}}t_{n-1,\alpha/2}Es(S) = \frac{2}{\sqrt{n}}t_{n-1,\alpha/2}\sqrt{\frac{2}{n-1}}\frac{\Gamma(n/2)}{\Gamma(n-1)/2)}\sigma$$

which can be made as small as we want by making a proper choice of n for a given σ and α.

Figure 21.2 : t Values such that there is an area α/2 in the right tall and α/2 in the left tall of the distribution.



Example 5: Let $X_1, X_2, \ldots, X_n$ be a random sample, from $N(\mu, \sigma^2)$. It is desired to obtain a confidence interval for σ^2 when μ is unknown.

Consider the interval (aS^2, bS^2), a, b > 0, $S^2 = (n-1)^{-1}\sum_{1}^{n}(X_i-\overline{X})^2$. We have

$$P\{aS^2 < \sigma^2 < bS^2\} \geq 1 - a$$

Notes

so that

$$P\left\{b^{-1} < \frac{S^2}{\sigma^2} < a^{-1}\right\} \geq 1-\alpha$$

It is known that

$$(n-1)S^2/\sigma^2 \sim \chi^2_{n-1}$$

We can therefore choose pairs of intervals (a, b) from the tables of the chi-square distribution. In particular we can choose a, b so that

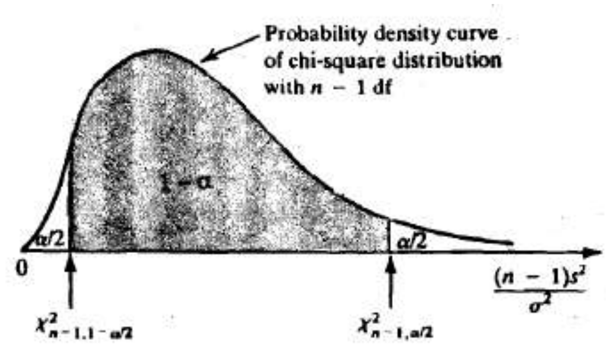
$$P\left\{\frac{S^2}{\sigma^2} \geq \frac{1}{a}\right\} = \alpha/2 = P\left\{\frac{S^2}{\sigma^2} \leq \frac{1}{b}\right\}.$$

Then $\frac{n-1}{a} = x^2_{n-1,\alpha/2}$ and $\frac{n-1}{b} = x^2_{n-1,1-\alpha/2}$ and the 1 – α level confidence interval for σ^2 when μ is unknown is

$$\left(\frac{(n-1)S^2}{\chi^2_{n-1,\alpha/2}}, \frac{(n-1)S^2}{\chi^2_{n-1,1-\alpha/2}}\right)$$

If however, μ is known then (n – 1) S^2 is replaced by $\sum_1^n (X_i - \mu)^2$ and the degrees of freedom of χ^2 is n instead of n – 1, for $\sum_1^n (X_i - \mu)^2/\sigma^2 \sim c^2_n$.

Figure 21.3 : Chi-square values such that area 1 – α/2 and α/2 are to their right.



Example 6: Let $X_1, \ldots, X_2$ and $Y_1, \ldots, Y_m$ denote respectively independent random samples from the two independent distributions having respectively the probability density functions $N(\mu_1, \sigma^2)$ and $N(\mu_2, \sigma^2)$. We wish to obtain a confidence interval for $\mu_1 - \mu_2$.

Consider the interval $\{(\overline{X} - \overline{Y}) - a, (\overline{X} - \overline{Y}) + b\}$. In order that this is a (1 - α) level confidence interval, we mbt have

$$P\{(\overline{X} - \overline{Y}) - a < \mu_1 - \mu_2 < (\overline{X} - \overline{Y}) + b\} \geq 1 - \alpha$$

Notes

which is the same as

$$P\{-b < (\overline{X}-\overline{Y})-(\mu_1-\mu_2) < a\} \geq 1-\alpha$$

or

$$P\left\{\frac{b}{\sigma\sqrt{\left(\frac{1}{n}+\frac{1}{m}\right)}} < \frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sigma\sqrt{\left(\frac{1}{n}+\frac{1}{m}\right)}} < \frac{a}{\sigma\sqrt{\left(\frac{1}{n}+\frac{1}{m}\right)}}\right\} \geq 1-\alpha$$

Here $\overline{X} = \frac{1}{n}\sum_1^n X_i$ and $\overline{Y} = \frac{1}{m}\sum_1^m Y_i$

Since $\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sigma\sqrt{\left(\frac{1}{n}+\frac{1}{m}\right)}} \sim N(0,1)$.

we can choose a and b to satisfy

$$P\left\{\frac{-b}{\sigma\sqrt{\left(\frac{1}{n}+\frac{1}{m}\right)}} < \frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sigma\sqrt{\left(\frac{1}{n}+\frac{1}{m}\right)}} < \frac{a}{\sigma\sqrt{\left(\frac{1}{n}+\frac{1}{m}\right)}}\right\} \geq 1-\alpha$$

provided that σ is known. There are infinitely many such pairs of values (a, b). In particular, an intuitively reasonable choice is a = b = c, say. In that case $c/\left\{s\left(\frac{1}{n}+\frac{1}{m}\right)^{1/2}\right\} = Z_{\alpha/2}$ and the confidence interval is

$$\left\{(\overline{X}-\overline{Y})-\sigma\left(\frac{1}{n}+\frac{1}{m}\right)^{1/2} Z_{\alpha/2}, (\overline{X}-\overline{Y})+\sigma\left(\frac{1}{n}+\frac{1}{m}\right)^{1/2} Z_{\alpha/2}\right\}$$

The length of the intaval is $2\sigma\left(\frac{1}{n}+\frac{1}{m}\right)^{1/2} Z_{\alpha/2}$. Given α and σ one can choose n and m to get a desired length confidence interval.

If σ^2 is unknown, we have from

$$P\{-b < (\overline{X}-\overline{Y})-(\mu_1-\mu_2) < a\} \geq 1-\alpha$$

that

$$P\left\{\frac{-b}{S\sqrt{\left(\frac{1}{n}+\frac{1}{m}\right)}} < \frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{S\sqrt{\left(\frac{1}{n}+\frac{1}{m}\right)}} < \frac{a}{S\sqrt{\left(\frac{1}{n}+\frac{1}{m}\right)}}\right\} \geq 1-\alpha$$

Notes

where $$\frac{\sum_{1}^{n}(X_i-\overline{X})2+\sum_{1}^{m}(Y_i-\overline{Y})^2}{(n+m-2)}=\frac{(n-1)S_x^2+(m-1)S_y^2}{n+m-2}$$

It is known that

$\dfrac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{S2\sqrt{\left(\frac{1}{n}+\frac{1}{m}\right)}}\sim t_{n+m-2}$. We can choose pairs of values (a, b) using Student's t-distribution with n + m – 2 degrees of freedom such that

$$\left\{\frac{-b}{S\sqrt{\left(\frac{1}{n}+\frac{1}{m}\right)}}<\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{S\sqrt{\left(\frac{1}{n}+\frac{1}{m}\right)}}<\frac{a}{S\sqrt{\left(\frac{1}{n}+\frac{1}{m}\right)}}\right\}=1-\alpha$$

In particular, an intuitively reasonable choice is a = b = c, say. Then

$$\frac{c}{S\sqrt{\left(\frac{1}{n}+\frac{1}{m}\right)}}=t_{n+m-2,\alpha/2}$$

and $\left\{(\overline{X}-\overline{Y})-S\left(\frac{1}{n}+\frac{1}{m}\right)^{1/2}tn+m-2,a/2,(\overline{X}-\overline{Y})+S\left(\frac{1}{n}+\frac{1}{m}\right)^{1/2}t_{n+m-2,\alpha/2}\right\}$

is a 1 – α level confidence interval for μ_1 – μ_2.

Example 7: Let $X_1, \ldots, X_n$, and $Y_1, \ldots, Y_m$, n, m > 2, denote respectively independent random samples from the two distributions having respectively the probability density fundions $N(\mu_1,\sigma_1^2)$ and N $N(\mu_2,\sigma_2^2)$. We wish to obtain a confidence interval for the ratio σ_2^2/σ_1^2 when μ_1 and μ_2 are unknown.

Consider the interval $(a\,S_2^2/S_1^2,\ bS_2^2/S_1^2)$ a , b > 0, where

$$S_1^2=\frac{1}{(n-1)}\sum_{1}^{n}(X_i-\overline{X})^2, S_2^2=\frac{1}{(m-1)}\sum_{1}^{m}(Y_i-\overline{Y})^2,$$

$$\overline{X}=\frac{1}{n}\sum_{1}^{n}X_i, \overline{Y}=\frac{1}{m}\sum_{1}^{m}Y_i.\ \text{We have}$$

$$P\left\{a\frac{S_2^2}{S_1^2}<\frac{\sigma_2^2}{\sigma_1^2}<b\frac{S_2^2}{S_1^2}\right\}\geq 1-\alpha$$

so that

$$P\left\{\frac{1}{b}<\frac{(S_2^2/S_1^2)}{(\sigma_2^2/\sigma_1^2)}<\frac{1}{a}\right\}\geq 1-\alpha$$

Notes

It is also known that if X and Y are independent χ^2 random variables with m and n degrees of freedom respectively, the random variable F = (X/m)/(Y/n) is said to have an F-distribution with (m, n) degrees of freedom. It is also known that if X has an F (m, n) distribution then l/X has an F (n, m) distribution, and $F_{m, n, 1-\alpha} = 1/F_{n, m, \alpha}$. Therefore

$$\frac{S_2^2/\sigma_2^2}{S_1^2/\sigma_1^2} = \frac{S_2^2/S_1^2}{\sigma_2^2/\sigma_1^2} \sim F_{(m-1),(n-1)}$$

We can therefore choose paris of values (a, b) from the tables of F-distribution. In particular, we can choose a and b so that

$$P\left\{\frac{(S_2^2/\sigma_2^2)}{(S_1^2/\sigma_1^2)} \geq \frac{1}{a}\right\} = \alpha/2 = P\left\{\frac{(S_2^2/\sigma_2^2)}{(S_1^2/\sigma_1^2)} \leq \frac{1}{b}\right\}$$

Then $\frac{1}{a} = F_{m,n,\alpha/2}$ and $\frac{1}{b} = F_{m,n,1-\alpha/2}$ and the 1 – α level confihnce interval for σ_2^2/σ_1^2 is

$$\left(\frac{S_2^2}{S_1^2} - F_{n,m,1-\alpha/2}, \frac{S_2^2}{S_1^2} - F_{n,m,1/\alpha/2}\right)$$

21.3 Summary

- We have already described with examples two procedures for testing statistical hypotheses. In this section we will employ Neyman-Pearson Lamma and likelihood ratio test for testing of hypothesis related to a normal population.
- In you have been briefly exposed to some notions of interval estimation of a parameter. In this section we discuss in detail the problem of obtaining interval estimates of parameters and describe, through examples, some methods of constructing interval extimates of parameters. We may remind you again that an interval estimate is also called a confidence interval or a confidence set. We first illustrate through small examples the need for constructing confidence intervals. Suppose X denotes the tensile strength of a copper wire. A potential user may desire to know the lower bound for the mean of X, so that he can use the wire if the average tensile strength is not less than say go. Similarly, if the random 'variable X measures the toxicity of a drug, a doctor may wish to have a knowledge t of the upper bound for the hean of X in order to prescribe this dmg. If the random variable X measures the waiting times at the emergency room of a large city hospital, one may be interested in the mean waiting time at this emergency room. In this case we wish to obtain both the lower and upper bounds for the waiting time.
- In this unit we are concerned with the problem of determining confidence intervals for a parameter. A formal definition of a confidence interval has been given in Section 15.6. However, for the sake of completeness we define some terms here.

21.4 Keywords

Confidence interval: Let $X_1, X_2, \ldots, X_n$ be a random sample from a population with density (or, mass) function f (x, θ), $\theta \in \Omega \subseteq R^1$. The object is to find statistics $r_L(X_1, \ldots, X_n)$ and $r_U(X_1, \ldots, X_n)$ such that

$P_\theta\{(r_L(X_1,\ldots, X_n) \leq \theta \leq r_U(X_1,\ldots, X_n)] \geq 1 - \alpha$ for all $\theta \in \Omega \subseteq R^1$. The interval $(r_L(\underline{X}), r_U(\underline{X}))$ is called a confidence interval.

Notes

21.5 Self Assessment

1. If the random variable X measures the toxicity of a drug, a doctor may wish to have a knowledge t of the .... for the hean of X in order to prescribe this dmg.
2. If the .... X measures the waiting times at the emergency room of a large city hospital, one may be interested in the mean waiting time at this emergency room.

21.6 Review Questions

1. Let $X_1, X_2, \ldots, X_n$ be a random sample from a normal population, N (μ, σ^2). We wish to obtain a $(1 - \alpha)$ level confidence interval for μ.
2. Let $X_1, X_2, \ldots, X_n$ be a random sample, from $N(\mu, \sigma^2)$. It is desired to obtain a confidence interval for σ^2 when μ is unknown.
3. Let $X_1, X_2, \ldots, X_n$ be a random sample from a normal population, N (μ, σ^2). We wish to obtain a $(1 - 2\alpha)$ level confidence interval for μ.
4. Let $X_1, \ldots, X_2$ and $Y_1, \ldots, Y_m$ denote respectively independent random samples from the two independent distributions having respectively the probability density functions $N(\mu_1, \sigma^2)$ and $N(\mu_2, \sigma^2)$. We wish to obtain a confidence interval for $\mu_1 - \mu_2$.
5. Let $X_1, X_2, \ldots, X_n$ be a random sample from a normal population, N (μ, σ^2). We wish to obtain a $(1 - 3\alpha)$ level confidence interval for μ.
6. Let $X_1, X_2, \ldots, X_n$ be a random sample from a normal population, N (μ, σ^2). We wish to obtain a $(1 - \alpha)^2$ level confidence interval for μ.

Answers: Self Assessment

1. upper bound
2. random variable

21.7 Further Readings

Books

Sheldon M. Ross, Introduction to Probability Models, Ninth Edition, Elsevier Inc., 2007.

Jan Pukite, Paul Pukite, Modeling for Reliability Analysis, IEEE Press on Engineering of Complex Computing Systems, 1998.

Unit 22: Correlation

Notes

CONTENTS

Objectives

Introduction

22.1 Definition of Correlation

22.1.1 Scope of Correlation Analysis

22.2 Scatter Diagram

22.3 Karl Pearson's Coefficient of Linear Correlation

22.4 Properties of Coefficient of Correlation

22.5 Probable Error of r

22.5.1 Uses of P.E.(r)

22.6 Correlation in a Bivariate Frequency Distribution

22.7 Merits and Limitations of Coefficient of Correlation

22.8 Spearman's Rank Correlation

22.9 Coefficient of Correlation by Concurrent Deviation Method

22.10 Summary

22.11 Keywords

22.12 Self Assessment

22.13 Review Questions

22.14 Further Readings

Objectives

After studying this unit, you will be able to:

- Definition of Correlation
- Discuss Scatter Diagram
- Explain Karl Pearson's Coefficient of Linear Correlation
- Discuss Properties of Coefficient of Correlation
- Describe Probable Error of r

Introduction

So far we have considered distributions relating to a single characteristics. Such distributions are known as Univariate Distribution. When various units under consideration are observed simultaneously, with regard to two characteristics, we get a Bivariate Distribution. For example, the simultaneous study of the heights and weights of students of a college. For such data also, we

Notes

can compute mean, variance, skewness etc. for each individual characteristics. In addition to this, in the study of a bivariate distribution, we are also interested in knowing whether there exists some relationship between two characteristics or in other words, how far the two variables, corresponding to two characteristics, tend to move together in same or opposite directions i.e. how far they are associated.

The knowledge of this type of relationship is useful for predicting the value of one variable given the value of the other. It also helps in understanding and analysis of various economic and business problems. It should be noted here that statistical relations are different from the exact mathematical relations. Given a statistical relation Y = a + bX, between two variables X and Y, we can only get a value of Y that we expect on the average for a given value of X. The study of relationship between two or more variables can be divided into two broad categories:

(i) To determine whether there exists some sort of association between the variables. If so, what is the degree of association or the magnitude of correlation between the two.

(ii) To determine the most suitable form of the relationship between the variables given that they are correlated.

The first category relates to the study of 'Correlation' which will be discussed in this chapter and the second relates to the study of 'Regression', to be discussed in next chapter.

22.1 Definition of Correlation

Various experts have defined correlation in their own words and their definitions, broadly speaking, imply that correlation is the degree of association between two or more variables. Some important definitions of correlation are given below:

(i) "*If two or more quantities vary in sympathy so that movements in one tend to be accompanied by corresponding movements in other(s) then they are said to be correlated.*"

— L.R. Connor

(ii) "*Correlation is an analysis of covariation between two or more variables.*"

— A.M. Tuttle

(iii) "*When the relationship is of a quantitative nature, the appropriate statistical tool for discovering and measuring the relationship and expressing it in a brief formula is known as correlation.*"

— Croxton and Cowden

(iv) "*Correlation analysis attempts to determine the 'degree of relationship' between variables*".

— Ya Lun Chou

Correlation Coefficient: It is a numerical measure of the degree of association between two or more variables.

22.1.1 Scope of Correlation Analysis

The existence of correlation between two (or more) variables only implies that these variables (i) either tend to increase or decrease together or (ii) an increase (or decrease) in one is accompanied by the corresponding decrease (or increase) in the other. The questions of the type, whether

Notes

changes in a variable are due to changes in the other, i.e., whether a cause and effect type relationship exists between them, are not answered by the study of correlation analysis. If there is a correlation between two variables, it may be due to any of the following situations:

(i) ***One of the variable may be affecting the other:*** A correlation coefficient calculated from the data on quantity demanded and corresponding price of tea would only reveal that the degree of association between them is very high. It will not give us any idea about whether price is affecting demand of tea or vice-versa. In order to know this, we need to have some additional information apart from the study of correlation. For example if, on the basis of some additional information, we say that the price of tea affects its demand, then price will be the cause and quantity will be the effect. The causal variable is also termed as independent variable while the other variable is termed as dependent variable.

(ii) ***The two variables may act upon each other:*** Cause and effect relation exists in this case also but it may be very difficult to find out which of the two variables is independent. For example, if we have data on price of wheat and its cost of production, the correlation between them may be very high because higher price of wheat may attract farmers to produce more wheat and more production of wheat may mean higher cost of production, assuming that it is an increasing cost industry. Further, the higher cost of production may in turn raise the price of wheat. For the purpose of determining a relationship between the two variables in such situations, we can take any one of them as independent variable.

(iii) ***The two variables may be acted upon by the outside influences:*** In this case we might get a high value of correlation between the two variables, however, apparently no cause and effect type relation seems to exist between them. For example, the demands of the two commodities, say X and Y, may be positively correlated because the incomes of the consumers are rising. Coefficient of correlation obtained in such a situation is called a spurious or nonsense correlation.

(iv) ***A high value of the correlation coefficient may be obtained due to sheer coincidence (or pure chance):*** This is another situation of spurious correlation. Given the data on any two variables, one may obtain a high value of correlation coefficient when in fact they do not have any relationship. For example, a high value of correlation coefficient may be obtained between the size of shoe and the income of persons of a locality.

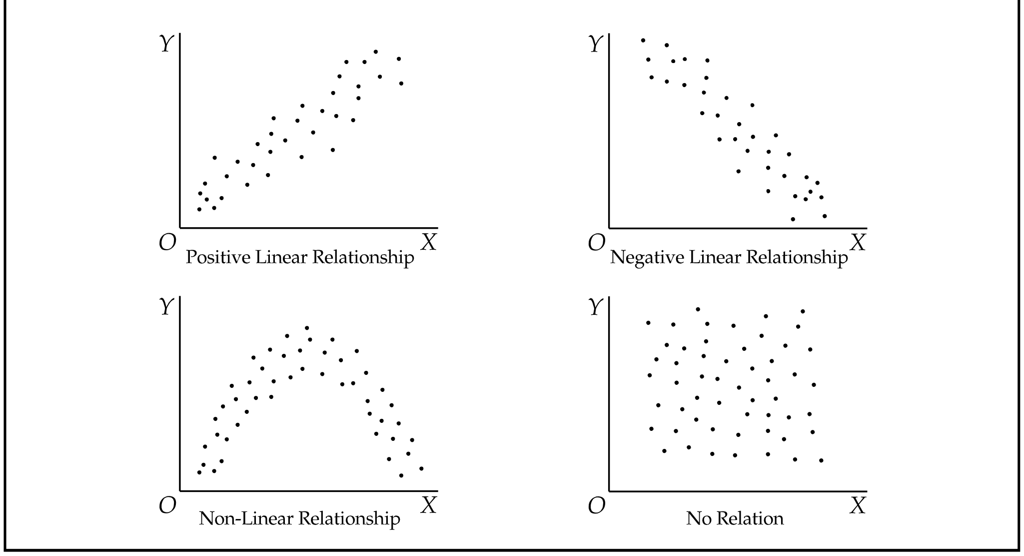
22.2 Scatter Diagram

Let the bivariate data be denoted by (X_i, Y_i), where i = 1, 2 ...... n. In order to have some idea about the extent of association between variables X and Y, each pair (X_i, Y_i), i = 1, 2......n, is plotted on a graph. The diagram, thus obtained, is called a Scatter Diagram.

Each pair of values (X_i, Y_i) is denoted by a point on the graph. The set of such points (also known as dots of the diagram) may cluster around a straight line or a curve or may not show any tendency of association. Various possible situations are shown with the help of given diagrams:

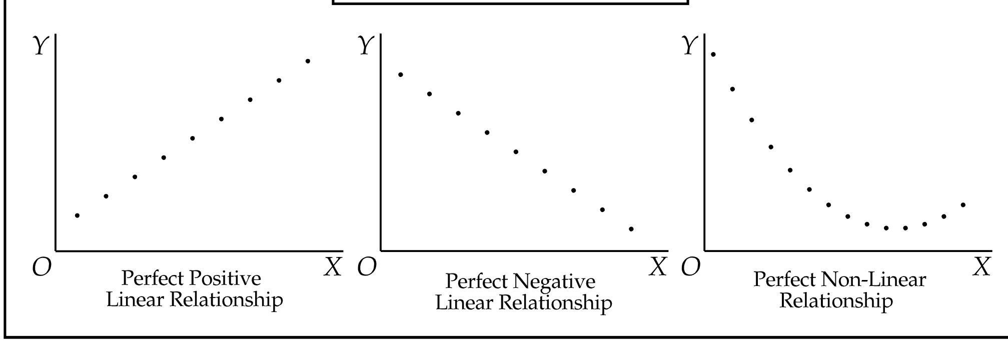
Notes

Figure 22.1



If all the points or dots lie exactly on a straight line or a curve, the association between the variables is said to be perfect. This is shown below:

Figure 22.2



A scatter diagram of the data helps in having a visual idea about the nature of association between two variables. If the points cluster along a straight line, the association between variables is linear. Further, if the points cluster along a curve, the corresponding association is non-linear or curvilinear. Finally, if the points neither cluster along a straight line nor along a curve, there is absence of any association between the variables.

It is also obvious from the above figure that when low (high) values of X are associated with low (high) value of Y, the association between them is said to be positive. Contrary to this, when low (high) values of X are associated with high (low) values of Y, the association between them is said to be negative.

This chapter deals only with linear association between the two variables X and Y. We shall measure the degree of linear association by the Karl Pearson's formula for the coefficient of linear correlation.

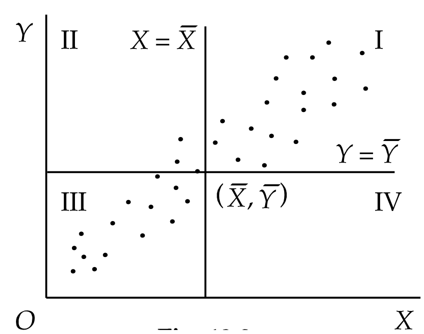
Notes

22.3 Karl Pearson's Coefficient of Linear Correlation

Let us assume, again, that we have data on two variables X and Y denoted by the pairs (X_i, Y_i), i = 1,2, ... n. Further, let the scatter diagram of the data be as shown in figure 22.3.

Let $\overline{X}$ and $\overline{Y}$ be the arithmetic means of X and Y respectively. Draw two lines $X = \overline{X}$ and $Y = \overline{Y}$ on the scatter diagram. These two lines, intersect at the point ($\overline{X}, \overline{Y}$) and are mutually perpendicular, divide the whole diagram into four parts, termed as I, II, III and IV quadrants, as shown.

Figure 22.3



As mentioned earlier, the correlation between X and Y will be positive if low (high) values of X are associated with low (high) values of Y. In terms of the above figure, we can say that when values of X that are greater (less) than $\overline{X}$ are generally associated with values of Y that are greater (less) than $\overline{Y}$, the correlation between X and Y will be positive. This implies that there will be a general tendency of points to concentrate in I and III quadrants. Similarly, when correlation between X and Y is negative, the point of the scatter diagram will have a general tendency to concentrate in II and IV quadrants.

Further, if we consider deviations of values from their means, i.e., $\left(X_i - \overline{X}\right)$ and $\left(Y_i - \overline{Y}\right)$, we note that:

(i) Both $\left(X_i - \overline{X}\right)$ and $\left(Y_i - \overline{Y}\right)$ will be positive for all points in quadrant I.

(ii) $\left(X_i - \overline{X}\right)$ will be negative and $\left(Y_i - \overline{Y}\right)$ will be positive for all points in quadrant II.

(iii) Both $\left(X_i - \overline{X}\right)$ and $\left(Y_i - \overline{Y}\right)$ will be negative for all points in quadrant III.

(iv) $\left(X_i - \overline{X}\right)$ will be positive and $\left(Y_i - \overline{Y}\right)$ will be negative for all points in quadrant IV.

It is obvious from the above that the product of deviations, i.e., $\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)$ will be positive for points in quadrants I and III and negative for points in quadrants II and IV.

Since, for positive correlation, the points will tend to concentrate more in I and III quadrants than in II and IV, the sum of positive products of deviations will outweigh the sum of negative products of deviations. Thus, $\sum\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)$ will be positive for all the n observations.

Similarly, when correlation is negative, the points will tend to concentrate more in II and IV quadrants than in I and III. Thus, the sum of negative products of deviations will outweigh the sum of positive products and hence $\sum\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)$ will be negative for all the n observations.

Further, if there is no correlation, the sum of positive products of deviations will be equal to the sum of negative products of deviations such that $\sum\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)$ will be equal to zero.

Notes

On the basis of the above, we can consider $\sum(X_i - \bar{X})(Y_i - \bar{Y})$ as an absolute measure of correlation. This measure, like other absolute measures of dispersion, skewness, etc., will depend upon (i) the number of observations and (ii) the units of measurements of the variables.

In order to avoid its dependence on the number of observations, we take its average, i.e., $\frac{1}{n}\sum(X_i - \bar{X})(Y_i - \bar{Y})$. This term is called covariance in statistics and is denoted as Cov(X,Y).

To eliminate the effect of units of measurement of the variables, the covariance term is divided by the product of the standard deviation of X and the standard deviation of Y. The resulting expression is known as the Karl Pearson's coefficient of linear correlation or the product moment correlation coefficient or simply the coefficient of correlation, between X and Y.

$$r_{XY} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} \quad .... (1)$$

or $$r_{XY} = \frac{\frac{1}{n}\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\frac{1}{n}\sum(X_i - \bar{X})^2}\sqrt{\frac{1}{n}\sum(Y_i - \bar{Y})^2}} \quad .... (2)$$

Cancelling $\frac{1}{n}$ from the numerator and the denominator, we get

$$r_{XY} = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum(X_i - \bar{X})^2}\sqrt{\sum(Y_i - \bar{Y})^2}} \quad .... (3)$$

Consider $\sum(X_i - \bar{X})(Y_i - \bar{Y}) = \sum(X_i - \bar{X})Y_i - \bar{Y}\sum(X_i - \bar{X})$

$= \sum X_i Y_i - \bar{X}\sum Y_i$ (second term is zero)

$= \sum X_i Y_i - n\bar{X}\bar{Y} \quad \left(\sum Y_i = n\bar{Y}\right)$

Similarly we can write $\sum(X_i - \bar{X})^2 = \sum X_i^2 - n\bar{X}^2$

and $\sum(Y_i - \bar{Y})^2 = \sum Y_i^2 - n\bar{Y}^2$

Substituting these values in equation (3), we have

$$r_{XY} = \frac{\sum X_i Y_i - n\bar{X}\bar{Y}}{\sqrt{\left[\sum X_i^2 - n\bar{X}^2\right]}\sqrt{\left[\sum Y_i^2 - n\bar{Y}^2\right]}} \quad .... (4)$$

$$r_{XY} = \frac{\sum X_i Y_i - n \cdot \frac{\sum X_i}{n} \times \frac{\sum Y_i}{n}}{\sqrt{\sum X_i^2 - n\left(\frac{\sum X_i}{n}\right)^2}\sqrt{\sum Y_i^2 - n\left(\frac{\sum Y_i}{n}\right)^2}}$$

Notes

$$= \frac{\sum X_i Y_i - \frac{(\sum X_i)(\sum Y_i)}{n}}{\sqrt{\sum X_i^2 - \frac{(\sum X_i)^2}{n}}\sqrt{\sum Y_i^2 - \frac{(\sum Y_i)^2}{n}}} \quad .... (5)$$

On multiplication of numerator and denominator by n, we can write

$$r_{XY} = \frac{n\sum X_i Y_i - (\sum X_i)(\sum Y_i)}{\sqrt{n\sum X_i^2 - (\sum X_i)^2}\sqrt{n\sum Y_i^2 - (\sum Y_i)^2}} \quad .... (6)$$

Further, if we assume $x_i = X_i - \overline{X}$ and $y_i = Y_i - \overline{Y}$, equation (2), given above, can be written as

$$r_{XY} = \frac{\frac{1}{n}\sum x_i y_i}{\sqrt{\frac{1}{n}\sum x_i^2}\sqrt{\frac{1}{n}\sum y_i^2}} \quad .... (7)$$

$$\text{or} \quad r_{XY} = \frac{\sum x_i y_i}{\sqrt{\sum x_i^2}\sqrt{\sum y_i^2}} \quad .... (8)$$

$$\text{or} \quad r_{XY} = \frac{1}{n}\frac{\sum x_i y_i}{\sigma_x \sigma_y} \quad .... (9)$$

Equations (5) or (6) are often used for the calculation of correlation from raw data, while the use of the remaining equations depends upon the forms in which the data are available. For example, if standard deviations of X and Y are given, equation (9) may be appropriate.

Example 1: Calculate the Karl Pearson's coefficient of correlation from the following pairs of values :

Values of X :	12	9	8	10	11	13	7
Values of Y :	14	8	6	9	11	12	3

Solution.

The formula for Karl Pearson's coefficient of correlation is

$$r_{XY} = \frac{n\sum X_i Y_i - (\sum X_i)(\sum Y_i)}{\sqrt{n\sum X_i^2 - (\sum X_i)^2}\sqrt{n\sum Y_i^2 - (\sum Y_i)^2}}$$

Notes

The values of different terms, given in the formula, are calculated from the following table :

X_i	Y_i	X_iY_i	X_i^2	Y_i^2
12	14	168	144	196
9	8	72	81	64
8	6	48	64	36
10	9	90	100	81
11	11	121	121	121
13	12	156	169	144
7	3	21	49	9
70	63	676	728	651

Here n = 7 (no. of pairs of observations)

$$r_{XY} = \frac{7 \times 676 - 70 \times 63}{\sqrt{7 \times 728 - (70)^2}\sqrt{7 \times 651 - (63)^2}} = 0.949$$

Example 2: Calculate the Karl Pearson's coefficient of correlation between X and Y from the following data:

No. of pairs of observations n = 8, $\sum\left(X_i - \bar{X}\right)^2 = 184$, $\sum\left(Y_i - \bar{Y}\right)^2 = 148$,

$\sum\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right) = 164$, $\bar{X} = 11$ and $\bar{Y} = 10$

Solution.

Using the formula, $r_{XY} = \dfrac{\sum\left(X - \bar{X}_i\right)\left(Y_i - \bar{Y}\right)}{\sqrt{\sum\left(X_i - \bar{X}\right)^2}\sqrt{\sum\left(Y_i - \bar{Y}\right)^2}}$, we get

$$r_{XY} = \frac{164}{\sqrt{184}\sqrt{148}} = 0.99$$

Example 3:

(a) The covariance between the length and weight of five items is 6 and their standard deviations are 2.45 and 2.61 respectively. Find the coefficient of correlation between length and weight.

(b) The Karl Pearson's coefficient of correlation and covariance between two variables X and Y is – 0.85 and – 15 respectively. If variance of Y is 9, find the standard deviation of X.

Solution.

(a) Substituting the given values in formula (1) for correlation, we get

$$r_{XY} = \frac{6}{2.45 \times 2.61} = 0.94$$

Notes

(b) Substituting the given values in the formula of correlation, we get

$$-0.85 = \frac{-15}{\sigma_X \times 3} \quad \text{or} \quad s_X = 5.88$$

22.4 Properties of Coefficient of Correlation

1. The coefficient of correlation is independent of the change of origin and scale of measurements.

   In order to prove this property, we change origin and scale of both the variables X and Y.

   Let $u_i = \frac{X_i - A}{h}$ and $v_i = \frac{Y_i - B}{k}$, where the constants A and B refer to change of origin and the constants h and k refer to change of scale. We can write

   $$X_i = A + hu_i, \quad \therefore \quad \bar{X} = A + h\bar{u}$$

   Thus, we have $X_i - \bar{X} = A + hu_i - A - h\bar{u} = h(u_i - \bar{u})$

   Similarly, $Y_i = B + kv_i, \quad \therefore \quad \bar{Y} = B + k\bar{v}$

   Thus, $Y_i - \bar{Y} = B + kv_i - B - k\bar{v} = k(v_i - \bar{v})$

   The formula for the coefficient of correlation between X and Y is

   $$r_{XY} = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum(X_i - \bar{X})^2}\sqrt{\sum(Y_i - \bar{Y})^2}}$$

   Substituting the values of $(X_i - \bar{X})$ and $(Y_i - \bar{Y})$, we get

   $$r_{XY} = \frac{\sum h(u_i - \bar{u})k(v_i - \bar{v})}{\sqrt{\sum h^2(u_i - \bar{u})^2}\sqrt{\sum k^2(v_i - \bar{v})^2}} = \frac{\sum(u_i - \bar{u})(v_i - \bar{v})}{\sqrt{\sum(u_i - \bar{u})^2}\sqrt{\sum(v_i - \bar{v})^2}}$$

   $$\therefore \quad r_{XY} = r_{uv}$$

   This shows that correlation between X and Y is equal to correlation between u and v, where u and v are the variables obtained by change of origin and scale of the variables X and Y respectively.

   This property is very useful in the simplification of computations of correlation. On the basis of this property, we can write a short-cut formula for the computation of r_{XY} :

   $$r_{XY} = \frac{n\sum u_i v_i - (\sum u_i)(\sum v_i)}{\sqrt{n\sum u_i^2 - (\sum u_i)^2}\sqrt{n\sum v_i^2 - (\sum v_i)^2}} \quad \text{.... (10)}$$

2. The coefficient of correlation lies between - 1 and + 1.

   To prove this property, we define

   $$x'_i = \frac{X_i - \bar{X}}{\sigma_X} \quad \text{and} \quad y'_i = \frac{Y_i - \bar{Y}}{\sigma_Y}$$

$$\therefore \qquad x_i'^2 = \frac{\left(X_i - \bar{X}\right)^2}{\sigma_X^2} \text{ and } y_i'^2 = \frac{\left(Y_i - \bar{Y}\right)^2}{\sigma_Y^2}$$

$$\text{or} \qquad \sum x_i'^2 = \frac{\sum\left(X_i - \bar{X}\right)^2}{\sigma_X^2} \text{ and } \sum y_i'^2 = \frac{\sum\left(Y_i - \bar{Y}\right)^2}{\sigma_Y^2}$$

From these summations we can write $\sum x_i'^2 = \sum y_i'^2 = n$

$$\text{Also, } r = \frac{\frac{1}{n}\sum\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{\sigma_X\sigma_Y} = \frac{1}{n}\cdot\sum\left(\frac{X_i - \bar{X}}{\sigma_X}\right)\left(\frac{Y_i - Y}{\sigma_Y}\right) = \frac{1}{n}\sum x'_i y'_i$$

Consider the sum $x_i' + y_i'$. The square of this sum is always a non-negative number, i.e., $(x_i' + y_i')^{2\ 3}$ 0.

Taking sum over all the observations and dividing by n, we get

$$\frac{1}{n}\sum\left(x'_i + y'_i\right)^2 \geq 0 \quad \text{or} \quad \frac{1}{n}\sum\left(x_i'^2 + y_i'^2 + 2x'_i y'_i\right) \geq 0$$

$$\text{or} \qquad \frac{1}{n}\sum x_i'^2 + \frac{1}{n}\sum y_i'^2 + \frac{2}{n}\sum x'_i y'_i \geq 0$$

or $\quad 1 + 1 + 2r \geq 0$ or $2 + 2r \geq 0$ or $r \geq -1$ .... (11)

Further, consider the difference $x_i' - y_i'$. The square of this difference is also non-negative, i.e., $(x_i' - y_i')^{2\ 3}$ 0.

Taking sum over all the observations and dividing by n, we get

$$\frac{1}{n}\sum\left(x'_i - y'_i\right)^2 \geq 0 \quad \text{or} \quad \frac{1}{n}\sum\left(x_i'^2 + y_i'^2 - 2x'_i y'_i\right) \geq 0$$

$$\text{or} \qquad \frac{1}{n}\sum x_i'^2 + \frac{1}{n}\sum y_i'^2 - \frac{2}{n}\sum x'_i y'_i \geq 0$$

or $\quad 1 + 1 - 2r \geq 0$ or $2 - 2r \geq 0$ or $r \leq 1$ .... (12)

Combining the inequalities (11) and (12), we get $-1 \leq r \leq 1$. Hence r lies between -1 and +1.

3. If X and Y are independent they are uncorrelated, but the converse is not true.

   If X and Y are independent, it implies that they do not reveal any tendency of simultaneous movement either in same or in opposite directions. In terms of figure 12.3, the dots of the scatter diagram will be uniformly spread in all the four quadrants. Therefore, $\sum\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)$ or Cov(X, Y) will be equal to zero and hence, $r_{XY} = 0$. Thus, if X and Y are independent, they are uncorrelated.

   The converse of this property implies that if $r_{XY} = 0$, then X and Y may not necessarily be independent. To prove this, we consider the following data :

X	1	2	3	4	5	6	7
Y	9	4	1	0	1	4	9

Notes

Here $SX_i = 28$, $SY_i = 28$ and $SX_iY_i = 112$.

$$\therefore \ \text{Cov}(X,Y) = \frac{1}{n}\left[\sum X_iY_i - \frac{(\sum X_i)(\sum Y_i)}{n}\right] = \frac{1}{7}\left[112 - \frac{28 \times 28}{7}\right] = 0 \text{ . Thus, } r_{XY} = 0$$

A close examination of the given data would reveal that although $r_{XY} = 0$, but X and Y are not independent. In fact they are related by the mathematical relation $Y = (X - 4)^2$.

Remarks: This property points our attention to the fact that r_{XY} is only a measure of the degree of linear association between X and Y. If the association is non-linear, the computed value of r_{XY} is no longer a measure of the degree of association between the two variables.

Example 4:

Calculate the Karl Pearson's coefficient of correlation from the following data:

Height of fathers (inches)	:	66	68	69	72	65	59	62	67	61	71
Height of sons (inches)	:	65	64	67	69	64	60	59	68	60	64

Solution.

Note: When there is no common factor, we can take h = k = 1 and define $u_i = X_i - A$ and $v_i = Y_i - B$.

Calculation of r

Height of fathers (X_i)	*Height of sons* (Y_i)	$u_i = X_i - 65$	$v_i = Y_i - 64$	u_iv_i	u_i^2	v_i^2
66	65	1	1	1	1	1
68	64	3	0	0	9	0
69	67	4	3	12	16	9
72	69	7	5	35	49	25
65	64	0	0	0	0	0
59	60	- 6	- 4	24	36	16
62	59	- 3	- 5	15	9	25
67	68	2	4	8	4	16
61	60	- 4	- 4	16	16	16
71	64	6	0	0	36	0
Total		10	0	111	176	108

Here n = 10. Using formula (10) for correlation, we get

$$= \frac{10 \times 111 - 10 \times 0}{\sqrt{10 \times 176 - (10)^2}\sqrt{10 \times 108 - 0^2}} = 0.83$$

Example 5:

(a) Calculate the Karl Pearson's coefficient of correlation from the following data:

(i) Sum of deviations of X values = 5

(ii) Sum of deviations of Y values = 4

(iii) Sum of squares of deviations of X values = 40

(iv) Sum of squares of deviations of Y values = 50

Notes

(v) Sum of the product of deviations of X and Y values = 32

(vi) No. of pairs of observations = 10

(b) Given the following, calculate the coefficient of correlation :

(i) Sum of squares of deviations of X values from mean = 136

(ii) Sum of squares of deviations of Y values from mean = 138

(iii) Sum of products of deviations of X and Y values from their means = 122.

Solution.

(a) Let $u_i = X_i - A$ and $v_i = Y_i - B$ be the deviations of X and Y values. We are given $Su_i = 5$, $Sv_i = 4$, $Su_i^2 = 40$, $Sv_i^2 = 50$, $Su_iv_i = 32$ and n = 10.

Substituting these values in formula (10), we get

$$r_{XY} = \frac{10 \times 32 - 5 \times 4}{\sqrt{10 \times 40 - 5^2}\sqrt{10 \times 50 - 4^2}} = 0.704$$

(b) Using formula (3) for correlation, we get $r = \frac{122}{\sqrt{136}\sqrt{138}} = 0.89$

Example 6: Calculate the coefficient of correlation between age group and rate of mortality from the following data:

Age group	:	0-20	20-40	40-60	60-80	80-100
Rate of Mortality	:	350	280	540	760	900

Solution.

Since class intervals are given for age, their mid-values shall be used for the calculation of r.

Table for calculation of r

Age group	*M.V. (X)*	*Rate of Mort. (Y)*	$u_i = \frac{X_i - 50}{20}$	$v_i = \frac{Y_i - 540}{10}$	u_iv_i	u_i^2	v_i^2
0 - 20	10	350	- 2	- 19	38	4	361
20 - 40	30	280	- 1	- 26	26	1	676
40 - 60	50	540	0	0	0	0	0
60 - 80	70	760	1	22	22	1	484
80 - 100	90	900	2	36	72	4	1296
Total			0	13	158	10	2817

Here n = 5. Using the formula (10) for correlation, we get

$$r_{XY} = \frac{5 \times 158 - 0 \times 13}{\sqrt{5 \times 10 - 0^2}\sqrt{5 \times 2817 - 13^2}} = 0.95$$

Example 7:

Deviations from assumed average of the two series are given below :

Deviations, X series : - 10, - 6, - 4, - 1, 0, + 2, + 1, + 5, + 7, + 11

Deviations, Y series : - 8, - 5, + 4, - 2, - 4, 0, + 2, 0, - 2, + 4

Find out Karl Pearson's coefficient of correlation.

Solution.

Here the values of $u_i = X_i - A$ and $v_i = X_i - B$ are given.

Table for calculation of r

u_i	- 10	- 6	- 4	- 1	0	2	1	5	7	11	5
v_i	- 8	- 5	4	- 2	- 4	0	2	0	- 2	4	- 11
$u_i v_i$	80	30	- 16	2	0	0	2	0	- 14	44	128
u_i^2	100	36	16	1	0	4	1	25	49	121	353
v_i^2	64	25	16	4	16	0	4	0	4	16	149

Here n = 10.

$$r_{XY} = \frac{10 \times 128 - 5 \times (-11)}{\sqrt{10 \times 353 - 5^2}\sqrt{10 \times 149 - 11^2}} = 0.609$$

Example 8:

From the following table, find the missing values and calculate the coefficient of correlation by Karl Pearson's method :

X :	6	2	10	4	?
Y :	9	11	?	8	7

Arithmetic means of X and Y series are 6 and 8 respectively.

Solution.

The missing value in X - series = 5 × 6 – (6 + 2 + 10 + 4) = 30 – 22 = 8

The missing value in Y - series = 5 × 8 – (9 + 11 + 8 + 7) = 40 – 35 = 5

Table for calculation of r

X	Y	$X - \overline{X}$	$(Y - \overline{Y})$	$(X - \overline{X})(Y - \overline{Y})$	$(X - \overline{X})^2$	$(Y - \overline{Y})^2$
6	9	0	1	0	0	1
2	11	- 4	3	- 12	16	9
10	5	4	- 3	- 12	16	9
4	8	- 2	0	0	4	0
8	7	2	- 1	- 2	4	1
Total				- 26	40	20

Using formula (3) for correlation, we get $r = \frac{-26}{\sqrt{40}\sqrt{20}} = -0.92$

Notes

Example 9:

Calculate Karl Pearson's coefficient of correlation for the following series :

Price (in Rs)	:	10	11	12	13	14	15	16	17	18	19
Demand (in kgs)	:	420	410	400	310	280	260	240	210	210	200

Solution.

Table for calculation of r

Price (X)	*Demand* (Y)	$u = X - 14$	$v = \frac{Y - 310}{10}$	uv	u^2	v^2
10	420	-4	11	-44	16	121
11	410	-3	10	-30	9	100
12	400	-2	9	-18	4	81
13	310	-1	0	0	1	0
14	280	0	-3	0	0	9
15	260	1	-5	-5	1	25
16	240	2	-7	-14	4	49
17	210	3	-10	-30	9	100
18	210	4	-10	-40	16	100
19	200	5	-11	-55	25	121
Total		5	-16	-236	85	706

$$r = \frac{-10 \times 236 + 5 \times 16}{\sqrt{10 \times 85 - 25}\sqrt{10 \times 706 - 256}} = -0.96$$

Example 10:

A computer while calculating the correlation coefficient between two variables, X and Y, obtained the following results :

$$n = 25, \Sigma X = 125, \Sigma X^2 = 650, \Sigma Y = 100, \Sigma Y^2 = 460, \Sigma XY = 508.$$

It was, however, discovered later at the time of checking that it had copied down two pairs of observations as $\begin{array}{c|c} X & Y \\ \hline 6 & 14 \\ 8 & 6 \end{array}$ in place of the correct pairs $\begin{array}{c|c} X & Y \\ \hline 8 & 12 \\ 6 & 8 \end{array}$. Obtain the correct value of r.

Solution.

First we have to correct the values of ΣX, ΣX^2......etc.

Corrected $\Sigma X = 125 - (6 + 8) + (8 + 6) = 125$

Corrected $\Sigma X^2 = 650 - (36 + 64) + (64 + 36) = 650$

Corrected $\Sigma Y = 100 - (14 + 6) + (12 + 8) = 100$

Corrected $\Sigma Y^2 = 460 - (196 + 36) + (144 + 64) = 436$

Corrected SXY = 508 - (84 + 48) + (96 + 48) = 520

$$r = \frac{25 \times 520 - 125 \times 100}{\sqrt{25 \times 650 - (125)^2}\sqrt{25 \times 436 - (100)^2}} = 0.67$$

Notes

22.5 Probable Error of r

It is an old measure to test the significance of a particular value of r without the knowledge of test of hypothesis. Probable error of r, denoted by P.E.(r) is 0.6745 times its standard error. The value 0.6745 is obtained from the fact that in a normal distribution $\bar{r} \pm 0.6745 \times \text{S.E.}$ covers 50% of the total distribution.

According to Horace Secrist "The probable error of correlation coefficient is an amount which if added to and subtracted from the mean correlation coefficient, gives limits within which the chances are even that a coefficient of correlation from a series selected at random will fall."

Since standard error of r, i.e., $\text{S.E.}_r = \frac{1-r^2}{\sqrt{n}}$, $\therefore \text{P.E.}(r) = 0.6745 \times \frac{1-r^2}{\sqrt{n}}$

22.5.1 Uses of P.E.(r)

(i) It can be used to specify the limits of population correlation coefficient ρ (rho) which are defined as r – P.E.(r) ≤ r ≤ r + P.E.(r), where ρ denotes correlation coefficient in population and r denotes correlation coefficient in sample.

(ii) It can be used to test the significance of an observed value of r without the knowledge of test of hypothesis. By convention, the rules are:

(a) If |r| < 6 P.E.(r), then correlation is not significant and this may be treated as a situation of no correlation between the two variables.

(b) If |r| > 6 P.E.(r), then correlation is significant and this implies presence of a strong correlation between the two variables.

(c) If correlation coefficient is greater than 0.3 and probable error is relatively small, the correlation coefficient should be considered as significant.

Example 11: Find out correlation between age and playing habit from the following information and also its probable error.

Age	:	15	16	17	18	19	20
No. of Students	:	250	200	150	120	100	80
Regular Players	:	200	150	90	48	30	12

Solution.

Let X denote age, p the number of regular players and q the number of students. Playing habit, denoted by Y, is measured as a percentage of regular players in an age group, i.e., Y = (p/q)×100.

Table for calculation of r

X	q	p	Y	$u = X - 17$	$v = Y - 40$	uv	u^2	v^2
15	250	200	80	-2	40	-80	4	1600
16	200	150	75	-1	35	-35	1	1225
17	150	90	60	0	20	0	0	400
18	120	48	40	1	0	0	1	0
19	100	30	30	2	-10	-20	4	100
20	80	12	15	3	-25	-75	9	625
Total				3	60	-210	19	3950

Notes

$$r_{XY} = \frac{-6 \times 210 - 3 \times 60}{\sqrt{6 \times 19 - 9}\sqrt{6 \times 3950 - 3600}} = -0.99$$

Probable error of r, i.e., $\text{P.E.}(r) = 0.6745 \times \frac{\left[1-(0.99)^2\right]}{\sqrt{6}} = 0.0055$

Example 12:

Test the significance of correlation for the values based on the number of observations (i) 10, and (ii) 100 and r = 0.4 and 0.9.

Solution.

(i) (a) Consider n = 10 and r = 0.4. Thus, $\text{P.E.}(r) = 0.6745 \times \frac{1-0.4^2}{\sqrt{10}} = 0.179$ and 6 P.E. = 6 × 0.179 = 1.074. Since $|r| < 6$ P.E., r is not significant.

(i) (b) Take n = 10 and r = 0.9. Thus, $\text{P.E.} = 0.6745 \times \frac{1-0.9^2}{\sqrt{10}} = 0.041$ and 6 P.E. = 6 × 0.041 = 0.246. Since $|r| > 6$ P.E., r is highly significant.

(ii) (a) Take n = 100 and r = 0.4. Thus, $6\text{P.E.} = 6 \times 0.6745 \frac{\left(1-0.4^2\right)}{\sqrt{100}} = 0.34$

Since $|r| > 6$ P.E., r is significant.

(ii) (b) Take n = 100 and r = 0.9. Thus, $6\text{P.E.} = 6 \times 0.6745 \frac{\left(1-0.9^2\right)}{\sqrt{100}} = 0.077$

Since $|r| > 6$ P.E., r is significant.

22.6 Correlation in a Bivariate Frequency Distribution

Let the two variables X and Y take respective values X_i, i = 1, 2, ... m and Y_j, j = 1, 2, ... n. These values, taken together, will make m´n pairs (X_i, Y_j). Let f_{ij} be the frequency of this pair. This frequency distribution can be presented in a tabular form as given below :

$Y \rightarrow$ / $X \downarrow$	Y_1	Y_2	$\cdots$	Y_j	$\cdots$	Y_n	*Total*
X_1	f_{11}	f_{12}	$\cdots$	f_{1j}	$\cdots$	f_{1n}	f_1
X_2	f_{21}	f_{22}	$\cdots$	f_{2j}	$\cdots$	f_{2n}	f_2
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
X_i	f_{i1}	f_{i2}		f_{ij}		f_{in}	f_i
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
X_m	f_{m1}	f_{m2}	$\cdots$	f_{mj}	$\cdots$	f_{mn}	f_m
Total	f_1'	f_2'	$\cdots$	f_j'	$\cdots$	f_n'	N

Here $\sum\sum f_{ij} = \sum f_i = \sum f_j' = N$ (the total frequency).

Notes

The formula for correlation can be written on the basis of the formula discussed earlier.

$$r_{XY} = \frac{N\sum\sum f_{ij}X_iY_j - \left(\sum f_iX_i\right)\left(\sum f_j'Y_j\right)}{\sqrt{N\sum f_iX_i^2 - \left(\sum f_iX_i\right)^2}\sqrt{N\sum f_j'Y_j^2 - \left(\sum f_j'Y_j\right)^2}}$$

When we make changes of origin and scale by making the transformations $u_i = \frac{X_i - A}{h}$ and $v_j = \frac{Y_j - B}{k}$, then we can write

$$r_{XY} = \frac{N\sum\sum f_{ij}u_iv_j - \left(\sum f_iu_i\right)\left(\sum f_j'v_j\right)}{\sqrt{N\sum f_iu_i^2 - \left(\sum f_iu_i\right)^2}\sqrt{N\sum f_j'v_j^2 - \left(\sum f_j'v_j\right)^2}}$$

Example 13:

Calculate Karl Pearson's coefficient of correlation from the following data :

Age(yrs) → *Marks* ↓	18	19	20	21	22
20 - 25	3	2			
15 - 20		5	4		
10 - 15			7	10	
5 - 10				3	2
0 - 5					4

Solution.

Let X_i denote the mid-value of the class interval of marks. Various values of X_i can be written as 22.5, 17.5, 12.5, 7.5 and 2.5.

Further, let $u_i = (X_i - 12.5) \div 5$. Various values of u_i would be 2, 1, 0, - 1 and - 2.

Similarly, let Y_j denote age. Various values of Y_j are 18, 19, 20, 21 and 22.

Assuming $v_j = Y_j - 20$, various values of v_j would be - 2, - 1, 0, 1 and 2.

We shall use the values of u_i and v_j in the computation of r.

Notes

Table for Calculation of r

u_i \ v_j	−2	−1	0	1	2	f_i	f_iu_i	$f_iu_i^2$	$f_{ij}u_iv_j$
2	3 (−12)	2 (−4)	...	...	...	5	10	20	−16
1	...	5 (−5)	4 (0)	...	...	9	9	9	−5
0	...	...	7 (0)	10 (0)	...	17	0	0	0
−1	...	...	...	3 (−3)	2 (−4)	5	−5	5	−7
−2	...	...	...	...	4 (−16)	4	−8	16	−16
f_j'	3	7	11	13	6	40	6	50	−44
$f_j'v_j$	−6	−7	0	13	12	12			
$f_j'v_j^2$	12	7	0	13	24	56			

Substituting various values in the formula for r, we get

$$r = \frac{40\times(-44)-6\times 12}{\sqrt{40\times 50-36}\sqrt{40\times 56-144}} = \frac{-1832}{\sqrt{1964}\sqrt{2096}} = -0.903$$

Examples 14:

Given the following data, compute the coefficient of correlation r, between X and Y.

Y → X ↓	30-50	50-70	70-90	Total
0-5	10	6	2	18
5-10	3	5	4	12
10-15	4	7	9	20
Total	17	18	15	50

Solution.

Note: Instead of doing the computation work in a single table, as done in example 13, it can be split into the following steps:

Taking mid-values of the class intervals, we have

Mid-values (X) : 2.5 7.5 12.5

Mid-values (Y) : 40 60 80

Let $u_i = \frac{X_i - 7.5}{5}$ and $v_i = \frac{Y_i - 60}{20}$

∴ various u values are : - 1 0 1

and various v values are : - 1 0 1

Notes

(i) Calculation of $SSf_{ij}u_iv_j$

u_i \ v_j	−1	0	1	Total
−1	10 [10]	6 [0]	2 [−2]	8
0	3 [0]	5 [0]	4 [0]	0
1	4 [−4]	7 [0]	9 [9]	5
Total	6	0	7	13

\ $Sf_{ij}u_iv_j = 13$

(ii) Calculation of Sf_iu_i and $Sf_iu_i^2$

u_i	f_i	f_iu_i	$f_iu_i^2$
−1	18	−18	18
0	12	0	0
1	20	20	20
Total	50	2	38

(iii) Calculation of $Sf_j'v_j$ and $Sf_j'v_j^2$

v_j	f_j'	$f_j'v_j$	$f_j'v_j^2$
−1	17	−17	17
0	18	0	0
1	15	15	15
Total	50	−2	32

Substituting these values in the formula of r, we have

$$r = \frac{50 \times 13 - 2 \times (-2)}{\sqrt{50 \times 38 - 4}\sqrt{50 \times 32 - 4}} = \frac{654}{\sqrt{1896}\sqrt{1596}} = 0.376$$

22.7 Merits and Limitations of Coefficient of Correlation

The only merit of Karl Pearson's coefficient of correlation is that it is the most popular method for expressing the degree and direction of linear association between the two variables in terms of a pure number, independent of units of the variables. This measure, however, suffers from certain limitations, given below :

1. Coefficient of correlation r does not give any idea about the existence of cause and effect relationship between the variables. It is possible that a high value of r is obtained although none of them seem to be directly affecting the other. Hence, any interpretation of r should be done very carefully.

2. It is only a measure of the degree of linear relationship between two variables. If the relationship is not linear, the calculation of r does not have any meaning.

3. Its value is unduly affected by extreme items.

4. If the data are not uniformly spread in the relevant quadrants (see - Fig 12.3), the value of r may give a misleading interpretation of the degree of relationship between the two variables. For example, if there are some values having concentration around a point in first quadrant and there is similar type of concentration in third quadrant, the value of r will be very high although there may be no linear relation between the variables.

5. As compared with other methods, to be discussed later in this chapter, the computations of r are cumbersome and time consuming.

Notes

22.8 Spearman's Rank Correlation

This is a crude method of computing correlation between two characteristics. In this method, various items are assigned ranks according to the two characteristics and a correlation is computed between these ranks. This method is often used in the following circumstances:

(i) When the quantitative measurements of the characteristics are not possible, e.g., the results of a beauty contest where various individuals can only be ranked.

(ii) Even when the characteristics is measurable, it is desirable to avoid such measurements due to shortage of time, money, complexities of calculations due to large data, etc.

(iii) When the given data consist of some extreme observations, the value of Karl Pearson's coefficient is likely to be unduly affected. In such a situation the computation of the rank correlation is preferred because it will give less importance to the extreme observations.

(iv) It is used as a measure of the degree of association in situations where the nature of population, from which data are collected, is not known.

The coefficient of correlation obtained on the basis of ranks is called 'Spearman's Rank Correlation' or simply the 'Rank Correlation'. This correlation is denoted by ρ (rho).

Let X_i be the rank of i th individual according to the characteristics X and Y_i be its rank according to the characteristics Y. If there are n individuals, there would be n pairs of ranks (X_i, Y_i), i = 1, 2, ... n. We assume here that there are no ties, i.e., no two or more individuals are tied to a particular rank. Thus, X_i's and Y_i's are simply integers from 1 to n, appearing in any order.

The means of X and Y, i.e., $\overline{X} = \overline{Y} = \frac{1+2+\cdots\cdots n}{n} = \frac{n(n+1)}{2n} = \frac{n+1}{2}$. Also,

$$\sigma_X^2 = \sigma_Y^2 = \frac{1}{n}[1^2 + 2^2 + \cdots + n^2] - \frac{(n+1)^2}{4} = \frac{1}{n}\left[\frac{n(n+1)(2n+1)}{6}\right] - \frac{(n+1)^2}{4} = \frac{n^2-1}{12}$$

Let d_i be the difference in ranks of the i th individual, i.e.,

$$d_i = X_i - Y_i = \left(X_i - \overline{X}\right) - \left(Y_i - \overline{Y}\right) \qquad \left(\because \overline{X} = \overline{Y}\right)$$

Squaring both sides and taking sum over all the observations, we get

$$\sum d_i^2 = \sum\left[\left(X_i - \overline{X}\right) - \left(Y_i - \overline{Y}\right)\right]^2$$

$$= \sum\left(X_i - \overline{X}\right)^2 + \sum\left(Y_i - \overline{Y}\right)^2 - 2\sum\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)$$

Dividing both sides by n, we get

$$\frac{1}{n}\sum d_i^2 = \frac{1}{n}\sum\left(X_i - \overline{X}\right)^2 + \frac{1}{n}\sum\left(Y_i - \overline{Y}\right)^2 - \frac{2}{n}\sum\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)$$

$$= \sigma_X^2 + \sigma_Y^2 - 2Cov(X,Y) = 2\sigma_X^2 - 2Cov(X,Y) \qquad \left(\because \sigma_X^2 = \sigma_Y^2\right)$$

$$= 2\sigma_X^2 - 2\rho\sigma_X\sigma_Y = 2\sigma_X^2 - 2\rho\sigma_X^2 = 2\sigma_X^2(1-\rho) \qquad \left(\because \rho = \frac{Cov(X,Y)}{\sigma_X\sigma_Y}\right)$$

Notes

From this, we can write $1-\rho = \frac{1}{n} \times \frac{\sum d_i^2}{2\sigma_X^2}$

$$or \quad \rho = 1-\frac{1}{n} \times \frac{\sum d_i^2}{2\sigma_X^2} = 1-\frac{1}{n} \times \frac{\sum d_i^2}{2} \times \frac{12}{n^2-1} = 1-\frac{6\sum d_i^2}{n\left(n^2-1\right)}$$

Note: This formula is not applicable in case of a bivariate frequency distribution.

Example 15:

The following table gives the marks obtained by 10 students in commerce and statistics. Calculate the rank correlation.

Marks in Statistics	:	35	90	70	40	95	45	60	85	80	50
Marks in Commerce	:	45	70	65	30	90	40	50	75	85	60

Solution.

Calculation Table

Marks in Statistics	*Marks in Commerce*	*Rank of Marks in*		$d_i = X_i - \overline{Y}_i$	d_i^2
		Statistics X	*Commerce Y*		
35	45	1	3	–2	4
90	70	9	7	2	4
70	65	6	6	0	0
40	30	2	1	1	1
95	90	10	10	0	0
45	40	3	2	1	1
60	50	5	4	1	1
85	75	8	8	0	0
80	85	7	9	–2	4
50	60	4	5	–1	1

From the above table, we have $\sum d_i^2 = 16$.

$\therefore$ Rank Correlation $\rho = \mathbf{1 - \frac{6\sum d_i^2}{n\left(n^2-1\right)} = 1 - \frac{6\times 16}{10\times 99} = 0.903}$

22.9 Coefficient of Correlation by Concurrent Deviation Method

This is another simple method of obtaining a quick but crude idea of correlation between two variables. In this method, only direction of change in the concerned variables are noted by comparing a value from its preceding value. If the value is greater than its preceding value, it is indicated by a '+' sign; if less, it is indicated by a '-' sign and equal values are indicated by '=' sign. All the pairs having same signs, i.e., either both the deviations are positive or negative or have equal sign ('='), are known as concurrent deviations and are indicated by '+' sign in a separate column designated as 'concurrences'. The number of such concurrences is denoted by C. Similarly, the remaining pairs are marked by '-' sign in another column designated as 'disagreements'. The coefficient of correlation, denoted by r_C, is given by the formula $r_C = \pm\sqrt{\pm\left(\frac{2C-D}{D}\right)}$, where C denotes the number of concurrences and D (= number of observations - 1) is the number of pairs of deviation.

Note:

(i) The sign of r_C is taken to be equal to the sign of $\left(\frac{2C - D}{D}\right)$.

(ii) When $\left(\frac{2C - D}{D}\right)$ is negative, we make it positive for the purpose of taking its square root. However, the computed value will have a negative sign.

(iii) The sign of r_C will be positive when $\left(\frac{2C - D}{D}\right)$ is positive.

(iv) This method gives same weights to smaller as well as to the larger deviations.

(v) This method is suitable only for the study of short term fluctuations because it does not take into account the changes in magnitudes of the values.

Example 18:

The following table gives the marks obtained by 11 students of a class in micro and macro-economics papers. Calculate the coefficient of correlation by concurrent deviation method.

Roll No.	:	1	2	3	4	5	6	7	8	9	10	11
Marks in Micro - economics	:	80	45	55	56	58	60	65	68	70	75	85
Marks in Macro - economics	:	82	56	50	48	60	62	64	65	70	74	90

Solution.

Let D_1 and D_2 denote deviations from the preceding marks in micro and macro economics respectively.

Calculation Table

Roll No.	*Marks in Micro - economics*	D_1	*Marks in Macro - economics*	D_2	*Concurr - ences*	*Disagree - ments*
1	80		82			
2	45	-	56	-	+	
3	55	+	50	-		-
4	56	+	48	-		-
5	58	+	60	+	+	
6	60	+	62	+	+	
7	65	+	64	+	+	
8	68	+	65	+	+	
9	70	+	70	+	+	
10	75	+	74	+	+	
11	85	+	90	+	+	
			Total		8	2

Here C = 8 and the no. of pairs of deviation D = 10.

Now, $\frac{2C - D}{D} = \frac{16 - 10}{10} = 0.6$ which is positive, $\therefore\ r_C = \sqrt{0.6} = 0.77$

This value indicates the presence of a very high positive correlation between the marks obtained in two papers.

Example 19:

Find out the coefficient of correlation by concurrent deviation method from the following information:

Number of pairs of deviations = 96

Number of concurrent deviations = 32

Solution.

We are given C = 32 and D = 96

Now, $\frac{2C-D}{D} = \frac{64-96}{96} = -\frac{1}{3}$ which is negative, $\therefore\ r_C = -\sqrt{\frac{1}{3}} = -0.577$

22.10 Summary

- ***One of the variable may be affecting the other:*** A correlation coefficient calculated from the data on quantity demanded and corresponding price of tea would only reveal that the degree of association between them is very high. It will not give us any idea about whether price is affecting demand of tea or vice-versa. In order to know this, we need to have some additional information apart from the study of correlation. For example if, on the basis of some additional information, we say that the price of tea affects its demand, then price will be the cause and quantity will be the effect. The causal variable is also termed as independent variable while the other variable is termed as dependent variable.
- ***The two variables may act upon each other:*** Cause and effect relation exists in this case also but it may be very difficult to find out which of the two variables is independent. For example, if we have data on price of wheat and its cost of production, the correlation between them may be very high because higher price of wheat may attract farmers to produce more wheat and more production of wheat may mean higher cost of production, assuming that it is an increasing cost industry. Further, the higher cost of production may in turn raise the price of wheat. For the purpose of determining a relationship between the two variables in such situations, we can take any one of them as independent variable.
- ***The two variables may be acted upon by the outside influences:*** In this case we might get a high value of correlation between the two variables, however, apparently no cause and effect type relation seems to exist between them. For example, the demands of the two commodities, say X and Y, may be positively correlated because the incomes of the consumers are rising. Coefficient of correlation obtained in such a situation is called a spurious or nonsense correlation.
- ***A high value of the correlation coefficient may be obtained due to sheer coincidence (or pure chance):*** This is another situation of spurious correlation. Given the data on any two variables, one may obtain a high value of correlation coefficient when in fact they do not have any relationship. For example, a high value of correlation coefficient may be obtained between the size of shoe and the income of persons of a locality.
- Let the bivariate data be denoted by (X_i, Y_i), where i = 1, 2 ...... n. In order to have some idea about the extent of association between variables X and Y, each pair (X_i, Y_i), i = 1, 2......n, is plotted on a graph. The diagram, thus obtained, is called a Scatter Diagram.

Notes

- Each pair of values (X_i, Y_i) is denoted by a point on the graph. The set of such points (also known as dots of the diagram) may cluster around a straight line or a curve or may not show any tendency of association.
- It can be used to specify the limits of population correlation coefficient ρ (rho) which are defined as $r - P.E.(r) \leq r \leq r + P.E.(r)$, where ρ denotes correlation coefficient in population and r denotes correlation coefficient in sample.
- It can be used to test the significance of an observed value of r without the knowledge of test of hypothesis. By convention, the rules are:
  - If $|r| < 6$ P.E.(r), then correlation is not significant and this may be treated as a situation of no correlation between the two variables.
  - If $|r| > 6$ P.E.(r), then correlation is significant and this implies presence of a strong correlation between the two variables.
  - If correlation coefficient is greater than 0.3 and probable error is relatively small, the correlation coefficient should be considered as significant.

22.11 Keywords

Correlation: It is an analysis of covariation between two or more variables.

Correlation Coefficient: It is a numerical measure of the degree of association between two or more variables.

Scatter diagram: A scatter diagram of the data helps in having a visual idea about the nature of association between two variables. If the points cluster along a straight line, the association between variables is linear.

22.12 Self Assessment

1. Fill in the blanks :
   (i) Coefficient of correlation is a measure of the strength of the .... relationship between two variables.
   (ii) Coefficient of correlation is .... of the change of origin and scale.
   (iii) Coefficient of correlation between sale of woolen garments and the day temperature is likely to be ....
   (iv) Coefficient of correlation lies between .... and ....
   (v) Correlation between number of accidents and number of babies born in different years is termed as .... correlation.
   (vi) If two variables X and Y are such that their difference (X – Y) is always equal to 25. The correlation between X and Y is .... and positive.
2. Examine the validity of the following statements giving necessary proofs and reasons for your answer:
   (i) If $r_{XY} = 0$, then X and Y are always independent.
   (ii) If the sum of squares of the difference in ranks of 8 pairs of observations is 126, then the rank correlation coefficient is 0.5.

Notes

(iii) If $u + 3x = 5$, $2y - v = 7$ and $r_{xy} = 0.12$, then $r_{uv} = 0.12$.

(iv) If $2x - u = 8$, $y - 3v = 10$ and $r_{xy} = 0.8$, then $r_{uv} = 0.8$.

(v) If $\sum d_i^2 = 33$ and $n = 10$ then $r = 0.8$.

22.13 Review Questions

1. (a) Define correlation between two variables. Distinguish between positive and negative correlation. Illustrate by using diagrams.

   (b) Define the concept of covariance. How do you interpret it?

2. Define correlation and discuss its significance in statistical analysis. Does it signify 'cause and effect' relationship between the two variables?

3. (a) What do you understand by the coefficient of linear correlation? Explain the significance and limitations of this measure in any statistical analysis.

   (b) Write down an expression for the Karl Pearson's coefficient of linear correlation. Why is it termed as the coefficient of linear correlation? Explain.

4. (a) Describe the method of obtaining the Karl Pearson's formula of coefficient of linear correlation. What do positive and negative values of this coefficient indicate?

   (b) Does a zero value of Karl Pearson's coefficient of correlation between two variables X and Y imply that X and Y are not related? Explain.

5. Define product moment coefficient of correlation. What are the advantages of the study of correlation?

6. Show that the coefficient of correlation, r, is independent of change of origin and scale.

7. Prove that the coefficient of correlation lies between - 1 and +1.

8. "If two variables are independent the correlation between them is zero, but the converse is not always true". Explain the meaning of this statement.

9. What is Spearman's rank correlation? What are the advantages of the coefficient of rank correlation over Karl Pearson's coefficient of correlation?

10. Distinguish between the Spearman's coefficient of rank correlation and Karl Pearson's coefficient of correlation. Explain the situations under which Spearman's coefficient of rank correlation can assume a maximum and a minimum value. Under what conditions will Spearman's formula and Karl Pearson's formula give equal results?

11. Explain the method of calculating coefficient of correlation by Concurrent Deviation Method.

12. Write short notes on:

    (i) Positive and negative correlation.

    (ii) Linear and non-linear correlation.

    (iii) Probable error of correlation.

    (iv) Scatter diagram.

Notes

13. Compute Karl Pearson's coefficient of correlation from the following data :

X	:	8	11	15	10	12	16
Y	:	6	9	11	7	9	12

14. Calculate Karl Pearson's coefficient of correlation between the marks obtained by 10 students in economics and statistics.

Roll No.	:	1	2	3	4	5	6	7	8	9	10
Marks in eco.	:	23	27	28	29	30	31	33	35	36	39
Marks in stat.	:	18	22	23	24	25	26	28	29	30	32

15. Find Karl Pearson's coefficient of correlation from the following data and interpret its value.

Wages (Rs)	:	100	101	103	102	100	99	97	98	96	95
Cost of Living (Rs)	:	98	99	99	97	95	92	95	94	90	91

16. Find the coefficient of correlation between X and Y. Assume 69 and 112 as working origins for X and Y respectively.

X	:	78	89	96	69	59	79	68	61
Y	:	125	137	156	112	107	136	123	108

17. The distribution of population (in thousand) and blind persons according to various age groups is given in the following table. Find out correlation between age and blindness.

Age groups	:	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
Population	:	100	60	40	36	24	11	6	3
No. of Blind	:	55	40	40	40	36	22	18	15

18. Find out the coefficient of correlation from the following data :

X	:	300	350	400	450	500	550	600	650	700
Y	:	1600	1500	1400	1300	1200	1100	1000	900	800

19. Calculate the coefficient of correlation of the following figures relating to the consumption of fertiliser (in metric tonnes) and the output of food grains (in metric tonnes) in a district. Comment on your result.

Chemical Fertiliser used	Output of food grains	Chemical Fertiliser used	Output of food grains
100	1000	170	1360
110	1050	180	1420
120	1080	190	1500
130	1150	200	1600
140	1200	210	1650
150	1220	220	1650
160	1300	230	1650

Answers: Self Assessment

Notes

1. (i) linear (ii) independent (iii) negative (iv) – 1, + 1 (v) spurious or nonsense (vi) perfect
2. (i) invalid (ii) invalid (iii) invalid (iv) valid (v) valid.

22.14 Further Readings

Books

Sheldon M. Ross, Introduction to Probability Models, Ninth Edition, Elsevier Inc., 2007.

Jan Pukite, Paul Pukite, Modeling for Reliability Analysis, IEEE Press on Engineering of Complex Computing Systems, 1998.

Notes

Unit 23: Regression Analysis

CONTENTS

Objectives

Introduction

23.1 Two Lines of Regression

23.1.1 Line of Regression of Y on X

23.1.2 Line of Regression of X on Y

23.1.3 Correlation Coefficient and the two Regression Coefficients

23.2 Regression Coefficient in a Bivariate Frequency Distribution

23.3 The Coefficient of Determination

23.3.1 The Coefficient of Non-Determination

23.4 Mean of the Estimated Values

23.5 Mean and Variance of 'ei' values

23.5.1 Standard Error of the Estimate

23.6 Summary of Formulae

23.7 Keywords

23.8 Self Assessment

23.9 Review Questions

23.10 Further Readings

Objectives

After studying this unit, you will be able to:

- Define Two Lines of Regression
- Explain Regression Coefficient in a Bivariate Frequency Distribution
- Discuss The Coefficient of Determination
- Describe Mean of the Estimated Values
- Explain Mean and Variance of 'ei' values

Introduction

If the coefficient of correlation calculated for bivariate data (X_i, Y_i), i = 1,2, ... n, is reasonably high and a cause and effect type of relation is also believed to be existing between them, the next logical step is to obtain a functional relation between these variables. This functional relation is known as regression equation in statistics. Since the coefficient of correlation is measure of the degree of linear association of the variables, we shall discuss only linear regression equation. This does not, however, imply the non-existence of non-linear regression equations.

The regression equations are useful for predicting the value of dependent variable for given value of the independent variable. As pointed out earlier, the nature of a regression equation is different from the nature of a mathematical equation, e.g., if Y = 10 + 2X is a mathematical equation then it implies that Y is exactly equal to 20 when X = 5. However, if Y = 10 + 2X is a regression equation, then Y = 20 is an average value of Y when X = 5.

The term regression was first introduced by Sir Francis Galton in 1877. In his study of the relationship between heights of fathers and sons, he found that tall fathers were likely to have tall sons and vice-versa. However, the mean height of sons of tall fathers was lower than the mean height of their fathers and the mean height of sons of short fathers was higher than the mean height of their fathers. In this way, a tendency of the human race to regress or to return to a normal height was observed. Sir Francis Galton referred this tendency of returning to the mean height of all men as regression in his research paper, "Regression towards mediocrity in hereditary stature". The term 'Regression', originated in this particular context, is now used in various fields of study, even though there may be no existence of any regressive tendency.

23.1 Two Lines of Regression

For a bivariate data (X_i, Y_i), i = 1,2, ... n, we can have either X or Y as independent variable. If X is independent variable then we can estimate the average values of Y for a given value of X. The relation used for such estimation is called regression of Y on X. If on the other hand Y is used for estimating the average values of X, the relation will be called regression of X on Y. For a bivariate data, there will always be two lines of regression. It will be shown later that these two lines are different, i.e., one cannot be derived from the other by mere transfer of terms, because the derivation of each line is dependent on a different set of assumptions.

23.1.1 Line of Regression of Y on X

The general form of the line of regression of Y on X is $Y_{Ci} = a + bX_i$, where Y_{Ci} denotes the average or predicted or calculated value of Y for a given value of $X = X_i$. This line has two constants, a and b. The constant a is defined as the average value of Y when X = 0. Geometrically, it is the intercept of the line on Y- axis. Further, the constant b, gives the average rate of change of Y per unit change in X, is known as the regression coefficient.

The above line is known if the values of a and b are known. These values are estimated from the observed data (X_i, Y_i), i = 1,2, ... n.

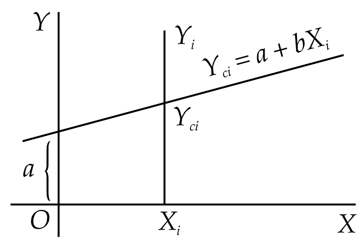
> *Note* It is important to distinguish between Y_{Ci} and Y_i. Where as Y_i is the observed value, Y_{Ci} is a value calculated from the regression equation.

Using the regression $Y_{Ci} = a + bX_i$, we can obtain Y_{C1}, Y_{C2}, ... Y_{Cn} corresponding to the X values X_1, X_2, ... X_n respectively. The difference between the observed and calculated value for a

Notes

particular value of X say X_i is called error in estimation of the i th observation on the assumption of a particular line of regression. There will be similar type of errors for all the n observations. We denote by $e_i = Y_i - Y_{Ci}$ (i = 1,2,.....n), the error in estimation of the i th observation. As is obvious from figure 23.1, e_i will be positive if the observed point lies above the line and will be negative if the observed point lies below the line. Therefore, in order to obtain a figure of total error, $e_i^{'s}$ are squared and added. Let S denote the sum of squares of these errors, i.e., $S = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \left(Y_i - Y_{Ci}\right)^2$.

Figure 23.1



Note The regression line can, alternatively, be written as a deviation of Y_i from Y_{ci} i.e. $Y_i - Y_{ci} = e_i$ or $Y_i = Y_{ci} + e_i$ or $Y_i = a + bX_i + e_i$. The component $a + bX_i$ is known as the deterministic component and e_i is random component.

The value of S will be different for different lines of regression. A different line of regression means a different pair of constants a and b. Thus, S is a function of a and b. We want to find such values of a and b so that S is minimum. This method of finding the values of a and b is known as the Method of Least Squares.

Rewrite the above equation as $S = S(Y_i - a - bX_i)^2$ ($\because Y_{Ci} = a + bX_i$).

The necessary conditions for minima of S are

(i) $\frac{\partial S}{\partial a} = 0$ and (ii) $\frac{\partial S}{\partial b} = 0$, where $\frac{\partial S}{\partial a}$ and $\frac{\partial S}{\partial b}$ are the partial derivatives of S w.r.t. a and b respectively.

$$\text{Now } \frac{\partial S}{\partial a} = -2\sum_{i=1}^{n}\left(Y_i - a - bX_i\right) = 0$$

$$or \quad \sum_{i=1}^{n}\left(Y_i - a - bX_i\right) = \sum_{i=1}^{n} Y_i - na - b\sum_{i=1}^{n} X_i = 0$$

$$or \quad \sum_{i=1}^{n} Y_i = na + b\sum_{i=1}^{n} X_i \qquad .... (1)$$

Notes

Also, $\frac{\partial S}{\partial b} = 2\sum_{i=1}^{n}\left(Y_i - a - bX_i\right)\left(-X_i\right) = 0$

$$or \quad -2\sum_{i=1}^{n}\left(X_iY_i - aX_i - bX_i^2\right) = \sum_{i=1}^{n}\left(X_iY_i - aX_i - bX_i^2\right) = 0$$

$$or \quad \sum_{i=1}^{n}X_iY_i - a\sum_{i=1}^{n}X_i - b\sum_{i=1}^{n}X_i^2 = 0$$

$$or \quad \sum_{i=1}^{n}X_iY_i = a\sum_{i=1}^{n}X_i + b\sum_{i=1}^{n}X_i^2 \qquad .... (2)$$

Equations (1) and (2) are a system of two simultaneous equations in two unknowns a and b, which can be solved for the values of these unknowns. These equations are also known as normal equations for the estimation of a and b. Substituting these values of a and b in the regression equation $Y_{Ci} = a + bX_i$, we get the estimated line of regression of Y on X.

Expressions for the Estimation of a and b.

Dividing both sides of the equation (1) by n, we have

$$\frac{\sum Y_i}{n} = \frac{na}{n} + \frac{b\sum X_i}{n} \quad \text{or} \quad \overline{Y} = a + b\overline{X} \qquad .... (3)$$

This shows that the line of regression $Y_{Ci} = a + bX_i$ passes through the point $\left(\overline{X}, \overline{Y}\right)$.

From equation (3), we have $a = \overline{Y} - b\overline{X}$ .... (4)

Substituting this value of a in equation (2), we have

$$\sum X_iY_i = \left(\overline{Y} - b\overline{X}\right)\sum X_i + b\sum X_i^2$$

$$= \overline{Y}\sum X_i - b\overline{X}\sum X_i + b\sum X_i^2 = n\overline{X}\overline{Y} - b.n\overline{X}^2 + b\sum X_i^2$$

$$\text{or} \quad \sum X_iY_i - n\overline{X}\overline{Y} = b\left(\sum X_i^2 - n\overline{X}^2\right)$$

$$\text{or} \quad b = \frac{\sum X_iY_i - n\overline{X}\overline{Y}}{\sum X_i^2 - n\overline{X}^2} \qquad .... (5)$$

Also, $\sum X_iY_i - n\overline{X}\overline{Y} = \sum\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)$ (See Chapter 12)

and $\sum X_i^2 - n\overline{X}^2 = \sum\left(X_i - \overline{X}\right)^2$

$$\therefore \quad b = \frac{\sum\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{\sum\left(X_i - \overline{X}\right)^2} \qquad .... (6)$$

$$\text{or} \quad b = \frac{\sum x_iy_i}{\sum x_i^2} \qquad .... (7)$$

where x_i and y_i are deviations of values from their arithmetic mean.

Notes

Dividing numerator and denominator of equation (6) by n we have

$$b=\frac{\frac{1}{n}\sum\left(X_i-\overline{X}\right)\left(Y_i-\overline{Y}\right)}{\frac{1}{n}\sum\left(X_i-\overline{X}\right)^2}=\frac{Cov(X,Y)}{\sigma_X^2} \quad \text{.... (8)}$$

The expression for b, which is convenient for use in computational work, can be written from equation (5) is given below:

$$b=\frac{\sum X_iY_i-n\frac{\sum X_i}{n}\cdot\frac{\sum Y_i}{n}}{\sum X_i^2-n\left(\frac{\sum X_i}{n}\right)^2}=\frac{\sum X_iY_i-\frac{\left(\sum X_i\right)\left(\sum Y_i\right)}{n}}{\sum X_i^2-\frac{\left(\sum X_i\right)^2}{n}}$$

Multiplying numerator and denominator by n, we have

$$b=\frac{n\sum X_iY_i-\left(\sum X_i\right)\left(\sum Y_i\right)}{n\sum X_i^2-\left(\sum X_i\right)^2} \quad \text{.... (9)}$$

To write the shortcut formula for b, we shall show that it is independent of change of origin but not of change of scale.

As in case of coefficient of correlation we define

$$u_i=\frac{X_i-A}{h} \quad \text{and} \quad v_i=\frac{Y_i-B}{k}$$

or $X_i = A + hu_i$ and $Y_i = B + kv_i$

$\therefore$ $\overline{X}=A+h\overline{u}$ and $\overline{Y}=B+k\overline{v}$

also $\left(X_i-\overline{X}\right)=h\left(u_i-\overline{u}\right)$ and $Y_i-\overline{Y}=k\left(v_i-\overline{v}\right)$

Substituting these values in equation (6), we have

$$b=\frac{hk\sum\left(u_i-\overline{u}\right)\left(v_i-\overline{v}\right)}{h^2\sum\left(u_i-\overline{u}\right)^2}=\frac{k\sum\left(u_i-\overline{u}\right)\left(v_i-\overline{v}\right)}{h\sum\left(u_i-\overline{u}\right)^2}$$

$$=\frac{k}{h}\left[\frac{n\sum u_iv_i-\left(\sum u_i\right)\left(\sum v_i\right)}{n\sum u_i^2-\left(\sum u_i\right)^2}\right] \quad \text{.... (10)}$$

(Note: if h = k they will cancel each other)

Consider equation (8), $b=\frac{Cov(X,Y)}{s_X^2}$

Writing $Cov(X,Y)=r.\sigma_X\sigma_Y$, we have $b=\frac{r.\sigma_X\sigma_Y}{\sigma_X^2}=r\cdot\frac{\sigma_Y}{\sigma_X}$

The line of regression of Y on X, i.e $Y_{Ci} = a + bX_i$ can also be written as

$$Y_{Ci} = \overline{Y} - b\overline{X} + bX_i \text{ or } Y_{Ci} - \overline{Y} = b\left(X_i - \overline{X}\right) \quad \text{.... (11)}$$

or $$\left(Y_{Ci} - \overline{Y}\right) = r \cdot \frac{\sigma_Y}{\sigma_X}\left(X_i - \overline{X}\right) \quad \text{.... (12)}$$

23.1.2 Line of Regression of X on Y

The general form of the line of regression of X on Y is $X_{Ci} = c + dY_i$, where X_{Ci} denotes the predicted or calculated or estimated value of X for a given value of $Y = Y_i$ and c and d are constants. d is known as the regression coefficient of regression of X on Y.

In this case, we have to calculate the value of c and d so that

$S' = \Sigma(X_i - X_{Ci})^2$ is minimised.

Figure 23.2

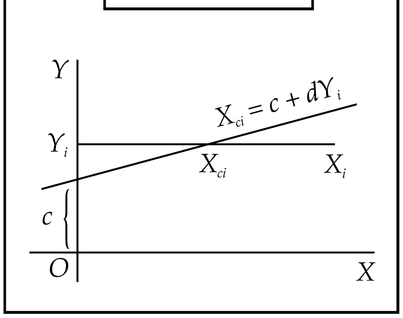
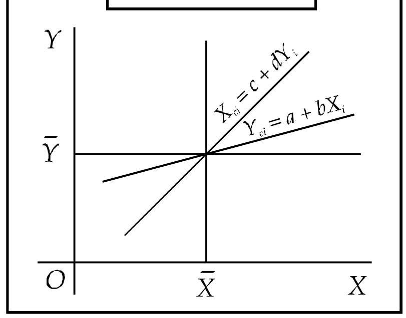


Figure 23.3



As in the previous section, the normal equations for the estimation of c and d are

$$\Sigma X_i = nc + d\Sigma Y_i \quad \text{.... (13)}$$

and $$\Sigma X_i Y_i = cSY_i + d\Sigma Y_i^2 \quad \text{.... (14)}$$

Dividing both sides of equation (13) by n, we have $\overline{X} = c + d\overline{Y}$.

This shows that the line of regression also passes through the point $\left(\overline{X}, \overline{Y}\right)$. Since both the lines of regression passes through the point $\left(\overline{X}, \overline{Y}\right)$, therefore $\left(\overline{X}, \overline{Y}\right)$ is their point of intersection as shown in Figure 23.3.

We can write $c = \overline{X} - d\overline{Y}$ .... (15)

As before, the various expressions for d can be directly written, as given below.

$$d = \frac{\sum X_i Y_i - n\overline{X}\overline{Y}}{\sum Y_i^2 - n\overline{Y}^2} \quad \text{.... (16)}$$

or $$d = \frac{\sum\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{\sum\left(Y_i - \overline{Y}\right)^2} \quad \text{.... (17)}$$

or $$d = \frac{\sum x_i y_i}{\sum y_i^2} \quad \text{.... (18)}$$

Notes

$$= \frac{\frac{1}{n}\sum\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{\frac{1}{n}\sum\left(Y_i - \bar{Y}\right)^2} = \frac{Cov(X,Y)}{\sigma_Y^2} \quad \text{.... (19)}$$

Also $$d = \frac{n\sum X_i Y_i - \left(\sum X_i\right)\left(\sum Y_i\right)}{n\sum Y_i^2 - \left(\sum Y_i\right)^2} \quad \text{.... (20)}$$

This expression is useful for calculating the value of d. Another short-cut formula for the calculation of d is given by

$$d = \frac{h}{k}\left[\frac{n\sum u_i v_i - \left(\sum u_i\right)\left(\sum v_i\right)}{n\sum v_i^2 - \left(\sum v_i\right)^2}\right] \quad \text{.... (21)}$$

where $u_i = \frac{X_i - A}{h}$ and $v_i = \frac{Y_i - B}{k}$

Consider equation (19)

$$d = \frac{Cov(X,Y)}{\sigma_Y^2} = \frac{r\sigma_X\sigma_Y}{\sigma_Y^2} = r\cdot\frac{\sigma_X}{\sigma_Y} \quad \text{.... (22)}$$

Substituting the value of c from equation (15) into line of regression of X on Y we have

$$X_{Ci} = \bar{X} - d\bar{Y} + dY_i \quad or \quad \left(X_{Ci} - \bar{X}\right) = d\left(Y_i - \bar{Y}\right) \quad \text{.... (23)}$$

$$or \quad \left(X_{Ci} - \bar{X}\right) = r\cdot\frac{\sigma_X}{\sigma_Y}\left(Y_i - \bar{Y}\right) \quad \text{.... (24)}$$

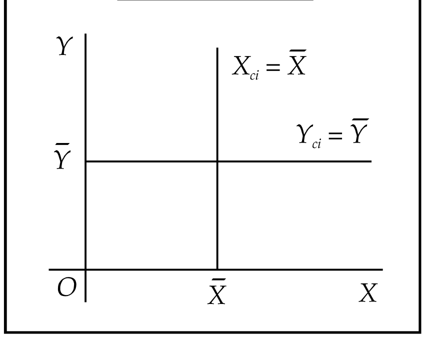
Remarks: It should be noted here that the two lines of regression are different because these have been obtained in entirely two different ways. In case of regression of Y on X, it is assumed that the values of X are given and the values of Y are estimated by minimising $\Sigma(Y_i - Y_{Ci})^2$ while in case of regression of X on Y, the values of Y are assumed to be given and the values of X are estimated by minimising $\Sigma(X_i - X_{Ci})^2$. Since these two lines have been estimated on the basis of different assumptions, they are not reversible, i.e., it is not possible to obtain one line from the other by mere transfer of terms. There is, however, one situation when these two lines will coincide. From the study of correlation we may recall that when r = ±1, there is perfect correlation between the variables and all the points lie on a straight line. Therefore, both the lines of regression coincide and hence they are also reversible in this case. By substituting r = ±1 in equation (12) or (24) it can be shown that the lines of regression in both the cases become

$$\left(\frac{Y_i - \bar{Y}}{\sigma_Y}\right) = \pm\left(\frac{X_i - \bar{X}}{\sigma_X}\right)$$

Notes

Further when r = 0, equation (12) becomes $Y_{Ci} = \overline{Y}$ and equation (24) becomes $X_{Ci} = \overline{X}$. These are the equations of lines parallel to X-axis and Y-axis respectively. These lines also intersect at the point $\left(\overline{X}, \overline{Y}\right)$ and are mutually perpendicular at this point, as shown in figure 23.4.

Figure 23.4



23.1.3 Correlation Coefficient and the two Regression Coefficients

Since $b = r \cdot \frac{\sigma_Y}{\sigma_X}$ and $d = r \cdot \frac{\sigma_X}{\sigma_Y}$, we have

$b.d = r\frac{\sigma_Y}{\sigma_X} \cdot r\frac{\sigma_X}{\sigma_Y} = r^2$ or $r = \sqrt{b.d}$. This shows that correlation coefficient is the geometric mean of the two regression coefficients.

Remarks:

The following points should be kept in mind about the coefficient of correlation and the regression coefficients :

(i) Since $r = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$, $b = \frac{Cov(X,Y)}{\sigma_X^2}$ and $d = \frac{Cov(X,Y)}{\sigma_Y^2}$, therefore the sign of r, b and d will always be same and this will depend upon the sign of Cov (X, Y).

(ii) Since bd = r^2 and 0 £ r^2 £ 1, therefore either both b and d are less than unity or if one of them is greater than unity, the other must be less than unity such that 0 £ b.d £ 1 is always true.

Example 1:

Obtain the two regression equations and find correlation coefficient between X and Y from the following data :

X	:	10	9	7	8	11
Y	:	6	3	2	4	5

Notes

Solution.

Calculation table

X	Y	XY	X^2	Y^2
10	6	60	100	36
9	3	27	81	9
7	2	14	49	4
8	4	32	64	16
11	5	55	121	25
45	20	188	415	90

(a) Regression of Y on X

$$b = \frac{n\sum XY - (\sum X)(\sum Y)}{n\sum X^2 - (\sum X)^2} = \frac{5 \times 188 - 45 \times 20}{5 \times 415 - (45)^2} = 0.8$$

Also, $\overline{X} = \frac{45}{5} = 9$ and $\overline{Y} = \frac{20}{5} = 4$

Now $a = \overline{Y} - b\overline{X} = 4 - 0.8 \times 9 = -3.2$

$\therefore$ Regression of Y on X is $Y_C = -3.2 + 0.8X$

(b) Regression of X on Y

$$d = \frac{n\sum XY - (\sum X)(\sum Y)}{n\sum Y^2 - (\sum Y)^2} = \frac{5 \times 188 - 45 \times 20}{5 \times 90 - (20)^2} = 0.8$$

Also, $c = \overline{X} - d\overline{Y} = 9 - 0.8 \times 4 = 5.8$

$\therefore$ The regression of X on Y is $X_C = 5.8 + 0.8Y$

(c) Coefficient of correlation $r = \sqrt{b.d} = \sqrt{0.8 \times 0.8} = 0.8$

Example 2:

From the data given below, find :

(a) The two regression equations.

(b) The coefficient of correlation between marks in economics and statistics.

(c) The most likely marks in statistics when marks in economics are 30.

Marks in Eco. : 25 28 35 32 31 36 29 38 34 32

Marks in Stat. : 43 46 49 41 36 32 31 30 33 39

Solution.

Notes

Calculation table

Marks in Eco. (*X*)	*Marks in Stat.* (*Y*)	$u = X - 31$	$v = Y - 41$	uv	u^2	v^2
25	43	– 6	2	– 12	36	4
28	46	– 3	5	– 15	9	25
35	49	4	8	32	16	64
32	41	1	0	0	1	0
31	36	0	– 5	0	0	25
36	32	5	– 9	– 45	25	81
29	31	– 2	– 10	20	4	100
38	30	7	– 11	– 77	49	121
34	33	3	– 8	– 24	9	64
32	39	1	– 2	– 2	1	4
Total		10	– 30	– 123	150	488

From the table, we have

$$\overline{X} = 31 + \frac{10}{10} = 32 \text{ and } \overline{Y} = 41 - \frac{30}{10} = 38.$$

(a) The lines of regression

(i) Regression of Y on X

$$b = \frac{n\sum uv - (\sum u)(\sum v)}{n\sum u^2 - (\sum u)^2} = \frac{-1230 + 300}{1500 - 100} = -0.66$$

$$a = \overline{Y} - b\overline{X} = 38 + 0.66 \times 32 = 59.26$$

∴ Regression equation is

$Y_C = 59.26 - 0.66X$

(ii) Regression of X on Y

$$d = \frac{n\sum uv - (\sum u)(\sum v)}{n\sum v^2 - (\sum v)^2} = \frac{-1230 + 300}{4880 - 900} = -0.23$$

$$c = \overline{X} - d\overline{Y} = 32 + 0.23 \times 38 = 40.88$$

∴ Regression equation is

$X_C = 40.88 – 0.23Y$

(b) Coefficient of correlation

$$r = \sqrt{b \cdot d} = -\sqrt{-0.66 \times -0.23} = -0.39$$

Note that r, b and d are of same sign.

(c) Since we have to estimate marks in statistics denoted by Y, therefore, regression of Y on X will be used. The most likely marks in statistics when marks in economics are 30, is given by

$Y_C = 59.26 - 0.66 \times 30 = 39.33$

Notes

Example 3:

Obtain the two lines of regression from the following data and estimate the blood pressure when age is 50 years. Can we also estimate the blood pressure of a person aged 20 years on the basis of this regression equation? Discuss.

Age (X) (in years) : 56 42 72 39 63 47 52 49 40 42 68 60

Blood Pressure (Y) : 127 112 140 118 129 116 130 125 115 120 135 133

Solution.

Calculation table

X	Y	$u = X - 52$	$v = Y - 125$	uv	u^2	v^2
56	127	4	2	8	16	4
42	112	- 10	- 13	130	100	169
72	140	20	15	300	400	225
39	118	- 13	- 7	91	169	49
63	129	11	4	44	121	16
47	116	- 5	- 9	45	25	81
52	130	0	5	0	0	25
49	125	- 3	0	0	9	0
40	115	- 12	- 10	120	144	100
42	120	- 10	- 5	50	100	25
68	135	16	10	160	256	100
60	133	8	8	64	64	64
	Total	6	0	1012	1404	858

From the table, we have

$$\overline{X} = 52 + \frac{6}{12} = 52.5 \quad \text{and} \quad \overline{Y} = 125$$

(a) Regression of Y on X

$$b = \frac{n\sum uv - (\sum u)(\sum v)}{n\sum u^2 - (\sum u)^2} = \frac{12 \times 1012 - 6 \times 0}{12 \times 1404 - (6)^2} = 0.72$$

Also $a = \overline{Y} - b\overline{X} = 125 - 0.72 \times 52.5 = 87.2$

∴ The line of regression of Blood pressure (Y) on Age (X) is

$Y_C = 87.2 + 0.72X$

(b) Regression of X on Y

$$d = \frac{n\sum uv - (\sum u)(\sum v)}{n\sum v^2 - (\sum v)^2} = \frac{12 \times 1012 - 6 \times 0}{12 \times 858 - 0} = 1.18$$

Also $c = \overline{X} - d\overline{Y} = 52.5 - 1.18 \times 125 = -95$

∴ Line of regression of Age (X) on Blood pressure (Y) is

$X_C = -95 + 1.18Y$

Notes

(c) (i) To estimate blood pressure (Y) for a given age, X = 50 years, we shall use regression of Y on X

$\therefore Y_C = 87.2 + 0.72 \times .\ 50 = 123.2$

(ii) The estimate of blood pressure when age is 20 years

$Y_C = 87.2 + 0.72 \times .\ 20 = 101.6$

It should be noted here that this estimate is wrong because the blood pressure of a normal person cannot be less than 110.

This result reflects the limitations of regression analysis with regard to estimation or prediction. It is important to note that the prediction, based on regression line, should be done only for those values of the variable that are not very far from the range of the observed data, used to derive the line of regression. The prediction from a regression line for a value of the variable that is far away from the observed data is likely to give inconsistent results like the one obtained above.

Example 4:

A panel of judges P and Q graded seven dramatic performances by independently awarding marks as follows :

Performance	:	1	2	3	4	5	6	7
Marks by P	:	46	42	44	40	43	41	45
Marks by Q	:	40	38	36	35	39	37	41

The eighth performance which Judge Q could not attend, was awarded 37 marks by Judge P. If Judge Q had also been present, how many marks would be expected to have been awarded by him to eighth performance?

Solution.

Let us denote marks awarded by the Judge P as X and marks awarded by the Judge Q as Y. Since we have to estimate marks that would have been awarded by Judge Q, we shall fit a line of regression of Y on X to the given data.

Calculation table

X	Y	$u = X - 43$	$v = X - 37$	uv	u^2	v^2
46	40	3	3	9	9	9
42	38	-1	1	-1	1	1
44	36	1	-1	-1	1	1
40	35	-3	-2	6	9	4
43	39	0	2	0	0	4
41	37	-2	0	0	4	0
45	41	2	4	8	4	16
Total		0	7	21	28	35

From the table, we have

$\overline{X} = 43$ and $\overline{Y} = 37 + \frac{7}{7} = 38$

Further, $b = \frac{n\sum uv - (\sum u)(\sum v)}{n\sum u^2 - (\sum u)^2} = \frac{7 \times 21 - 0}{7 \times 28 - 0} = 0.75$

Also $a = \overline{Y} - b\overline{X} = 38 - 0.75 \times 43 = 5.75$

Notes

$\therefore$ $Y_C = 5.75 + 0.75X$ is the fitted line of regression.

Estimate of Y when X = 37

$Y_C = 5.75 + 0.75 \times 37 = 33.5$ marks

$\therefore$ It is expected that the Judge Q would have awarded 33.5 marks to the eighth performance.

Example 5:

Find out the regression coefficients of Y on X, X on Y and correlation coefficient between X and Y on the basis of the following data :

$\Sigma XY = 350$, $\Sigma X = 50$, $\Sigma Y = 60$, n = 10, Variance of X = 4 and Variance of Y = 9.

Solution.

Regression coefficient of Y on X is given by

$$b = \frac{\frac{\sum XY}{n} - \left(\frac{\sum X}{n}\right)\left(\frac{\sum Y}{n}\right)}{\sigma_X^2} = \frac{\frac{350}{10} - \left(\frac{50}{10}\right)\left(\frac{60}{10}\right)}{4} = 1.25$$

Regression coefficient of X on Y is given by

$$d = \frac{\frac{\sum XY}{n} - \left(\frac{\sum X}{n}\right)\left(\frac{\sum Y}{n}\right)}{\sigma_Y^2} = \frac{35 - 30}{9} = 0.55$$

Coefficient of correlation between X and Y is given by

$$r = \sqrt{1.25 \times 0.55} = 0.83$$

Example 6:

The following results were worked out from scores in statistics and mathematics in a certain examination :

	Scores in Statistics (X)	*Scores in Mathematics (Y)*
Mean	39.5	47.5
Standard Deviation	10.8	17.8

Karl Pearson's correlation coefficient between X and Y = 0.42. Find both the regression lines. Use these lines to estimate the value of Y when X = 50 and the value of X when Y = 30.

Solution.

(a) Regression of Y on X

Regression coefficient $b = r \cdot \frac{\sigma_Y}{\sigma_X} = 0.42 \times \frac{17.8}{10.8} = 0.69$

and $a = \overline{Y} - b\overline{X} = 47.5 - 0.69 \times 39.5 = 20.24$

$\therefore$ The line of regression of Y on X is $Y_C = 20.24 + 0.69X$, and the predicted value of Y when X = 50, is given by

$Y_C = 20.24 + 0.69 \times 50 = 54.74.$

(b) Regression of X on Y

Regression coefficient $d = r \cdot \frac{\sigma_X}{\sigma_Y} = 0.42 \times \frac{10.8}{17.8} = 0.25$

and $c = \overline{X} - d\overline{Y} = 39.5 - 0.25 \times 47.5 = 27.62$

$\therefore$ The line of regression of X on Y is $X_C = 27.62 + 0.25Y$

and the predicted value of X when Y = 30 is given by

$X_C = 27.62 + 0.25 \times 30 = 35.12$

Example 7:

For a bivariate data, you are given the following information :

$\Sigma(X - 58) = 46$ $\quad\Sigma(X - 58)^2 = 3086$

$\Sigma(Y - 58) = 9$ $\quad\Sigma(Y - 58)^2 = 483$

$\Sigma(X - 58)(Y - 58) = 1095.$

Number of pairs of observations = 7. You are required to determine (i) the two regression equations and (ii) the coefficient of correlation between X and Y.

Solution.

Let u = X - 58 and v = Y - 58. In terms of our notations, we are given $\Sigma u = 46$, $\Sigma u^2 = 3086$, $\Sigma v = 9$, $\Sigma v^2 = 483$, $\Sigma uv = 1095$ and $n = 7$.

Now $\overline{X} = 58 + \frac{46}{7} = 64.7$ and $\overline{Y} = 58 + \frac{9}{7} = 59.29$

(a) For regression equation of Y on X, we have

$$b = \frac{7 \times 1095 - 46 \times 9}{7 \times 3086 - (46)^2} = 0.37$$

and $a = \overline{Y} - b\overline{X} = 59.29 - 0.37 \times 64.57 = 35.40$

$\therefore$ The line of regression of Y on X is given by

$Y_C = 35.40 + 0.37X$

(b) For regression equation of X on Y, we have

$$d = \frac{7 \times 1095 - 46 \times 9}{7 \times 483 - (9)^2} = 2.20$$

and $c = \overline{X} - d\overline{Y} = 64.57 - 2.2 \times 59.29 = -65.87$

$\therefore$ The line of regression of X on Y is given by

$X_C = -65.87 + 2.2Y$

Notes

(c) The coefficient of correlation

$$r = \sqrt{b \cdot d} = \sqrt{0.37 \times 2.2} = 0.90$$

Example 8:

Find the means of X and Y variables and the coefficient of correlation between them from the following two regression equations :

3Y - 2X - 10 = 0

2Y - X - 50 = 0

Solution.

(a) The means of X and Y

We know that both the lines of regression intersect at the point $(\overline{X}, \overline{Y})$. The simultaneous solution of the given equations will give the mean values of X and Y as

$\overline{X} = 130$ and $\overline{Y} = 90$ respectively.

(b) Correlation Coefficient

Let us assume that the first equation be regression of Y on X. Rewriting this equation as 3Y = 2X + 10 or $Y = \frac{2}{3}X + \frac{10}{3}$.

$\therefore$ The corresponding regression coefficient, $b = \frac{2}{3}$

Further, assuming the second equation as regression of X on Y, we can rewrite this equation as X = 2Y - 50.

$\therefore$ The regression coefficient, d = 2

Since b.d $= \frac{2}{3} \cdot 2 = \frac{4}{3}$ > 1, therefore, our assumptions regarding the two regression lines are wrong.

Now we reverse these assumptions and assume that the first equation is regression of X on Y and second the regression of Y on X.

$\therefore$ The first equation can be written as 2X = 3Y - 10 or $X = \frac{3}{2}Y - 5$, so that the corresponding regression coefficient is $d = \frac{3}{2}$. Further, the second equation can be written as 2Y = X + 50 or $Y = \frac{1}{2}X + 25$, so that the corresponding regression coefficient is $b = \frac{1}{2}$. Since b.d $= \frac{3}{2} \times \frac{1}{2} = \frac{3}{4} < 1$, our assumption is correct.

Also r^2 = b.d = $\frac{3}{4}$ $\therefore$ $r = \sqrt{\frac{3}{4}} = 0.87$

Notes

23.2 Regression Coefficient in a Bivariate Frequency Distribution

As in case of calculation of correlation coefficient (see § 12.6), we can directly write the formula for the two regression coefficients for a bivariate frequency distribution as given below :

$$b = \frac{N\sum\sum f_{ij}X_iY_j - \left(\sum f_iX_i\right)\left(\sum f_j'Y_j\right)}{N\sum f_iX_i^2 - \left(\sum f_iX_i\right)^2}$$

or, if we define $u_i = \frac{X_i - A}{h}$ and $v_j = \frac{Y_j - B}{k}$,

$$b = \frac{k}{h}\left[\frac{N\sum\sum f_{ij}u_iv_j - \left(\sum f_iu_i\right)\left(\sum f_j'v_j\right)}{N\sum f_iu_i^2 - \left(\sum f_iu_i\right)^2}\right]$$

Similarly, $d = \frac{N\sum\sum f_{ij}X_iY_j - \left(\sum f_iX_i\right)\left(\sum f_j'Y_j\right)}{N\sum f_j'Y_j^2 - \left(\sum f_j'Y_j\right)^2}$

or $d = \frac{h}{k}\left[\frac{N\sum f_{ij}u_iv_j - \left(\sum f_iu_i\right)\left(\sum f_j'v_j\right)}{N\sum f_j'v_j^2 - \left(\sum f_j'v_j\right)^2}\right]$

Example 12:

By calculating the two regression coefficients obtain the two regression lines from the following data:

$Y \rightarrow$ $X \downarrow$	0 - 5	5 - 10	10 - 15
0 - 10	2	5	7
10 - 20	1	3	2
20 - 30	8	4	0

Solution.

The mid points of X-values are 5, 15, 25.

Let $u = \frac{X - 15}{10}$, $\therefore$ Corresponding u-values become - 1, 0, 1

Similarly, the mid-points of Y-values are 2.5, 7.5, 12.5

Let $v = \frac{Y - 7.5}{5}$, $\therefore$ Corresponding v-values become - 1, 0, 1

Notes

Calculation Table

$u \backslash v$	−1	0	1	f_i	$f_i u_i$	$f_i u_i^2$	$f_i u_i v_j$
−1	2 [2]	5 [0]	7 [−7]	14	−14	14	−5
0	1 [0]	3 [0]	2 [0]	6	0	0	0
1	8 [−8]	4 [0]	0 [0]	12	12	12	−8
f'_j	11	12	9	32	−2	26	−13
$f'_j v'_j$	−11	0	9	−2			
$f_j v_j^2$	11	0	9	20			

From the table N = 32 (total frequency)

(a) Regression of Y on X

Regression Coefficient (here h = 10 and k = 5)

$$b = \left[\frac{-32 \times 13 - 2 \times 2}{32 \times 26 - 4}\right] \times \frac{5}{10} = \frac{-416 - 4}{832 - 4} \times \frac{1}{2} = -\,0.25$$

Also, $\overline{X} = 15 + \frac{10(-2)}{32} = 14.73$ and $\overline{Y} = 7.5 + \frac{5(-2)}{32} = 7.19$

$\therefore\ a = \overline{Y} - b\overline{X} = 7.19 + 0.25 \times 14.73 = 10.87$

Hence, the regression of Y on X becomes $Y_C = 10.87 - 0.25X$

(b) Regression of X on Y

Regression coefficient $d = \left[\frac{-420}{32 \times 20 - 4}\right] \times \frac{10}{5} = -1.32$

Also, $c = \overline{X} - d\overline{Y} = 14.73 + 1.32 \times 7.19 = 24.22$

Hence, the regression of X on Y becomes $X_C = 24.22 - 1.32Y$

23.3 The Coefficient of Determination

We recall that in the line of regression $Y_C = a + bX$, X is used to estimate the value of Y. Further, the estimate of Y, independently of X, is given by a constant. Let this constant be A. Thus, we can write $Y_C = A$.

Given the observations $Y_1, Y_2, \ldots\ldots Y_n$, A will be the best estimate of Y if $S = \sum_{i=1}^{n}(Y_i - A)^2$ is minimum.

The necessary condition for minimum of S is $\frac{\partial S}{\partial A} = 0$.

i.e., $2\sum(Y_i - A) = 0$ *or* $\sum Y_i - nA = 0$ *or* $A = \overline{Y}$.

Notes

$\therefore$ The best estimate (an estimate having minimum sum of squares of errors) of Y, independently of X, is given by $Y_C = \overline{Y}$.

Remarks: If X and Y are independent variables, the two lines of regression are $Y_C = \overline{Y}$ and $X_C = \overline{X}$.

Very often, when we use X for the estimation of Y, we are interested in knowing how far the use of X enables us to explain the variations in Y values from $\overline{Y}$ or, in other words, how much of the variations in Y, from $\overline{Y}$, are being explained by the regression equation $Y_{Ci} = a + bX_i$? To answer this question, we write

$$Y_i - \overline{Y} = Y_i - Y_{Ci} + Y_{Ci} - \overline{Y} \quad \text{(Subtracting and adding } Y_{Ci}\text{)}$$

or $$Y_i - \overline{Y} = \left(Y_i - Y_{Ci}\right) + \left(Y_{Ci} - \overline{Y}\right)$$

Squaring both sides and taking sum over all the observations, we have

$$\sum\left(Y_i - \overline{Y}\right)^2 = \sum\left(Y_i - Y_{Ci}\right)^2 + \sum\left(Y_{Ci} - \overline{Y}\right)^2 + 2\sum\left(Y_i - Y_{Ci}\right)\left(Y_{Ci} - \overline{Y}\right) \quad ....(1)$$

Consider the product term

$$2\sum\left(Y_i - Y_{Ci}\right)\left(Y_{Ci} - \overline{Y}\right) = 2\sum\left[\left\{Y_i - \overline{Y} - b\left(X_i - \overline{X}\right)\right\}\left\{b\left(X_i - \overline{X}\right)\right\}\right]$$

$$= 2b\sum\left(Y_i - \overline{Y}\right)\left(X_i - \overline{X}\right) - 2b^2\sum\left(X_i - \overline{X}\right)^2$$

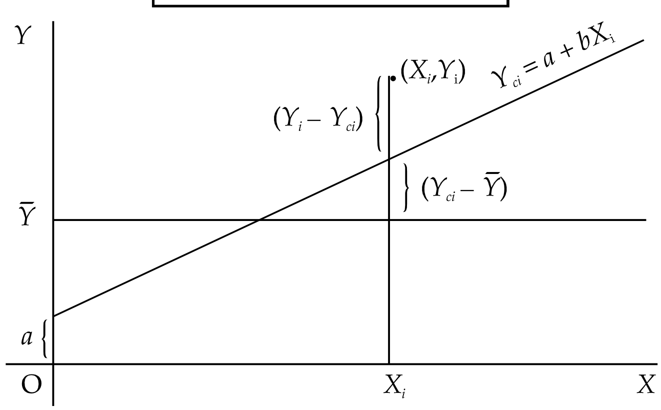
$$= 2b^2\sum\left(X_i - \overline{X}\right)^2 - 2b^2\sum\left(X_i - \overline{X}\right)^2 = 0$$

Thus, equation (1) becomes

$$\sum\left(Y_i - \overline{Y}\right)^2 = \sum\left(Y_i - Y_{Ci}\right)^2 + \sum\left(Y_{Ci} - \overline{Y}\right)^2 \quad .... (2)$$

From the above figure, we note that $Y_{Ci} - \overline{Y}$ is the deviation of the estimated value from $\overline{Y}$. This deviation has occurred because X and Y are related by the regression equation $Y_{Ci} = a + bX_i$, so that the estimate of Y is Y_{Ci} when $X = X_i$. Similar type of deviations would occur for other values of X. Thus, the magnitude of the term $\sum\left(Y_{Ci} - \overline{Y}\right)^2$ gives the strength of the relationship, $Y_{Ci} = a + bX_i$, between X and Y or, equivalently, the variations in Y that are explained by the regression equation.

Figure 23.5



Notes

The other term $Y_i - Y_{Ci}$ gives the deviation of i th observed value from the regression line and thus the magnitude of the term $\sum(Y_i - Y_{Ci})^2$ gives the variations in Y about the line of regression. These variations are also known as unexplained variations in Y.

Adding the two types of variations, we get the magnitude of total variations in Y. Thus, equation (2) can also be written as

Total variations in Y = Unexplained variations in Y + Explained variations in Y.

Dividing both sides of equation (2) by $\sum(Y_i - \bar{Y})^2$, we have

$$1 = \frac{\sum(Y_i - Y_{Ci})^2}{\sum(Y_i - \bar{Y})^2} + \frac{\sum(Y_{Ci} - \bar{Y})^2}{\sum(Y_i - \bar{Y})^2} \qquad \text{.... (3)}$$

or 1 = Proportion of unexplained variations + Proportion of variations explained by the regression equation.

The proportion of variation explained by regression equation is called the coefficient of determination.

Thus, the coefficient of determination $= \dfrac{\sum(Y_{Ci} - \bar{Y})^2}{\sum(Y_i - \bar{Y})^2}$

$$= \frac{b^2\sum(X_i - \bar{X})^2}{\sum(Y_i - \bar{Y})^2} = \frac{\left[\sum(X_i - \bar{X})(Y_i - \bar{Y})\right]^2}{\sum(X_i - \bar{X})^2\sum(Y_i - \bar{Y})^2} = r^2$$

This result shows that the coefficient of determination is equal to the square of the coefficient of correlation, i.e., r^2 gives the proportion of variations explained by each regression equation.

Remarks:

(i) It should be obvious from the above that it is desirable to calculate the coefficient of correlation prior to the fitting of a regression line. If r^2 is high enough, the fitted line will explain a greater proportion of the variations in the dependent variable. A low value of r^2 would, however, indicate that the proposed fitting of regression would not be of much use.

(ii) The expression for the coefficient of determination for regression of X on Y can be written in a similar way. Here we can write $r^2 = \dfrac{\sum(X_{Ci} - \bar{X})^2}{\sum(X_i - \bar{X})^2}$.

23.3.1 The Coefficient of Non-Determination

The proportion of unexplained variations is also termed as the coefficient of non-determination. It is denoted by k^2, where $k^2 = (1 - r^2)$. The square root of k^2 is termed as the coefficient of alienation, i.e., $k = \sqrt{(1 - r^2)}$.

Notes

Example 13:

Comment on the following statements :

(i) The two regression coefficients of bivariate data are 0.7 and 1.4.

(ii) A correlation coefficient r = 0.8, between the two variables X and Y, implies a relationship twice as close as r = 0.4.

Solution.

(i) This statement implies that $r^2 = 0.7 \times 1.4 = 0.98$, i.e., a linear regression fitted to the data would explain 98% of the variations in the dependent variable.

(ii) The given statement is wrong. Since r = 0.8 implies that a regression fitted to the data would explain 64% of the variations in the dependent variable while r = 0.4 implies that the proportion of such variations is only 16%. Thus, r = 0.8 implies a relation that is four times as close as r = 0.4.

Example 14:

The correlation coefficient between two variables is found to be 0.8. Explain the meaning of this statement.

Solution.

The given statement implies that :

(i) Two variables are highly correlated.

(ii) There is positive association between them, i.e., an increase in value of one is accompanied by the increase in value of the other and vice-versa.

(iii) A linear regression fitted to the data would explain 64% of the variations in the dependent variable.

23.4 Mean of the Estimated Values

We may recall that Y_C and X_C are the estimated values from the regressions of Y on X and X on Y respectively.

Consider the regression equation $Y_{Ci} - \overline{Y} = b(X_i - \overline{X})$.

Taking sum over all the observations, we get

$$\sum\left(Y_{Ci} - \overline{Y}\right) = b\sum\left(X_i - \overline{X}\right) = 0$$

$$\Rightarrow \quad \sum Y_{Ci} - n\overline{Y} = 0 \quad \text{or} \quad \frac{\sum Y_{Ci}}{n} = \overline{Y}_C = \overline{Y} \qquad \text{.... (1)}$$

Similarly, it can be shown that $\overline{X}_C = \overline{X}$.

This implies that the mean of the estimated values is also equal to the mean of the observed values.

Notes

23.5 Mean and Variance of 'e_i' values

(i) Mean of e_i values

We know that $e_i = Y_i - Y_{Ci}$.

Taking sum over all the observations, we have

$$\sum e_i = \sum\left(Y_i - Y_{Ci}\right) = \sum Y_i - \sum Y_{Ci} = 0 \quad \text{[from equation (1)]}$$

$\therefore$ Mean of e_i values is equal to zero.

(ii) Variance of e_i values

The variance of e_i values, in case of regression of Y on X, is given by

$$S_{Y.X}^2 = \frac{1}{n}\sum\left(e_i - 0\right)^2 = \frac{1}{n}\sum\left(Y_i - Y_{Ci}\right)^2 \qquad \text{.... (2)}$$

[Note that $\sum\left(Y_i - Y_{Ci}\right)^2$ is the magnitude of unexplained variation in Y]

$$S_{Y.X}^2 = \frac{1}{n}\sum\left[\left(Y_i - \overline{Y}\right) - b\left(X_i - \overline{X}\right)\right]^2$$

$$= \frac{\sum\left(Y_i - \overline{Y}\right)^2}{n} + \frac{b^2\sum\left(X_i - \overline{X}\right)^2}{n} - \frac{2b\sum\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{n}$$

$$= \sigma_Y^2 + b^2\sigma_X^2 - 2b \cdot b\sigma_X^2 = \sigma_Y^2 - b^2\sigma_X^2$$

$$= \sigma_Y^2 - r^2\sigma_Y^2 = \sigma_Y^2\left(1 - r^2\right)$$

Similarly, it can be shown that the mean of e'_i ($= X_i - X_{Ci}$) values, in case of regression of X on Y, is also equal to zero. Further, their variance, i.e.,

$$S_{X.Y}^2 = \sigma_X^2\left(1 - r^2\right)$$

Alternatively equation (2) can be written as

$$S_{Y.X}^2 = \frac{1}{n}\sum\left(Y_i - Y_{ci}\right)Y_i = \frac{1}{n}\left[\sum Y_i^2 - a\sum Y_i - b\sum X_i Y_i\right]$$

Similarly, we can write

$$S_{X.Y}^2 = \frac{1}{n}\left[\sum X_i^2 - c\sum X_i - d\sum X_i Y_i\right]$$

Remarks:

The above expressions for the variance are based on the following:

$$\Sigma(Y_i - Y_{ci})^2 = \Sigma(Y_i - Y_{ci})(Y_i - Y_{ci})$$

$$= \Sigma(Y_i - Y_{ci})Y_i - \Sigma(Y_i - Y_{ci})Y_{ci}$$

It can be shown that the last term is zero.

Notes

$$\Sigma(Y_i - Y_{ci})Y_{ci} = \Sigma[(Y_i - \bar{Y}) - b(X_i - \bar{X})][\bar{Y} + b(X_i - \bar{X})]$$

$$= \bar{Y}\Sigma(Y_i - \bar{Y}) - b\bar{Y}\Sigma(X_i - \bar{X}) + b\Sigma(X_i - \bar{X})(Y_i - \bar{Y}) - b^2\Sigma(X_i - \bar{X})^2$$

$$= 0 - 0 + b^2\Sigma(X_i - \bar{X})^2 - b^2\Sigma(X_i - \bar{X})^2 = 0$$

23.5.1 Standard Error of the Estimate

The standard error of the estimate of regression is given by the positive square root of the variance of e_i values.

The standard error of the estimate of regression of Y on X or simply the standard error of the estimate of Y is given as, $S_{Y.X} = \sigma_Y\sqrt{1-r^2}$.

Similarly, $S_{X.Y} = \sigma_X\sqrt{1-r^2}$ is the standard error of the estimate X.

According to the theory of estimation, to be discussed in Chapter 21, an unbiased estimate of the variance of e_i values is given by

$$s^2_{Y.X} = \frac{\sum e_i^2}{n-2} = \frac{n}{n-2}\cdot\frac{\sum e_i^2}{n} = \frac{n}{n-2}\cdot\sigma_Y^2\left(1-r^2\right)$$

∴ The standard errors of the estimate of Y and that of X are written as

$$s_{Y.X} = \sigma_Y\sqrt{\frac{n}{(n-2)}\left(1-r^2\right)} \text{ and } s_{X.Y} = \sigma_X\sqrt{\frac{n}{(n-2)}\left(1-r^2\right)} \text{ respectively.}$$

Example 15:

From the following data, compute (i) the coefficient of correlation between X and Y, (ii) the standard error of the estimate of Y :

$\sum x^2 = 24 \quad \sum y^2 = 42 \quad \sum xy = 30 \quad N = 10$, where $x = X - \bar{X}$ and $y = Y - \bar{Y}$.

Solution.

The coefficient of correlation between X and Y is given by

$$r = \frac{\sum xy}{\sqrt{\sum x^2}\sqrt{\sum y^2}} = \frac{30}{\sqrt{24}\sqrt{42}} = 0.94$$

The standard error of the estimate of Y is given by (n < 30)

$$s_{Y.X} = \sqrt{\frac{\left(1-r^2\right)\sum y^2}{n-2}} = \sqrt{\frac{\left(1-0.94^2\right)\times 42}{8}} = 0.79$$

Example 16: For 100 items, it is given that the regression equations of Y on X and X on Y are 8X – 10Y + 66 = 0 and 40X – 18Y = 214 respectively. Compute the arithmetic means of X and Y and the coefficient of determination. If the standard deviation of X is given to be 3, compute the standard error of the estimate of Y.

Notes

Solution.

(a) The means of X and Y

Since the lines of regression pass through the point $(\overline{X}, \overline{Y})$, the simultaneous solution of the given regression equations would give the mean values of X and Y as $\overline{X} = 13, \overline{Y} = 17$

(b) The coefficient of determination

We assume that 8X - 10Y + 66 = 0 is the regression of Y on X and 40X - 18Y = 214 is the regression of X on Y. Thus, the respective regression coefficients b and d are given by $\frac{8}{10}$ and $\frac{18}{40}$.

$\therefore$ The coefficient of determination r^2 = b.d $= \frac{8}{10} \times \frac{18}{40} = 0.36$

(c) The standard error of the estimate of Y

We know that $s_{Y.X} = \sigma_Y \sqrt{1-r^2}$. To find s_Y we use the relation $b = r \cdot \frac{\sigma_Y}{\sigma_X}$.

Also $r^2 = \frac{9}{25}$ \ $r = \frac{3}{5}$ Thus, $\sigma_Y = \frac{b.\sigma_X}{r} = \frac{8}{10} \times \frac{5}{3} \times 3 = 4$

Hence, $s_{Y.X} = 4\sqrt{1-0.36} = 3.2$

23.6 Summary of Formulae

I. Regression of Y on X

1. Regression coefficient $b = \frac{\sum XY - n\overline{X}\,\overline{Y}}{\sum X^2 - n\overline{X}^2} = \frac{n\sum XY - (\sum X)(\sum Y)}{n\sum X^2 - (\sum X)^2}$

   Also b $= \frac{Cov(X,Y)}{\sigma_X^2} = r \cdot \frac{\sigma_Y}{\sigma_X}$

2. Change of scale and origin

   If $u = \frac{X-A}{h}$ and $v = \frac{Y-B}{h}$, then $b = \frac{k}{h}\left[\frac{n\sum uv - (\sum u)(\sum v)}{n\sum u^2 - (\sum u)^2}\right]$.

3. Constant term $a = \overline{Y} - b\overline{X}$

4. Alternative form of regression equation

   $$Y_C - \overline{Y} = \left(X - \overline{X}\right) \quad or \quad Y_C - \overline{Y} = r \cdot \frac{\sigma_Y}{\sigma_X}\left(X - \overline{X}\right)$$

Notes

5. Regression coefficient in bivariate frequency distribution

$$b = \frac{k}{h}\left[\frac{N\sum\sum f_{ij}u_i v_j - \left(\sum f_i u_i\right)\left(\sum f'_j v_j\right)}{N\sum f_i u_i^2 - \left(\sum f_i u_i\right)^2}\right]$$

6. Standard Error of the estimate

$$s_{Y.X} = \sigma_Y\sqrt{1-r^2} \quad \text{for large n (i.e., n > 30)}$$

$$= \sqrt{\frac{\sum\left(Y_i - \bar{Y}\right)^2\left(1-r^2\right)}{n-2}} \quad \text{for small n}$$

II. Regression of X on Y

1. Regression Coefficient $d = \frac{\sum XY - n\bar{X}\bar{Y}}{\sum Y^2 - n\bar{Y}^2} = \frac{n\sum XY - \left(\sum X\right)\left(\sum Y\right)}{n\sum Y^2 - \left(\sum Y\right)^2}$

$$= \frac{Cov(X,Y)}{\sigma_Y^2} = r\cdot\frac{\sigma_X}{\sigma_Y}$$

2. Change of scale and origin

If $u = \frac{X-A}{h}$ and $v = \frac{Y-B}{h}$, then $d = \frac{h}{k}\left[\frac{n\sum uv - (\sum u)(\sum v)}{n\sum v^2 - (\sum v)^2}\right]$.

3. Constant term $c = \bar{X} - d\bar{Y}$

4. Alternative form of regression equation

$X_C - \bar{X} = d\left(Y - \bar{Y}\right)$ *or* $X_C - \bar{X} = r.\frac{\sigma_X}{\sigma_Y}\left(Y - \bar{Y}\right)$

5. Regression coefficient in a bivariate frequency distribution

$$d = \frac{h}{k}\left[\frac{N\sum\sum f_{ij}u_i v_j - \left(\sum f_i u_i\right)\left(\sum f'_j v_j\right)}{N\sum f'_j v_j^2 - \left(\sum f'_j v_j\right)^2}\right]$$

6. Standard error of the estimate

$$s_{X.Y} = \sigma_X\sqrt{1-r^2} \quad \text{for large n (i.e., n > 30)}$$

$$= \sqrt{\frac{\sum\left(X_i - \bar{X}\right)^2\left(1-r^2\right)}{n-2}} \quad \text{for small n}$$

Notes

III. Relation of r with b and d

$$b \times d = r \cdot \frac{\sigma_Y}{\sigma_X} \cdot r \cdot \frac{\sigma_X}{\sigma_Y} = r^2$$

or $r = \sqrt{b \times d}$

23.7 Keywords

Coefficient of correlation: If the coefficient of correlation calculated for bivariate data (X_i, Y_i), i = 1,2, ... n, is reasonably high and a cause and effect type of relation is also believed to be existing between them, the next logical step is to obtain a functional relation between these variables.

Term regression: The term regression was first introduced by Sir Francis Galton in 1877.

Independent variable: For a bivariate data (X_i, Y_i), i = 1,2, ... n, we can have either X or Y as independent variable.

23.8 Self Assessment

1. Fill in the blanks :
   (i) The two regression coefficients are of .... sign.
   (ii) If a regression coefficient is negative then the correlation between the variables would also be ....
   (iii) The coefficient of determination is a real number lying between .... and ....
   (iv) Regression analysis is used to study .... between the variables.
   (v) If correlation between two variables is zero, the two regression lines are .... to each other and if it is equal to ± 1, the two lines are the ....
   (vi) The smaller is the angle between the two lines of regression, the .... is correlation between the variables.
   (vii) If r ≠ ± 1, the two regression lines are ....

23.9 Review Questions

1. Distinguish between correlation and regression. Discuss least square method of fitting regression.
2. What do you understand by linear regression ? Why there are two lines of regression? Under what condition(s) can there be only one line ?
3. Define the regression of Y on X and of X on Y for a bivariate data (X_i, Y_i), i = 1, 2, ... n. What would be the values of the coefficient of correlation if the two regression lines (a) intersect at right angle and (b) coincide?
4. (a) Show that the proportion of variations explained by a regression equation is r^2
   (b) What is the relation between Total Sum of Squares (TSS), Explained Sum of Squares (ESS) and Residual Sum of squares (RSS). Use this relationship to prove that the coefficient of correlation has a value between –1 and +1.

   Hint: See § 23.3

Notes

5. Write a note on the standard error of the estimate.
6. " The regression line gives only a 'best estimate' of the quantity in question. We may assess the degree of uncertainty in this estimate by calculating its standard error ". Explain.
7. Given a scatter diagram of a bivariate data involving two variables X and Y. Find the conditions of minimisation of $\sum(Y - Y_C)^2$ and hence derive the normal equations for the linear regression of Y on X. What sum is to be minimised when X is regressed on Y? Write down the normal equation in this case.
8. Explain, fully, the meaning of regression of one variable Y on another variable X. Discuss the method of least squares for fitting a linear regression of the form Y = a + bX. Write down the normal equations and show that $b = r \cdot \frac{\sigma_Y}{\sigma_X}$, where the symbols have their usual meaning.
9. Show that the coefficient of correlation is the geometric mean of the two regression coefficients.
10. What is the method of least squares ? Show that the two lines of regression obtained by this method are irreversible except when r = ± 1. Explain.
11. Show that, in principle, there are always two lines of regression for a bivariate data. Prove that the coefficient of correlation between two variables is either + 1 or - 1 when the two lines are identical and is zero when they are perpendicular.
12. Fit a linear regression of Y on X to the following data :

X	:	1	2	3	4	5	6	7	8
Y	:	65	80	45	86	178	205	200	250

13. Obtain the two lines of regression from the following data and show them on a graph. Also construct a scatter diagram of the data.

Age of husband (X) (*in years*)	:	23	27	28	28	28	30	30	33	35	38
Age of wife (Y) (*in years*)	:	18	20	22	27	21	29	27	29	28	29

14. The following table gives the information on the years of education (X) of nine farmers and annual yields per acre (Y) on their farms :

X	:	0	2	4	6	8	10	12	14	16
Y	:	4	4	6	10	10	8	12	8	6

(a) Find the regression equation of yield per acre on education and give an economic interpretation to it.

(b) What is the magnitude of the 'explained variation' in the dependent variable? Find the coefficient of correlation from it.

Notes

15. The following table gives the data relating to purchases and sales. Obtain the two regression equations by the method of least squares and estimate the likely sales when purchases equal 100.

Purchases	:	62	72	98	76	81	56	76	92	88	49
Sales	:	112	124	131	117	132	96	120	136	97	85

Answers: Self Assessment

1. (i) same (ii) negative (iii) 0 and 1 (iv) dependence (v) perpendicular, coincident (vi) more (vii) irreversible.

23.10 Further Readings

Books

Sheldon M. Ross, Introduction to Probability Models, Ninth Edition, Elsevier Inc., 2007.

Jan Pukite, Paul Pukite, Modeling for Reliability Analysis, IEEE Press on Engineering of Complex Computing Systems, 1998.

Unit 24: Sampling Distributions

Notes

CONTENTS

Objectives

Introduction

24.1 Distinction between Parameter and Statistic

24.2 Sampling Distribution of Sample Mean

24.2.1 Nature of the Sampling Distribution of Mean

24.2.2 Sampling Distribution of the Difference Between two Sample Means

24.3 Sampling Distribution of the Number of Successes

24.3.1 Sampling Distribution of Proportion of Successes

24.3.2 Sampling Distribution of the Difference of two Proportions

24.4 Summary

24.5 Keywords

24.6 Self Assessment

24.7 Review Questions

24.8 Further Readings

Objectives

After studying this unit, you will be able to:

- Distinction Between Parameter and Statistic
- Sampling Distribution of Sample Mean
- Sampling Distribution of the Number of Successes

Introduction

A theoretical probability distribution is constructed on the basis of the specification of the conditions of a random experiment. In contrast to this, if the construction of the probability distribution is based upon the random experiment of obtaining a sample from a population, the resulting distribution is termed as a sampling distribution.

As we know that the main aim of obtaining a sample from a population is to draw certain conclusions about it. The process of drawing such conclusions, known as 'Statistical Inference', is based upon the rules or the framework provided by various sampling distributions.

It may be recalled here that simple random sampling is a procedure of obtaining a sample of size n from a population of size N such that each combination of n units has an equal chance of being selected as a sample. This definition also implies that every unit of the population has an equal chance of being selected in the sample.

Notes

The above definition of random sampling holds in both the situations, i.e., in simple random sampling with replacement *(srswr)* and in simple random sampling with out replacement (srswor).

24.1 Distinction between Parameter and Statistic

Let $P_1, P_2 \ldots\ldots P_N$ denote the observations on N units of a population and $X_1, X_2 \ldots\ldots X_n$ be a simple random sample of size n from it.

A parameter is a measure computed from the observation of the population. For example :

Population Mean $(\mu) = \dfrac{P_1 + P_2 + \cdots\cdots + P_N}{N}$,

Population Variance $(\sigma^2) = \dfrac{1}{N}\sum(P_i - \mu)^2$, etc. are parameters.

In a similar way, a statistics is a measure computed from the observations of a sample. For example:

Sample Mean $(\overline{X}) = \dfrac{X_1 + X_2 + \cdots\cdots + X_n}{n}$,

Sample Variance $(S^2) = \dfrac{1}{n}\sum(X_i - \overline{X})^2$, etc. are statistic.

Formally, a parameter is any function of population values while a statistic is a function of sample values.

Very often, the values of various parameters are unknown and these are estimated by the corresponding statistic. For example, sample mean $\overline{X}$ is used as an estimator of population mean *m,* sample standard deviation S is used as an estimator of population standard deviation *s*, etc. The difference between a statistic and the corresponding parameter is known as sampling error. For example, the sampling error in estimation of *m* is $\overline{X}$ - *m*. It may be noted that the sampling error is an error caused by pure chance factors.

When we take a random sample $X_1, X_2 \ldots\ldots X_n$ from a population $P_1, P_2 \ldots\ldots P_N$, the first sample observation X_1 could be any one of the N population observations $P_1, P_2 \ldots\ldots P_N$. We know that the probability of selection of any one of the population observation is $\dfrac{1}{N}$ and therefore, we can regard X_1 as a random variable which can take values $P_1, P_2 \ldots\ldots P_N$ each with probability $\dfrac{1}{N}$.

Further, $E(X_1) = \dfrac{1}{N}\cdot P_1 + \dfrac{1}{N}\cdot P_2 + \cdots + \dfrac{1}{N}\cdot P_N = \dfrac{1}{N}\sum P_i = \mu$ and

Variance of $X_1 = E(X_1 - m)^2$

$$= \frac{1}{N}\left[(P_1 - \mu)^2 + (P_2 - \mu)^2 + \cdots + (P_N - \mu)^2\right] = \frac{1}{N}\sum(P_i - \mu)^2 = \sigma^2$$

In a similar way, $X_2, X_3 \ldots\ldots X_n$ are all random variables, each with mean *m* and variance *s*^{2}. The magnitude of covariance between any two of these variables, say X_i and X_j, will depend upon whether the sampling is with or without replacement.

In the case of sampling with replacement, $X_1, X_2 \ldots\ldots X_n$ would be statistically independent and the $Cov(X_i, X_j) = 0$ for $i \neq j$.

Notes

In the case of sampling without replacement, we can write

$$\text{Cov}(X_i, X_j) = E(X_i - m)(X_j - m) = \sum_{r=1}^{N} \sum_{\substack{s=1,\\ s\neq r}}^{N} (P_r - \mu)(P_s - \mu) \cdot p_{rs},$$

where p_{rs} is the joint probability that the rth unit of population is drawn at the ith draw and the sth unit of population is drawn at the jth draw. We note that $p_{rs} = \frac{1}{N(N-1)}$. Thus, we have

$$Cov(X_i, X_j) = \sum_{r=1}^{N} \sum_{\substack{s=1,\\ s\neq r}}^{N} (P_r - \mu)(P_s - \mu) \cdot \frac{1}{N(N-1)}$$

$$= \frac{1}{N(N-1)} \sum_{r=1}^{N} (P_r - \mu) \sum_{\substack{s=1,\\ s\neq r}}^{N} (P_s - \mu)$$

$$= \frac{1}{N(N-1)} \sum_{r=1}^{N} (P_r - \mu) \left[\sum_{s=1}^{N} (P_s - \mu) - (P_r - \mu) \right]$$

$$= \frac{1}{N(N-1)} \sum_{r=1}^{N} (P_r - \mu) \left[0 - (P_r - \mu) \right]$$

$$= -\frac{1}{N(N-1)} \sum_{r=1}^{N} (P_r - \mu)^2 = -\frac{1}{N(N-1)} \cdot N\sigma^2 = -\frac{\sigma^2}{(N-1)}$$

24.2 Sampling Distribution of Sample Mean

We know that $\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$. In the previous section we have shown that if the sample is random, then each of the X_i's are random variable with mean m and variance s^2. Since $\bar{X}$ is a linear combination of these random variables, therefore, it is also a random variable with mean equal to $E(\bar{X}) = \frac{1}{n}\left[E(X_1) + E(X_2) + \cdots + E(X_n)\right] = \frac{1}{n} \cdot n\mu = \mu$ and variance equal to

$$Var(\bar{X}) = E(\bar{X} - \mu)^2 = E\left[\frac{X_1 + X_2 + \cdots + X_n}{n} - \mu\right]^2$$

$$= E\left[\frac{(X_1 + X_2 + \cdots + X_n) - n\mu}{n}\right]^2 = \frac{1}{n^2} E\left[\sum (X_i - \mu)\right]^2$$

$$= \frac{1}{n^2} E\left[\sum (X_i - \mu)^2 + \sum\sum_{i\neq j} (X_i - \mu)(X_j - \mu)\right]$$

$$= \frac{1}{n^2}\left[\sum E(X_i - \mu)^2 + \sum\sum_{i\neq j} E(X_i - \mu)(X_j - \mu)\right]$$

$$= \frac{1}{n^2}\left[n\sigma^2 + \sum_{i \neq j}\sum Cov\left(X_i, X_j\right)\right]$$

Case I. If the sample is drawn with replacement, then X_1, X_2 ...... X_n are independent random variates and hence, $Cov(X_i, X_j) = 0$. Thus, we have

$$Var\left(\overline{X}\right) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

Case II. If the sample is drawn without replacement, then

$$Cov\left(X_i, X_j\right) = -\frac{\sigma^2}{N-1}, \text{ therefore,}$$

$$Var\left(\overline{X}\right) = \frac{1}{n^2}\left[n\sigma^2 - n(n-1)\frac{\sigma^2}{N-1}\right] = \frac{N-n}{N-1}\cdot\frac{\sigma^2}{n}$$

We note that if $N \to \infty$ (i.e., population becomes large), $\frac{N-n}{N-1} \to 1$ and therefore, in this case also, $Var\left(\overline{X}\right) = \frac{\sigma^2}{n}$.

Remarks:

1. The standard deviation of a statistic is termed as standard error. The standard error of $\overline{X}$, to be written in abbreviated form as $S.E.\left(\overline{X}\right)$, is equal to $\frac{\sigma}{\sqrt{n}}$, when sampling is with replacement and it is equal to $\frac{\sigma}{\sqrt{n}}\cdot\sqrt{\frac{N-n}{N-1}}$, when sampling is without replacement.
2. $S.E.\left(\overline{X}\right)$ is inversely related to the sample size.
3. The term $\sqrt{\frac{N-n}{N-1}}$ is termed as finite population correction (fpc). We note that fpc tends to become closer and closer to unity as population size becomes larger and larger.
4. As a general rule, fpc may be taken to be equal to unity when sample size is less than 5% of population size, i.e., $n < 0.05N$.

Example 1: Construct a sampling distribution of the sample mean for the following population when random samples of size 2 are taken from it (a) with replacement and (b) without replacement. Also find the mean and standard error of the distribution in each case.

Population Unit	:	1	2	3	4
Observation	:	22	24	26	28

Solution.

The mean and standard deviation of population are

$$\mu = \frac{22+24+26+28}{4} = 25 \text{ and}$$

$$\sigma = \sqrt{\frac{(22)^2+(24)^2+(26)^2+(28)^2}{4} - (25)^2} = \sqrt{5} = 2.236 \text{ respectively.}$$

(a) When random samples of size 2 are drawn, we have $4^2 = 16$ samples, shown below :

Sample No.	*Sample Values*	$\overline{X}$
1	22, 22	22
2	22, 24	23
3	22, 26	24
4	22, 28	25
5	24, 22	23
6	24, 24	24
7	24, 26	25
8	24, 28	26
9	26, 22	24
10	26, 24	25
11	26, 26	26
12	26, 28	27
13	28, 22	25
14	28, 24	26
15	28, 26	27
16	28, 28	28

Since all of the above samples are equally likely, therefore, the probability of each value of $\overline{X}$ is $\frac{1}{16}$. Thus, we can write the sampling distribution of $\overline{X}$ as given below:

$\overline{X}$	22	23	24	25	26	27	28	*Total*
$p(\overline{X})$	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{3}{16}$	$\frac{4}{16}$	$\frac{3}{16}$	$\frac{2}{16}$	$\frac{1}{16}$	1

The mean of $\overline{X}$, i.e.,

$$\mu_{\overline{X}} = E(X) = 22\times\frac{1}{16}+23\times\frac{2}{16}+24\times\frac{3}{16}+25\times\frac{4}{16}+26\times\frac{3}{16}+ 27\times\frac{2}{16}+28\times\frac{1}{16} = 25$$

Further, $S.E.(\overline{X}) = \sigma_{\overline{X}} = \sqrt{E(\overline{X}^2) - \left[E(\overline{X})\right]^2}$, where

$$E(\overline{X}^2) = \frac{1}{16}\left(22^2+23^2\times2+24^2\times3+25^2\times4+26^2\times3+27^2\times2+28^2\right)$$

$$= 627.5$$

Thus, $\sigma_{\overline{X}} = \sqrt{627.5-25^2} = \sqrt{2.5}$ which is equal to $\frac{\sigma}{\sqrt{n}}$.

(b) When random samples of size 2 are drawn without replacement, we have ${}^4C_2$ samples,

Notes

shown below:

Sample No.	*Sample Values*	$\bar{X}$
1	22, 24	23
2	22, 26	24
3	22, 28	25
4	24, 26	25
5	24, 28	26
6	26, 28	27

Since all the samples are equally likely, the probability of each value of $\bar{X}$ is $\frac{1}{6}$. Thus, we can write the sampling distribution of $\bar{X}$ as

$\bar{X}$	23	24	25	26	27	*Total*
$p(\bar{X})$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	1

Further, $\mu_{\bar{X}} = E(\bar{X}) = \frac{1}{6}[23 + 24 + 25 \times 2 + 26 + 27] = 25.$

To find $S.E.(\bar{X})$, we first find $E(\bar{X}^2)$ given by

$$E(\bar{X}^2) = \frac{1}{6}\left[23^2 + 24^2 + 2 \times 25^2 + 26^2 + 27^2\right] = \frac{3760}{6} = 626.67.$$

Thus, $\sigma_{\bar{X}} = \sqrt{626.67 - 25^2} = \sqrt{1.67} = 1.292.$

Alternatively, $\sigma_{\bar{X}} = \sqrt{\frac{N-n}{N-1} \cdot \frac{\sigma^2}{n}} = \sqrt{\frac{4-2}{3} \times \frac{5}{2}} = \sqrt{1.67} = 1.292.$

24.2.1 Nature of the Sampling Distribution of Mean

It can be deduced that when a random sample $X_1, X_2 \ldots\ldots X_n$ is obtained from a normal population with mean *m* and standard deviation *s*, then each of the X_i's are also distributed normally with mean *m* and standard deviation *s*.

By the use of additive (or reproductive) property of normal distribution, it follows that the distribution of $\bar{X}$, a linear combination of $X_1, X_2 \ldots\ldots X_n$, will also be normal. As shown in the previous section, the mean and standard error of the distribution would be *m* and $\frac{\sigma}{\sqrt{n}}$ respectively.

Remarks: Since normal population is often a large population, the fpc is always taken equal to unity.

The nature of the sampling distribution of $\bar{X}$, when parent population is not normal, is provided by Central Limit Theorem. This theorem states that:

If $X_1, X_2 \ldots\ldots X_n$ is a random sample of size n from a non-normal population of size N with mean *m* and standard deviation *s*, then the sampling distribution of $\bar{X}$ will approach normal distribution

Notes

with mean *m* and standard error $\frac{\sigma}{\sqrt{n}}\left(or\sqrt{\frac{N-n}{N-1}\cdot\frac{\sigma^2}{n}}\right)$ as n becomes larger and larger.

Remarks: As a general rule, when n ³ 30, the sampling distribution of $\overline{X}$ is taken to be normal for practical purposes.

Application of the Sampling Distribution

Decisions by various government and non-government agencies are made on the basis of sample results. For example, a sales manager may take a sample of quantities purchased of its product to predict sales. A government agency may take a sample of residents to assess the effect of a certain welfare program etc. Thus, in order to draw reliable conclusions, we must have a sound knowledge regarding the sample. An extremely common and quite useful knowledge about the sample is given by the sampling distribution of the relevant statistic.

An important application of sampling distribution is to determine the probability of the statistic lying in a given interval.

24.2.2 Sampling Distribution of the Difference Between two Sample Means

Let there be two populations of sizes N_1 and N_2, means m_1 and m_2 and standard deviations s_1 and s_2 respectively. Let $\overline{X}_1$ be the mean of the random sample of size n_1 obtained from the first population and $\overline{X}_2$ be the mean of the random sample of size n_2 obtained from the second population. Thus, we can regard $\overline{X}_1$ and $\overline{X}_2$ as two independent random variables with means m_1 and m_2 and standard errors as

$$\frac{\sigma_1}{\sqrt{n_1}}\left(or\sqrt{\frac{N_1-n_1}{N_1-1}\cdot\frac{\sigma_1^2}{n_1}}\right) \text{ and } \frac{\sigma_2}{\sqrt{n_2}}\left(or\sqrt{\frac{N_2-n_2}{N_2-1}\cdot\frac{\sigma_2^2}{n_2}}\right) \text{ respectively.}$$

Further, their difference, $\overline{X}_1 - \overline{X}_2$, will also be a random variable with mean $= E\left(\overline{X}_1 - \overline{X}_2\right) = E\left(\overline{X}_1\right) - E\left(\overline{X}_2\right) = m_1 - m_2$ and standard error

$$= \sqrt{Variance\left(\overline{X}_1 - \overline{X}_2\right)} = \sqrt{Var\left(\overline{X}_1\right) + Var\left(\overline{X}_2\right)}$$

$$= \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \text{ when both the samples are drawn using srswr) or}$$

$$= \sqrt{\frac{N_1-n_1}{N_1-1}\cdot\frac{\sigma_1^2}{n_1} + \frac{N_2-n_2}{N_2-1}\cdot\frac{\sigma_2^2}{n_2}} \text{ (when both the samples are drawn using srswor).}$$

Remarks:

1. When both the populations are normal, then $\overline{X}_1 - \overline{X}_2$ will be distributed normally with mean $m_1 - m_2$ and standard error $\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$.

Notes

2. Using Central Limit Theorem, the above result will also hold for a non-normal population when both n_1 and n_2 > 30 and fpc is approximately equal to unity, i.e., $n_i < 0.05 N_i$ (for i = 1, 2).

24.3 Sampling Distribution of the Number of Successes

Let *p* denote the proportion of successes in population, i.e.,

$$\pi = \frac{Number\ of\ successes\ in\ population}{Total\ number\ of\ units\ in\ population}$$

Let us take a random sample of n units from this population and let X denote the number of successes in the sample. Thus, X is a random variable with mean n*p* and standard error

$$\sqrt{n\pi(1-\pi)}\ \left(or \sqrt{\frac{N-n}{N-1}\cdot n\pi(1-\pi)}\right)$$

If sampling is done with replacement, then X is a binomial variate with mean np and standard error $\sqrt{n\pi(1-\pi)}$. Using central limit theorem, we can say that the distribution of the number of successes will approach a normal variate with mean n*p* and standard error $\sqrt{n\pi(1-\pi)}$ or $\sqrt{\frac{N-n}{N-1}\cdot n\pi(1-\pi)}$ for sufficiently large sample. The sample size is said to be sufficiently large if both n π and n(1 - π) are greater than 5.

24.3.1 Sampling Distribution of Proportion of Successes

Let $p = \frac{X}{n}$ be the proportion of successes in sample. Since X is a random variable, therefore, p is also a random variable with mean

$$E(p) = \frac{E(X)}{n} = \frac{n\pi}{n} = \pi \text{ and standard error}$$

$$= \sqrt{\frac{1}{n^2}Var(X)} = \sqrt{\frac{n\pi(1-\pi)}{n^2}} = \sqrt{\frac{\pi(1-\pi)}{n}} \quad \text{(when srswr)}$$

or $$= \sqrt{\frac{N-n}{N-1}\cdot\frac{\pi(1-\pi)}{n}} \quad \text{(when srswor)}$$

As in the previous section, the sampling distribution of p will also be normal if both n π and n(1 - π) are greater than 5.

Example 2: There are 500 mangoes in a basket out of which 80 are defective. If obtaining a defective mango is termed as a success, determine the mean and standard error of the proportion of successes in a random sample of 10 mangoes, drawn (a) with replacement and (b) without replacement.

Solution.

Notes

It is given that $\pi = \frac{80}{500} = \frac{4}{25}$. Therefore, $E(p) = \pi = \frac{4}{25}$ and

(a) S.E.(p) $= \sqrt{n\pi(1-\pi)} = \sqrt{10 \times \frac{4}{25} \times \frac{21}{25}} = 1.159$ (srswr)

(b) S.E.(p) $= \sqrt{\frac{500-80}{499} \times 10 \times \frac{4}{25} \times \frac{21}{25}} = 1.063$ (srswor)

Example 3: 20% under graduates of a large university are found to be smokers. A sample of 100 students is selected at random. Construct the sampling distribution of the number of smokers. Also find the probability that the number of smokers in the sample is greater than 25.

Solution.

It is given that $\pi = \frac{20}{100} = \frac{1}{5}$. Since sample size, n = 100, is large, the number of successes X will be distributed normally with mean $100 \times \frac{1}{5} = 20$ and standard error $\sqrt{100 \times \frac{1}{5} \times \frac{4}{5}} = 4$.

Further, $P(X > 25) = P\left(z > \frac{25-20}{4}\right) = P(z > 1.25) = 0.1056.$

24.3.2 Sampling Distribution of the Difference of two Proportions

Let p_1 be proportion of successes in a random sample of size n_1 from a population with proportion of successes = p_1 and p_2 be the proportion of successes in a random sample of size n_2 from second population with proportion of successes = p_2. Assuming that the sample sizes are large, we can write

$$p_1 \sim N\left(\pi_1, \sqrt{\frac{\pi_1(1-\pi_1)}{n_1}}\right) \text{ and } p_2 \sim N\left(\pi_2, \sqrt{\frac{\pi_2(1-\pi_2)}{n_2}}\right)$$

Thus, their difference (p_1 - p_2) will be distributed normally with mean = $\pi_1 - \pi_2$ and standard error

$$\sqrt{\frac{\pi_1(1-\pi_1)}{n_1} + \frac{\pi_2(1-\pi_2)}{n_2}}.$$

Note: The above result will hold when we ignore fpc and the sample size, n_1 and n_2, is greater than 5 divided by the minimum of π_1, $(1 - \pi_1)$, π_2 and $(1 - \pi_2)$.

Some other Sampling Distributions

We have seen that the sampling distributions of mean and proportion of successes are normal.

Notes

Apart from normal distribution, there are certain other probability distributions that are useful in sampling theory. These distributions are:

1. Chi - square $\left(\chi^2\right)$ distribution.
2. Student's t - distribution.
3. Snedecor's F - distribution.

24.4 Summary

- Let $P_1, P_2 \ldots\ldots P_N$ denote the observations on N units of a population and $X_1, X_2 \ldots\ldots X_n$ be a simple random sample of size n from it.

  A parameter is a measure computed from the observation of the population. For example:

  Population Mean $(\mu) = \dfrac{P_1 + P_2 + \cdots\cdots + P_N}{N}$,

  Population Variance $\left(\sigma^2\right) = \dfrac{1}{N}\sum\left(P_i - \mu\right)^2$, etc. are parameters.

  In a similar way, a statistics is a measure computed from the observations of a sample. For example:

  Sample Mean $\left(\overline{X}\right) = \dfrac{X_1 + X_2 + \cdots\cdots + X_n}{n}$,

  Sample Variance $\left(S^2\right) = \dfrac{1}{n}\sum\left(X_i - \overline{X}\right)^2$, etc. are statistic.

- The standard deviation of a statistic is termed as standard error. The standard error of $\overline{X}$, to be written in abbreviated form as $S.E.\left(\overline{X}\right)$, is equal to $\dfrac{\sigma}{\sqrt{n}}$, when sampling is with replacement and it is equal to $\dfrac{\sigma}{\sqrt{n}} \cdot \sqrt{\dfrac{N-n}{N-1}}$, when sampling is without replacement.

- $S.E.\left(\overline{X}\right)$ is inversely related to the sample size.

- The term $\sqrt{\dfrac{N-n}{N-1}}$ is termed as finite population correction (fpc). We note that fpc tends to become closer and closer to unity as population size becomes larger and larger.

24.5 Keywords

Theoretical probability: A theoretical probability distribution is constructed on the basis of the specification of the conditions of a random experiment.

Parameter: A parameter is any function of population values while a statistic is a function of sample values.

24.6 Self Assessment

1. State whether the following statements are True or False:

   (i) Mean of the sample means is equal to population mean.

   (ii) Random variable of a sampling distribution is called a statistic.

   (iii) The sampling distribution of $\overline{X}$ is normal if the drawn samples are of size 20.

   (iv) When population is large, the finite population correction (fpc) is negligible, i.e., approximately equal to zero.

   (v) In order that a statistic t follows a t - distribution, the sample should have been obtained from a normal population.

24.7 Review Questions

1. Explain the concept of sampling distribution of a statistics.

2. Find the mean and standard error of sample mean in (a) Simple random sampling with replacement, (b) Simple random sampling without replacement.

3. Distinguish between:

   (a) Parameter and Statistic.

   (b) Sampling distribution and Probability distribution.

   (c) Standard deviation and Standard error.

4. (a) Distinguish between sampling with replacement and sampling without replacement. How many random samples of size n can be drawn from a population consisting N items if the sampling is done (i) with replacement, (ii) without replacement?

   (b) What is the variance of the sample mean if sampling is done (i) with replacement (ii) without replacement?

   (c) Under what conditions do the answers in (b) approach each other?

5. If X_i (i = 1, 2, ... n) are n independent normal variates with respective mean μ_i and standard deviation σ_i, then show that the variate u = $\sum X_i$ is normally distributed with mean $\sum_{mi}$ and variance $\sum \sigma_i^2$.

Answers: Self Assessment

1. (i) T (ii) T (iii) F (iv) F (v) T

24.8 Further Readings

Books

Sheldon M. Ross, Introduction to Probability Models, Ninth Edition, Elsevier Inc., 2007.

Jan Pukite, Paul Pukite, Modeling for Reliability Analysis, IEEE Press on Engineering of Complex Computing Systems, 1998.

Notes

Unit 25: Chi - Sqaure (χ^2) Distribution

CONTENTS

Objectives

Introduction

25.1 Chi - Square Distribution

25.1.1 Sampling Distribution of Variance

25.2 Summary

25.3 Keywords

25.4 Self Assessment

25.5 Review Questions

25.6 Further Readings

Objectives

After studying this unit, you will be able to:

- Discuss Chi - Square (χ^2) Distribution
- Describe some examples related to Chi - Square

Introduction

When sampling is done with replacement, each unit of the population has a probability of its selection equal to $\frac{1}{N}$. Further, there are N^n possible samples that are equally likely, and therefore, the probability of selection of each sample is $\frac{1}{N^n}$.

When sampling is done without replacement, the units are either drawn one by one, without replacement, or all the n units are selected in one attempt. We know that the permutations of N objects taking n at a time is ${}^N P_n$ and this becomes the number of ordered samples. Corresponding to this, the number of unordered samples will be ${}^N C_n$, each with probability $\frac{1}{{}^N C_n}$. In this case also, the probability of selection of a unit at any draw is $\frac{1}{N}$. For example, the probability of selection of a unit at the first draw = $\frac{1}{N}$, the probability of its selection at the second draw is $\frac{N-1}{N} \times \frac{1}{N-1} = \frac{1}{N}$ and so on, the probability of its selection at the rth draw is

$$\frac{N-1}{N} \cdot \frac{N-2}{N-1} \cdot \;\ldots\ldots\; \frac{N-r+1}{N-r+2} \cdot \frac{1}{N-r+1} = \frac{1}{N}$$

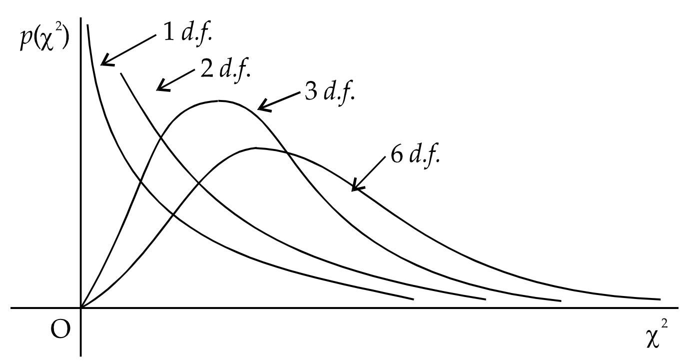
Notes

25.1 Chi - Square $\left(\chi^2\right)$ Distribution

We know that if X is a random variate distributed normally with mean *m* and standard deviation *s*, then $z = \frac{X-\mu}{\sigma}$ is a standard normal variate. Square of z, i.e., $z^2 = \frac{(X-\mu)^2}{\sigma^2}$ if distributed as χ^2 - variate with one degree of freedom and is written as χ_1^2. Further, the value of χ_1^2, a squared value, will lie between 0 to ∞, for z lying between - ∞ to ∞. Since most of the z-values are close to zero, the probability density of χ^2 will be highest near zero. The χ^2 distribution with one degree of freedom is shown in Figure 25.1.

Generalising the above result, we can say that if X_1, X_2 ...... X_n are n independent normal variates each with mean m_i and standard deviations σ_i, i = 1, 2, ... n, respectively, then the sum of squares $\sum z_i^2 = \sum \frac{(X_i - \mu_i)^2}{\sigma_i^2}$ is a χ^2 variate with n degrees of freedom, i.e., χ_n^2. Thus, we can say that χ_n^2 is sum of squares of n independent standard normal variate.

Figure 25.1



Features of χ^2 Distribution

1. The distribution has only one parameter, i.e., number of degrees of freedom or d.f. (in abbreviated form) which is a positive integer.
2. We may note that as the d.f. increases, the height of the probability density function decreases. The distribution is positively skewed and the skewness decreases asd.f. increases. For large values of d.f., the distribution approaches normal distribution. The curves for various d.f. are shown in figure 20.1.
3. The mean of χ_n^2, i.e., $E\left(\chi_n^2\right) = n$ and its variance = 2n, where n = d.f.

Notes

4. Additive property

The sum of two independent χ^2 variates is also a χ^2 variate with degrees of freedom equal to the sum of their individual degrees of freedom.

If χ_n^2 and χ_m^2 are two independent random variates with n and m degrees of freedom respectively, then $\chi_n^2 + \chi_m^2$ is also a χ^2 variate with n + m degrees of freedom.

Remarks:

1. The degrees of freedom is defined as the number of independent random variables. If n is the number of variables and k is the number of restrictions on them, the degrees of freedom are said to be n - k.

2. On the basis of the definition of degrees of freedom, given above, we can say that $\sum_{i=1}^{n}\left(\frac{X_i - \bar{X}}{\sigma}\right)^2$ is a χ^2 variate with (n - 1) degrees of freedom. It may be pointed out here that one degree of freedom is reduced because for a given value of $\bar{X}$, the number of independent variables is (n - 1).

25.1.1 Sampling Distribution of Variance

Using χ^2 -distribution, we can construct the sampling distribution of $S^2 = \frac{1}{n}\sum\left(X_i - \bar{X}\right)^2$.

Let X_1, X_2 ...... X_n be a random sample of size n from a normal population with mean *m* and variance *s*^{2}. We can write

$$X_i - \mu = \left(X_i - \bar{X}\right) + \left(\bar{X} - \mu\right)$$

Squaring both sides and taking sum over all the n observations, we get

$$\sum_{i=1}^{n}\left(X_i - \mu\right)^2 = \sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 + \sum_{i=1}^{n}\left(\bar{X} - \mu\right)^2 + 2\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(\bar{X} - \mu\right)$$

$$= \sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 + n\left(\bar{X} - \mu\right)^2 + 2\left(\bar{X} - \mu\right)\sum_{i=1}^{n}\left(X_i - \bar{X}\right)$$

We note that the last term is zero. Therefore, we have

$$\sum_{i=1}^{n}\left(X_i - \mu\right)^2 = \sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 + n\left(\bar{X} - \mu\right)^2$$

Dividing both sides by *s*^{2}, we get

$$\frac{\sum_{i=1}^{n}\left(X_i - \mu\right)^2}{\sigma^2} = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}{\sigma^2} + \frac{n\left(\bar{X} - \mu\right)^2}{\sigma^2}$$

Notes

or $$\frac{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2}{\sigma^2}=\frac{\sum_{i=1}^{n}\left(X_i-\mu\right)^2}{\sigma^2}-\frac{\left(\bar{X}-\mu\right)^2}{\sigma^2/n}=\chi_n^2-\chi_1^2=\chi_{n-1}^2$$

Thus, $\frac{\sum\left(X_i-\bar{X}\right)^2}{\sigma^2}$ *or* $\frac{nS^2}{\sigma^2}$ is a χ^2 -variate with (n - 1) d.f.

Mean and Standard Error of S²

Since the random variable $\frac{nS^2}{\sigma^2}$ is a χ^2 -variate with (n - 1) d.f.,

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

Thus, we have $E\left(S^2\right)=\frac{n-1}{n}\cdot\sigma^2$

Further, if we define $s^2=\frac{1}{n-1}\sum\left(X_i-\bar{X}\right)^2$ so that $s^2=\frac{n}{n-1}\cdot S^2$, we have

$E\left(s^2\right)=\frac{n}{n-1}\cdot E\left(S^2\right)=\frac{n}{n-1}\cdot\frac{n-1}{n}\cdot\sigma^2=\sigma^2$ (See Remarks 2 below).

To find variance of S^2, we make use of the fact that variance of $\frac{nS^2}{\sigma^2}$ is 2(n - 1). This implies that

$$E\left[\frac{nS^2}{\sigma^2}-(n-1)\right]^2=2(n-1) \text{ or } \frac{n^2}{\sigma^4}E\left(S^2-\frac{n-1}{n}\cdot\sigma^2\right)^2=2(n-1)$$

$$\therefore\ E\left[S^2-E\left(S^2\right)\right]^2=\frac{2(n-1)}{n^2}\cdot\sigma^4 \text{ or } Var\left(S^2\right)=\frac{2(n-1)}{n^2}\cdot\sigma^4$$

Further, variance of s^2 = variance of $\left(\frac{n}{n-1}\cdot S^2\right)$. This gives

$$Var\left(s^2\right)=\frac{n^2}{(n-1)^2}\cdot Var\left(S^2\right)=\frac{n^2}{(n-1)^2}\times\frac{2(n-1)}{n^2}\cdot\sigma^4=\frac{2}{n-1}\cdot\sigma^4$$

Remarks:

1. The distributions of c^2 and S^2 are based upon the assumption that the parent population is normal. If the parent population is not normal, it is not possible to comment upon the nature of the distribution of the above statistics.

Notes

2. It will be discussed in the following chapter that when expected value of a statistic equals the value of parameter, it is said to be an unbiased estimate of the parameter.

Problem 1

The Acme Battery Company has developed a new cell phone battery. On average, the battery lasts 60 minutes on a single charge. The standard deviation is 4 minutes.

Suppose the manufacturing department runs a quality control test. They randomly select 7 batteries. The standard deviation of the selected batteries is 6 minutes. What would be the chi-square statistic represented by this test?

Solution

We know the following:

- The standard deviation of the population is 4 minutes.
- The standard deviation of the sample is 6 minutes.
- The number of sample observations is 7.

To compute the chi-square statistic, we plug these data in the chi-square equation, as shown below.

$x^2 = [(n – 1) * s^2] / \sigma^2$

$x^2 = [(7 – 1) * 6^2] / 4^2 = 13.5$

where x^2 is the chi-square statistic, n is the sample size, s is the standard deviation of the sample, and σ is the standard deviation of the population.

Problem 2

Let’s revisit the problem presented above. The manufacturing department ran a quality control test, using 7 randomly selected batteries. In their test, the standard deviation was 6 minutes, which equated to a chi-square statistic of 13.5.

Suppose they repeated the test with a new random sample of 7 batteries. What is the probability that the standard deviation in the new test would be greater than 6 minutes?

Solution

We know the following:

- The sample size *n* is equal to 7.
- The degrees of freedom are equal to *n* - 1 = 7 - 1 = 6.
- The chi-square statistic is equal to 13.5 (see Example 1 above).

Given the degrees of freedom, we can determine the cumulative probability that the chi-square statistic will fall between 0 and any positive value. To find the cumulative probability that a chi-square statistic falls between 0 and 13.5, insert the values in formula then result is the cumulative probability: 0.96.

This tells us that the probability that a standard deviation would be less than or equal to 6 minutes is 0.96. This means (by the subtraction rule) that the probability that the standard deviation would be *greater than* 6 minutes is 1 – 0.96 or .04.

Notes

25.2 Summary

- We know that if X is a random variate distributed normally with mean *m* and standard deviation *s*, then $z = \frac{X-\mu}{\sigma}$ is a standard normal variate. Square of z, i.e., $z^2 = \frac{(X-\mu)^2}{\sigma^2}$ if distributed as χ^2 - variate with one degree of freedom and is written as χ_1^2. Further, the value of χ_1^2, a squared value, will lie between 0 to ∞, for z lying between - ∞ to ∞. Since most of the z-values are close to zero, the probability density of χ^2 will be highest near zero. The χ^2 distribution with one degree of freedom is shown in Figure 25.1.

  Generalising the above result, we can say that if X_1, X_2 ...... X_n are n independent normal variates each with mean m_i and standard deviations σ_i, i = 1, 2, ... n, respectively, then the sum of squares $\sum z_i^2 = \sum \frac{(X_i - \mu_i)^2}{\sigma_i^2}$ is a χ^2 variate with n degrees of freedom, i.e., χ_n^2.

  Thus, we can say that χ_n^2 is sum of squares of n independent standard normal variate.

- Since the random variable $\frac{nS^2}{\sigma^2}$ is a χ^2 -variate with (n - 1) d.f.,

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

  Thus, we have $E\left(S^2\right) = \frac{n-1}{n}\cdot\sigma^2$

  Further, if we define $s^2 = \frac{1}{n-1}\sum\left(X_i - \overline{X}\right)^2$ so that $s^2 = \frac{n}{n-1}\cdot S^2$, we have

  $E\left(s^2\right) = \frac{n}{n-1}\cdot E\left(S^2\right) = \frac{n}{n-1}\cdot\frac{n-1}{n}\cdot\sigma^2 = \sigma^2$ (See Remarks 2 below).

25.3 Keywords

Standard normal variate: if X is a random variate distributed normally with mean *m* and standard deviation *s*, then $z = \frac{X-\mu}{\sigma}$ is a standard normal variate.

Distribution: The distribution has only one parameter, i.e., number of degrees of freedom or d.f. (in abbreviated form) which is a positive integer.

25.4 Self Assessment

1. Fill in the blanks:

   (i) The positive square root of variance of a sampling distribution is known as ....

   (ii) The standard error of $\overline{X}$ varies ...... with standard deviation and ...... with sample size.

   (iii) The sampling distribution of proportion would be approximately normal when n is greater than or equal to ....

   (iv) The mean and variance of a χ^2-variate depend upon its ....

25.5 Review Questions

1. We are given the fact that 30% of all patients admitted to a medical clinic fail to pay their bills and the bills are eventually forgiven. If the clinic treats 2000 different patients over a period of one year, what is the expected number of bills that would have to be forgiven. If X is the number of forgiven bills in the group of 2000 patients, find the variance and standard deviation of X. What can you say about the probability that X will exceed 700?

2. A random sample of 10 observations is to be taken from a normal population with variance equal to 16. What is the probability of obtaining a sample with variance greater than 20?

   Hint : Use χ^2 - distribution.

3. Two independent random samples of sizes 15 and 12 are taken from a normal population. Find the probability that the ratio of their variances is greater than 3. Assume that the variance of the sample of size 15 is greater than the variance of the other.

Answers: Self Assessment

1. (i) standard error (ii) directly, inversely (iii) 50 (iv) parameter

25.6 Further Readings

Books

Sheldon M. Ross, Introduction to Probability Models, Ninth Edition, Elsevier Inc., 2007.

Jan Pukite, Paul Pukite, Modeling for Reliability Analysis, IEEE Press on Engineering of Complex Computing Systems, 1998.

Unit 26: χ^2 - Test Hypothesis

Notes

CONTENTS

Objectives

Introduction

26.1 Test of Hypothesis concerning Correlation Coefficient

26.1.1 Test of Hypothesis Concerning Significance of Correlation Coefficient

26.1.2 Test of Hypothesis concerning Correlation Coefficient using Fisher's Z test

26.1.3 Test Concerning Equality of Correlations in two Populations

26.2 Uses of c2 test

26.2.1 c 2 - test as a Goodness of Fit

26.2.2 c2 - test as a Test for Independence of Two Attributes

26.3 Summary

26.4 Keywords

26.5 Self Assessment

26.6 Review Questions

26.7 Further Readings

Objectives

After studying this unit, you will be able to:

- Discuss Test of Hypothesis Concerning Significance of Correlation Coefficient
- Describe Test of Hypothesis concerning Correlation Coefficient using Fisher's Z test
- Explain Test Concerning Equality of Correlations in two Populations

Introduction

In last unit you have studied about hypothesis concerning standard deviation. In this unit you will go through χ^2 - test hypothesis.

26.1 Test of Hypothesis concerning Correlation Coefficient

Let ρ be coefficient of linear correlation in a bivariate normal population and r be its estimator based on a sample of n observations (X_i, Y_i).

26.1.1 Test of Hypothesis Concerning Significance of Correlation Coefficient

Here we have to test whether ρ is different from zero. Accordingly, H_0 and H_a are $\rho = 0$ and $\rho \neq 0$ respectively.

Notes

For small samples, the test statistic can be obtained from the sampling distribution of b. We note that if $r = 0$, then b would also be zero.

Therefore, $\frac{b}{S.E.(b)} = r \cdot \frac{S_Y}{S_X} \cdot \frac{1}{S.E.(b)} = r \cdot \frac{S_Y}{S_X} \cdot \frac{S_X}{S_Y}\sqrt{\frac{n-2}{1-r^2}} = r\sqrt{\frac{n-2}{1-r^2}}$ will follow t - distribution with

(n - 2) d.f. Hence, $r\sqrt{\frac{n-2}{1-r^2}}$ can be taken as the test statistic. We note that $S.E.(r) = \sqrt{\frac{1-r^2}{n-2}}$.

Therefore, 100(1 - a)% confidence limits of r can be written as $r \pm t_{a/2}$ S.E.(r).

Example 1: A random sample of 11 pairs of observations from a bivariate normal population gave r = 0.29. Test the significance of correlation in population.

Solution.

We have to test $H_0 : \rho = 0$ against $H_a : \rho \neq 0$.

$$t_{cal} = |0.29|\sqrt{\frac{9}{1-0.29^2}} = 0.91.$$

The value of t from tables at 5% level of significance and 9 d.f. is 2.26. Thus, there is no evidence against H_0.

26.1.2 Test of Hypothesis concerning Correlation Coefficient using Fisher's Z test

This test is applicable whether n is small or large. If r is correlation in sample, then its Fisher's Z transformation is given by $Z = \frac{1}{2}\log_e \frac{1+r}{1-r}$.

Further, if r is correlation in population, its Fisher's Z transformation, denoted by x, is given by

$$\xi = \frac{1}{2}\log_e \frac{1+\rho}{1-\rho}$$

Fisher has shown that the sampling distribution of Z is approximately normal with mean x and standard error $\frac{1}{\sqrt{n-3}}$. Thus, $(Z-\xi)\sqrt{n-3} \sim N(0,1)$.

Note: Since the values of Z and x are defined using e as the base of the logarithms, it is necessary to convert them into logarithms with base 10 for calculation purposes. Accordingly, we write

$$Z = \frac{1}{2}\log_e \frac{1+r}{1-r} = \frac{1}{2}\log_{10} \frac{1+r}{1-r} \times \log_e 10 = \frac{1}{2}\log_{10} \frac{1+r}{1-r} \times \frac{1}{\log_{10} e}$$

$$= \frac{1}{2} \times 2.3026 \times \log_{10} \frac{1+r}{1-r} = 1.1513 \log_{10} \frac{1+r}{1-r}$$

Similarly, we have $\xi = 1.1513 \log_{10} \frac{1+\rho}{1-\rho}$

Example 2: If the correlation of 10 pairs of observations (X, Y) is 0.96, test the hypothesis that correlation in population is 0.99.

Solution.

We have to test $H_0 : \rho = 0.99$ against $H_a : \rho \neq 0.99$

Further,

$$Z = 1.1513 \log_{10} \frac{1+0.96}{1-0.96} = 1.1513 \log_{10} 49 = 1.1513 \times 1.6902 = 1.9459 \text{ and}$$

$$\xi = 1.1513 \log_{10} \frac{1+0.99}{1-0.99} = 1.1513 \log_{10} 199 = 1.1513 \times 2.2989 = 2.6464$$

The test statistic is $z = |Z - \xi|\sqrt{n-3} = |1.9459 - 2.6464|\sqrt{7} = 1.8563.$

Since this value is less than 1.96, there is no evidence against H_0 at 5% level of significance.

26.1.3 Test Concerning Equality of Correlations in two Populations

Let there be two independent random samples of sizes n_1 and n_2 from two normal populations with correlations ρ_1 and ρ_2 respectively. Let r_1 and r_2 be the correlations computed from the respective samples.

If Z_1, Z_2, ξ_1 and ξ_2 denote Fisher's transformation of r_1, r_2, r_1 and ρ_2 respectively, then $Z_1 \sim N\left(\xi_1, \frac{1}{\sqrt{n_1-3}}\right)$ and $Z_2 \sim N\left(\xi_2, \frac{1}{\sqrt{n_2-3}}\right)$

$$\therefore\ Z_1 - Z_2 \sim N\left(\xi_1 - \xi_2, \sqrt{\frac{1}{n_1-3} + \frac{1}{n_2-3}}\right)$$

$$\text{or } \frac{Z_1 - Z_2}{\sqrt{\frac{1}{n_1-3} + \frac{1}{n_2-3}}} \sim N(0,1) \text{ under } H_0 : \rho_1 = \rho_2$$

Example 3: The correlation coefficients 0.89 and 0.85 were computed from two independent samples of sizes 12 and 16 respectively. Test whether they can be regarded to have come from two bivariate populations with different correlation coefficients?

Solution.

We shall test $H_0 : \rho_1 = \rho_2$ against $H_a : \rho_1 \neq \rho_2$.

Now $Z_1 = 1.1513 \log_{10} \frac{1.89}{0.11} = 1.1513 \log_{10} 17.18 = 1.1513 \times 1.2350 = 1.42$

and $Z_2 = 1.1513 \log_{10} \frac{1.85}{0.15} = 1.1513 \log_{10} 12.33 = 1.1513 \times 1.0911 = 1.26$

Notes

$\therefore$ The test statistic is $z = \frac{|1.42 - 1.26|}{\sqrt{\frac{1}{9}+\frac{1}{13}}} = 0.16 \times \sqrt{\frac{9 \times 13}{22}} = 0.369.$

Since this value is less than 1.96, there is no evidence against H_0 at 5% level of significance. Thus, the given samples provide no evidence of different correlations in two populations.

26.2 Uses of χ^2 test

In addition to the use of χ^2 in tests of hypothesis concerning the standard deviation, it is used as a test of goodness of fit and as a test of independence of two attributes. These tests are explained in the following sections.

26.2.1 χ^2 - test as a Goodness of Fit

χ^2 - test can be used to test, how far the fitted or the expected frequencies are in agreement with the observed frequencies. We know that for large values of n, the sampling distribution of X, the number of successes, is normal with mean np and variance np(1 - π). Thus, $z = \frac{X - n\pi}{\sqrt{n\pi(1-\pi)}} \sim N(0,1).$

Further, square of z is a c^2 - variate with one degree of freedom. We can write

$$z^2 = \frac{(X-n\pi)^2}{n\pi(1-\pi)} = (X-n\pi)^2\left[\frac{(1-\pi)+\pi}{n\pi(1-\pi)}\right] \qquad (\because 1 = 1-\pi+\pi)$$

$$= (X-n\pi)^2\left[\frac{1}{n\pi}+\frac{1}{n(1-\pi)}\right] = \frac{(X-n\pi)^2}{n\pi}+\frac{(X-n\pi)^2}{n(1-\pi)} \qquad \text{.... (1)}$$

We can write $\frac{(X-n\pi)^2}{n(1-\pi)} = \frac{(X-n+n-n\pi)^2}{n(1-\pi)} = \frac{\left[(X-n)+n(1-\pi)\right]^2}{n(1-\pi)}$

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

Similarly $\frac{(X-n\pi)^2}{n\pi} = \frac{\left[X-E(X)\right]^2}{E(X)}$.

Thus, equation (1) can be written as $z^2 = \frac{\left[X-E(X)\right]^2}{E(X)} + \frac{\left[(n-X)-E(n-X)\right]^2}{E(n-X)}$

Here X denotes the observed number of successes and (n - X) the observed number of failures.

Let O_1, E_1 denote the observed and expected number of successes respectively and O_2, E_2 denote the observed and expected number of failures respectively.

$\therefore \quad z^2 = \frac{(O_1-E_1)^2}{E_1} + \frac{(O_2-E_2)^2}{E_2}$ is a χ^2 - variate with 1 d.f.

Notes

Also we note that $O_1 + O_2 = E_1 + E_2 = n$.

The above result can be generalised for a manifold classification. If a population is divided into k mutually exclusive classes with observed and expected frequencies as $O_1, O_2, \ldots O_k$ and $E_1, E_2, \ldots E_k$ respectively, then $\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$ is a c^2 - variate with (k - 1) d.f. Here again we have

$\sum_{i=1}^{k} O_i = \sum_{i=1}^{k} E_i = N$ (total frequency).

Example 40: 300 digits were chosen from a table of numbers and the following frequency distribution was obtained :

Digit	:	0	1	2	3	4	5	6	7	8	9
Frequency	:	26	28	33	32	28	37	33	30	30	23

Test the hypothesis that the digits are uniformly distributed over the table.

Solution.

When H_0 is true, the expected frequency of each digit would be 30.

$$\therefore \quad \chi^2 = \frac{1}{30} \sum O_i^2 - N$$

$$= \frac{1}{30}\left(26^2 + 28^2 + 33^2 + 32^2 + 28^2 + 37^2 + 33^2 + 30^2 + 30^2 + 23^2\right) - 300 = 4.8$$

The value of χ^2 from table for 5% level of significance and 9 d.f. is 16.92. Since the calculated value is less than tabulated, there is no evidence against H_0. Thus, the distribution of numbers over the table may be treated as uniform.

Example 41: A sample analysis of examination results of 200 M.B.A.'s was made. It was found that 46 students had failed, 68 secured a third division, 62 secured a second division and the rest were placed in the first division. Are these figures commensurate with the general examination result which is in the ratio of 2 : 3 : 3 : 2 for the various categories, respectively? (Given : Table value of chi-square for 3 d.f. at 5% level of significance is 7.81.)

Solution.

H_0: The students in various categories are distributed in the ratio 2 : 3 : 3 : 2.

The expected number of students, under the assumption that H_0 is true, are :

expected number of failures $= \frac{2}{(2+3+3+2)} \times 200 = 40$,

expected number of third divisioners $= \frac{3}{10} \times 200 = 60$,

expected number of second divisioners $= \frac{3}{10} \times 200 = 60$ and

expected number of first divisioners $= \frac{2}{10} \times 200 = 40$.

Notes

Thus, we have $\chi^2 = \frac{(46-40)^2}{40} + \frac{(68-60)^2}{60} + \frac{(62-60)^2}{60} + \frac{(24-40)^2}{40} = 8.44.$

Since this value is greater than the tabulated value, 7.81, for 3 d.f. and 5% level of significance, H_0 is rejected.

Example 42: A survey of 320 families with 5 children each revealed the following distribution:

No. of boys	:	5	4	3	2	1	0
No. of girls	:	0	1	2	3	4	5
No. of families	:	14	56	110	88	40	12

Is the result consistent with the hypothesis that male and female births are equally probable?

Solution.

Assuming that H_0 (i.e., male and female births are equally probable) is true, the expected number of families having r boys (or equivalently 5 - r girls) is given by $E_r = 320 \times {}^5C_r \left(\frac{1}{2}\right)^5 = 10 \times {}^5C_r$. On substituting r = 5, 4, 3, 2, 1, 0, the respective expected frequencies are 10, 50, 100, 100, 50 and 10.

$$\therefore\ \chi^2 = \frac{(14-10)^2}{10} + \frac{(56-50)^2}{50} + \frac{(110-100)^2}{100} + \frac{(88-100)^2}{100} + \frac{(40-50)^2}{50} + \frac{(12-10)^2}{10} = 7.16.$$

The value from table for 5 d.f. at 5% level of significance is 11.07, which is greater than the calculated value. Thus, there is no evidence against H_0.

Example 43:

The record for a period of 180 days, showing the number of electricity failures per day in Delhi are shown in the following table :

No.of failures	:	0	1	2	3	4	5	6	7
No.of days	:	12	39	47	40	20	17	3	2

Determine, by using c^2 - test, whether the number of failures can be regarded as a Poisson variate?

Solution.

We have to test H_0 : No. of failures is a Poisson variate against H_a : No. of failures is not a Poisson variate.

The mean of the Poisson distribution is

$$m = \frac{0\times 12 + 1\times 39 + 2\times 47 + 3\times 40 + 4\times 20 + 5\times 17 + 6\times 3 + 7\times 2}{180} = 2.5$$

Notes

The computations of χ^2 are done in the following table :

No.of families	*Expected freq.*(E_i)	*Observed freq.*(O_i)	$\frac{(O_i - E_i)^2}{E_i}$
0	$180 \times e^{-2.5} = 14.76$	12	0.52
1	$E_0 \times 2.5 = 36.94$	39	0.11
2	$E_1 \times 2.5/2 = 46.17$	47	0.01
3	$E_2 \times 2.5/3 = 38.48$	40	0.06
4	$E_3 \times 2.5/4 = 24.05$	20	0.68
5	$E_4 \times 2.5/5 = 12.02$	17	2.06
6 or more	*by difference* = 7.58	5	0.88
Total	180		$\chi^2 = 4.32$

The value of χ^2 from table at 5% level of significance and 5 d.f. is 11.07. Since the calculated value is less than the tabulated value, there is no evidence against H_0.

26.2.2 χ^2 - test as a Test for Independence of Two Attributes

Let us assume that a population is classified into m mutually exclusive classes, A_1, A_2, ... A_m, according to an attribute A and each of these m classes are further classified into n mutually exclusive classes, like A_iB_1, A_iB_2, ... A_iB_n , etc., according to another attribute B.

If O_{ij} is the observed frequency of A_iB_j , i.e., $(A_iB_j) = O_{ij}$, the above classification can be expressed in form of following table, popularly known as contingency table.

$B \rightarrow$ $A \downarrow$	B_1	B_2	$\cdots\cdots$	B_n	*Total*
A_1	O_{11}	O_{12}	$\cdots\cdots$	O_{1n}	(A_1)
A_2	O_{21}	O_{22}	$\cdots\cdots$	O_{2n}	(A_2)
$\vdots$	$\vdots$	$\vdots$	$\cdots\cdots$	$\vdots$	
A_m	O_{m1}	O_{m2}	$\cdots\cdots$	O_{mn}	(A_m)
Total	(B_1)	(B_2)	$\cdots\cdots$	(B_n)	N

Assuming that A and B are independent, we can compute the expected frequencies of each cell,

i.e., $E_{ij} = \frac{(A_i)(B_j)}{N}$. Thus, $\chi^2 = \sum_{i=1}^{m}\sum_{j=1}^{n}\frac{(O_{ij} - E_{ij})^2}{E_{ij}}$ would be a c^2 - variate with (m - 1)(n - 1) d.f.

Remarks : The expected frequencies of some cells may be obtained by the application of the above formula while the remaining cell frequencies can be obtained by subtraction. The minimum number of cell frequencies, that must be computed by the use of the formula, is equal to the degrees of freedom of the c^2 statistic.

Example 44: The employees in 4 different firms are distributed in three skill categories shown in the following table. Test the hypothesis that there is no relationship between the firm and the type of labour. Let the level of significance be 5%.

Notes

Firm → / Type of labour ↓	A	B	C	D
Skilled	24	24	23	49
Semi-skilled	32	60	37	51
Manual	24	56	40	80

Solution.

H_0 : There is no relation between the firm and the nature of labour.

Calculation of Expected Frequencies

Firm → / labour ↓	A	B	C	D	Total
Skilled	$\frac{80\times120}{500}$ $=19.2$	$\frac{140\times120}{500}$ $=33.6$	$\frac{100\times120}{500}$ $=24.0$	$\frac{180\times120}{500}$ $=43.2$	120
Semi-skilled	$\frac{80\times180}{500}$ $=28.8$	$\frac{140\times180}{500}$ $=50.4$	$\frac{100\times180}{500}$ $=36.0$	$\frac{180\times180}{500}$ $=64.8$	180
Manual	$\frac{80\times200}{500}$ $=32.0$	$\frac{140\times200}{500}$ $=56.0$	$\frac{100\times200}{500}$ $=40.0$	$\frac{180\times200}{500}$ $=72.0$	200
Total	80	140	100	180	500

We note that the totals of corresponding rows or columns are same for the observed as well as the expected frequencies.

From the observed and the expected frequencies, we get c^2 = 12.81. Further, the value of c^2 from the table for (4 - 1)(3 - 1) = 6 d.f. and 5% level of significance is 12.59. Since the calculated value is greater than the tabulated value H_0 is rejected.

Example 44: Samples of household income were taken from four cities. Test whether the cities are homogeneous with regard to the distribution of income?

Cities → / Income(Rs) ↓	A	B	C	D	Total
Under 3000	10	15	15	10	50
3000-5000	5	10	15	10	40
Over 5000	15	15	10	20	60
Total	30	40	40	40	150

Solution.

H_0 : Various cities are homogeneous with regard to the distribution of income.

Computation of Expected Frequencies

Notes

Cities → *Income(Rs)* ↓	*A*	*B*	*C*	*D*	*Total*
Under 3000	10.00	13.33	13.33	13.33	50
3000-5000	8.00	10.67	10.67	10.67	40
Over 5000	12.00	16.00	16.00	16.00	60
Total	30	40	40	40	150

Note that the expected frequencies for city A, under various income groups, are computed as $\frac{30 \times 50}{150} = 10.00, \frac{30 \times 40}{150} = 8.00$ and $\frac{30 \times 60}{150} = 12.00.$ Other frequencies have also been computed in a similar manner.

Using the observed and expected frequencies, the value of c^2 = 8.28.

Further, the value of X^2 from tables for 6 d.f. at 5% level of significance is

Since the calculated value is less than the tabulated value, there is no evidence against H_0.

The value of χ^2 for a 2 × 2 Contingency table

For a 2 × 2 contingency table,

a	b	$a+b$
c	d	$c+d$
$a+c$	$b+d$	$a+b+c+d=N$

, the value of χ^2 can be directly computed with the use of the following formula :

$$\chi^2 = \frac{N(ad-bc)^2}{(a+b)(a+c)(b+d)(c+d)}$$

Yate's correction for continuity

We know that c^2 is a continuous random variate but the frequencies of various cells of a contingency table are integers. When N is large, the distribution of $\sum \frac{(O-E)^2}{E}$ is approximately c^2. However, the corrections for continuity are required when N is small. Yates has suggested the following corrections for continuity in a 2 × 2 contingency table :

If ad > bc, reduce a and d by $\frac{1}{2}$ and increase b and c by $\frac{1}{2}$. Similarly, If ad < bc, increase a and d by $\frac{1}{2}$ and decrease b and c by $\frac{1}{2}$.Thus, the contingency tables in the two situations become

$$\begin{array}{c|c} a-\frac{1}{2} & b+\frac{1}{2} \\ \hline c+\frac{1}{2} & d-\frac{1}{2} \end{array} \quad \text{and} \quad \begin{array}{c|c} a+\frac{1}{2} & b-\frac{1}{2} \\ \hline c-\frac{1}{2} & d+\frac{1}{2} \end{array}$$

respectively.

Notes

The value of c^2 can now be obtained as $\chi^2 = \dfrac{N\left(|ad-bc| - \dfrac{N}{2}\right)^2}{(a+b)(a+c)(b+d)(c+d)}$.

Brand and Snedecor formula for a 2 ´ r Contingency table

For a 2 ´×r contingency table,

$A \rightarrow$ $B \downarrow$	A_1	A_2	$\cdots$	A_r	*Total*
B_1	a_1	a_2	$\cdots$	a_r	a
B_2	b_1	b_2	$\cdots$	b_r	b
Total	n_1	n_2	$\cdots$	n_r	N

, the value of c^2 can be directly computed by the use of the following formula :

$$\chi^2 = \frac{N^2}{ab}\left(\sum_{i=1}^{r}\frac{a_i^2}{n_i} - \frac{a^2}{N}\right) \text{ with (r - 1) d.f.}$$

Example 46: In a recent diet survey, the following results were obtained in an Indian city:

No.of families	*Hindus*	*Muslims*	*Total*
Tea takers	1236	164	1400
Non-tea takers	564	36	600
Total	1800	200	2000

Discuss whether there is any significant difference between the two communities in the matter of taking tea? Use 5% level of significance.

Solution.

The null hypothesis to be tested can be written as H_0 : There is no difference between the two communities in the matter of taking tea.

Using the direct formula, we have $\chi^2 = \dfrac{2000(1236 \times 36 - 164 \times 564)^2}{1400 \times 1800 \times 200 \times 600} = 15.24.$

The value of c^2 from table for 1 d.f. and 5% level of significance is 3.84. Since the calculated value is greater than the tabulated value, H_0 is rejected.

Example 47: A certain drug is claimed to be effective in curing cold. In an experiment on 160 persons with cold, half of them were given the drug and the remaining half were given sugar pills. The patients' reactions to the treatment are recorded in the following table :

	Helped	*Harmed*	*No effect*	*Total*
Drug	52	10	18	80
Sugar pills	44	10	26	80
Total	96	20	44	160

Test the hypothesis that the drug is no better than the sugar pills for curing cold.

Notes

Solution.

H_0: The drug is not effective in curing cold

Using the Brandt and Snedecor formula, we have

$$\chi^2 = \frac{160 \times 160}{80 \times 80}\left(\frac{52^2}{96} + \frac{10^2}{20} + \frac{18^2}{44} - \frac{80^2}{160}\right) = 2.12.$$

This value is less than the tabulated value (= 5.99) for 2 d.f. and 5% level of significance. Thus, there is no evidence against H_0.

26.3 Summary

- Here we have to test whether ρ is different from zero. Accordingly, H_0 and H_a are ρ = 0 and ρ ≠ 0 respectively.

  For small samples, the test statistic can be obtained from the sampling distribution of b. We note that if $r = 0$, then b would also be zero.

  Therefore, $\frac{b}{S.E.(b)} = r \cdot \frac{S_Y}{S_X} \cdot \frac{1}{S.E.(b)} = r \cdot \frac{S_Y}{S_X} \cdot \frac{S_X}{S_Y}\sqrt{\frac{n-2}{1-r^2}} = r\sqrt{\frac{n-2}{1-r^2}}$ will follow t - distribution with (n - 2) d.f. Hence, $r\sqrt{\frac{n-2}{1-r^2}}$ can be taken as the test statistic. We note that $S.E.(r) = \sqrt{\frac{1-r^2}{n-2}}$. Therefore, 100(1 - a)% confidence limits of r can be written as $r \pm t_{a/2}$ S.E.(r).

- Let there be two independent random samples of sizes n_1 and n_2 from two normal populations with correlations ρ_1 and ρ_2 respectively. Let r_1 and r_2 be the correlations computed from the respective samples.

  If Z_1, Z_2, ξ_1 and ξ_2 denote Fisher's transformation of r_1, r_2, r_1 and ρ_2 respectively, then

  $$Z_1 \sim N\left(\xi_1, \frac{1}{\sqrt{n_1 - 3}}\right) \text{ and } Z_2 \sim N\left(\xi_2, \frac{1}{\sqrt{n_2 - 3}}\right)$$

  $$\therefore\ Z_1 - Z_2 \sim N\left(\xi_1 - \xi_2, \sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}\right)$$

  $$\text{or } \frac{Z_1 - Z_2}{\sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}} \sim N(0,1) \text{ under } H_0 : \rho_1 = \rho_2$$

26.4 Keywords

Fisher's test: It is applicable whether n is small or large. If r is correlation in sample, then its Fisher's Z transformation is given by $Z = \frac{1}{2}\log_e \frac{1+r}{1-r}$.

χ^2 - test: It can be used to test, how far the fitted or the expected frequencies are in agreement with the observed frequencies.

Notes

26.5 Self Assessment

1. Let ρ be coefficient of .... in a bivariate normal population and r be its estimator based on a sample of n observations (X_i, Y_i).
2. Let there be two .... samples of sizes n_1 and n_2 from two normal populations with correlations ρ_1 and ρ_2 respectively. Let r_1 and r_2 be the correlations computed from the respective samples.
3. In addition to the use of χ^2 in tests of hypothesis concerning the standard deviation, it is used as a test of .... and as a test of independence of two attributes. These tests are explained in the following sections.
4. .... can be used to test, how far the fitted or the expected frequencies are in agreement with the observed frequencies.

26.6 Review Questions

1. We want to decide whether a cubic die is balanced or not. For this purpose the die is thrown 300 times and various outcomes are recorded. If the observed frequencies of the six faces, namely 1, 2, 3, 4, 5 and 6 are 35, 40, 32, 60, 68 and 65 respectively, can we conclude that the die is unbiased?

   Hint : The expected frequency of each face under H_0 is 50.

2. Four coins are tossed 320 times and the number of heads obtained were recorded as follows :

No. of heads	:	0	1	2	3	4
Frequency	:	15	102	108	68	27

   Can we regard all the coins as unbiased?

   Hint : Find expected frequencies of the number of heads on the assumption that the coins are unbiased.

3. Three dice were thrown 80 times and the number of times 2, 4 or 6 was obtained, were recorded as given below:

No. of dice showing 2, 4 or 6	:	0	1	2	3
Frequency	:	8	28	32	12

   Test the hypothesis that all the three dice are fair.

   Hint : Under H_0, the probability of success, i.e., getting 2 or 4 or 6 on a die is 0.5. If r denotes the number of dice giving successes in a throw, we have $p(r) = {}^4C_r 0.5^4$.

4. The health department of municipal corporation of a city believes that 14% persons of the city are smokers as well as drinkers, 30% are drinkers while 40% are smokers. In a random sample of 150 persons, it was found that 24 persons were smokers as well as drinkers, 21 were only drinkers and 36 were only smokers. Do the above data support the belief of the department. Use 5% level of significance.

   Hint : Use the figures of belief, given in percentages, to find the expected frequencies.

Notes

5. A normal distribution was fitted to the distribution of new business brought by 100 insurance agents with the following results:

New business ('000Rs)	:	10-20	20-30	30-40	40-50	50-60
Observed Frequency	:	10	20	33	22	15
Expected Frequency	:	9	22	32	25	12

Test the goodness of fit of the distribution.

Hint : The degrees of freedom of the χ^2 statistic would be 4.

Answers: Self Assessment

1. linear correlation
2. independent random
3. goodness of fit
4. χ^2 - test

26.7 Further Readings

Books

Sheldon M. Ross, Introduction to Probability Models, Ninth Edition, Elsevier Inc., 2007.

Jan Pukite, Paul Pukite, Modeling for Reliability Analysis, IEEE Press on Engineering of Complex Computing Systems, 1998.

Notes

Unit 27: T - Distributions

CONTENTS

Objectives

Introduction

27.1 The Student's T-Distribution

27.2 T test

27.3 Summary

27.4 Keywords

27.4 Self Assessment

27.5 Review Questions

27.6 Further Readings

Objectives

After studying this unit, you will be able to:

- Discuss T - Distribution
- Explain example of T - Distribution

Introduction

In last unit you have studied about chi-square. This unit will provide you information related to T - Distribution.

27.1 The Student's T-Distribution

Let X_1, X_2 ...... X_n be n independent random variables from a normal population with mean *m* and standard deviation *s* (unknown).

When *s* is not known, it is estimated by s, the sample standard deviation $\left(s = \sqrt{\frac{1}{n-1}\Sigma\left(X_i - \bar{X}\right)^2} \right)$.

In such a case we would like to know the exact distribution of the statistic $\frac{\bar{X}-\mu}{s/\sqrt{n}}$ and the answer to this is provided by t - distribution.

W.S. Gosset defined t statistic as $t = \frac{\bar{X}-\mu}{s/\sqrt{n}}$ which follows t - distribution with (n - 1) degrees of freedom.

Features of t- distribution

Notes

1. Like χ^2 - distribution, t - distribution also has one parameter n = n - 1, where n denotes sample size. Hence, this distribution is known if n is known.
2. Mean of the random variable t is zero and standard deviation is $\sqrt{\dfrac{\nu}{\nu-2}}$, for n > 2.
3. The probability curve of t - distribution is symmetrical about the ordinate at t = 0. Like a normal variable, the t variable can take any value from - ∞ to ∞.
4. The distribution approaches normal distribution as the number of degrees of freedom become large.
5. The random variate t is defined as the ratio of a standard normal variate to the square root of χ^2 - variate divided by its degrees of freedom.

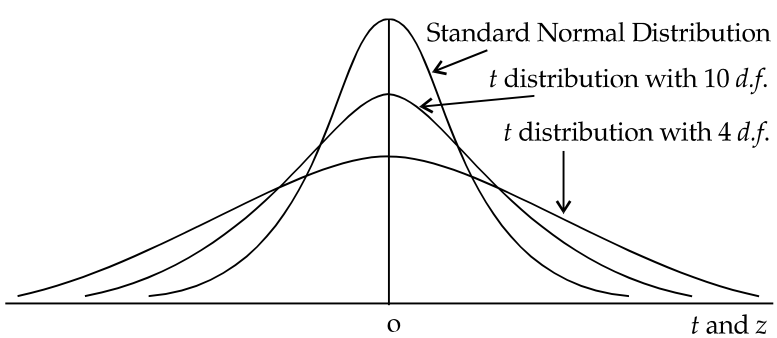
To show this we can write $t = \dfrac{\bar{X}-\mu}{s/\sqrt{n}} = \dfrac{(\bar{X}-\mu)\sqrt{n}}{s}$

Dividing numerator and denominator by s, we get

$$t = \frac{\dfrac{(\bar{X}-\mu)\sqrt{n}}{\sigma}}{\dfrac{s}{\sigma}} = \frac{\dfrac{(\bar{X}-\mu)}{\sigma/\sqrt{n}}}{\sqrt{s^2/\sigma^2}} = \frac{\dfrac{(\bar{X}-\mu)}{\sigma/\sqrt{n}}}{\sqrt{\dfrac{1}{n-1}\cdot\dfrac{\sum(X_i-\bar{X})^2}{\sigma^2}}}$$

$$= \frac{\dfrac{(\bar{X}-\mu)}{\sigma/\sqrt{n}}}{\sqrt{\dfrac{\chi^2_{n-1}}{n-1}}} = \frac{\textit{Standard Normal Variate}}{\sqrt{\chi^2\textit{-variate}}}$$

Figure 27.1



Notes

27.2 T test

Acme Corporation manufactures light bulbs. The CEO claims that an average Acme light bulb lasts 300 days. A researcher randomly selects 15 bulbs for testing. The sampled bulbs last an average of 290 days, with a standard deviation of 50 days. If the CEO's claim were true, what is the probability that 15 randomly selected bulbs would have an average life of no more than 290 days?

> *Note* There are two ways to solve this problem, using the T Distribution Calculator. Both approaches are presented below. Solution A is the traditional approach. It requires you to compute the t score, based on data presented in the problem description. Then, you use the T Distribution Calculator to find the probability. Solution B is easier. You simply enter the problem data into the T Distribution Calculator. The calculator computes a t score "behind the scenes", and displays the probability. Both approaches come up with exactly the same answer.

Solution A

The first thing we need to do is compute the t score, based on the following equation:

$$t = [\ x - \mu\ ] \ /\ [\ s \ /\ \text{sqrt}(\ n\ )\ ]$$

$$t = (\ 290 - 300\ ) \ /\ [\ 50 \ /\ \text{sqrt}(\ 15)\ ] = -10 \ /\ 12.909945 = -\ 0.7745966$$

where x is the sample mean, ì is the population mean, s is the standard deviation of the sample, and n is the sample size.

Now, we are ready to use the T Distribution Calculator. Since we know the t score, we select "T score" from the Random Variable dropdown box. Then, we enter the following data:

- The degrees of freedom are equal to 15 – 1 = 14.
- The t score is equal to – 0.7745966.

The calculator displays the cumulative probability: 0.226. Hence, if the true bulb life were 300 days, there is a 22.6% chance that the average bulb life for 15 randomly selected bulbs would be less than or equal to 290 days.

Solution B:

This time, we will work directly with the raw data from the problem. We will not compute the t score; the T Distribution Calculator will do that work for us. Since we will work with the raw data, we select "Sample mean" from the Random Variable dropdown box. Then, we enter the following data:

- The degrees of freedom are equal to 15 - 1 = 14.
- Assuming the CEO's claim is true, the population mean equals 300.
- The sample mean equals 290.
- The standard deviation of the sample is 50.

The calculator displays the cumulative probability: 0.226. Hence, there is a 22.6% chance that the average sampled light bulb will burn out within 290 days.

Problem 2

Notes

Suppose scores on an IQ test are normally distributed, with a mean of 100. Suppose 20 people are randomly selected and tested. The standard deviation in the sample group is 15. What is the probability that the average test score in the sample group will be at most 110?

Solution:

To solve this problem, we will work directly with the raw data from the problem. We will not compute the t score; the T Distribution Calculator will do that work for us. Since we will work with the raw data, we select "Sample mean" from the Random Variable dropdown box. Then, we enter the following data:

- The degrees of freedom are equal to 20 - 1 = 19.
- The population mean equals 100.
- The sample mean equals 110.
- The standard deviation of the sample is 15.

We enter these values into the T Distribution Calculator. The calculator displays the cumulative probability: 0.996. Hence, there is a 99.6% chance that the sample average will be no greater than 110.

27.3 Summary

- Let X_1, X_2 ...... X_n be n independent random variables from a normal population with mean *m* and standard deviation *s* (unknown).

  When *s* is not known, it is estimated by s, the sample standard deviation $\left(s = \sqrt{\frac{1}{n-1}\sum\left(X_i - \overline{X}\right)^2}\right)$. In such a case we would like to know the exact distribution of the statistic $\frac{\overline{X}-\mu}{s/\sqrt{n}}$ and the answer to this is provided by t - distribution.

  W.S. Gosset defined t statistic as $t = \frac{\overline{X}-\mu}{s/\sqrt{n}}$ which follows t - distribution with (n - 1) degrees of freedom.

- Like χ^2 - distribution, t - distribution also has one parameter *n* = n - 1, where n denotes sample size. Hence, this distribution is known if *n* is known.
- Mean of the random variable t is zero and standard deviation is $\sqrt{\frac{\nu}{\nu-2}}$, for *n* > 2.
- The probability curve of t - distribution is symmetrical about the ordinate at t = 0. Like a normal variable, the t variable can take any value from - ∞ to ∞.
- The distribution approaches normal distribution as the number of degrees of freedom become large.
- The random variate t is defined as the ratio of a standard normal variate to the square root of χ^2 - variate divided by its degrees of freedom.

Notes

27.4 Keywords

Mean: Mean of the random variable t is zero and standard deviation is $\sqrt{\frac{\nu}{\nu-2}}$, for $n > 2$.

T - distribution: The probability curve of t - distribution is symmetrical about the ordinate at t = 0. Like a normal variable, the t variable can take any value from - ∞ to ∞.

27.4 Self Assessment

1. State whether the following statements are true or false:

   (i) Both, t and χ^2 distributions depend only one parameter.

   (ii) Total number of samples of size 4, with replacement, from a population of 15 units is 1365.

   (iii) F - statistic is equal the ratio of two χ^2 variates.

   (iv) In sampling with replacement if N = n, the standard error of $\overline{X}$ is equal to zero.

   (v) χ^2 - distribution depends upon two parameters.

27.5 Review Questions

1. A population consists of 4 families consisting of 2, 3, 4 and 5 children. By considering all possible random samples of size two, with replacement, find mean and standard error of $\overline{X}$. Show that S.E. of $\overline{X}$ depends upon the sample size.

2. If X_1, X_2, X_3 is a simple random sample of size three from a large population with mean 5 and variance 4, evaluate the expected value and standard error of the statistics $T = (2X_1 + X_2 - 3X_3)$.

3. If X_1, X_2 and X_3 constitute a random sample of size 3 from a normal population with mean μ and the variance σ^2, find the efficiency of $\frac{X_1 + 2X_2 + X_3}{4}$ relative to $\frac{X_1 + X_2 + X_3}{3}$.

4. The diameter of a component produced on a semi-automatic machine is known to be distributed normally with mean of 10 mm. and a standard deviation of 0.1 mm. If we pick up a random sample of size 25, what is the probability that the sample mean will be between 9.95 and 10.05 mm?

5. It is known that 10% of the bolts manufactured by a factory are defective. If a random sample of 100 bolts is chosen at random from a day's production, construct the sampling distribution of (i) the number of defective bolts, (ii) the proportion of defective bolts.

6. The mean and standard deviation of per capita consumption of wheat in rural and urban areas of Delhi are estimated to be 450 gms, 75 gms and 410 gms, 100 gms respectively. Assuming that the per capita consumption of wheat is distributed normally, construct the sampling distribution of the difference between two sample means obtained from random samples of sizes 80 and 60 from rural and urban populations respectively.

Notes

Answers: Self Assessment

1. (i) T (ii) F (iii) T (iv) F (v) T

27.6 Further Readings

Books

Sheldon M. Ross, Introduction to Probability Models, Ninth Edition, Elsevier Inc., 2007.

Jan Pukite, Paul Pukite, Modeling for Reliability Analysis, IEEE Press on Engineering of Complex Computing Systems, 1998.

Notes

Unit 28: F-distribution

CONTENTS

Objectives

Introduction

28.1 Snedecor's F- Distribution

28.2 Summary

28.3 Keywords

28.4 Self Assessment

28.5 Review Questions

28.6 Further Readings

Objectives

After studying this unit, you will be able to:

- Define F - distribution
- Discuss F - distribution examples

Introduction

In the last unit you studied about samples distribution and T - Distribution. This unit provides you information related to F - distribution.

28.1 Snedecor's F- Distribution

Let there be two independent random samples of sizes n_1 and n_2 from two normal populations with variances $s_1{}^2$ and $s_2{}^2$ respectively. Further, let $s_1^2 = \frac{1}{n_1 - 1}\sum\left(X_{1i} - \overline{X}_1\right)^2$ and $s_2^2 = \frac{1}{n_2 - 1}\sum\left(X_{2i} - \overline{X}_2\right)^2$ be the variances of the first sample and the second samples respectively.

Then F - statistic is defined as the ratio of two χ^2 - variates. Thus, we can write

$$F = \frac{\dfrac{\chi^2_{n_1-1}}{n_1-1}}{\dfrac{\chi^2_{n_2-1}}{n_2-1}} = \frac{\dfrac{(n_1-1)s_1^2}{\sigma_1^2}\Big/(n_1-1)}{\dfrac{(n_2-1)s_2^2}{\sigma_2^2}\Big/(n_2-1)} = \frac{\dfrac{s_1^2}{\sigma_1^2}}{\dfrac{s_2^2}{\sigma_2^2}}$$

Features of F- distribution

Notes

1. This distribution has two parameters n_1 (= n_1 - 1) and n_2 (= n_2 - 1).

2. The mean of F - variate with n_1 and n_2 degrees of freedom is $\frac{\nu_2}{\nu_2 - 2}$ and standard error is

$$\left(\frac{\nu_2}{\nu_2 - 2}\right)\sqrt{\frac{2(\nu_1 + \nu_2 - 2)}{\nu_1(\nu_2 - 4)}}.$$

We note that the mean will exist if $\nu_2 > 2$ and standard error will exist if $\nu_2 > 4$. Further, the mean > 1.

3. The random variate F can take only positive values from 0 to ∞. The curve is positively skewed, as shown in Fig. 20.3

4. For large values of ν_1 and ν_2, the distribution approaches normal distribution. This behaviour is shown in the following figure.

5. If a random variate follows t-distribution with ν degrees of freedom, then its square follows F-distribution with 1 and n d.f. i.e. $t^2_\nu = F_{1,\nu}$

6. F and c^2 are also related as $F_{\nu_1, \nu_2} = \frac{(\chi^2_{\nu_1})}{\nu_1}$ as $\nu_2 \to \infty$

Example 1: Suppose you randomly select 7 women from a population of women, and 12 men from a population of men. The table below shows the standard deviation in each sample and in each population.

Population	Population standard deviation	Sample standard deviation
Women	30	35
Men	50	45

Compute the f statistic.

Solution A: The f statistic can be computed from the population and sample standard deviations, using the following equation:

$$f = [\, s_1^2/\sigma_1^2 \,] \,/\, [\, s_2^2/\sigma_2^2 \,]$$

where σ_1 is the standard deviation of population 1, s_1 is the standard deviation of the sample drawn from population 1, σ_2 is the standard deviation of population 2, and s_1 is the standard deviation of the sample drawn from population 2.

As you can see from the equation, there are actually two ways to compute an f statistic from these data. If the women's data appears in the numerator, we can calculate an f statistic as follows:

$$f = (35^2 / 30^2) / (45^2 / 50^2) = (1225 / 900) / (2025 / 2500) = 1.361 / 0.81 = 1.68$$

For this calculation, the numerator degrees of freedom ν_1 are 7 - 1 or 6; and the denominator degrees of freedom ν_2 are 12 - 1 or 11.

On the other hand, if the men's data appears in the numerator, we can calculate an f statistic as follows:

$$f = (45^2 / 50^2) / (35^2 / 30^2) = (2025 / 2500) / (1225 / 900) = 0.81 / 1.361 = 0.595$$

Notes

For this calculation, the numerator degrees of freedom v_1 are 12 – 1 or 11; and the denominator degrees of freedom v_2 are 7 – 1 or 6.

When you are trying to find the cumulative probability associated with an f statistic, you need to know v_1 and v_2. This point is illustrated in the next example.

Example 2: Find the cumulative probability associated with each of the f statistics from Example 1, above.

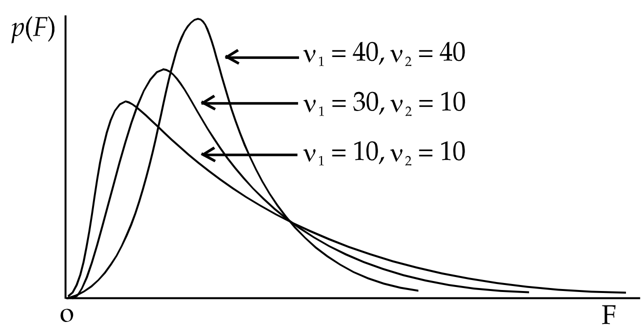
Solution: To solve this problem, we need to find the degrees of freedom for each sample. Then, we will use the F Distribution Calculator to find the probabilities.

- The degrees of freedom for the sample of women is equal to n – 1 = 7 – 1 = 6.
- The degrees of freedom for the sample of men is equal to n – 1 = 12 – 1 = 11.

Therefore, when the women's data appear in the numerator, the numerator degrees of freedom v_1 is equal to 6; and the denominator degrees of freedom v_2 is equal to 11. And, based on the computations shown in the previous example, the f statistic is equal to 1.68. We plug these values into the F Distribution Calculator and find that the cumulative probability is 0.78.

On the other hand, when the men's data appear in the numerator, the numerator degrees of freedom v_1 is equal to 11; and the denominator degrees of freedom v_2 is equal to 6. And, based on the computations shown in the previous example, the f statistic is equal to 0.595. We plug these values into the F Distribution Calculator and find that the cumulative probability is 0.22.

Figure 28.3



28.2 Summary

- Let there be two independent random samples of sizes n_1 and n_2 from two normal populations with variances s_1^2 and s_2^2 respectively. Further, let $s_1^2 = \frac{1}{n_1 - 1}\sum\left(X_{1i} - \bar{X}_1\right)^2$ and $s_2^2 = \frac{1}{n_2 - 1}\sum\left(X_{2i} - \bar{X}_2\right)^2$ be the variances of the first sample and the second samples respectively. Then F - statistic is defined as the ratio of two χ^2 - variates. Thus, we can write

$$F = \frac{\dfrac{\chi^2_{n_1-1}}{n_1-1}}{\dfrac{\chi^2_{n_2-1}}{n_2-1}} = \frac{\dfrac{(n_1-1)s_1^2}{\sigma_1^2}/(n_1-1)}{\dfrac{(n_2-1)s_2^2}{\sigma_2^2}/(n_2-1)} = \frac{\dfrac{s_1^2}{\sigma_1^2}}{\dfrac{s_2^2}{\sigma_2^2}}$$

Notes

- This distribution has two parameters n_1 (= n_1 - 1) and n_2 (= n_2 - 1).
- The mean of F - variate with n_1 and n_2 degrees of freedom is $\frac{\nu_2}{\nu_2 - 2}$ and standard error is

  $$\left(\frac{\nu_2}{\nu_2 - 2}\right)\sqrt{\frac{2(\nu_1 + \nu_2 - 2)}{\nu_1(\nu_2 - 4)}}.$$

  We note that the mean will exist if $\nu_2 > 2$ and standard error will exist if $\nu_2 > 4$. Further, the mean > 1.
- The random variate F can take only positive values from 0 to ∞. The curve is positively skewed, as shown in Fig. 20.3
- For large values of ν_1 and ν_2, the distribution approaches normal distribution. This behaviour is shown in the following figure.
- If a random variate follows t-distribution with ν degrees of freedom, then its square follows F-distribution with 1 and n d.f. i.e. $t^2_\nu = F_{1,\nu}$

28.3 Keywords

The random variate: The random variate F can take only positive values from 0 to ∞. The curve is positively skewed.

F-distribution: If a random variate follows t-distribution with ν degrees of freedom, then its square follows F-distribution with 1 and n d.f. i.e. $t^2_\nu = F_{1,\nu}$

28.4 Self Assessment

1. Fill in the Blanks:
   (1) The mean and standard error of a F-variate depend upon its ...... parameters.
   (ii) The sum of squares of standard normal variates is a ...... variate.
   (iii) If N = 8 and n = 3, the number of samples without replacement is equal to ....
   (iv) The ratio of two sample variances follows F - distribution when the variances of their parent population are ....
   (v) In sampling without replacement if N = n, the standard error of $\overline{X}$ is equal to ....
   (vi) Both χ^2 and F-distributions are ...... skewed distributions.

28.5 Review Questions

1. Define F distribution and discuss feature of F distribution.
2. Two random samples of sizes 100 and 150 are drawn from two different normal populations. Find mean and standard error of the statistic F.
3. Suppose you randomly select 8 women from a population of women, and 10 men from a population of men. The table below shows the standard deviation in each sample and in each population.

Notes

Population	Population standard deviation	Sample standard deviation
Women	40	45
Men	60	35

Compute the f statistic.

4. Find the cumulative probability associated with each of the f statistics. Suppose you randomly select 6 women from a population of women, and 10 men from a population of men. The table below shows the standard deviation in each sample and in each population.

Population	Population standard deviation	Sample standard deviation
Women	30	35
Men	50	45

5. Suppose you randomly select 8 women from a population of women, and 12 men from a population of men. The table below shows the standard deviation in each sample and in each population.

Population	Population standard deviation	Sample standard deviation
Women	30	35
Men	60	35

Compute the f statistic.

Answers: Self Assessment

1. (i) two (ii) χ^2 (iii) 56 (iv) equal (v) zero (vi) positively.

28.6 Further Readings

Books

Sheldon M. Ross, Introduction to Probability Models, Ninth Edition, Elsevier Inc., 2007.

Jan Pukite, Paul Pukite, Modeling for Reliability Analysis, IEEE Press on Engineering of Complex Computing Systems, 1998.

Unit 29: Estimation of Parameters: Criteria for Estimates

Notes

CONTENTS

Objectives

Introduction

29.1 Theory of Estimation

29.2 Point Estimation

29.2.1 Unbiasedness

29.2.2 Consistency

29.2.3 Efficiency

29.2.4 Sufficiency

29.2.5 Methods of Point Estimation

29.3 Interval Estimation

29.3.1 Determination of an Approximate Sample Size for a Given Degree of Accuracy

29.3.2 Confidence Interval for Population Standard Deviation

29.4 Summary

29.5 Keywords

29.6 Self Assessment

29.7 Review Questions

29.8 Further Readings

Objectives

After studying this unit, you will be able to:

- Discuss Theory of Estimation
- Explain Point Estimation (Properties of Good Estimators)
- Describe Interval Estimation

Introduction

Estimation: It is a procedure by which sample information is used to estimate the numerical magnitude of one or more parameters of the population. A function of sample values is called an estimator (or statistic) while its numerical value is called an estimate. For example is an estimator of population mean m. On the other hand if for a sample, the estimate of population mean is said to be 50.

Notes

29.1 Theory of Estimation

Let X be a random variable with probability density function (or probability mass function) f(X ; θ_1, θ_2, ... θ_k), where θ_1, θ_2, ... θ_k are k parameters of the population.

Given a random sample X_1, X_2, ... X_n from this population, we may be interested in estimating one or more of the k parameters θ_1, θ_2, ... θ_k. In order to be specific, let X be a normal variate so that its probability density function can be written as N(X : μ, σ). We may be interested in estimating m or s or both on the basis of random sample obtained from this population.

It should be noted here that there can be several estimators of a parameter, e.g., we can have any of the sample mean, median, mode, geometric mean, harmonic mean, etc., as an estimator of population mean μ. Similarly, we can use either $S = \sqrt{\frac{1}{n}\sum\left(X_i - \bar{X}\right)^2}$ or $s = \sqrt{\frac{1}{n-1}\sum\left(X_i - \bar{X}\right)^2}$ as an estimator of population standard deviation s. This method of estimation, where single statistic like Mean, Median, Standard deviation, etc. is used as an estimator of population parameter, is known as Point Estimation. Contrary to this it is possible to estimate an interval in which the value of parameter is expected to lie. Such a procedure is known as Interval Estimation. The estimated interval is often termed as Confidence Interval.

29.2 Point Estimation

As mentioned above, there can be more than one estimators of a population parameter. Therefore, it becomes necessary to determine a good estimator out of a number of available estimators. We may recall that an estimator, a function of random variables X_1, X_2, ... X_n, is a random variable. Therefore, we can say that a good estimator is one whose distribution is more concentrated around the population parameter. R. A. Fisher has given the following properties of a good estimators. These are:

(i) Unbiasedness (ii) Consistency (iii) Efficiency (iv) Sufficiency.

29.2.1 Unbiasedness

An estimator t (X_1, X_2, ... X_n) is said to be an unbiased estimator of a parameter q if E(t) = θ.

If E(t) ≠ q, then t is said to be a biased estimator of θ. The magnitude of bias = E(t) – θ. We have seen in § 20.2 that $E\left(\bar{X}\right) = \mu$, therefore, $\bar{X}$ is said to be an unbiased estimator of population mean m. Further, refer to § 20.4.1, we note that $E\left(S^2\right) = \frac{n-1}{n}\cdot\sigma^2$, where $S^2 = \frac{1}{n}\sum\left(X_i - \bar{X}\right)^2$. Therefore, S^2 is a biased estimator of s^2. The magnitude of bias $= \left(\frac{n-1}{n} - 1\right)\sigma^2 = -\frac{1}{n}\sigma^2$.

Contrary to this, if we define $s^2 = \frac{1}{n-1}\sum\left(X_i - \bar{X}\right)^2$, we have seen in § 20.4.1 that $E(s^2) = \sigma^2$. Thus, s^2 is an unbiased estimator of s^2. Also from § 20.3.1 we note that E(p) = π, therefore, p is an unbiased estimator of π.

29.2.2 Consistency

It is desirable to have an estimator, with a probability distribution, that comes closer and closer to the population parameter as the sample size is increased. An estimator possessing this property is called a consistent estimator. An estimator $t_n(X_1, X_2, ... X_n)$ is said to be consistent if its probability distribution converges to θ as $n \to \infty$.

Symbolically, we can write $P(t_n \to \theta) = 1$ as $n \to \infty$. Alternatively, t_n is said to be a consistent estimator of q if $E(t_n) \to q$ and $Var(t_n) \to 0$, as $n \to \infty$.

We may note that $\overline{X}$ is a consistent estimator of population mean m because $E(\overline{X}) = \mu$ and

$Var(\overline{X}) = \frac{\sigma^2}{n} \to 0$ as $n \to \infty$.

Note: An unbiased estimator is necessarily a consistent estimator.

29.2.3 Efficiency

Let t_1 and t_2 be two estimators of a population parameter *q* such that both are either unbiased or consistent. To select a good estimator, from t_1 and t_2, we consider another property that is based upon its variance.

If t_1 and t_2 are two estimators of a parameter *q* such that both of them are either unbiased or consistent, then t_1 is said to be more efficient than t_2 if $Var(t_1) < Var(t_2)$. The efficiency of an estimator is measured by its variance.

For a normal population, we know that both the sample mean and median are unbiased estimator of population mean. However, their respective variances are $\frac{\sigma^2}{n}$ and $\frac{\pi}{2} \cdot \frac{\sigma^2}{n}$, where σ^2 is population variance. Since $\frac{\sigma^2}{n} < \frac{\pi}{2} \cdot \frac{\sigma^2}{n}$, therefore, sample mean is said to be efficient estimator of population mean.

Remarks: The precision of an estimator = 1/ S. E. of estimator.

An estimator having minimum variance among all the estimators of a population parameter is termed as Most Efficient Estimator or Best Estimator. If an estimator is unbiased and best, then it is termed as Best Unbiased Estimator. Further, if the best unbiased estimator is a linear function of the sample observations, it is termed as Best Linear Unbiased Estimator (BLUE). It may be pointed out here that sample mean is best linear unbiased estimator of population mean.

Cramer Rao Inequality:

This inequality gives the minimum possible value of the variance of an unbiased estimator. If t is an unbiased estimator of parameter *q* of a continuous population with probability density function f(X, *q*), then

$$Var(t) \geq \frac{1}{nE\left(\frac{\partial \log f(X, \theta)}{\partial \theta}\right)^2}$$

29.2.4 Sufficiency

An estimator t is said to be a sufficient estimator of parameter θ if it utilises all the information given in the sample about θ. For example, the sample mean $\overline{X}$ is a sufficient estimator of μ because no other estimator of μ can add any further information about μ.

Notes

Let X_1, X_2, ... X_n be a random sample of n independent observations from a population with p.d.f. (or p.m.f.) given by f(X; θ_1, θ_2), where q_1 and q_2 are two parameters. The joint probability distribution of X_1, X_2, ... X_n, denoted by L(X; θ_1, θ_2) is given by :

$$L(X;\theta_1,\theta_2) = f(X_1;\theta_1,\theta_2) \times f(X_2;\theta_1,\theta_2) \times \ldots\ldots \times f(X_n;\theta_1,\theta_2)$$

An estimator t is said to be sufficient for q_1 if the conditional p.d.f. (or p.m.f.) of X_1, X_2, ... X_n given t is independent of q_1, i.e.,

$$\frac{f(X_1;\theta_1,\theta_2)\times f(X_2;\theta_1,\theta_2)\times \;....\; \times f(X_n;\theta_1,\theta_2)}{g(t,\theta_1)} = h(X_1, X_2,\; ....\; X_n)$$, where g(t, q_1) is p.d.f.

(or p.m.f.) of t and h is a function of sample values that is independent of θ_1. We may note that each of the functions g(t, θ_1) and h(X_1, X_2, ... X_n) may or may not be function of θ_2.

Alternatively, we can write the sufficiency condition as

$f(X_1;\theta_1,\theta_2) \times f(X_2;\theta_1,\theta_2) \times \ldots\ldots \times f(X_n;\theta_1,\theta_2) = g(t, q_1) \times h(X_1, X_2, \ldots\ldots X_n)$, which implies that if the joint p.d.f. (or p.m.f.) of X_1, X_2, ... X_n can be written as a function of t and θ_1 multiplied by a function independent of θ_1, then t is sufficient estimator of θ_1.

Sufficient estimators are the most desirable but are not very commonly available. The following points must be noted about sufficient estimators:

1. A sufficient estimator is always consistent.
2. A sufficient estimator is most efficient if an efficient estimator exists.
3. A sufficient estimator may or may not be unbiased.

Example 1: If X_1, X_2, ... X_n is a sample of n independent observations from a normal population with mean *m* and variance s^2, show that $\bar{X}$ is a sufficient estimator of *m* but $S^2 = \frac{1}{n}\sum(X_i - \bar{X})^2$ is not sufficient estimator of s^2.

Solution.

The probability density function of a normal variate is given by

$$f(X;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(X-\mu)^2}$$

Thus, the joint probability density function of X_1, X_2, ... X_n is given by

$$f(X_1;\mu,\sigma)\times f(X_2;\mu,\sigma)\times \;....\; \times f(X_n;\mu,\sigma) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2}$$

We can write $X_i - \mu = (X_i - \bar{X}) + (\bar{X} - \mu)$.

Squaring both sides and taking sum over n observations, we get

$$\sum(X_i-\mu)^2 = \sum(X_i-\bar{X})^2 + \sum(\bar{X}-\mu)^2 + 2\sum(X_i-\bar{X})(\bar{X}-\mu)$$

$$= \sum(X_i-\bar{X})^2 + n(\bar{X}-\mu)^2 + 2(\bar{X}-\mu)\sum(X_i-\bar{X})$$

Notes

$$= \sum \left(X_i - \bar{X}\right)^2 + n\left(\bar{X} - \mu\right)^2 \qquad \text{(last term is zero)}$$

$$= nS^2 + n\left(\bar{X} - \mu\right)^2$$

Therefore, we can write $-\frac{1}{2\sigma^2}\sum\left(X_i - \mu\right)^2 = -\frac{n}{2\sigma^2}S^2 - \frac{n}{2\sigma^2}\left(\bar{X} - \mu\right)^2$.

Hence $f\left(X_1;\mu,\sigma\right) \times f\left(X_2;\mu,\sigma\right) \times \; .... \; \times f\left(X_n;\mu,\sigma\right)$

$$= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\frac{n}{2\sigma^2}S^2 - \frac{n}{2\sigma^2}\left(\bar{X}-\mu\right)^2} = e^{-\frac{n}{2\sigma^2}\left(\bar{X}-\mu\right)^2} \times \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\frac{n}{2\sigma^2}S^2}$$

$$= g\left(\bar{X},\mu,\sigma\right) \times h\left(S^2,\sigma\right)$$

Since h is independent of m, therefore $\bar{X}$ is a sufficient estimator of μ. However, S^2 is not sufficient estimator of σ^2 because g is not independent of σ.

Further, if we define $S^2 = \frac{1}{n}\sum\left(X_i - \mu\right)^2$, then

$$f\left(X_1;\mu,\sigma\right) \times f\left(X_2;\mu,\sigma\right) \times \; .... \; \times f\left(X_n;\mu,\sigma\right) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\frac{n}{2\sigma^2}S^2}$$

Thus, the newly defined S^2 becomes a sufficient estimator of σ^2. We note that $h(X_1, X_2, ... X_n) = 1$ in this case.

The above result suggests that if *m* is known, then we should use $S^2 = \frac{1}{n}\sum\left(X_i - \mu\right)^2$ rather than $S^2 = \frac{1}{n}\sum\left(X_i - \bar{X}\right)^2$ because former is better estimator of s^2.

29.2.5 Methods of Point Estimation

Given various criteria of a good estimator, the next logical step is to obtain an estimator possessing some or all of the above properties.

There are several methods of obtaining a point estimator of the population parameter. For example, we can use the method of maximum likelihood, method of least squares, method of minimum variance, method of minimum χ^2, method of moments, etc. We shall, however, use the most popular method of maximum likelihood.

Method of Maximum Likelihood

Let $X_1, X_2, ... X_n$ be a random sample of n independent observations from a population with probability density function (or p.m.f.) f(X; θ), where θ is unknown parameter for which we desire to find an estimator.

Since $X_1, X_2, ... X_n$ are independent random variables, their joint probability function or the probability of obtaining the given sample, termed as likelihood function, is given by

$$L = f(X_1 ; \theta) . f(X_2 ; \theta) . ... f(X_n ; \theta) = \prod_{i=1}^{n} f\left(X_i ; \theta\right) .$$

We have to find that value of *q* for which L is maximum. The conditions for maxima of L are : $\frac{dL}{d\theta} = 0$ and $\frac{d^2L}{d\theta^2} < 0$. The value of *q* satisfying these conditions is known as Maximum Likelihood Estimator (MLE).

Generalising the above, if L is a function of k parameters θ_1, θ_2, ... θ_k, the first order conditions for maxima of L are: $\frac{\partial L}{\partial \theta_1} = \frac{\partial L}{\partial \theta_1} =$ ...... $\frac{\partial L}{\partial \theta_k} = 0$.

This gives k simultaneous equations in k unknowns θ_1, θ_2, ... θ_k, and can be solved to get k maximum likelihood estimators.

Sometimes it is convenient to work using logarithm of L. Since log L is a monotonic transformation of L, the maxima of L and maxima of log L occur at the same value.

Properties of Maximum Likelihood Estimators

1. The maximum likelihood estimators are consistent.
2. The maximum likelihood estimators are not necessarily unbiased. If a maximum likelihood estimator is biased, then by slight modifications it can be converted into an unbiased estimator.
3. If a maximum likelihood estimator is unbiased, then it will also be most efficient.
4. A maximum likelihood estimator is sufficient provided sufficient estimator exists.
5. The maximum likelihood estimators are invariant under functional transformation, i.e., if t is a maximum likelihood estimator of θ, then f(t) would be maximum likelihood estimator of f(θ).

Example 2: Obtain a maximum likelihood estimator of *p* (the proportion of successes) in a population with p.m.f. given by $f\left(X;\pi\right) = {}^nC_X \pi^X \left(1-\pi\right)^{n-X}$, where X denotes the number of successes in a sample of n trials.

Solution.

Since ${}^nC_X \pi^X \left(1-\pi\right)^{n-X}$ is the probability of X successes out of n trials, therefore, this is also the likelihood function. Thus, we can write $L = {}^nC_X \pi^X \left(1-\pi\right)^{n-X}$.

Taking logarithm of both sides, we get

$$\log L = \log{}^n C_X + X \log \pi + \left(n - X\right) \log\left(1-\pi\right)$$

Differentiating w.r.t. *p*, we get

$$\frac{d \log L}{d\pi} = 0 + \frac{X}{\pi} - \frac{n-X}{1-\pi} = 0 \quad \text{for maxima of L.}$$

or $X(1- p) - (n - X)p = 0$

Notes

This gives $\hat{\pi} = \frac{X}{n}$, where $\hat{\pi}$ denotes an estimator of p.

It can also be shown that $\frac{d^2 \log L}{d\pi^2} < 0$ when $\hat{\pi} = \frac{X}{n}$.

Example 3: Obtain the maximum likelihood estimator of the parameter m of the Poisson distribution.

Solution.

Let $X_1, X_2, \ldots\ldots X_n$ be a random sample of n independent observations from the given population. Therefore, we can write

$$L = \frac{e^{-m}.m^{X_1}}{X_1!} \times \frac{e^{-m}.m^{X_2}}{X_2!} \times \ \ldots\ldots \ \times \frac{e^{-m}.m^{X_n}}{X_n!} = \frac{e^{-nm}.m^{\sum X_i}}{\prod(X_i!)}$$

Taking logarithm of both sides, we get

$$\log L = -nm + \sum X_i \log m - \sum \log(X_i!)$$

Differentiating w.r.t. m, we get

$$\frac{d \log L}{dm} = -n + \frac{\sum X_i}{m} = 0 \ \Rightarrow \ \hat{m} = \frac{\sum X_i}{n} = \bar{X}$$

Thus, sample mean is MLE of parameter m.

Example 4: For a normal population with parameter μ and σ^2, obtain the maximum likelihood estimators of the parameters.

Solution.

The probability density function of normal distribution is

$$f(X;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(X-\mu)^2}{\sigma^2}}$$

Given a random sample of n independent observations, the likelihood function L is given by

$$L = \prod_{i=1}^{n} f(X_i;\mu,\sigma).$$

Taking logarithm of both sides, we get

$$\log L = \sum \log\left(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(X-\mu)^2}{\sigma^2}}\right) = \sum \log \frac{1}{\sigma\sqrt{2\pi}} - \frac{1}{2\sigma^2}\sum(X_i - \mu)^2$$

Notes

$$= \sum\left(-\log\sigma - \frac{1}{2}\log 2\pi\right) - \frac{1}{2\sigma^2}\sum\left(X_i - \mu\right)^2$$

$$= -n\log\sigma - \frac{n}{2}\log 2\pi - \frac{1}{2\sigma^2}\sum\left(X_i - \mu\right)^2 \qquad \text{.... (1)}$$

(i) MLE of *m*

$$\frac{\partial \log L}{\partial \mu} = \frac{1}{2\sigma^2}\cdot 2\sum\left(X_i - \mu\right) = 0 \text{ or } \sum\left(X_i - \mu\right) = 0 \Rightarrow \hat{\mu} = \frac{\sum X_i}{n} = \bar{X}$$

(ii) MLE of s^2

Rewriting equation (1) as a function of s^2, we get

$$\log L = -\frac{n}{2}\log\sigma^2 - \frac{n}{2}\log 2\pi - \frac{1}{2\sigma^2}\sum\left(X_i - \mu\right)^2$$

$$\therefore \frac{\partial \log L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{\sum\left(X_i - \mu\right)^2}{2\sigma^4} = 0 \text{ or } -n\sigma^2 + \sum\left(X_i - \mu\right)^2 = 0$$

$$\Rightarrow \hat{\sigma}^2 = \frac{\sum\left(X_i - \mu\right)^2}{n}$$

29.3 Interval Estimation

Using point estimation, it is possible to provide a single quantity as an estimator of a parameter. Any point estimator, even if it satisfies all the characteristics of a good estimator, has a limitation that it provides no information about the magnitude of errors due to sampling. This problem is taken care of by the method of interval estimation, that gives a range of the estimator of the parameter.

The method of interval estimation is based upon the sampling distribution of an estimator. The standard error of the estimator is used in the construction of an interval so that the probability of the parameter lying within the interval can be specified.

Given a random sample of n observations X_1, X_2, ... X_n, we can find two values l_1 and l_2 such that the probability of population parameter *q* lying between l_1 and l_2 is (say) *h*. Using symbols, we can write P(l_1 £ *q* £ l_2) = *h*.

Such an interval is termed as a Confidence Interval for *q* and the two limits l_1 and l_2 are termed as Confidential or Fiducial Limits. The percentage probability or confidence is termed as the Level of Confidence or Confidence Coefficient of the interval. For example, the level of confidence of the above interval is 100*h*%. The level of confidence implies that if a large number of random samples are taken from a population and confidence intervals are constructed for each, then 100*h*% of these intervals are expected to contain the population parameter *q*. Alternatively, a 100 *h*% confidence interval implies that we are 100 *h*% confident that the population parameter *q* lies between l_1 and l_2.

As compared to point estimation, the interval estimation is better because it takes into account the variability of the estimator in addition to its single value and thus, provides a range of values. Unlike point estimation, interval estimation indicates that estimation is an uncertain process.

Notes

The methods of construction of confidence intervals in various situations are explained through the following examples.

Confidence Interval for Population Mean

Example 5: Construct 95% and 99% confidence intervals for mean of a normal population.

Solution.

Let X_1, X_2, ... X_n be a random sample of size n from a normal population with mean m and standard deviation s.

We know that sampling distribution of $\overline{X}$ is normal with mean *m* and standard error $\frac{\sigma}{\sqrt{n}}$.

Therefore, $z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$ will be a standard normal variate.

From the tables of areas under standard normal curve, we can write

$$P[-1.96 \le z \le 1.96] = 0.95 \text{ or } P[-1.96 \le \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \le 1.96] = 0.95 \qquad \text{.... (1)}$$

The inequality $-1.96 \le \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$ can be written as

$$-1.96 \frac{\sigma}{\sqrt{n}} \le \bar{X} - \mu \text{ or } \mu \le \bar{X} + 1.96 \frac{\sigma}{\sqrt{n}} \qquad \text{.... (2)}$$

Similarly, from the inequality $\frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \le 1.96$, we can write

$$\mu \ge \bar{X} - 1.96 \frac{\sigma}{\sqrt{n}} \qquad \text{.... (3)}$$

Combining (2) and (3), we get

$$\bar{X} - 1.96 \frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X} + 1.96 \frac{\sigma}{\sqrt{n}}$$

Thus, we can write equation (1) as

$$P\left(\bar{X} - 1.96 \frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X} + 1.96 \frac{\sigma}{\sqrt{n}} \right) = 0.95$$

This gives us a 95% confidence interval for the parameter m. The lower limit of μ is $\bar{X} - 1.96 \frac{\sigma}{\sqrt{n}}$ and the upper limit is $\bar{X} + 1.96 \frac{\sigma}{\sqrt{n}}$. The probability of m lying between these limits is 0.95 and therefore, this interval is also termed as 95% confidence interval for μ.

In a similar way, we can construct a 99% confidence interval for *m* as

$$P\left(\bar{X} - 2.58 \frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X} + 2.58 \frac{\sigma}{\sqrt{n}} \right) = 0.99$$

Notes

Thus, the 99% confidence limits for *m* are $\bar{X} \pm 2.58 \frac{\sigma}{\sqrt{n}}$.

Remarks: When *s* is unknown and n < 30, we use t value instead of 1.96 or 2.58 and use S in place of *s*.

Confidence Interval for Population Proportion

Example 6: Obtain the 95% confidence limits for the proportion of successes in a binomial population.

Solution.

Let the parameter *p* denote the proportion of successes in population. Further, p denotes the proportion of successes in n (≥ 50) trials. We know that the sampling distribution of p will be approximately normal with mean *p* and standard error $\sqrt{\frac{\pi(1-\pi)}{n}}$.

Since *p* is not known, therefore, its estimator p is used in the estimation of standard error of p, i.e., $S.E.(p) = \sqrt{\frac{p(1-p)}{n}}$

Thus, the 95% confidence interval for p is given by

$$P\left(p - 1.96\sqrt{\frac{p(1-p)}{n}} \le \pi \le p + 1.96\sqrt{\frac{p(1-p)}{n}}\right) = 0.95$$

This gives the 95% fiducial limits as $p \pm 1.96\sqrt{\frac{p(1-p)}{n}}$.

Example 7: In a newspaper article of 1600 words in Hindi, 64% of the words were found to be of Sanskrit origin. Assuming that the simple sampling conditions hold good, estimate the confidence limits of the proportion of Sanskrit words in the writer's vocabulary.

Solution.

Let *p* be the proportion of Sanskrit words in the writer's vocabulary. The corresponding proportion in the sample is given as p = 0.64.

$$\therefore \quad S.E.(p) = \sqrt{\frac{0.64 \times 0.36}{1600}} = \frac{0.48}{40} = 0.012$$

We know that almost whole of the distribution lies between 3*s* limits. Therefore, the confidence interval is given by

$$P[p - 3S.E.(p) \le p \le p + 3\,S.E.(p)] = 0.9973$$

Thus, the 99.73% confidence limits of *p* are 0.604 (= 0.64 - 3 × 0.012) and 0.676 (= 0.64 + 3 × 0.012) respectively.

Hence, the proportion of Sanskrit words in the writer's vocabulary are between 60.4% to 67.6%.

Notes

Example 8: A random sample of 500 pineapples was taken from a large consignment and 65 were found to be bad. Estimate the proportion of bad pineapples in the consignment and obtain the standard error of the estimator. Deduce that the percentage of bad pineapples in the consignment almost certainly lies between 8.5 and 17.5.

Solution.

Let p be the proportion of bad pine apples in the large consignment. Its estimate based on the sample is $\hat{p} = \frac{65}{500} = 0.13$ with $S.E.(\hat{p}) = \sqrt{\frac{0.13 \times 0.87}{500}} = 0.015$

Thus, the 99.73% confidence limits of p are $0.13 \pm 3 \times 0.015$, i.e., 0.085 and 0.175. Hence, the proportion of bad pineapples in the given consignment almost certainly lies between 8.5% and 17.5%.

Remarks: The width of a confidence interval can be controlled in two ways:

(i) By adjusting the sample size: More is the sample size the narrower will be the interval.

(ii) By adjusting the level of confidence: Lower the level of confidence the narrower will be the interval.

29.3.1 Determination of an Approximate Sample Size for a Given Degree of Accuracy

Let us assume that we want to find the size of a sample to be taken from the population such that the difference between sample mean and the population mean would not exceed a given value, say Î, with a given level of confidence. In other words, we want to find n such that

$$P\left(\left|\overline{X} - \mu\right| \le \in\right) = 0.95 \text{ (say)} \qquad \text{.... (1)}$$

Assuming that the sampling distribution of $\overline{X}$ is normal with mean *m* and $S.E._{\overline{X}} = \frac{\sigma}{\sqrt{n}}$, we can write

$$P\left(-1.96 \le \frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \le 1.96\right) = 0.95 \text{ or } P\left(\left|\frac{\overline{X} - \mu}{\sigma / \sqrt{n}}\right| \le 1.96\right) = 0.95$$

$$\text{or } P\left(\left|\overline{X} - \mu\right| \le 1.96 \cdot \frac{\sigma}{\sqrt{n}}\right) = 0.95 \qquad \text{.... (2)}$$

Comparing (1) and (2), we get

$$\in = 1.96 \cdot \frac{\sigma}{\sqrt{n}} \text{ or } n = \left(\frac{1.96\sigma}{\in}\right)^2 = \frac{3.84\sigma^2}{\in^2}$$

Remarks:

1. The sample size required with a maximum error of estimation, Î and with a given level of confidence is $n = \frac{z^2\sigma^2}{\in^2}$, where z is the value of standard normal variate for a given level of confidence and σ^2 is the variance of population.

Notes

2. For a given level of confidence and σ^2, n is inversely related to $\in^2$, the square of the maximum error of estimation. This implies that to reduce $\in$ to $\frac{\in}{k}$, the size of the sample must be k^2 times the original sample size.
3. The lesser the magnitude of Î, the more precise will be the interval estimate.

Example 9: What should be the sample size for estimating mean of a normal population if the probability that sample mean differs from population mean by not more than 30% of standard deviation is 0.99.

Solution.

Let n be the size of the sample. It is given that

$$P\left(\left|\bar{X}-\mu\right| \le 0.30\sigma\right) = 0.99 \qquad \text{.... (1)}$$

Assuming that the sampling distribution of $\overline{X}$ is normal with mean *m* and $S.E._{\bar{X}} = \frac{\sigma}{\sqrt{n}}$, we can write

$$P\left(\left|\bar{X}-\mu\right| \le 2.58\frac{\sigma}{\sqrt{n}}\right) = 0.99 \text{ (from table of areas)} \qquad \text{.... (2)}$$

Comparing (1) and (2), we get

$$0.30\sigma = 2.58\frac{\sigma}{\sqrt{n}} \Rightarrow n = \left(\frac{2.58}{0.30}\right)^2 = 73.96 \text{ or } 74 \text{ (approx.)}$$

Example 10: A survey of middle class families of Delhi is proposed to be conducted for the estimation of average monthly consumption (in Rs) per family. What should be the size of the sample so that the average consumption is estimated within a range of Rs 300 with 95% level of confidence. It is known that the standard deviation of the consumption in population is Rs 1,600.

Solution.

Let n denote the size of the sample to be drawn. With usual notations, we want to find n such that

$$P\left(\left|\bar{X}-\mu\right| \le 300\right) = 0.95 \qquad \text{.... (1)}$$

Assuming that the sampling distribution of $\overline{X}$ is normal with mean *m* and $S.E._{\bar{X}} = \frac{\sigma}{\sqrt{n}}$, we can write $P\left(\left|\bar{X}-\mu\right| \le 1.96\frac{\sigma}{\sqrt{n}}\right) = 0.95$

or $$P\left(\left|\bar{X}-\mu\right| \le \frac{1.96 \times 1600}{\sqrt{n}}\right) = 0.95 \qquad \text{.... (2)}$$

Comparing (1) and (2), we get

$$300 = \frac{1.96 \times 1600}{\sqrt{n}} \quad \text{or} \quad n = \left(\frac{1.96 \times 1600}{300}\right)^2 = 109.3$$

Since this value is greater than 109, therefore, the size of the sample should be 110.

29.3.2 Confidence Interval for Population Standard Deviation

Let $S = \sqrt{\frac{1}{n}\sum\left(X_i - \bar{X}\right)^2}$ be the sample standard deviation of a random sample of size n drawn from a normal population with standard deviation *s*. It can be shown that the sampling distribution of S is approximately normal, for large values of n, with mean *s* and standard error $\frac{\sigma}{\sqrt{2n}}$. Thus,

$z = \frac{S - \sigma}{\sigma / \sqrt{2n}}$ can be taken as a standard normal variate.

Example 11: A random sample of 50 observations gave a value of its standard deviation equal to 24.5. Construct a 95% confidence interval for population standard deviation σ.

Solution.

It is given that S = 24.5 and n = 50 (large). We know that $S.E.(S) = \frac{\sigma}{\sqrt{2n}}$. Since *s* is not known, we use its estimate based on sample. Thus, we can write $S.E.(S) = \frac{S}{\sqrt{2n}} = \frac{24.5}{\sqrt{100}} = 2.45$.

Hence 95% confidence interval for *s* is given by

$24.5 - 1.96 \times 2.45 \leq \sigma \leq 24.5 + 1.96 \times 2.45$ or $19.7 \leq \sigma \leq 29.3$

Note More examples on confidence intervals are given later with the questions on test of significance.

29.4 Summary

- Let X be a random variable with probability density function (or probability mass function) $f(X ; \theta_1, \theta_2, ... \theta_k)$, where $\theta_1, \theta_2, ... \theta_k$ are k parameters of the population.

  Given a random sample $X_1, X_2, ... X_n$ from this population, we may be interested in estimating one or more of the k parameters $\theta_1, \theta_2, ... \theta_k$. In order to be specific, let X be a normal variate so that its probability density function can be written as N(X : μ, σ). We may be interested in estimating m or s or both on the basis of random sample obtained from this population.

  It should be noted here that there can be several estimators of a parameter, e.g., we can have any of the sample mean, median, mode, geometric mean, harmonic mean, etc., as an

Notes

estimator of population mean μ. Similarly, we can use either $S = \sqrt{\frac{1}{n}\sum\left(X_i - \bar{X}\right)^2}$ or

$s = \sqrt{\frac{1}{n-1}\sum\left(X_i - \bar{X}\right)^2}$ as an estimator of population standard deviation s. This method of estimation, where single statistic like Mean, Median, Standard deviation, etc. is used as an estimator of population parameter, is known as Point Estimation. Contrary to this it is possible to estimate an interval in which the value of parameter is expected to lie. Such a procedure is known as Interval Estimation. The estimated interval is often termed as Confidence Interval.

- The maximum likelihood estimators are consistent.
- The maximum likelihood estimators are not necessarily unbiased. If a maximum likelihood estimator is biased, then by slight modifications it can be converted into an unbiased estimator.
- If a maximum likelihood estimator is unbiased, then it will also be most efficient.
- A maximum likelihood estimator is sufficient provided sufficient estimator exists.
- The maximum likelihood estimators are invariant under functional transformation, i.e., if t is a maximum likelihood estimator of θ, then f(t) would be maximum likelihood estimator of f(θ).

29.5 Keywords

Estimation: It is a procedure by which sample information is used to estimate the numerical magnitude of one or more parameters of the population.

Cramer Rao Inequality: This inequality gives the minimum possible value of the variance of an unbiased estimator.

Estimator: An estimator t is said to be a sufficient estimator of parameter θ if it utilises all the information given in the sample about θ.

29.6 Self Assessment

1. State whether the following statements are True or False:

   (i) Sample mean is an unbiased estimator of population mean.

   (ii) Sample standard deviation is an unbiased estimator of population standard deviation.

   (iii) An estimator whose variance tends to zero as sample size tends to infinity is called a consistent estimator.

   (iv) An efficient estimator may or may not be unbiased.

   (v) A sufficient estimator is always consistent.

   (vi) The width of the confidence interval depends upon the level of significance as well as on the sample size.

29.7 Review Questions

Notes

1. A random sample of 400 farms in certain year revealed that the average yield per acre of sugarcane was 925 kgs with a standard deviation of 88 kgs.

   (a) Determine the 95% confidence interval for the population mean.

   (b) What should be the size of the sample if the width of 95% confidence interval estimate of *m* is not more than 15?

   Hint : (a) See example 5, (b) $\in = 15/2$.

2. A random sample of 100 sale receipts of a firm showed that its average sales per customer are Rs 250 with a standard deviation of Rs 50 (assume that there is one receipt for each customer).

   (a) Determine the 99% confidence interval for the mean sales.

   (b) How does the width of the confidence interval change if sample size is 400 instead?

   (c) How many sale receipts should be included in the sample in order that a 98% confidence interval has a maximum error of estimation equal to Rs 10.

   Hint: (a) z = 2.58 and since s is not known, use S as its estimate. (b) Sample size is inversely related to the width of Confidence interval. (c) z = 2.33.

3. A survey revealed that 30% of the persons of a state are suffering from a particular disease. How many persons should be included in the sample so that the maximum width of the 95% confidence interval of proportion of persons suffering from the disease is 0.15 units?

   Hint : $n = \frac{z^2 pq}{\in^2}$.

4. A random sample of size 64 has been drawn from a population with standard deviation 20. The mean of the sample is 80. (i) Calculate 95% confidence limits for the population mean. (ii) How does the width of the confidence interval changes if the sample size is 256 instead?

   Hint : σ is given to be 20.

5. In a random sample of 100 articles taken from a large batch of articles, 10 are found to be defective. Obtain a 95% confidence interval for the true proportion of defectives in the batch.

   Hint : See example 6.

6. A random sample of size 10 from a normal population gives the values 64, 72, 65, 70, 68, 71, 65, 62, 66, 67. If it is known that the standard error of the sample mean is $\sqrt{0.7}$, find 95% confidence limits for the population mean. Also find the population variance.

   Hint : $\frac{\sigma}{\sqrt{n}} = \sqrt{0.7}$.

Answers: Self Assessment

1. (i) T (ii) F (iii) F (iv) T (v) T (vi) T

Notes

29.8 Further Readings

Books

Sheldon M. Ross, Introduction to Probability Models, Ninth Edition, Elsevier Inc., 2007.

Jan Pukite, Paul Pukite, Modeling for Reliability Analysis, IEEE Press on Engineering of Complex Computing Systems, 1998.

Unit 30: Method of Least Square

Notes

CONTENTS

Objectives

Introduction

30.1 Method of Least Squares

30.1.1 Fitting of Linear Trend

30.1.2 Fitting of Parabolic Trend

30.1.3 Fitting of Exponential Trend

30.1.4 Merits and Demerits of Least Squares Method

30.2 Summary

30.3 Keywords

30.4 Self Assessment

30.5 Review Questions

30.6 Further Readings

Objectives

After studying this unit, you will be able to:

- Discuss Method of Least Squares
- Describe Method of Selected Points and Method of Semi-Averages
- Explain Seasonal Variations

Introduction

A series of observations, on a variable, recorded after successive intervals of time is called a time series. The successive intervals are usually equal time intervals, e.g., it can be 10 years, a year, a quarter, a month, a week, a day, an hour, etc. The data on the population of India is a time series data where time interval between two successive figures is 10 years. Similarly figures of national income, agricultural and industrial production, etc., are available on yearly basis.

It should be noted here that the time series data are bivariate data in which one of the variables is time. This variable will be denoted by t. The symbol Yt will be used to denote the observed value, at point of time t, of the other variable. If the data pertains to n periods, it can be written as (t, Yt), t = 1, 2, ... n.

Notes

30.1 Method of Least Squares

This is one of the most popular methods of fitting a mathematical trend. The fitted trend is termed as the best in the sense that the sum of squares of deviations of observations, from it, are minimised. We shall use this method in the fitting of following trends:

1. Linear Trend
2. Parabolic Trend
3. Exponential Trend

30.1.1 Fitting of Linear Trend

Given the data (Y_t, t) for n periods, where t denotes time period such as year, month, day, etc., we have to find the values of the two constants, a and b, of the linear trend equation $Y_t = a + bt$.

Using the least square method, the normal equation for obtaining the values of a and b are :

$$\Sigma Y_t = na + b\Sigma t \text{ and}$$

$$\Sigma tY_t = a\Sigma t + b\Sigma t^2$$

Let X = t - A, such that ΣX = 0, where A denotes the year of origin.

The above equations can also be written as

$$\Sigma Y = na + b\Sigma X$$

$$\Sigma XY = a\Sigma X + b\Sigma X^2$$

(Dropping the subscript t for convenience).

Since SX = 0, we can write $a = \frac{\sum Y}{n}$ and $b = \frac{\sum XY}{\sum X^2}$

Note: The procedure for calculation of the two constants is slightly different for even and odd number of observations. This distinction will become obvious from the following two examples.

Example:

Fit a straight line trend to the following data and estimate the likely profit for the year 1986. Also calculate various trend values.

Year	:	1977	1978	1979	1980	1981	1982	1983
Profit (*in lacs of Rs*)	:	60	72	75	65	80	85	95

Solution.

Notes

Calculation Table

Years (t)	Y	$X = t - 1980$	XY	X^2	*Trend Values*
1977	60	-3	-180	9	61.42
1978	72	-2	-144	4	66.28
1979	75	-1	-75	1	71.14
1980	65	0	0	0	76.00
1981	80	1	80	1	80.86
1982	85	2	170	4	85.72
1983	95	3	285	9	90.58
Total	532	0	136	28	

From the table we can write $a = \frac{532}{7} = 76$ (n = 7, the no. of observations)

and $b = \frac{136}{28} = 4.86$

Thus, the fitted line of trend is Y = 76 + 4.86X

Note: It is very important to provide the following details for any trend equations:

(i) The year of origin, (ii) unit of X and (iii) the nature of Y values such as annual figures, monthly figures or monthly averages, quarterly figures or quarterly averages, etc. Thus, the appropriate way of writing the trend equation would be : Y = 76 + 4.86X, where (i) year of origin = 1st July 1980 (the year in which X = 0), (ii) unit of X = 1 year and (iii) $Y^{'s}$ are annual figures of profits.

Calculation of trend values

Trend value of a particular year is obtained by substituting the associated value of X in the trend equation. For example, X = - 3 for 1977, therefore, trend for 1977 is Y = 76 + 4.86 × (- 3) = 61.42

Alternatively, trend values can be calculated as follows:

We know that a is the trend value in the year of origin and b gives the rate of change per unit of time. Thus, the trend for 1980 = 76, for 1979 = 76 - 4.86 = 71. 14, for 1978 = 71.14 - 4.86 = 66.28 and for 1977 = 66.28 - 4.86 = 61.42, etc. Similarly, trend for 1981 = 76 + 4.86 = 80.86, for 1982 = 80.86 + 4.86 = 85.72, etc.

Prediction of trend for a year

Using the trend equation we can predict a trend value for a year which doesn't belong to the observed data. To predict the value for 1986, the associated value of X = 6. Substituting this in the trend equation we get Y = 76 + 6 × 4.86 = Rs 105.16 lacs.

Remarks: The prediction of trend is only valid for periods that are not too far from the observed data.

Example 8: Fit a straight line trend, by the method of least squares, to the following data. Assuming that the same rate of change continues, what would be the predicted sales for 1993?

Year	:	1987	1988	1989	1990	1991	1992
Sales (*in* '000 *Rs*)	:	15	17	20	21	23	24

Notes

Solution.

We note that n is even in the given example.

Calculation Table

Year (t)	*Sales* (Y)	$d = t - 1989.5$	$X = 2d$	XY	X^2	*Trend Values*
1987	15	−2.5	−5	−75	25	15.45
1988	17	−1.5	−3	−51	9	17.27
1989	20	−0.5	−1	−20	1	19.09
1990	21	0.5	1	21	1	20.91
1991	23	1.5	3	69	3	22.73
1992	24	2.5	5	120	5	24.55
Total	120		0	64	70	

From the above table, we can write

$$a = \frac{120}{6} = 20 \text{ and } b = \frac{64}{70} = 0.91$$

∴ The fitted trend line is Y = 20 + 0.91 X

Year of origin : Middle of 1989 and 1990 or 1st Jan. 1990

Unit of X : $\frac{1}{2}$ year (Since X changes by 2 units in one year)

Nature of Y values : Annual figures of sales.

Calculation of trend values

Trend for 1989 = 20 - 0.91 = 19.09

Trend for 1988 = 19.09 - 2 × 0.91 = 17.27

Trend for 1990 = 20 + 0.91 = 20.91

Trend for 1991 = 20.91 + 2 × 0.91 = 22.73, etc.

To predict the sales for 1993, we note that X = 7

Thus, the predicted sales = 20 + 7 × 0.91 = Rs 26.37 (thousand).

Shifting of Origin of a Trend Equation

Let Y = a + bX be the equation of linear trend, with 1985 as the year of origin and unit of X equal to 1 year.

To shift origin of the above equation, say to 1990, we proceed as follows : The associated value of X for 1990 is 5. Thus, the trend for 1990 = a + 5b. We know that a linear trend equation is given by Y = trend value in the year of origin + bX. Thus, we can write the trend equation, with origin at 1990, as Y = a + 5b + bX = a + b (X + 5). This implies that the required equation can be obtained by replacing X by X + 5 in the original trend equation.

Similarly, the trend equation with 1984 as origin can be written as Y = a + b (X - 1) = (a - b) + bX.

Further, if the unit of X is given to be half year, the trend equation with 1990 as the year of origin can be written as Y = a + b (X + 10) = (a + 10b) + bX.

Example 9:

Given the following trend equations:

(a) Y = 50 + 3X, with 1985 as the year of origin and unit of X = 1 year. Shift the origin to 1991.

(b) Y = 100 + 2.5 X, with origin at the middle of 1987 and 1988 and unit of X = $\frac{1}{2}$ year. Shift the origin to (i) 1988 and (ii) 1992.

Solution.

(a) Replacing X by X + 6, in the trend equation, we get

Y = 50 + 3(X + 6) = 68 + 3X, the required trend equation.

(b) (i) For shifting origin to 1988 (i.e., middle of 1988), we have to replace X by X + 1. (note that X = 1 for $\frac{1}{2}$ year)

$\therefore$ Y = 100 + 2.5(X + 1) = 102.5 + 2.5X

(ii) Replace X by X + 9, to get the required equation

$\therefore$ Y = 100 + 2.5(X + 9) = 122.5 + 2.5X

Conversion of Annual Trend Equation into Monthly trend Equation

Usually a trend is fitted to the annual figures because the fitting of a monthly trend is time consuming. However, monthly trend equations are often obtained from annual trend equations.

Let the annual trend equation be Y = a + bX, where Y denotes annual figures and the unit of X = 1 year.

To obtain the monthly trend equation, we have to convert the constants a and b into monthly values.

Thus, when a denotes an annual value, $\frac{a}{12}$ would give the value of the corresponding constant for the monthly equation.

Further, the value of b denotes the annual change in Y per unit of X, i.e., per year. Therefore $\frac{b}{12}$ would be the monthly (average) change in Y per year. Thus, the equation $Y = \frac{a}{12} + \frac{b}{12}X$, denotes a monthly average equation, where Y denotes monthly average for the year and unit of X = 1 year.

In a similar way, the value $\frac{b}{12 \times 12} = \frac{b}{144}$ would denote the monthly change in Y per month. Thus, $Y = \frac{a}{12} + \frac{b}{144}X,$ is the monthly trend equation, where Y denotes monthly figures and the unit of X = 1 month.

Notes

A quarterly trend equation can also be obtained in a similar way. We can write $Y = \frac{a}{4} + \frac{b}{4}X,$ as the quarterly average equation and $Y = \frac{a}{4} + \frac{b}{16}X,$ as the quarterly trend equation.

Example 10: The equation for yearly sales (in '000 Rs) of a commodity with 1st July, 1971, as origin is Y = 91.6 + 28.8X.

(i) Determine the trend equation to give monthly trend values with 15th January, 1972, as origin.

(ii) Calculate the trend values for March, 1972 to August, 1972.

Solution.

(i) The monthly trend equation with 1st July, 1971, as origin is given by $Y = \frac{91.6}{12} + \frac{28.8}{144}X$ = 7.63 + 0.2X, where unit of X is one month.

To shift the origin to 15th January, 1972, we replace X by X + 6.5 in the above equation. Note that the associated value of X for 15th January, 1972, is 6.5. Thus, the required equation is Y = 7.63 + 0.2(X + 6.5) = 8.93 + 0.2X

(ii) Calculation of trend values

Trend value for March, 1972 = 8.93 + 0.2 × 2 = Rs 9.33

Trend value for April, 1972 = 9.33 + 0.2 = Rs 9.53

Trend value for May, 1972 = 9.53 + 0.2 = Rs 9.73

Trend value for June, 1972 = 9.73 + 0.2 = Rs 9.93

Trend value for July, 1972 = 9.93 + 0.2 = Rs 10.13

Trend value for August, 1972 = 10.13 + 0.2 = Rs 10.33.

Example 11:

Convert the following into annual trend equation :

Y = 350 + 3X with origin = I - II Quarter, 1986, unit of X = one quarter and Y denotes quarterly production.

Solution.

Important Note : To convert a quarterly (or monthly) equation into an annual equation, it is necessary to first shift the origin to the middle of the year.

In the given example, since the middle of the year lies a quarter ahead, we shall replace X by X + 1 in the above equation. Thus, the quarterly equation with middle of the year as origin is Y = 350 + 3(X +1) = 353 + 3X.

Then, the annual trend equation can be written as

Y = 353 × 4 + 3 × 16X = 1412 + 48X

Notes

Example 12: Convert the following annual trend equation, for the production of cloth in a factory, into monthly average equation and predict the monthly averages for 1988 and 1989.

Y = 96 + 7.2X, with origin = 1986, unit of X = 1 year and Y denotes annual cloth production in '000 metres.

Solution.

The average monthly equation is given by

$Y = \frac{96}{12} + \frac{7.2}{12}X$ = 8 + 0.6 X, where origin = 1986, unit of X = 1 year and Y denotes monthly average production in the year.

The predicted values of Y are $8 + 0.6 \times 2 = 9.2$ thousand metres for 1988 and 9.2 + 0.6 = 9.8 thousand metres for 1989.

30.1.2 Fitting of Parabolic Trend

The mathematical form of a parabolic trend is given by $Y_t = a + bt + ct^2$ or $Y = a + bt + ct^2$ (dropping the subscript for convenience). Here a, b and c are constants to be determined from the given data.

Using the method of least squares, the normal equations for the simultaneous solution of a, b, and c are :

$$\Sigma Y = na + b\Sigma t + c\Sigma t^2$$

$$\Sigma tY = a\Sigma t + b\Sigma t^2 + c\Sigma t^3$$

$$\Sigma t^2Y = a\Sigma t^2 + b\Sigma t^3 + c\Sigma t^4$$

By selecting a suitable year of origin, i.e., define X = t - origin such that ΣX = 0, the computation work can be considerably simplified. Also note that if ΣX = 0, then ΣX^3 will also be equal to zero. Thus, the above equations can be rewritten as:

$$SY = na + cSX^2 \qquad .... \text{(i)}$$

$$SXY = bSX^2 \qquad .... \text{(ii)}$$

$$SX^2Y = aSX^2 + cSX^4 \qquad .... \text{(iii)}$$

From equation (ii), we get $b = \frac{\sum XY}{\sum X^2}$ .... (iv)

Further, from equation (i), we get $a = \frac{\sum Y - c\sum X^2}{n}$ .... (v)

And from equation (iii), we get $c = \frac{n\sum X^2Y - \left(\sum X^2\right)\left(\sum Y\right)}{n\sum X^4 - \left(\sum X^2\right)^2}$ .... (vi)

Thus, equations (iv), (v) and (vi) can be used to determine the values of the constants a, b and c.

Notes

Example 13: Fit a parabolic trend $Y = a + bt + ct^2$ to the following data, where t denotes years and Y denotes output (in thousand units).

t	:	1981	1982	1983	1984	1985	1986	1987	1988	1989
Y	:	2	6	7	8	10	11	11	10	9

Also compute the trend values. Predict the value for 1990.

Solution.

Calculation Table

t	Y	$X = t - 1985$	XY	X^2Y	X^2	X^3	X^4	*Trend Values*
1981	2	−4	−8	32	16	−64	256	2.28
1982	6	−3	−18	54	9	−27	81	5.02
1983	7	−2	−14	28	4	−8	16	7.22
1984	8	−1	−8	8	1	−1	1	8.88
1985	10	0	0	0	0	0	0	10.00
1986	11	1	11	11	1	1	1	10.58
1987	11	2	22	44	4	8	16	10.62
1988	10	3	30	90	9	27	81	10.12
1989	9	4	36	144	16	64	256	9.08
Total	74	0	51	411	60	0	708	

From the above table, we can write

$$b = \frac{51}{60} = 0.85$$

$$c = \frac{9 \times 411 - 60 \times 74}{9 \times 708 - (60)^2} = -0.27$$

$$a = \frac{74 - (-0.27) \times 60}{9} = 10.0$$

∴ The fitted trend equation is $Y = 10.0 + 0.85X - 0.27X^2$,

with origin = 1985 and unit of X = 1 year.

Various trend values are calculated by substituting appropriate values of X in the above equation. These values are shown in the last column of the above table.

The predicted value for 1990 is given by

$$Y = 10.0 + 0.85 \times 5 - 0.27 \times 25 = 7.5$$

Example 14: The prices of a commodity during 1981-86 are given below. Fit a second degree parabola to the following data. Calculate the trend values and estimate the price of the commodity in 1986.

Year	:	1981	1982	1983	1984	1985	1986
Price	:	110	114	120	138	152	218

Solution.

Notes

Calculation Table

$Year\ (t)$	$Price\ (Y)$	$X = 2(t - 1983.5)$	XY	X^2Y	X^2	X^4	$Trend\ Values$
1981	110	−5	−550	2750	25	625	114.40
1982	114	−3	−342	1026	9	81	109.12
1983	120	−1	−120	120	1	1	116.08
1984	138	1	138	138	1	1	135.28
1985	152	3	456	1368	9	81	166.72
1986	218	5	1090	5450	25	625	210.40
	852	0	672	10852	70	1414	

From the above table, we get

$$b = \frac{672}{70} = 9.6\ ,\ c = \frac{6 \times 10852 - 70 \times 852}{6 \times 1414 - (70)^2} = 1.53 \text{ and } a = \frac{852 - 1.53 \times 70}{6} = 124.15$$

$\therefore$ The equation of parabolic trend is $Y = 124.15 + 9.6X + 1.53X^2$, with year of origin = 1983.5 or 1st January, 1984 and the unit of X = $\frac{1}{2}$ year.

The calculated trend values are shown in the last column of the above table.

The price of the commodity in 1986 is obtained by substituting X = 5, in the above equation.

Thus, $Y = 124.15 + 9.5 \times 5 + 1.53 \times 25 = 210.4$

30.1.3 Fitting of Exponential Trend

The general form of an exponential trend is $Y = a.b^t$, where a and b are constants to be determined from the observed data.

Taking logarithms of both sides, we have $\log Y = \log a + t \log b$.

This is a linear equation in log Y and t and can be fitted in a similar way as done in case of linear trend. Let A = log a and B = log b, then the above equation can be written as log Y = A + Bt.

The normal equations, based on the principle of least squares are

$\Sigma \log Y = nA + B\,\Sigma t$

and $\Sigma t \log Y = A\Sigma t + B\,\Sigma t^2$.

By selecting a suitable origin, i.e., defining X = t - origin, such that SX = 0, the computation work can be simplified. The values of A and B are given by $A = \frac{\sum \log Y}{n}$ and $B = \frac{\sum X \log Y}{\sum X^2}$ respectively.

Thus, the fitted trend equation can be written as log Y = A + BX

or $Y = \text{Antilog}\,[A + BX] = \text{Antilog}\,[\log a + X \log b]$

$= \text{Antilog}\,[\log a.b^X] = a.b^X$.

Notes

Example 15:

Fit a simple exponential trend to the following data and calculate the trend values. Also estimate the trend for 1992.

Year	:	1985	1986	1987	1988	1989
Sales (*Rs Crores*)	:	100	105	112	120	130

Solution.

Calculation Table

Year (t)	*Sales* (Y)	$X = t - 1987$	$\log Y$	$X\log Y$	X^2	*log of Trend Values*	*Trend Values*
1985	100	−2	2.0000	−4.0000	4	1.9955	98.97
1986	105	−1	2.0212	−2.0212	1	2.0241	105.71
1987	112	0	2.0492	0.0000	0	2.0527	112.90
1988	120	1	2.0792	2.0792	1	2.0813	120.59
1989	130	2	2.1139	4.2278	4	2.1099	128.80
Total		0	10.2635	0.2858	10		

From the above table, we get

$$A = \frac{10.2635}{5} = 2.0527 \quad \text{and} \quad B = \frac{0.2858}{10} = 0.0286$$

Further, a = antilog 2.0527 = 112.90 and b = antilog 0.0286 = 1.07

Thus, the fitted trend equation is $Y = 112.90(1.07)^X$

Origin : 1st July, 1987, unit of X = 1 year.

The trend values, computed by the equation Y = antilog [2.0527 + 0.0286X], are written in the last column of the above table. Further, the trend for 1992 is obtained by substituting X = 5, in the above equation.

$\therefore$ Y = antilog[2.0527 + 0.0286 × 5] = antilog[2.1957] = 156.93.

Remarks: The exponential trend equation plotted on a semilogarithmic graph is a straight line.

Example 16:

Fit an exponential trend $Y = a.b^t$ to the following data :

Census Year (t)	:	1941	1951	1961	1971	1981	1991
Population of India (*in Crores*)	:	31.9	36.1	43.9	54.8	68.3	84.4

Predict the population for 2001.

Solution.

Notes

Calculation Table

Census Year t	Population Y	$X = \frac{(t-1966)}{5}$	$\log Y$	$X \log Y$	X^2
1941	31.9	−5	1.5038	−7.5190	25
1951	36.1	−3	1.5575	−4.6725	9
1961	43.9	−1	1.6425	−1.6425	1
1971	54.8	1	1.7388	1.7388	1
1981	68.3	3	1.8344	5.5032	9
1991	84.4	5	1.9263	9.6315	25
Total		0	10.2033	3.0395	70

From the above table, we get $A = \frac{10.2033}{6} = 1.70$ and $B = \frac{3.0395}{70} = 0.043$

Further, a = antilog 1.70 = 50.12 and b = antilog 0.043 = 1.10

Thus, the fitted trend equation is $Y = 50.12(1.10)^X$,

Origin : 1st July, 1966 and unit of X = 5 years.

The trend values can be computed by the equation Y = antilog [1.70 + 0.043X]. Further, the prediction of population for 2001 is obtained by substituting X = 7, in the above equation.

$\therefore$ Y = antilog[1.70 + 0.043 × 7] = antilog[2.001] = 100.2 crores

30.1.4 Merits and Demerits of Least Squares Method

Merits

1. Given the mathematical form of the trend to be fitted, the least squares method is an objective method.

2. Unlike the moving average method, it is possible to compute trend values for all the periods and predict the value for a period lying outside the observed data.

3. The results of the method of least squares are most satisfactory because the fitted trend satisfies the two important properties, i.e., (i) $S(Y_o - Y_t) = 0$ and (ii) $S(Y_o - Y_t)^2$ is minimum. Here Y_o denotes the observed value and Y_t denotes the calculated trend value.

   The first property implies that the position of fitted trend equation is such that the sum of deviations of observations above and below this is equal to zero. The second property implies that the sum of squares of deviations of observations, about the trend equation, are minimum.

Demerits

1. As compared with the moving average method, it is a cumbersome method.

2. It is not flexible like the moving average method. If some observations are added, then the entire calculations are to be done once again.

3. It can predict or estimate values only in the immediate future or past.

4. The computation of trend values, on the basis of this method, doesn't take into account the other components of a time series and hence not reliable.

Notes

5. Since the choice of a particular trend is arbitrary, the method is not, strictly, objective.
6. This method cannot be used to fit growth curves, the pattern followed by the most of the economic and business time series.

30.2 Summary

- Given the data (Y_t, t) for n periods, where t denotes time period such as year, month, day, etc., we have to find the values of the two constants, a and b, of the linear trend equation $Y_t = a + bt$.

  Using the least square method, the normal equation for obtaining the values of a and b are:

  $\Sigma Y_t = na + b\Sigma t$ and

  $\Sigma tY_t = a\Sigma t + b\Sigma t^2$

  Let X = t - A, such that ΣX = 0, where A denotes the year of origin.

  The above equations can also be written as

  $\Sigma Y = na + b\Sigma X$

  $\Sigma XY = a\Sigma X + b\Sigma X^2$

  (Dropping the subscript t for convenience).

  Since SX = 0, we can write $a = \frac{\sum Y}{n}$ and $b = \frac{\sum XY}{\sum X^2}$

- Unlike the moving average method, it is possible to compute trend values for all the periods and predict the value for a period lying outside the observed data.
- The results of the method of least squares are most satisfactory because the fitted trend satisfies the two important properties, i.e., (i) $S(Y_o - Y_t) = 0$ and (ii) $S(Y_o - Y_t)^2$ is minimum. Here Y_o denotes the observed value and Y_t denotes the calculated trend value.

  The first property implies that the position of fitted trend equation is such that the sum of deviations of observations above and below this is equal to zero. The second property implies that the sum of squares of deviations of observations, about the trend equation, are minimum.
- It is not flexible like the moving average method. If some observations are added, then the entire calculations are to be done once again.
- It can predict or estimate values only in the immediate future or past.
- The computation of trend values, on the basis of this method, doesn't take into account the other components of a time series and hence not reliable.

30.3 Keywords

The fitted trend is termed as the best in the sense that the sum of squares of deviations of observations, from it, are minimised.

Parabolic trend: The mathematical form of a parabolic trend is given by $Y_t = a + bt + ct^2$ or $Y = a + bt + ct^2$ (dropping the subscript for convenience). Here a, b and c are constants to be determined from the given data.

30.4 Self Assessment

Notes

1. Fill in the blanks:
   (i) Series of figures arranged in chronological order is known as ....
   (ii) .... is that irreversible movement which continues in the same direction for a considerable period of time.
   (iii) The trend equation fitted by the method of least squares is known as the equation of .... fit.
   (iv) In case of .... trend, the successive observations differ by a constant number.
   (v) In the case of an exponential trend, the successive observations differ by a constant ....
   (vi) In the case of linear trend Y = a + bX, a is termed as the .... value in the year of ....

30.5 Review Questions

1. Determine the trend and short-term fluctuations, assuming additive model, from the following data by calculating 3 yearly moving averages. The figures of profit are in Rs '000.

Years :	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990
Profits :	34	46	52	55	58	61	58	61	64	55

2. Calculate the long-term trend and short-term oscillations, assuming multiplicative model, with a three-year period from the following data on output (in tonnes) of tea.

Year	*Output*	*Year*	*Output*
1969	1632	1973	2620
1970	1557	1974	3120
1971	1652	1975	3236
1972	2100	1976	3562

3. Construct a four-year moving average from the following data on the consumption (in '000 bales) of imported cotton in India.

Year :	1920	1930	1940	1950	1960	1970	1980
Consumption :	129	131	106	91	95	84	93

4. Determine trend values by method of moving average if the observations, given below, are known to have a business cycle of 4 years.

Year :	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992
Values :	41	61	55	48	53	67	62	60	67	73	78	76	84

5. Assuming five-yearly cycle, determine trend of bank clearings (in Rs crores) by moving average method:

Years :	1	2	3	4	5	6	7	8	9	10	11	12
Bank Clearings :	53	79	76	66	69	94	105	87	79	104	97	92

6. Find trend of the following series using a three-year weighted moving average with weights 1, 2, 1.

Year :	1	2	3	4	5	6	7
Value :	2	4	5	7	8	10	13

Notes

7. From the following data calculate the 4-yearly moving average and determine the trend values. Find short-term fluctuations, assuming multiplicative model, and indicate their composition. Plot the original data and the trend values on a graph.

Year	:	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987
Value	:	50.0	36.5	43.0	44.5	38.9	38.1	32.6	41.7	41.1	33.8

Answers: Self Assessment

1. (i) time series (ii) trend (iii) best (iv) linear (v) proportion (vi) trend, origin.

30.6 Further Readings

Books

Sheldon M. Ross, Introduction to Probability Models, Ninth Edition, Elsevier Inc., 2007.

Jan Pukite, Paul Pukite, Modeling for Reliability Analysis, IEEE Press on Engineering of Complex Computing Systems, 1998.

Unit 31: Hypothesis Testing

Notes

CONTENTS

Objectives

Introduction

31.1 Test of Hypothesis

31.2 Tests of Hypothesis Concerning Mean

31.2.1 Test of Hypothesis Concerning Population Mean (s being known)

31.2.2 Test of Hypothesis Concerning Population Mean (s being unknown)

31.2.3 Test of Hypothesis Concerning Equality of two Population Means

31.2.4 Paired t - Test

31.3 Tests of Hypothesis concerning Proportion

31.3.1 Test of Hypothesis that Population Proportion is p0

31.3.2 Test of Hypothesis Concerning Equality of Proportions

31.5 Summary

31.6 Keywords

31.7 Self Assessment

31.8 Review Questions

31.9 Further Readings

Objectives

After studying this unit, you will be able to:

- Discuss Hypothesis Testing
- Explain Hypothesis Concerning Mean

Introduction

A hypothesis is a preconceived idea about the nature of a population or about the value of its parameters. The statements like the distribution of heights of students of a university is normally distributed, the number of road accidents per day in Delhi is 10, etc., are some examples of a hypothesis.

The test of a hypothesis is a procedure by which we test the validity of a given statement about a population. This is done on the basis of a random sample drawn from it.

The hypothesis to be tested is termed as Null Hypothesis, denoted by H_0. This hypothesis asserts that there is no difference between population and sample in the matter under consideration. For example, if H_0 is that population mean $\mu = \mu_0$, then we regard the random sample to have been obtained from a population with mean m_0.

Corresponding to any H_0, we always define an Alternative Hypothesis. This hypothesis, denoted by H_a, is alternate to H_0, i.e., if H_0 is false then H_a is true and vice-versa.

Notes

31.1 Test of Hypothesis

In order to illustrate the procedure of testing a null hypothesis, let us assume that the life of electric bulbs of a company is distributed normally with standard deviation of 150 hours and we want to test the null hypothesis that the mean life of bulbs is 1600 hours against the alternative hypothesis that the mean life is not 1600 hours.

Assuming that H_0 is true, we can construct a sampling distribution of $\overline{X}$, the mean life of bulbs in the sample. If a random sample of 100 bulbs is taken from this population, we know that the distribution of $\overline{X}$ will be normal with mean m = 1600 hours and standard error, $S.E._{\overline{X}} = \frac{150}{10} = 15$ hours. Further, we know that for a normal distribution

$$P\left(-2 \le \frac{\overline{X} - 1600}{15} \le 2\right) = 0.9544$$

or $$P\left(1600 - 2 \times 15 \le \overline{X} \le 1600 + 2 \times 15\right) = 0.9544$$

or $$P\left(1570 \le \overline{X} \le 1630\right) = 0.9544$$

This result shows that the likelihood of getting a random sample, from the given population, with mean lying between 1570 and 1630 hours is 95.44% or equivalently, the likelihood of getting a random sample having its mean either less than 1570 or more than 1630 hours is only 4.56%. Thus, a random sample with its mean lying outside these limits is highly unlikely under the assumption that null hypothesis is true.

However, if the mean computed from the drawn sample is found to lie outside these limits, it may imply that either null hypothesis is false or the rare event, with probability = 4.56%, has occurred.

Thus, if we decide to reject the null hypothesis whenever the computed sample mean falls outside the above limits, the probability of our decision being wrong is only 4.56% or 0.0456.

Two Types of Errors

The decision of acceptance or rejection of a null hypothesis is made on the basis of a sample from a population and hence, an element of uncertainty is always involved in making such decisions. Two types of errors are likely to be committed in the procedure of testing a hypothesis. These are Type I and Type II errors. Type I error is committed when a true null hypothesis is rejected. The probability of this error is termed as the Level of Significance of the test and will be denoted by *a*. The probability of committing an error is also termed as its size. Note that size of type I error, i.e., *a* = 0.0456, in the above example.

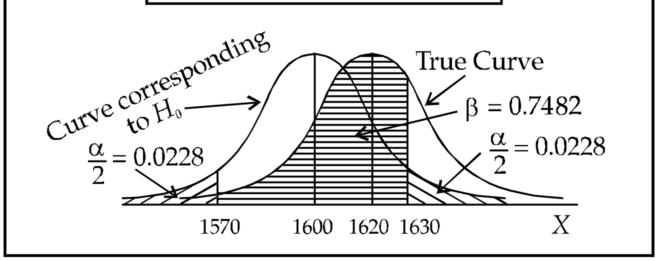
Contrary to this, type II error is committed when a false null hypothesis is accepted. The probability of type II error is denoted by *b*. To understand the meaning of type II error, we assume that the true value of *m* is 1620 instead of the hypothesised value of 1600 hours. If the standard deviation is same, the value of *b* is given by P (1570 £ $\overline{X}$ £ 1630) when *m* = 1620 or P

$\left(\frac{1570 - 1620}{15} \le z \le \frac{1630 - 1620}{15}\right)$ = P(–3.33 £ Z £ 0.67) = 0.4996 + 0.2486 = 0.7482

Notes

The two types of errors are shown by the following figure.

Figure 31.1



It is obvious, from the above figure, that it is not possible to simultaneously control both types of errors because a decrease in probability of committing one type of error is accompanied by the increase in probability of committing the other type of error. Further, we may note that farther the true value of parameter from the hypothesised value, smaller would be the size of type II error, *b*. The graph of various values of *m* against *b* is known as the Operating Characteristic Surve.

In the procedure of testing a hypothesis, the probability or size of type I error, i.e., *a* is specified in advance. Usually we take *a* = 0.05 (i.e., 5%) or 0.01 (i.e., 1%). Also see remarks (1) given at the end of this section.

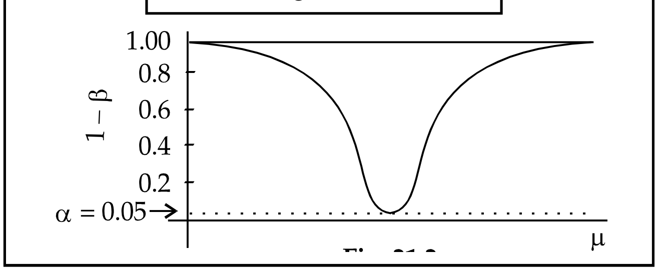
Power of a Test

The power of a test is defined as the probability of rejecting a false null hypothesis. Since *b* is the probability of accepting a false hypothesis, the power of test is given by 1 - *b*. More precisely, we can write

Power of a test = P [Rejecting H_0/H_0 is false] = 1 – *b*

Since the value of β depends upon the true value of population parameter (μ in the above example), the relationship between various values of m and 1 - β is termed as power function, as shown in Figure 31.2.

Figure 31.2



Critical Region and One Tailed versus Two Tailed Tests

Let $H_0 : \mu = \mu_0$ against $H_a : \mu \neq \mu_0$, where μ_0 denotes some specified value of population mean m. For example, μ_0 = 1600, in the example considered above.

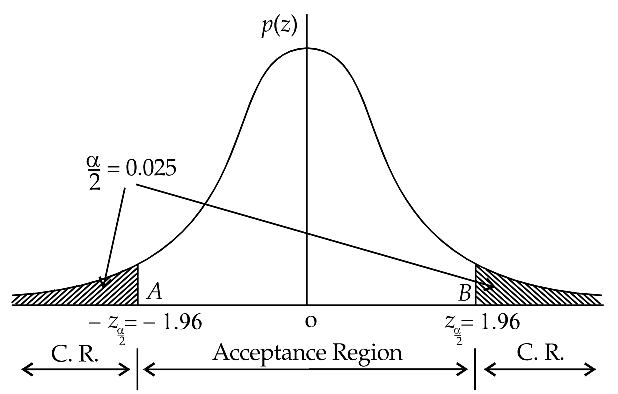
Notes

If we decide to have a = 0.05, we know that for a standard normal variate P[- 1.96 ≤ z ≤ 1.96] = 1 - 0.05 = 0.95, the procedure of testing of hypothesis can be outlined as:

Reject H_0 if the computed value of z from the sample $\left(i.e., \; z_{cal} = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}}\right)$ lies outside the interval (- 1.96, 1.96) and accept it otherwise.

In terms of figure, the portion of z axis covering the interval (- 1.96, 1.96), i.e, A to B is termed as the Acceptance Region and its remaining portions, which lie to the left of point A and to the right of point B, are termed as the Region of Rejection or Critical Region (C.R.).

Figure 31.3

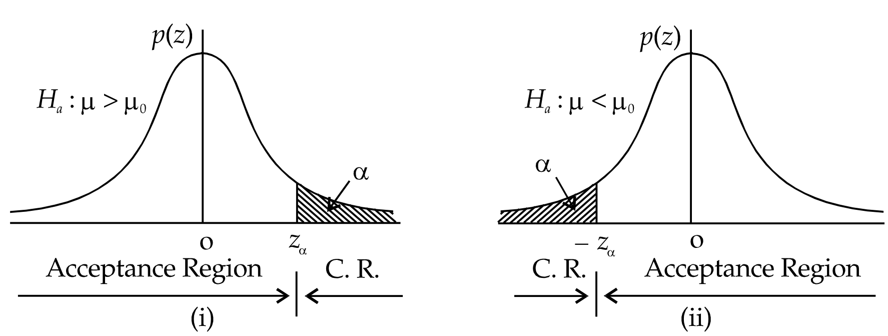


The specification of the critical region for a test depends upon the nature of the alternative hypothesis and the value of α. For example, $H_a : \mu \neq \mu_0$, this implies that m may be less or greater than μ_0. Thus, the critical region is to be specified on both tails of the curve with each part corresponding to half of the value of a. A test having critical region at both the tails of the probability curve is termed as a two tailed test.

Further, if $H_a : \mu > \mu_0$ or $\mu < \mu_0$, the critical region is to be specified only at one tail of the probability curve and the corresponding test is termed as a one tailed test. These situations are shown in the following figures.

The values of the random variable separating the acceptance region from critical region are termed as critical value(s). For example, $z_{a/2}$ and z_a, shown above, are critical values. Similarly, for a normal distribution the critical values for a two tailed test are - 1.96 and 1.96 for α = 0.05 or - 2.58 and 2.58 for α = 0.01 and the corresponding value for a one tailed test is ± 1.645 or ± 2.33 depending upon whether a = 0.05 or 0.01.

Figure 31.4



Remarks:

1. Out of the two types of errors, the type I error is considered to be more serious. Consequently, the probability of type I error is fixed at a low value (often 0.05 or lower). Thus, when the computed value of a statistic falls in the critical region, implying thereby that the probability of H_0 being true is low or equivalently the probability of H_0 being false is high, we reject H_0. However, if the computed value of statistics lies in the acceptance region, it would not be appropriate to say that the probability of H_0 being true is very high because the probability of accepting a false H_0 (the value of *b*) may also be high. Thus, accepting H_0 only implies that the sample information does not provide any evidence of H_0 being false. Because of this nature of the tests of hypothesis, the conclusion "accept H_0" is often replaced by "do not reject H_0" or "there is no evidence against H_0 on the basis of available sample information", etc.

2. The tests of hypothesis are also known as the Tests of Significance. We know that if the sample result is highly unlikely, H_0 is rejected because the sample result is significantly different from the hypothesised value. Alternatively, it implies that the observed difference between the computed and the hypothesised value is not attributable due to chance or fluctuations of sampling.

31.2 Tests of Hypothesis Concerning Mean

These tests can be divided into two broad categories depending upon whether *s*, the population standard deviation, is known or not.

31.2.1 Test of Hypothesis Concerning Population Mean (s being known)

This test is applicable when the random sample $X_1, X_2, \ldots\ldots X_n$ is drawn from a normal population. We can write

$H_0 : \mu = \mu_0$ (specified) against $H_a : \mu \neq \mu_0$ (two tailed test)

The test statistic $\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$. Let the value of this statistic calculated from sample be denoted

as $z_{cal} = \left|\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right|$. The decision rule would be :

Reject H_0 at 5%(say) level of significance if $z_{cal} > 1.96$. Otherwise, there is no evidence against H_0 at 5% level of significance.

Example 12: A manufacturer claims that the average mileage of scooters of his company is 40 kms/litre. A random sample of 20 scooters of the company showed an average mileage of 42 kms/litre. Test the claim of the manufacturer on the assumption that the mileage of scooter is normally distributed with a standard deviation of 2 kms/litre.

Solution.

Here, we have to test $H_0 : \mu = 40$ against $H_a : \mu \neq 40$.

$$z_{cal} = \left|\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right| = \left|\frac{42-40}{2/\sqrt{20}}\right| = 4.47.$$

Since $z_{cal} > 1.96$, is rejected at 5% level of significance.

Notes

Remarks:

1. If the manufacturer claims that the average mileage is more than 40 kms/litre rather than equal to 40 kms/litre, we have to use a one tailed test. Now we shall test H_0 : μ = 40 against H_a : μ > 40 and z_{cal} would be defined as $z_{cal} = \dfrac{\bar{X} - \mu}{\sigma / \sqrt{n}}$. Since this value is also equal to 4.47 and lies in the critical region, we reject at 5% level of significance. This implies that the claim of the manufacturer may be taken as correct.

2. In one tailed tests the alternative hypothesis is expressed as a strict inequality and the null hypothesis as a weak inequality or simply equality.

3. The decision rule can also be specified in terms of prob or p-value of the observed sample result. The p-value is the smallest level of significance at which the null hypothesis can be rejected. We define p-value

$$= 2P\left(z \geq \left| \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \right| \right) \text{, for a two tailed test,}$$

$$= P\left(z \geq \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \right) \text{, when } H_a : m > m_0 \text{ and}$$

$$= P\left(z \leq \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \right) \text{, when } H_a : m < m_0$$

   The decision rule is : If p-value < α, reject H_0.

   In the above example p -value is approximately equal to zero when H_a is either μ ≠ 40 or μ > 40, therefore H_0 is rejected. However, if H_a is taken as μ < 40, the p -value is almost equal to unity and consequently H_0 would be accepted.

4. As per the central limit theorem, even if the parent population is not normal, the sampling distribution of z will be approximately normal when n > 30.

Example 13: A filling machine at a soft drink factory is designed to fill bottles of 200 ml with a standard deviation of 10 ml. A sample of 50 bottles was selected at random from the filled bottles and the volume of soft drink was computed to be 198 ml per bottle. Test the hypothesis that the mean volume of soft drink per bottle is not less than 200 ml.

Solution.

Here n > 30, therefore, the sampling distribution of mean volume of soft drink per bottle will be normal.

We have to test H_0 : μ ≥ 200 against H_a : μ < 200.

It is given that $\bar{X} = 198$ and σ = 10.

Thus, the test static is $z_{cal} = \dfrac{\bar{X} - \mu_0}{\sigma / \sqrt{n}} = \dfrac{198 - 200}{10 / \sqrt{50}} = -1.41$

Since this value is greater than - 1.645, z_{cal} lies in the acceptance region. Hence, there is no evidence against H_0 at 5% level of significance.

Notes

Remarks:

Alternatively, a null hypothesis can be tested by computing critical sample mean $\bar{X}_C$ for a given standard error and the level of significance.

(i) Let H_0: $\mu = \mu_0$ against H_a: $\mu \neq \mu_0$

If a = 0.05, then $\bar{X}_C = \mu_0 \pm 1.96 \frac{\sigma}{\sqrt{n}}$

If $\mu_0 - 1.96 \frac{\sigma}{\sqrt{n}} < \bar{X} < \mu + 1.96 \frac{\sigma}{\sqrt{n}}$, we accept H_0.

(ii) Let H_0: $\mu \leq \mu_0$ against H_a: $\mu > \mu_0$ (Right tailed test)

If a = 0.05 then $\bar{X}_C = \mu_0 + 1.645 \frac{\sigma}{\sqrt{n}}$

If $\bar{X} > \bar{X}_C$, we reject H_0.

In the above example,

H_0: $\mu \geq 200$ against $\mu < 200$ (Left tailed test)

$\therefore \bar{X}_C = 200 - 1.645 \times \frac{10}{\sqrt{50}} = 197.67$

It is given that $\bar{X} = 198$. Since $\bar{X} > \bar{X}_C$, we accept H_0 at 5% level of significance.

31.2.2 Test of Hypothesis Concerning Population Mean (σ being unknown)

When *s* is not known, we use its estimate computed from the given sample. Here, the nature of the sampling distribution of $\bar{X}$ would depend upon sample size n. There are the following two possibilities:

(i) If parent population is normal and n < 30 (popularly known as small sample case), use t - test. The unbiased estimate of *s* in this case is given by $s = \sqrt{\frac{\sum\left(X_i - \bar{X}\right)^2}{n-1}}$.

Also, like normal test, the hypothesis may be one or two tailed.

(ii) If n ³ 30 (large sample case), use standard normal test. The unbiased estimate of *s* in this case can be taken as $S = \sqrt{\frac{\sum\left(X_i - \bar{X}\right)^2}{n}}$, since the difference between n and n - 1 is negligible for large values of n. Note that the parent population may or may not be normal in this case.

Example 14: The yield of alfalfa from six test plots is 2.75, 5.25, 4.50, 2.50, 4.25 and 3.25 tonnes per hectare. Test at 5% level of significance whether this supports the contention that true average yield for this kind of alfalfa is 3.50 tonnes per hectare.

Notes

Solution.

We note that s is not given and n = 6 (< 30), $\therefore$ t - test is applicable.

Using sample information we have

$$\bar{X} = \frac{2.75 + 5.25 + 4.50 + 2.50 + 4.25 + 3.25}{6} = 3.75.$$

To calculate s, we define $u_i = \frac{X_i - 3.75}{0.25} = (X_i - 3.75) \times 4$

X_i	2.75	5.25	4.50	2.50	4.25	3.25
u_i	– 4	6	3	– 5	2	– 2
u_i^2	16	36	9	25	4	4

From the above table $\sum u_i^2 = 94$. Therefore, $s = 0.25\sqrt{\frac{94}{6-1}} = 1.085$

We have to test H_0 : m = 3.50 against H_a : m ¹ 3.50.

The test statistic $\frac{\bar{X} - \mu_0}{s/\sqrt{n}}$ ~ t - distribution with (n - 1) d.f.

Thus, $t_{cal} = \left|\frac{3.75 - 3.50}{1.085/\sqrt{6}}\right| = 0.564$

Further, the critical value of t, from table at 5% level of significance and with 5 d.f. is 2.571. Since t_{cal} is less than this value, there is no evidence against at 5% level of significance.

Example 15: Daily sales figures of 40 shopkeepers showed that their average sales and standard deviation were Rs 528 and Rs 600 respectively. Is the assertion that daily sales on the average is Rs 400, contradicted at 5% level of significance by the sample?

Solution.

Since n > 30, standard normal test is applicable. It is given that n = 40, $\bar{X}$ = 528 and S = 600.

We have to test H_0 : μ = 400 against H_a : μ ≠ 400.

$$z_{cal} = \left|\frac{528 - 400}{600/\sqrt{40}}\right| = 1.35.$$

Since this value is less than 1.96, there is no evidence against H_0 at 5% level of significance. Hence, the given assertion is not contradicted by the sample.

31.2.3 Test of Hypothesis Concerning Equality of two Population Means

Notes

If random samples are obtained from each of the two normal populations, refer to § 20.2.2, the sampling distribution of the difference of their means is given by

$$\bar{X}_1 - \bar{X}_2 \sim N\left(\mu_1 - \mu_2, \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right).$$

Case I. If σ_1 and σ_2 are known, use standard normal test.

(a) To test $H_0 : \mu_1 = \mu_2$ against $H_a : \mu_1 \neq \mu_2$ (two tailed test), the test statistic is

$$z_{cal} = \frac{\left|\left(\bar{X}_1 - \bar{X}_2\right) - \left(\mu_1 - \mu_2\right)\right|}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{\left|\bar{X}_1 - \bar{X}_2\right|}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \text{ under } H_0.$$

This value is compared with 1.96 (2.58) for 5% (1%) level of significance.

(b) To test $H_0 : \mu_1 \leq \mu_2$ against $H_a : \mu_1 > \mu_2$ (one tailed test), the test statistic is $z_{cal} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$,

and the critical value for 5% (1%) level of significance is 1.645 (2.33).

(c) To test $H_0 : \mu_1 \geq \mu_2$ against $H_a : \mu_1 < \mu_2$ (one tailed test), the test statistic, i.e., z_{cal} is same as in (b) above, however, the critical value for 5% (or 1%) level of significance is - 1.645 (or - 2.33).

Case II. If σ_1 and σ_2 are not known, their estimates based on samples are used. This category of tests can be further divided into two sub-groups.

1. Small Sample Tests (when either n_1 or n_2 or both are less than or equal to 30). To test $H_0 : \mu_1 = \mu_2$, we use t - test. The respective estimates of σ_1 and σ_2 are given by

$$s_1 = \sqrt{\frac{\sum\left(X_{1i} - \bar{X}_1\right)^2}{n_1 - 1}} = S_1\sqrt{\frac{n_1}{n_1 - 1}} \text{ and } s_2 = \sqrt{\frac{\sum\left(X_{2i} - \bar{X}_2\right)^2}{n_2 - 1}} = S_2\sqrt{\frac{n_2}{n_2 - 1}}$$

This test is more restrictive because it is based on the assumption that the two samples are drawn from independent normal populations with equal standard deviations, i.e., $\sigma_1 = \sigma_2 = \sigma$ (say). The pooled estimate of σ, denotes by s, is defined as

$$s = \sqrt{\frac{\sum\left(X_{1i} - \bar{X}_1\right)^2 + \sum\left(X_{2i} - \bar{X}_2\right)^2}{n_1 + n_2 - 2}} = \sqrt{\frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2}} = \sqrt{\frac{\left(n_1 - 1\right)s_1^2 + \left(n_2 - 1\right)s_2^2}{n_1 + n_2 - 2}}$$

(a) To test $H_0 : m_1 = m_2$ against $H_a : m_1 {}^1 m_2$ (two tailed test), the test statistic is

$$t_{cal} = \frac{\left|\bar{X}_1 - \bar{X}_2\right|}{\sqrt{\frac{s^2}{n_1} + \frac{s^2}{n_2}}} = \frac{\left|\bar{X}_1 - \bar{X}_2\right|}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{\left|\bar{X}_1 - \bar{X}_2\right|}{s} \times \sqrt{\frac{n_1 n_2}{n_1 + n_2}}$$, which follows t - distribution with $(n_1 + n_2 - 2)$ d.f.

Notes

This value is compared with the value of t from tables, to be denoted as $t_{a/2}(n_1 + n_2 - 2)$, at 100*a*% level of significance with $(n_1 + n_2 - 2)$ d.f.

(b) To test $H_0 : \mu_1 \leq \mu_2$ against $H_a : \mu_1 > \mu_2$ (one tailed test), the test statistic is $t_{cal} = \frac{\left(\bar{X}_1 - \bar{X}_2\right)}{s} \times \sqrt{\frac{n_1 n_2}{n_1 + n_2}}$. This value is compared with $t_a(n_1 + n_2 - 2)$ from tables.

(c) To test $H_0 : \mu_1 \geq \mu_2$ against $H_a : \mu_1 < \mu_2$ (one tailed test), the test statistic, i.e., t_{cal} is same as in (b) above. This value is compared with $- t_a(n_1 + n_2 - 2)$.

2. Large Sample Tests (when both n_1 and n_2 is greater than 30)

In this case s_1 and s_2 are estimated by their respective sample standard deviations S_1 and S_2.

The test statistics for two and one tailed tests are $z_{cal} = \frac{\left|\bar{X}_1 - \bar{X}_2\right|}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$ and $z_{cal} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$

respectively. The remaining procedure is same as in case I above.

Remarks:

1. 100(1 - *a*)% confidence limits for m_1 - m_1 are given by $\bar{X}_1 - \bar{X}_2 \pm z_{\alpha/2} S.E._{\left(\bar{X}_1 - \bar{X}_2\right)}$.

   If $\bar{X}_1 - \bar{X}_2$ ~ t - distribution, $z_{a/2}$ is replaced by $t_{a/2}(n_1 + n_2 - 2)$.

2. If the two sample are drawn from populations with same standard deviations, i.e., $s_1 = s_2 = s$ (say), then $S.E._{\left(\bar{X}_1 - \bar{X}_2\right)} = \sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ for problems covered under case I and

   $S.E._{\left(\bar{X}_1 - \bar{X}_2\right)} = S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ for problems covered under case II, large sample tests. S is a pooled estimate of *s*, is given by

$$S = \sqrt{\frac{\sum\left(X_{1i} - \bar{X}_1\right)^2 + \sum\left(X_{2i} - \bar{X}_2\right)^2}{n_1 + n_2}} = \sqrt{\frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2}}$$

Example 16: An investigation of the relative merits of two kinds of flashlight batteries showed that a random sample of 100 batteries of brand X lasted on the average 36.5 hours with a standard deviation of 1.8 hours, while a random sample of 80 batteries of brand Y lasted on the average 36.8 hours with a standard deviation of 1.5 hours. Use a level of significance of 1% to test whether the observed difference between average life times is significant.

Solution.

Let X and Y denote the life time of flashlight batteries of type X and type Y respectively and let μ_X and μ_Y be their respective population means.

It is given that $\bar{X} = 36.5$, $S_X = 1.8$, $n_X = 100$, $\bar{Y} = 36.8$, $S_Y = 1.5$, $n_Y = 80$.

We have to test $H_0 : \mu_X = \mu_Y$ against $H_a : \mu_X \neq \mu_Y$.

Since sample sizes are large (> 30), it is a large sample case.

The test statistic is $$z_{cal} = \frac{|36.5 - 36.8|}{\sqrt{\frac{1.8^2}{100} + \frac{1.5^2}{80}}} = \frac{0.3}{0.246} = 1.219$$

Since this value is less than 2.58, there is no evidence against H_0 at 1% level of significance and thus, the observed difference between average life times cannot be regarded as significant.

Example 17: Measurements performed on random samples of two kinds of cigarettes yielded the following results on their nicotine content (in mgs)

Brand A : 21.4, 23.6, 24.8, 22.4, 26.3

Brand B : 22.4, 27.7, 23.5, 29.1, 25.8

Assuming that the nicotine content is distributed normally, test the hypothesis that brand B has a higher nicotine content than brand A.

Solution.

We have to test $H_0 : \mu_A \geq \mu_B$ against $H_a : \mu_A < \mu_B$.

Note that the rejection of H_0 would imply that brand B has a higher nicotine content than brand A.

The means of the two samples are

$$\bar{X}_A = \frac{21.4 + 23.6 + 24.8 + 22.4 + 26.3}{5} = 23.7$$

and $$\bar{X}_B = \frac{22.4 + 27.7 + 23.5 + 29.1 + 25.8}{5} = 25.7.$$

Also $$\sum \left(X_{Ai} - \bar{X}_A\right)^2 = 14.96 \ and \sum \left(X_{Bi} - \bar{X}_B\right)^2 = 31.30$$

The pooled estimate of s is $$s = \sqrt{\frac{14.96 + 31.30}{5 + 5 - 2}} = 2.40$$

Thus, the test statistic is $$t_{cal} = \frac{(23.7 - 25.7)}{2.40} \times \sqrt{\frac{5 \times 5}{5 + 5}} = -1.318.$$

The critical value of t at 5% level of significance and 8 d.f. is - 1.86. Since t_{cal} is greater than this value, it lies in the region of acceptance and hence, there is no evidence against at 5% level of significance. Thus, the nicotine content in brand B is not higher than in brand A.

Example 18: Two salesmen A and B are working in a certain district. From a sample survey conducted by the head office, the following results were obtained. State whether there is any significant difference in the average sales between the two salesmen:

Notes

	A	B
No. of Sales	20	18
Average Sales (in Rs)	170	205
Standard deviation (in Rs)	20	25

Solution.

Since $n_1, n_2 < 30$, it is a small sample case.

We have to test $H_0 : \mu_A = \mu_B$ against $H_a : \mu_A \neq \mu_B$.

Assuming that the two samples have come from the same population with S.D. *s*, we find its pooled estimate as

$$s = \sqrt{\frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2}} = \sqrt{\frac{20 \times 20^2 + 18 \times 25^2}{36}} = 23.12$$

Also $t_{cal} = \frac{|170 - 205|}{23.12}\sqrt{\frac{20 \times 18}{20 + 18}} = 4.66.$ This value is highly significant, therefore, H_0 is rejected at 5% level of significance.

Example 19: The mean life of a random sample of 10 light bulbs was found to be 1456 hours with a S.D. of 423 hours. A second sample of 17 bulbs chosen at random from a different batch showed a mean life of 1280 hours with S.D. of 398 hours. Is there a significant difference between the mean life of the two batches?

Solution.

Note that the two samples have been obtained from the same population with unknown *s*.

We have to test $H_0 : m_1 = m_2$ against $H_a : m_1$ ¹ m_2.

It is given that $\bar{X}_1 = 1456$, $S_1 = 423$, $n_1 = 10$, $\bar{X}_2 = 1280$, $S_2 = 398$, $n_2 = 17$.

The pooled estimate of *s* is $s = \sqrt{\frac{10 \times 423^2 + 17 \times 398^2}{10 + 17 - 2}} = 423.42$

Therefore $t_{cal} = \frac{|1456 - 1280|}{423.42} \times \sqrt{\frac{10 \times 17}{10 + 17}} = 1.04$

The value of t from table at 5% level of significance and with 25 d.f. is 2.06. Since t_{cal} is less than this value, there is no evidence against H_0. Hence, the observed difference in mean life of bulbs of the two batches can be regarded as due to fluctuations of sampling.

When the Hypothesized Difference is not Zero

Let H_0: m_1 £ m_2 + k against H_a: $m_1 > m_2 + k$, where k is constant. The above can also be written as.

H_0: $m_1 - m_2$ £ k against H_a: $m_1 - m_2 > k$

Thus we can write

$$\bar{X}_1 - \bar{X}_2 \sim N\left(k, \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$$

Notes

or $Z_{cal} = \dfrac{\left|\overline{X}_1 - \overline{X}_2 - k\right|}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$ under H_0.

In a similar way, we can write the expressions for t_{cal} under different situations.

Example 20: A sample of 100 electric bulbs of 'Philips' gave a mean life of 1500 hours with a standard deviation of 60 hours. Another sample of 100 electric bulbs of "HMT" gave a mean life of 1615 hours with a standard deviation of 80 hours. Can we conclude that the mean life of 'HMT' bulbs is greater then that of 'Philips' bulbs by 100 hours?

Let $\overline{X}_1 = 1615, S_1 = 80, n_1 = 100, \overline{X}_2 = 1500, S_2 = 60, n_2 = 100$.

We can write

$H_0: \mu_1 \leq \mu_2 + 100$ against $H_a: \mu_1 > \mu_2 + 100$

$$Z_{cal} = \frac{\left|1615 - 1500 - 100\right|}{\sqrt{\dfrac{80^2}{100} + \dfrac{60^2}{100}}} = 1.5$$

Since $Z_{cal} < 1.645$, we accept H_0 at 5% and say that the difference in mean life of 'HMT' bulbs and that of 'Philips' bulbs is less than or equal to 100 hours.

31.2.4 Paired t - Test

This test is used in situations where there is a pairing of observations (X_{1i}, X_{2i}), like marks obtained by students of a class in two subjects, performance of the patients before and after the administration of a drug, etc. We define $d_i = X_{1i} - X_{2i}$, the difference in the observations for the i th item.

Then, we compute $\bar{d} = \dfrac{\sum d_i}{n}$ and $s_d = \sqrt{\dfrac{\Sigma\left(d_i - \bar{d}\right)^2}{n-1}} = \sqrt{\dfrac{\Sigma d_i^2 - n\bar{d}^2}{n-1}}$

As before, we can test $H_0 : \mu_1 = \mu_2$ against $H_a : \mu_1 \neq \mu_2$ (two tailed test) or $H_0 : \mu_1 \leq (\text{or} \geq) \mu_2$ against $H_a : \mu_1 > (\text{or} <) \mu_2$ (one tailed test).

The test statistic $t = \dfrac{\left|\bar{d}\right|}{s_d / \sqrt{n}} = \dfrac{\left|\bar{d}\right|\sqrt{n}}{s_d} \sim t$-distribution with (n - 1) d.f.

Example 21: Eleven students of B.Com. (Hons) were given a test in economic analysis. They were imparted a month's special coaching and a second test was held at the end of it. The result were as follows :

Student No.	:	1	2	3	4	5	6	7	8	9	10	11
Marks in 1st Test	:	36	40	36	34	46	32	38	46	40	38	42
Marks in 2nd Test	:	40	44	40	40	46	40	34	48	38	44	36

Do the marks give an evidence that the students have benefited by extra coaching?

Notes

Solution.

We have to test $H_0 : \mu_1 = \mu_2$ against $H_a : \mu_1 \neq \mu_2$.

Note that H_0 implies that students have not benefited by the extra coaching.

Let X_1 and X_2 denote the marks in 1st and 2nd tests respectively.

Calculation of $\overline{d}$ and s_d

Student No.	:	1	2	3	4	5	6	7	8	9	10	11
d_i	:	−4	−4	−4	−6	0	−8	4	−2	2	−6	6
d_i^2	:	40	44	40	40	46	40	34	48	38	44	36

From the above table, we can write $\sum d_i = -22$ and $\sum d_i^2 = 244$

Thus, $\overline{d} = \frac{-22}{11} = -2$ and $s_d = \sqrt{\frac{244 - 11 \times 4}{10}} = 4.47$

Further, $t_{cal} = \frac{|-2 \times 11|}{4.47} = 1.48$. The value of t at 5% level of significance and 10 d.f. is 2.228. Therefore, the sample information provides no evidence that students have benefited by extra coaching.

Example 22: A random sample of heights of 20 students gave a mean of 68 inches with S.D. of 3 inches. Test the hypothesis that mean height in population is 70 inches under the assumption that the heights are normally distributed. Also construct a 95% confidence interval for the population mean.

Solution.

We have to test $H_0 : \mu_1 = 70$ against $H_a : \mu_1 \neq 70$.

It is given that n = 20, $\overline{X} = 68$ and S = 3.

The unbiased estimate of s.d. is $s = S\sqrt{\frac{n}{n-1}} = 3\sqrt{\frac{20}{19}} = 3.08$.

$$\therefore \; S.E._{\overline{X}} = \frac{s}{\sqrt{n}} = \frac{3.08}{\sqrt{20}} = 0.688.$$

Alternatively, we can directly write

$$S.E._{\overline{X}} = \frac{s}{\sqrt{n}} = S\sqrt{\frac{n}{n-1}} \times \frac{1}{\sqrt{n}} = \frac{S}{\sqrt{n-1}} = \frac{3}{\sqrt{19}} = 0.688.$$

Thus, $t_{cal} = \frac{|68 - 70| \times \sqrt{19}}{3} = 2.906$

This value is greater than 2.093, the value of t from tables at 5% level of significance and 19 d.f. Thus, H_0 is rejected.

Notes

The 100(1 - a)% confidence limits for m are $\bar{X} \pm t_{\alpha/2} S.E._{\bar{X}}$.

Thus, the 95% confidence limits for m are given by $68 \pm 2.093 \times \frac{3}{\sqrt{19}} = 68 \pm 1.44$, i.e., 66.56 and 69.44 inches.

Example 23: Ten individuals are chosen at random from a normal population and their weights (in kgs) are found to be 63, 63, 66, 67, 68, 69, 70, 70, 71, 71. In the light of this data, discuss the suggestion that the mean height in the population is 66 inches.

Solution.

We have to test $H_0 : \mu = 66$ against $H_a : \mu \neq 66$.

From the given data, we can compute $\bar{X} = 67.8$ and $s = 3.01$.

$$\therefore \quad t_{cal} = \frac{\left|(67.8 - 66.0)\sqrt{10}\right|}{3.01} = 1.89.$$

This value is less than 2.262, the value of t from tables for 9 d.f. at 5% level of significance. Thus, there is no evidence against H_0.

31.3 Tests of Hypothesis concerning Proportion

Like the tests concerning sample mean, the null hypothesis to be tested would be either $\pi = \pi_0$, i.e., the proportion of successes in population is π_0 or $\pi_1 = \pi_2$, i.e., two populations have the same proportion of successes. These tests are based upon the sampling distribution of p, the proportion of successes in sample and the sampling distribution of $p_1 - p_2$, the difference between two sample proportions.

31.3.1 Test of Hypothesis that Population Proportion is π_0

The null hypothesis to be tested is $H_0 : \pi = \pi_0$ against $H_a : \pi \neq \pi_0$ for a two tailed test and $\pi >$ or $< \pi_0$ for a one tailed test. The test statistic is

$$z_{cal} = \frac{p - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} = (p - \pi_0)\sqrt{\frac{n}{\pi_0(1-\pi_0)}}$$

Remarks: The 100(1 - *a*)% confidence limits for *p* are $p \pm z_{a/2} S.E.(p)$.

Example 24: A wholesaler in apples claims that only 4% of the apples supplied by him are defective. A random sample of 600 apples contained 36 defective apples. Test the claim of the wholesaler.

Solution.

We have to test $H_0 : \pi \leq 0.04$ against $H_a : \pi > 0.04$.

It is given that $p = \frac{36}{600} = 0.06$ and n = 600.

Notes

$$\therefore \qquad z_{cal} = (0.06 - 0.04)\sqrt{\frac{600}{0.04 \times 0.96}} = 2.5$$

This value is highly significant in comparison to 1.645, therefore, H_0 is rejected at 5% level of significance.

Example 25: The manufacturer of a spot remover claims that his product removes at least 90% of all spots. What can be concluded about his claim at the level of significance $a = 0.05$, if the spot remover removed only 174 of the 200 spots chosen at random from the spots on clothes brought to a dry cleaning establishment?

Solution.

We have to test $H_0 : \pi \geq 0.9$ against $H_a : p < 0.9$.

It is given that $p = \frac{174}{200} = 0.82$ and n = 200.

$$\therefore \qquad z_{cal} = (0.82 - 0.90)\sqrt{\frac{200}{0.9 \times 0.1}} = -3.77$$

Since this value is less than - 1.645, H_0 is rejected at 5% level of significance. Thus, the sample evidence does not support the claim of the manufacturer.

Example 26: 470 heads were obtained in 1,000 throws of an unbiased coin. Can the difference between the proportion of heads in sample and their proportion in population be regarded as due to fluctuations of sampling?

Solution.

We have to test $H_0 : \pi = 0.5$ against $H_a : \pi \neq 0.5$.

It is given that $p = \frac{470}{1000} = 0.47$ and n = 1000.

$$\therefore \qquad z_{cal} = |0.47 - 0.50|\sqrt{\frac{1000}{0.5 \times 0.5}} = 1.897.$$

Since this value is less than 1.96, the coin can be regarded as fair and thus, the difference between sample and population proportion of heads are only due to fluctuations of sampling.

31.3.2 Test of Hypothesis Concerning Equality of Proportions

The null hypothesis to be tested is $H_0 : \pi_1 = \pi_2$ against $H_a : \pi_1 \neq \pi_2$ for a two tailed test and $\pi_1 >$ or $< \pi_2$ for a one tailed test.

The test statistic is $z_{cal} = (p_1 - p_2)\sqrt{\frac{n_1 n_2}{\pi(1-\pi)(n_1 + n_2)}}$ under the assumption that $\pi_1 = \pi_2 = \pi$, where π is known. Often population proportion p is unknown and it is estimated on the basis of samples. The pooled estimate of π, denoted by p, is given by $p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$.

Thus, the test statistic becomes $z_{cal} = (p_1 - p_2)\sqrt{\frac{n_1 n_2}{p(1-p)(n_1+n_2)}}$.

Remarks: $100(1 - \alpha)\%$ confidence limits of $(\pi_1 - \pi_2)$ can be written as

$$(p_1 - p_2) \pm z_{a/2} S.E.(p_1 - p_2)$$

Example 27: In a random sample of 600 persons from a large city, 450 are found to be smokers. In another sample of 900 persons from another large city, 450 are smokers. Do the data indicate that the cities are significantly different with respect to the prevalence of smoking? Let the level of significance be 5%.

Solution.

We have to test $H_0 : \pi_1 = \pi_2$ against $H_a : \pi_1 \neq \pi_2$.

It is given that $n_1 = 600$, $n_2 = 900$, $X_1 = X_2 = 450$.

$$\therefore \quad p_1 = \frac{X_1}{n_1} = \frac{450}{600} = 0.75 \text{ and } p_2 = \frac{X_2}{n_2} = \frac{450}{900} = 0.50$$

The pooled estimate of p, i.e., $p = \frac{450+450}{600+900} = 0.6$

Thus, $z_{cal} = |0.75 - 0.50|\sqrt{\frac{600 \times 900}{0.6 \times 0.4 \times 1500}} = 9.682$

This value is highly significant, therefore, H_0 is rejected. Thus, the given samples indicate that the two cities are significantly different with regard to the prevalence of smoking.

Example 28: A company is considering two different television advertisements for the promotion of a new product. Management believes that advertisement A is more effective than advertisement B. Two test market areas with virtually identical consumer characteristics are selected ; advertisement A is used in one area and advertisement B is used in the other area. In a random sample of 60 customers who saw the advertisement A, 18 tried the product. In a random sample of 100 customers who saw advertisement B, 22 tried the product. Does this indicate that advertisement A is more effective than advertisement B, if a 5% level of significance is used?

Solution.

We have to test $H_0 : \pi_A \leq \pi_B$ against $H_a : \pi_A > \pi_B$.

It is given that $n_A = 60$, $X_A = 18$, $n_B = 100$ and $X_B = 22$.

Thus, $p_A = \frac{18}{60} = 0.30$ and $p_B = \frac{22}{100} = 0.22$.

Also, the pooled estimate of p, i.e., $p = \frac{18+22}{160} = 0.25$.

$$\therefore \quad z_{cal} = (0.30 - 0.22)\sqrt{\frac{60 \times 100}{0.25 \times 0.75 \times 160}} = 1.131$$

Notes

Since this value is less than 1.645, there is no evidence against H_0 at 5% level of significance. Thus, the sample information provides no indication that advertisement A is more effective than advertisement B.

Remarks:

As in the variable case, we can also test the hypothesis $\pi_1 = \pi_2 + k$. Since $\pi_1 \neq \pi_2$, pooling of proportions is not allowed for the computations of standard error of $p_1 - p_2$. The standard error in this case is

$$\sqrt{\frac{p_1(1-p_1)}{n_1}+\frac{p_2(1-p_2)}{n_2}}$$

31.5 Summary

- A hypothesis is a preconceived idea about the nature of a population or about the value of its parameters. The statements like the distribution of heights of students of a university is normally distributed, the number of road accidents per day in Delhi is 10, etc., are some examples of a hypothesis.
- The test of a hypothesis is a procedure by which we test the validity of a given statement about a population. This is done on the basis of a random sample drawn from it.
- The hypothesis to be tested is termed as Null Hypothesis, denoted by H_0. This hypothesis asserts that there is no difference between population and sample in the matter under consideration. For example, if H_0 is that population mean $\mu = \mu_0$, then we regard the random sample to have been obtained from a population with mean m_0.
- Corresponding to any H_0, we always define an Alternative Hypothesis. This hypothesis, denoted by H_a, is alternate to H_0, i.e., if H_0 is false then H_a is true and vice-versa.
- If the manufacturer claims that the average mileage is more than 40 kms/litre rather than equal to 40 kms/litre, we have to use a one tailed test. Now we shall test $H_0 : \mu = 40$ against $H_a : \mu > 40$ and z_{cal} would be defined as $z_{cal} = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$. Since this value is also equal to 4.47 and lies in the critical region, we reject at 5% level of significance. This implies that the claim of the manufacturer may be taken as correct.
- In one tailed tests the alternative hypothesis is expressed as a strict inequality and the null hypothesis as a weak inequality or simply equality.
- The decision rule can also be specified in terms of prob or p-value of the observed sample result. The p-value is the smallest level of significance at which the null hypothesis can be rejected. We define p-value

31.6 Keywords

Hypothesis: A hypothesis is a preconceived idea about the nature of a population or about the value of its parameters.

Power of a test: The power of a test is defined as the probability of rejecting a false null hypothesis. Since *b* is the probability of accepting a false hypothesis, the power of test is given by 1 - *b*. More precisely, we can write

Power of a test = P [Rejecting H_0/H_0 is false] = 1 – *b*

Notes

31.7 Self Assessment

1. Fill in the blanks:

   (i) The reciprocal of standard error of an estimator is ....

   (ii) ...... tailed test is used when H_0 is $\theta \geq$ or $\leq \theta_0$.

   (iii) For testing $H_0 : \mu = \mu_0$ or $\mu_1 = \mu_2$ (σ known), we always use ...... normal test.

   (iv) When $\sigma_1 = \sigma_2 = \sigma$ is not known, we compute its ...... estimate from sample.

   (v) The χ^2 - test is used to test $H_0 : \sigma = \sigma_0$ only in case of a ...... sample.

   (vi) The test of hypothesis regarding equality of standard deviations makes use of ...... statistics.

   (vii) The test of goodness of fit or of independence is always a ...... tailed test.

   (viii) When sample (from normal population) sizes are small and s_1 and s_2 are not known, the sampling distribution of the difference of sample means follows t - distribution under the assumption that ....

   (ix) The existence of a strong linear relationship between two variables implies that the regression coefficient is ....

   (x) Yate's correction for continuity is needed when sample size is ....

31.8 Review Questions

1. Certain motor oil is packed in tins holding 5 litres each. The filling machine can maintain this but with a S.D. of 0.15 litre. Two samples of 36 tins each are taken from the production line. If the sample means are 5.20 and 4.95 litres respectively, can we be 99% sure that the sample have come from a population of 5 liters?

   Hint : Check whether the two sample means lie in the interval $5 \pm \frac{2.58 \times 0.15}{6}$ or not.

2. The Industrial Placement Unit of Unisex Polytechnic believes that the average salary paid to the students during their industrial year is Rs 2,800. A sample of 17 of its own students reveals that their average salary is Rs 2,860 with a S.D. of Rs 105. Does this evidence suggest that the countrywide average salary is higher than Rs 2,800? Let the level of significance be 5%.

   Hint : Use one tailed test.

4. In a survey of buying habits, 400 women shoppers are chosen at random in super market A located in a certain section of the city. Their average weekly food expenditure is Rs 250 with a S.D. of Rs 40. For 400 women shoppers chosen at random in super market B in another section of the city, the average weekly food expenditure is Rs 220 with a S.D. of Rs 55. Test at 1% level of significance whether the average weekly food expenditure of the population of shoppers are equal?

   Hint : Apply two tailed test to test the hypothesis regarding equality of means. Also note that both the samples are large.

Notes

5. Samples of two types of electric bulbs were tested for length of life (in hours) and the following data were obtained :

	Type I	*Type II*
Sample size	9	8
Mean of the sample	1235	1125
S.D. of the sample	30	35

Test, at 5% level of significance, whether the difference in sample means is significant?

Hint : Use small sample test, i.e., t-test.

6. A company selects 9 salesmen at random and their sales figures for the previous month are recorded. These salesmen then undergo a course devised by a business consultant and their sales figures for the following month are compared as shown in the following table. Has the training course caused an improvement in the salesmen's ability? Let the level of significance be 5%.

Previous Month	75	90	94	85	100	90	69	70	64
Following Month	77	101	93	92	105	88	73	76	68

Hint : Use paired t-test to test $H_0 : \mu_1 \geq \mu_2$ against $H_a : \mu_1 < \mu_2$.

7. A trader wants to compare the delivery times for two suppliers A and B. The trader wishes to continue with his current supplier A if his mean delivery time is less than or equal to that of supplier B, otherwise will switch over to B. He has obtained the following two independent samples for the above purpose :

Supplier A : $n_1 = 40$, $\overline{X}_1 = 10$ days , $S_1 = 3$ days

Supplier B : $n_2 = 30$, $\overline{X}_2 = 8$ days , $S_2 = 4$ days.

Answers: Self Assessment

1. (i) precision (ii) one (iii) standard (iv) pooled (v) small (vi) F (vii) one (viii) $\sigma_1 = \sigma_2$ (ix) significant (x) small.

31.9 Further Readings

Books

Sheldon M. Ross, Introduction to Probability Models, Ninth Edition, Elsevier Inc., 2007.

Jan Pukite, Paul Pukite, Modeling for Reliability Analysis, IEEE Press on Engineering of Complex Computing Systems, 1998.

Unit 32: Hypothesis Concerning Standard Deviation

Notes

CONTENTS

Objectives

Introduction

32.1 Test of Hypothesis Concerning Population Standard Deviation ($n \leq 30$)

32.2 Test of Hypothesis Concerning Population Standard Deviation (Large Sample)

32.3 Test of Hypothesis Concerning the Equality of Standard Deviations (Small Samples)

32.4 Test of Hypothesis Concerning Equality of Standard Deviations (large samples)

32.5 Summary

32.6 Keywords

32.7 Self Assessment

32.8 Review Questions

32.9 Further Readings

Objectives

After studying this unit, you will be able to:

- Discuss Hypothesis concerning population standard deviation ($n \leq 30$)
- Describe Hypothesis concerning population for large sample.

Introduction

In last unit you have studied about hypothesis testing. In this unit you will be studying about hypothesis concerning standard deviation.

These tests can be divided into two broad categories depending upon whether the size of the sample is large or small.

32.1 Test of Hypothesis Concerning Population Standard Deviation ($n \leq 30$)

Refer to § 20.4.1, the statistic $\frac{\sum\left(X_i - \bar{X}\right)^2}{\sigma^2}$ *or* $\frac{nS^2}{\sigma^2}$ is a χ^2 - variate with (n - 1) degrees of freedom.

Under $H_0 : \sigma = \sigma_0$ (or $\sigma^2 = \sigma_0^2$), $\frac{nS^2}{\sigma_0^2}$ would be a χ^2 - variate with (n - 1) degrees of freedom.

Notes

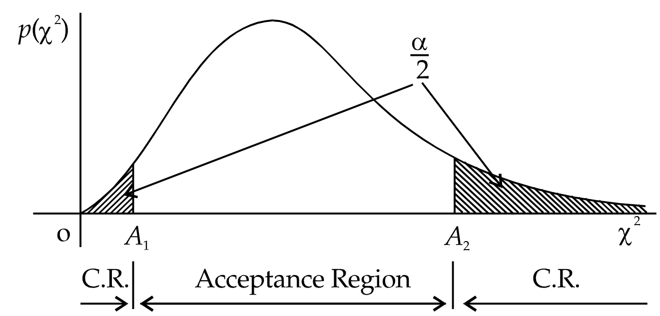
Example 1: A random sample of 20 bulbs from a large lot revealed a standard deviation of 150 hours. Assuming that the life of bulbs follow normal distribution, test the hypothesis that the standard deviation of the population is 130 hours.

Solution.

We have to test $H_0 : \sigma = 130$ against $H_a : \sigma \neq 130$ (two tailed test).

The test statistics, under H_0 is $\chi^2_{cal} = \frac{20 \times 150^2}{130^2} = 26.63$.

Figure 32.1



From the table of χ^2 at 5% level of significance and 19 degrees of freedom, the critical values are $A_1 = 8.91$ and $A_2 = 32.9$. Since χ^2_{cal} lies in the acceptance region, there is no evidence against H_0.

Remarks: To write (1 - *a*)% confidence interval for s^2, we write

$$P(A_1 \leq c^2 \leq A_2) = 1 - a \quad \text{or} \quad P\left(A_1 \leq \frac{nS^2}{\sigma^2} \leq A_2\right) = 1 - \alpha$$

The inequality $A_1 \leq \frac{nS^2}{\sigma^2}$ can be written as $\sigma^2 \leq \frac{nS^2}{A_1}$. Similarly, we can write $\frac{nS^2}{A_2} \leq \sigma^2$. Thus, the (1 - *a*)% confidence interval for s^2 is given by

$$P\left(\frac{nS^2}{A_2} \leq \sigma^2 \leq \frac{nS^2}{A_1}\right) = 1 - \alpha.$$

Example 2: The standard deviation of a random sample of 25 units, taken from a normal population with $s = 8.5$, was calculated to be 10.8. Test the hypothesis that the observed value of standard deviation is significantly higher than the population standard deviation.

Solution.

We have to test $H_0 : \sigma = 8.5$ against $H_a : \sigma > 8.5$. (one tailed test)

The test statistic is $\chi^2_{cal} = \frac{25 \times 10.8^2}{8.5^2} = 40.36.$

χ^2 from tables at 5% level of significance and 24 d.f. is 36.4. Since this value is less than the calculated value, H_0 is rejected. Thus, the observed value of standard deviation is significantly higher than the population standard deviation.

32.2 Test of Hypothesis Concerning Population Standard Deviation (Large Sample)

It can be shown that for large samples (n > 30), the sampling distribution of S is approximately normal with mean s and standard error $\frac{\sigma}{\sqrt{2n}}$. Thus,

$$z = \frac{(S-\sigma)\sqrt{2n}}{\sigma} \sim N(0,1).$$

Alternatively, using Fisher's approximation, we can say that when n > 30, the statistic $\sqrt{2\chi^2}$ follows a normal distribution with mean $\sqrt{2n}$ and standard error unity. Thus $z = \sqrt{2\chi^2} - \sqrt{2n}$ can be taken as standard normal variate for sufficiently large values of n.

Example 3: In a random sample of 300 units, the standard deviation was found to be 8.5. Can it reasonably be regarded as to have come from a population with standard deviation equal to 9.0?

Solution.

We have to test $H_0 : \sigma = 9.0$ against $H_a : \sigma \neq 9.0$ (two tailed test).

It is given that S = 8.5 and n = 300 (large).

Thus, the test statistic is $z_{cal} = \frac{|8.5 - 9.0|\sqrt{600}}{9.0} = 1.36.$

Since this value is less than 1.96, there is no evidence against H_0 at 5% level of significance.

Note: The same value of z is obtained by the use of the statistic $z = \sqrt{2\chi^2} - \sqrt{2n}$.

We can write

$$z_{cal} = \left|\sqrt{\frac{2nS^2}{\sigma^2}} - \sqrt{2n}\right| = \left|\sqrt{\frac{2 \times 300 \times 8.5^2}{9.0^2}} - \sqrt{600}\right| = 1.36$$

If s is unknown it is estimated by S. The 95% confidence limits for s are $S \pm 1.96\frac{S}{\sqrt{2n}}$ or $S\left(1 \pm \frac{1.96}{\sqrt{2n}}\right)$.

Notes

Example 4: The standard deviation of a random sample of size 81 was found to be 12. Test the hypothesis that population standard deviation is greater than 10.

Solution.

We have to test H_0 : *s* £ 10 against H_a : *s* > 10.

$$z = \frac{(12-10)}{10}\sqrt{2 \times 81} = 2.55.$$

Since this value is greater than 1.645, H_0 is rejected. Hence, the sample information supports the contention that *s* is greater than 10.

32.3 Test of Hypothesis Concerning the Equality of Standard Deviations (Small Samples)

We have to test $H_0 : \sigma_1 = \sigma_2$ against $\sigma_1 > \sigma_2$. Refer to § 20.6, the statistic $F = \frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2}$, would become $\frac{s_1^2}{s_2^2}$ under H_0, follows F - distribution with n_1 (= n_1 - 1) and n_2 (= n_2 - 1) degrees of freedom.

Remarks:

1. We can write $s_1^2 = \frac{1}{n_1 - 1}\sum\left(X_{1i} - \bar{X}_1\right)^2 = \frac{n_1}{n_1 - 1}S_1^2 = \frac{1}{n_1 - 1}\left(\sum X_{1i}^2 - \frac{\sum X_{1i}^2}{n_1}\right)$ and

$$s_2^2 = \frac{1}{n_2 - 1}\sum\left(X_{2i} - \bar{X}_2\right)^2 = \frac{n_2}{n_2 - 1}S_2^2 = \frac{1}{n_2 - 1}\left(\sum X_{2i}^2 - \frac{\sum X_{2i}^2}{n_2}\right).$$

2. In the variance ratio $F = \frac{s_1^2}{s_2^2}$, we take, by convention the largest of the two sample variance as σ_1^2. Thus, this test is always a one tailed test with critical region at the right hand tail of the F - curve.

3. The 100(1 - *a*)% confidence limits for the variance ratio $\frac{\sigma_1^2}{\sigma_2^2}$, are given by

$$P\left[\frac{s_1^2}{s_2^2}\cdot\frac{1}{F_{\alpha/2}} \le \frac{\sigma_1^2}{\sigma_2^2} \le \frac{s_1^2}{s_2^2}\cdot\frac{1}{F_{1-\alpha/2}}\right] = 1 - \alpha.$$

Example 5: Two independent samples of sizes 10 and 12 from two normal populations have their mean square deviations about their respective means equal to 12.8 and 15.2 respectively. Test the equality of variances of the two populations.

Solution.

Notes

We have to test $H_0 : \sigma_1 = \sigma_2$ against $\sigma_1 > \sigma_2$.

It is given that $S_1^2 = 15.2$, $S_2^2 = 12.8$, $n_1 = 12$ and $n_2 = 10$.

The unbiased estimates of respective population variances are

$$s_1^2 = \frac{12}{11} \times 15.2 = 16.58 \ \ and \ \ s_2^2 = \frac{10}{9} \times 12.8 = 14.22.$$

Thus, $F_{cal} = \dfrac{16.58}{14.22} = 1.166.$

The value of F from tables at 5% level of significance with 11 and 9 d.f. is 3.10. Since this value is greater than F_{cal}, there is no evidence against H_0.

Example 6: The increase in weight (in 100 gms) due to food A and food B given to two independent samples of children was recorded as follows. Test whether (i) mean weights and (ii) standard deviations of the two samples are equal.

Sample I : 6, 12, 10, 14, 12, 12, 10, 7, 5, 7.

Sample II : 9, 11, 8, 5, 6, 12, 7, 13, 10.

Solution.

We shall first test $H_0 : \sigma_1 = \sigma_2$ against $\sigma_1 > \sigma_2$.

The means of the samples are $\bar{X}_1 = \dfrac{95}{10} = 9.5 \ and \ \bar{X}_2 = \dfrac{81}{9} = 9.0$, respectively.

We can write $s_k^2 = \dfrac{n_k}{n_k - 1}\left(\dfrac{\sum X_{ki}^2}{n_k} - \bar{X}_k^2\right) = \dfrac{\sum X_{ki}^2}{n_k - 1} - \dfrac{n_k}{n_k - 1}\bar{X}_k^2$ (k = 1, 2)

Thus, we have $s_1^2 = \dfrac{987}{9} - \dfrac{10}{9} \times 9.5^2 = 9.39$ and $s_2^2 = \dfrac{789}{8} - \dfrac{9}{8} \times 9^2 = 7.50.$

Further, the test statistic is $F = \dfrac{9.39}{7.50} = 1.25.$

The critical value of F at 5% level of significance and (9,8) d.f. is 3.39, therefore, there is no evidence against H_0. Hence, s_1 and s_2 may be treated as equal.

To test $H_0 : \mu_1 = \mu_2$ against Ha ; $\mu_1 \neq \mu_2$, we note that samples are small, t-test is to be used. Since $\sigma_1 = \sigma_2 = s$ (say), its unbiased estimate is

$$s = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} = \sqrt{\frac{9 \times 9.39 + 8 \times 7.50}{10 + 9 - 2}} = 2.92.$$

The test statistic is $t_{cal} = \dfrac{\left|\bar{X}_1 - \bar{X}_2\right|}{s}\sqrt{\dfrac{n_1 n_2}{n_1 + n_2}} = \dfrac{|9.5 - 9.0|}{2.92}\sqrt{\dfrac{10 \times 9}{10 + 9}} = 0.37.$

Notes

The critical value of t at 5% level of significance and 17 d.f. is 2.11. Since this value is greater than the calculated, there is no evidence against H_0. Thus, we conclude that the two samples may be regarded to have drawn from a population with same means and same standard deviations.

32.4 Test of Hypothesis Concerning Equality of Standard Deviations (Large Samples)

It can be shown that when sample sizes are large, i.e., $n_1, n_2 > 30$, the sampling distribution of the statistic $S_1 - S_2$ is approximately normal with mean $s_1 - s_2$ and standard error $\sqrt{\frac{\sigma_1^2}{2n_1} + \frac{\sigma_2^2}{2n_2}}$.

Therefore $$z = \frac{(S_1 - S_2) - (\sigma_1 - \sigma_2)}{\sqrt{\frac{\sigma_1^2}{2n_1} + \frac{\sigma_2^2}{2n_2}}} \sim N(0,1)$$

or $$z = \frac{S_1 - S_2}{\sigma\sqrt{\frac{1}{2n_1} + \frac{1}{2n_2}}}$$ under $H_0 : \sigma_1 = \sigma_2 = \sigma$.

Very often σ is not known and is estimated on the basis of sample. The pooled estimate of σ is $S = \frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2}$. Thus, the test statistic becomes

$$z_{cal} = \frac{S_1 - S_2}{S\sqrt{\frac{1}{2n_1} + \frac{1}{2n_2}}} = \frac{S_1 - S_2}{S} \times \sqrt{\frac{2n_1 n_2}{n_1 + n_2}}.$$

Example 7: The standard deviation of a random sample of the heights of 500 individuals from country A was found to be 2.58 inches and that of 600 individuals from country B was found to be 2.35 inches. Do the data indicate that the standard deviation of heights in country A is greater than that in country B?

Solution.

We have to test $H_0 : \sigma_1 = \sigma_2$ against $H_a : \sigma_1 > \sigma_2$.

It is given that $S_1 = 2.58$, $n_1 = 500$, $S_2 = 2.35$ and $n_2 = 600$.

The pooled estimate of s is $$S = \sqrt{\frac{500 \times 2.58^2 + 600 \times 2.35^2}{1100}} = 2.46$$

The test statistic is $$z_{cal} = \frac{2.58 - 2.35}{2.46} \times \sqrt{\frac{600000}{1100}} = 2.17$$

Since this value is greater than 1.645, H_0 is rejected at 5% level of significance. Thus, the sample evidence indicates that the standard deviation of heights in country A is greater.

32.5 Summary

Notes

- We can write $s_1^2 = \frac{1}{n_1 - 1}\sum\left(X_{1i} - \bar{X}_1\right)^2 = \frac{n_1}{n_1 - 1}S_1^2 = \frac{1}{n_1 - 1}\left(\sum X_{1i}^2 - \frac{\sum X_{1i}^2}{n_1}\right)$ and

  $$s_2^2 = \frac{1}{n_2 - 1}\sum\left(X_{2i} - \bar{X}_2\right)^2 = \frac{n_2}{n_2 - 1}S_2^2 = \frac{1}{n_2 - 1}\left(\sum X_{2i}^2 - \frac{\sum X_{2i}^2}{n_2}\right).$$

- In the variance ratio $F = \frac{s_1^2}{s_2^2}$, we take, by convention the largest of the two sample variance as σ_1^2. Thus, this test is always a one tailed test with critical region at the right hand tail of the F - curve.

- The 100(1 - a)% confidence limits for the variance ratio $\frac{\sigma_1^2}{\sigma_2^2}$, are given by

  $$P\left[\frac{s_1^2}{s_2^2}\cdot\frac{1}{F_{\alpha/2}} \le \frac{\sigma_1^2}{\sigma_2^2} \le \frac{s_1^2}{s_2^2}\cdot\frac{1}{F_{1-\alpha/2}}\right] = 1 - \alpha.$$

32.6 Keywords

F - distribution: If $H_0 : \sigma_1 = \sigma_2$ against $\sigma_1 > \sigma_2$. Refer to § 20.6, the statistic $F = \frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2}$, would become $\frac{s_1^2}{s_2^2}$ under H_0, follows F - distribution with n_1 (= n_1 - 1) and n_2 (= n_2 - 1) degrees of freedom.

32.7 Self Assessment

Fill in the blanks:

1. These tests can be divided into two broad categories depending upon whether the .... of the sample is large or small.

2. If $H_0 : \sigma_1 = \sigma_2$ against $\sigma_1 > \sigma_2$. Refer to § 20.6, the statistic $F = \frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2}$, would become $\frac{s_1^2}{s_2^2}$ under H_0, follows .... with n_1 (= n_1 - 1) and n_2 (= n_2 - 1) degrees of freedom.

3. In the .... $F = \frac{s_1^2}{s_2^2}$, we take, by convention the largest of the two sample variance as σ_1^2. Thus, this test is always a one tailed test with critical region at the right hand tail of the F - curve.

Notes

4. The 100(1 - a)% .... for the variance ratio $\frac{\sigma_1^2}{\sigma_2^2}$, are given by

$$P\left[\frac{s_1^2}{s_2^2}\cdot\frac{1}{F_{\alpha/2}}\le\frac{\sigma_1^2}{\sigma_2^2}\le\frac{s_1^2}{s_2^2}\cdot\frac{1}{F_{1-\alpha/2}}\right]=1-\alpha.$$

32.8 Review Questions

1. Test the hypothesis that σ = 8, given that S = 10 for a random sample of size 51. Also construct 95% confidence interval for σ.

   Hint : Use a two tailed normal test.

2. A random sample of size 10 from a normal population gave the following observations : 169, 173, 171, 177, 161, 163, 174, 168, 172, 165.

   Test the hypothesis that population variance is 25.

   Hint : Use a two tailed χ^2 test.

3. The following two samples are drawn from two normal populations. Test at 5% level of significance whether their variance can be regarded as equal?

   Sample I : 60, 65, 71, 74, 76, 82, 85, 57.

   Sample II : 61, 66, 67, 85, 78, 63, 85, 86, 88, 91.

   Hint : Use F - test.

4. Can the following two samples obtained from two normal populations, be regarded to have same variances?

Sample No.	Sample Size	Sample Variance
1	15	20
2	25	35

   Test at 10% level of significance.

   Hint : Use F - test.

5. Two independent random samples, one of 12 observations with mean 15 and sum of squares of deviations from mean equal to 135 and another of 16 observations with mean 22 and sum of squares of deviations from mean equal to 250, were obtained from two normal populations. Test at 5% level of significance whether the two samples can be regarded to have come form the same population?

   Hint : Test $\sigma_1 = \sigma_2$ and $\mu_1 = \mu_2$ an in example 34.

6. The following figures relate to the number of units produced per shift by two workers A and B for a number of days:

   A : 19, 22, 24, 27, 24, 18, 20, 19 and 25.

   B : 26, 37, 40, 35, 30, 30, 40, 26, 30, 35 and 45.

   Can it be inferred that A is more stable worker compared to B? Answer using 5% level of significance.

   Hint : Use F - test.

Notes

7. In one sample of 10 observations from a normal population, the sum of squares of the deviations of sample values from their mean is 100.4 and in another sample of 12 observations from another normal population, the sum of squares of the deviations of sample values from their mean is 115.5. Test at 5% level whether the two normal populations have the same variance?

   Hint : Use F - test.

8. In a test given to two groups of students, the marks obtained were as follows:

   Group A : 18, 20, 36, 50, 49, 36, 34, 49, 41.

   Group B : 29, 28, 26, 35, 30, 44, 46.

   Assuming that the marks obtained follows normal distribution, examine at 5% level of significance whether the two groups of students can be regarded to have come from populations with same standard deviation?

   Hint : Use F - test.

Answers: Self Assessment

1. size   2. F - distribution   3. variance ratio   4. confidence limits

32.9 Further Readings

Books

Sheldon M. Ross, Introduction to Probability Models, Ninth Edition, Elsevier Inc., 2007.

Jan Pukite, Paul Pukite, Modeling for Reliability Analysis, IEEE Press on Engineering of Complex Computing Systems, 1998.

LOVELY PROFESSIONAL UNIVERSITY
Jalandhar-Delhi G.T. Road (NH-1)
Phagwara, Punjab (India)-144411
For Enquiry: +91-1824-300360
Fax.: +91-1824-506111
Email: odl@lpu.co.in