STA 217

Probability 2

FEDERAL UNIVERSITY OF TECHNOLOGY, MINNA

CENTRE FOR OPEN DISTANCE AND e-LEARNING

FEDERAL UNIVERSITY OF TECHNOLOGY, MINNA
NIGER STATE, NIGERIA

CENTRE FOR OPEN DISTANCE AND e-LEARNING (CODeL)

B.TECH. COMPUTER SCIENCE PROGRAMME

COURSE TITLE

PROBABILITY II

COURSE CODE

STA 217

COURSE CODE

STA 217

COURSE UNIT

3

Course Coordinator

Bashir Mohammad (Ph.D.)
Computer Science Department
Federal University of Technology, (FUT) Minna
Minna, Niger State, Nigeria.

Course Development Team

STA 217: PROBABILITY II

Course Coordinator	L. A. NAFIU (Ph.D.) Y. YAKUBU Federal University of Technology, Minna, Nigeria
Subject Matter Experts	U. Y. ABUBAKER (Ph.D.) A. ISAH (Ph.D.)
ODL Experts	Amosa Isiaka GAMBARI (Ph.D.) Nicholas E. ESEZOBOR
Instructional System Designers	FALODE, Oluwole Caleb (Ph.D.) Bushrah Temitope OJOYE (Mrs.)
Language Editors	Chinenye Priscilla UZOCHUKWU Mubarak Jamiu ALABEDE
Centre Director	Abiodun Musa AIBINU (Ph.D.) Centre for Open Distance & e-Learning FUT Minna, Nigeria.

STA 217 Study Guide

Introduction

STA217: Probability II is a 3-credit unit course for students studying towards acquiring a Bachelor of Technology in Mathematics and Statistics and other related disciplines. The course is divided into 4 modules and 13 study units. It will first take a brief review of the concepts of Probality. This course will then go ahead to deal with different types of probability distributions.

The course guide therefore gives you an overview of what the course; STA217 is all about, the textbooks and other materials to be referenced, what you expect to know in each unit, and how to work through the course material.

Recommended Study Time

This course is a 3-credit unit course having 15study units. You are therefore enjoined to spend at least 3 hours in studying the content of each study unit.

What You Are About to Learn in This Course

The overall aim of this course, STA217 is to introduce you to the basic concepts of Probbaility and its distributions to enable students understand the basics in the application in the society.

Course Aim

Aim of this course is to introduce students to the basics and concepts of probablity. It is believed the knowledge will enable the student to understand various applications of probablities.

Course Objectives

It is important to note that each unit has specific objectives. Students should study them carefully before proceeding to subsequent units. Therefore, it may be useful to refer to these objectives in the course of your study of the unit to assess your progress. You should always look at the unit objectives after completing a unit. In this way, you can be sure that you have done what is required of you by the end of the unit. However, below are overall objectives of this course. On completing this course, you should be able to:

i. understand combinatorial techniques
ii. state and prove axioms of probability
iii. comprehend different types of distribution of probabilities
iv. show case some practical issues of probablities

Working Through This Course

To complete this course, you are required to study all the units, the recommended textbooks, and other relevant materials. Each unit contains some self assessment exercises and tutor marked assignments, and at some point, in this course, you are required to submit the tutor marked assignments. There is also a final examination at the end of this course. Stated below are the components of this course and what you have to do.

Course Materials

The major components of the course are:

i. Course Guide
ii. Study Units
iii. Text Books
iv. Assignment File
v. Presentation Schedule

Study Units

There are 15 study units and 4 modules in this course. They are:

Module One	Unit 1: Introduction to Probability Unit 2: Statistical Experiments and Events Unit 3: Concept of probability Unit 4: Conditional Probability and Statistical Independence
Module Two	Unit 1: Combinatorics Unit 2: Random Variables and their Distributions Unit 3: Expected Value and Variance
Module Three	Unit 1: Discrete Probability Distribution Unit 2: Continuous Density Functions Unit 3: The Normal Distribution Unit 4: Joint Probability Functions
Module Four	Unit 1: Law of Large Numbers Unit 2: Introduction to the Central Limit Theorem Unit 3: Introduction to Statistical Inference Unit 4: Moments and Moment Generating Functions

Recommended Texts

These texts and especially the internet resource links will be of enormous benefit to you in learning this course:

http://www.stats.gla.ac.uk/steps/glossary/probability.html#baythm
http://www.stanford.edu/class/polisci100a/newprob2.pdf
http://en.wikipedia.org/wiki/Random_variable
http://people.stern.nyu.edu/wgreene/Statistics/DiscreteRandomVariablesCollection.pdf

http://www.stats.gla.ac.uk/steps/glossary/probability_distributions.html
http://faculty.palomar.edu/mmumford/120/notes/Chap5.pdf
http://stattrek.com/probability-distributions/negative-binomial.aspx?tutorial=stat
8.http://www.google.com.ng/url?sa=t&rct=j&q=random%20variables%20and%20probability%20distributions&source=web&cd=5&cad=rja&ved=0CFEQFjAE&url=http%3A%2F%2Fwww.stat.ufl.edu%2F~winner%2Fsta4321%2Fchapter3.ppt&ei=Y3M0UPGgIMLX0QX92YCwBg&usg=AFQjCNGbkSpHwrcWrGic7H2_rfk5ZXKLcg
Harry Frank and Steven C. Althoen (1995): Statistics: concepts and applications. Cambridge Edition.
W. J. Decoursey (2003): Statistics and probability for engineering applications. Newness.
Sheldon M. Ross (1997): Introduction to probability models, sixth edition. Academic Press. New York
Mario Lefebre (2000): Applied probability and statistics. Springer
Nitis Nikhopadhyay (2005): Probability and Statistical Inference.
Deborah Rumsey (2007): Intermediate statistics for dummies. Wiley publishing Inc.
Michael W. Trosset (2005): An introduction to statistical inference and its applications. Wiley.
James E. Gentle (2007): Matrix Algebra: Theory, computations and applications in statistics. Springer.
John Geweke (2003): Contemporary Bayesian Econometrics and Statistics.
Douglas C. Montgomery (2005): Design and Analysis of experiments. 3rd edition.
Jun Shao (1998**): M**athematical Statistics. Springer Texts in Statistics. Springer – Verlag, New York, Inc.
Wolter, K. M. (1985): Introduction to variance estimation. New York. Springer-Verlag.

Assignment File

The assignment file will be given to you in due course. In this file, you will find all the detailsof the work, you must submit to your tutor for marking. The marks you obtain for these assignments will count towards the final mark for the course. Altogether, there are tutor marked assignments for this course.

Presentation Schedule

The presentation schedule included in this course guide provides you with important dates for completion of each tutor marked assignment. You should therefore endeavour to meet the deadlines.

Assessment

There are two aspects to the assessment of this course. First, there are tutor marked assignments; and second, the written examination. Therefore, you are expected to take note of the facts, information and problem solving gathered during the course. The tutor marked assignments must be submitted to your tutor for formal assessment, in accordance to the deadline given. The work submitted will count for 40% of your total course mark.

At the end of the course, you will need to sit for a final written examination. This examination will account for 60% of your total score.

Tutor Marked Assignment (TMA)

There are TMAs in this course. You need to submit all the TMAs. The best 10 will therefore be counted. When you have completed each assignment, send them to your tutor as soon as possible and make certain that it gets to your tutor on or before the stipulated deadline. If for any reason, you cannot complete your assignment on time, contact your tutor before the assignment is due to discuss the possibility of extension. Extension will not be granted after the deadline, unless on extraordinary cases.

Final Examination and Grading

The final examination for STA217 will last for a period of 2 hours and have a value of 60% of the total course grade. The examination will consist of questions which reflect the Self Assessment Questions and tutor marked assignments that you have previously encountered. Furthermore, all areas of the course will be examined. It would be better to use the time between finishing the last unit and sitting for the examination, to revise the entire course. You might find it useful to review your TMAs and comment on them before the examination. The final examination covers information from all parts of the course.

Pratical Strategies for Working Through This Course

Read the course guide thoroughly

Organize a study schedule. Refer to the course overview for more details. Note the time you are expected to spend on each unit and how the assignment relates to the units. Important details, e.g. details of your tutorials and the date of the first day of the semester are available. You need to gather together all this information in one place such as a diary, a wall chart calendar or an organizer. Whatever method you choose, you should decide on and write in your own dates for working on each unit.

Once you have created your own study schedule, do everything you can to stick to it. The major reason that students fail is that they get behind with their course works. If you get into difficulties with your schedule, please let your tutor know before it is too late for help.

Turn to Unit 1 and read the introduction and the objectives for the unit.

Assemble the study materials. Information about what you need for a unit is given in the table of content at the beginning of each unit. You will almost always need both the study unit you are working on and one of the materials recommended for further readings, on your desk at the same time.

Work through the unit, the content of the unit itself has been arranged to provide a sequence for you to follow. As you work through the unit, you will be encouraged to read from your set books

Keep in mind that you will learn a lot by doing all your assignments carefully. They have been designed to help you meet the objectives of the course and will help you pass the examination.

Review the objectives of each study unit to confirm that you have achieved them. If you are not certain about any of the objectives, review the study material and consult your tutor.

When you are confident that you have achieved a unit's objectives, you can start on the next unit. Proceed unit by unit through the course and try to pace your study so that you can keep yourself on schedule.

When you have submitted an assignment to your tutor for marking, do not wait for its return before starting on the next unit. Keep to your schedule. When the assignment is returned, pay particular attention to your tutor's comments, both on the tutor marked assignment form and also written on the assignment. Consult you tutor as soon as possible if you have any questions or problems.

After completing the last unit, review the course and prepare yourself for the final examination. Check that you have achieved the unit objectives (listed at the beginning of each unit) and the course objectives (listed in this course guide).

Tutors and Tutorials

There are 8 hours of tutorial provided in support of this course. You will be notified of the dates, time and location together with the name and phone number of your tutor as soon as you are allocated a tutorial group. Your tutor will mark and comment on your assignments, keep a close watch on your progress and on any difficulties, you might encounter and provide assistance to you during the course. You must mail your tutor marked assignment to your tutor well before the due date. At least two working days are required for this purpose. They will be marked by your tutor and returned to you as soon as possible.

Do not hesitate to contact your tutor by telephone, e-mail or discussion board if you need help. The following might be circumstances in which you would find help necessary: contact your tutor if:

i. You do not understand any part of the study units or the assigned readings.
ii. You have difficulty with the self test or exercise.
iii. You have questions or problems with an assignment, with your tutor's comments on an assignment or with the grading of an assignment.

You should endeavour to attend the tutorials. This is the only opportunity to have face to face contact with your tutor and ask questions which are answered instantly. You can raise any problem encountered in the course of your study. To gain the maximum benefit from the course tutorials, have some questions handy before attending them. You will learn a lot from participating actively in discussions.

GOODLUCK!

Table of Contents

Module 1

Introduction to Probability

Unit 1

Introduction to Probability

Content

1.0 Introduction

The world is replete with prediction of likely of events happening at some time. For instance, meteorologists seek to predict what the weather of a particular place or location will be in the next few minutes or days.

Looking at the topic from another perspective, if you roll a pair of dice, what are the chances of rolling a double six? What is the likelihood of winning a state lottery? The subject of probability was invented to give precise answers to questions like these. In other word, Probability is the mathematical study of chance and random processes. It is now an indispensable tool for making decisions in virtually all areas of human endeavours.

To recap, this unit will introduce you to the basic idea portrayed by probability as a subject. The subsequent units will give more insights into what statistical events are, and how they are generated from set theory and combinational method.

2.0 Learning Outcome

At the end of this unit, you should be able to:

i. give your own understanding of the term ‘probability’
ii. mention various areas of human activities where probability find applications
iii. understand how important the concept of probability is in decision-making processes.

3.0 Learning Outcomes

3.1 Explanation of the Term ‘Probability’

The mathematical theory of probability was first discussed in 1654 by Pascal and Fermat. The subject was invented to give precise answers to questions of likelihood or otherwise of events to occur. For the purpose of simplicity, probability is a measure of the ratio of an item of interest to the total items (outcomes) under consideration. Mathematically, probability is defined as follows:

Let *S* be the sample space of an experiment in which all outcomes are equally likely and let *E* be an event. The probability of *E*, written as *P(E*), is

$$P(E) = \frac{\text{Number of elements in } E}{\text{Number of elements in } S}$$

Notice that $0 \leq n \leq (E) \leq n(S)$, so the probability *P(E*) of an event is a number between 0 and 1, that is,

$$0 \leq P(E) \leq 1.$$

Probability Concepts

We distinguish the following 3 concepts of probability – Equally likelihood or Classical or Apriori concept, Relative frequency concept and the Axiomatic concept.

Equally likelihood

If any experiment can result in Ns equally likely ways (when they have the same chance or same probability) and mutually exclusive number of ways out of which a particular event; say, A say N_A of these associated with it, then we define the probability of the event A as follows:

$$P(A) = \frac{\text{No of sample points in A}}{\text{The total no of sample points in the experiment}} = \frac{n_A}{n_S}$$

e.g. if a fair and honest coin is thrown 3 times, then the possible outcomes,

S, are = {HHH, HHT, HTH, TTT, TTH, THT, THH} i.e. Ns =8

Let {A exactly 2 heads} i.e. A = {HHT, HTH, THH}, $N_A = 3$

=> p (exactly 2 heads), = N_A/N_S ,= 3/8.

It therefore follows that the probability of none occurrence of A

$$p(\text{not } A) = \frac{NS-NA}{NS} = 1-\frac{NA}{NS} \quad 1 - p(A)$$

Self Assessment Exercises

1. What are the three (3) Probability Concepts

Self Assessment Answers

i. Equally likelihood or Classical or Apriori concept,
ii. Relative frequency concept and
iii. the Axiomatic concept.

3.2 Applications of Probability

The concept of probability finds applications in diverse areas such as business, manufacturing, psychology, genetics, and in many sub-disciplines of science. Probability is used to determine the effectiveness of new medicines, assess a fair price for an insurance policy, decide on the likelihood of a candidate winning an election, determine the opinion of many people on a certain topic (without interviewing everyone), and answer many other questions that involve a measure of uncertainty. It is also an indispensable tool in decision-making processes both in private organizations and government parastatal.

4.0 Conclusion

This module took you through the introduction to the subject of probability and how it may be defined. It went further to look at various areas where probability finds applications.

5.0 Summary

Probability has been described as the mathematical study of chance and random processes. The laws of probability are essential for understanding genetics, opinion polls, pricing stock options, setting odds in horseracing and games of chance, and many other fields.

6.0 Tutor-Marked Assignment (TMA)

1. (a) Explain what you understand by the term ‘probability’.

   (b) Itemize the various areas where the concept of probability finds applications.

2. In what way(s) do you think the concept of probability is relevant to managerial level of private organizations and government parastatals?

7.0 References

Harry Frank and Steven C. Althoen (1995): Statistics: Concepts and Applications. Cambridge Edition.

W. J. Decoursey (2003): Statistics and Probability for Engineering Applications. Newness.

Unit 2

Statistical Experiments and Events

Content

1.0 Introduction

This unit will introduce you to the methods of performing statistical experiments, events determination, representation of events with Venn diagram as well as how to specify mutually exclusive and independent events.

Though events can be described in terms of other events through the use of the standard set theory, this would be discussed in the subsequent units together with the corresponding Venn diagrams that illustrate them.

2.0 Learning Outcome

At the end of this unit, you should be able to:

i. Perform statistical experiments
ii. Determine simple and compound events
iii. Represent events using Venn diagrams
iv. Specify mutually exclusive and independent events

3.0 Learning Outcomes

3.1 Statistical Experiment

A**Statistical Experiment**is any operation that is performed according to a well-defined set of rules and which when repeated generates a number of outcomes that cannot be predetermined. Some examples of simple statistical experiments include a toss of two coins at a time, choosing 100 people from a large population and testing them for a certain characteristic of interest, etc.

A single performance of an experiment is called a ***trial***. For instance, a toss of a coin once, selecting n items from a large population of items, etc.

The possible result of a trial is called an ***outcome***. For instance, in a toss of two coins, an outcome could be any one of

$$HH\ HT\ TH\ TT$$

The set of all possible outcomes of one trial of an experiment is called a **SampleSpace**, denoted by S . For instance, in a toss of two coins, the sample space is

$$S = HH\ HT\ TH\ TT$$

If we roll a six-sided die, and the number that shows on top is observed, a sample space of the experiment is the 6-element set

$$S = \{1, 2, 3, 4, 5, 6\}$$

Certain specified sets of outcomes (subsets) are called ***events***. For example, for the above sample space, we can define A to be the event that the result of the roll is a number divisible by 3 and B to be the event that the result of the roll is an even number, then

$$A = \{3, 6\}$$

$$B = (2, 4, 6\}.$$

Self Assessment Exercises

1. What is a statistical experiment?
2. What is a trial?
3. An outcome is ………………………………………………………………………..
4. The set of all possible outcomes in one trial of an experiment is called a …………...
5. Subsets or sets of outcomes are called ……………………………………….

Self Assessment Answers

3.2 Simple and Compound Events

An event is called *simple* if it cannot be decomposed any further. That is, if it is a set containing only one element of the sample space. But if it can be decomposed into a set of simple events (i.e., can be expressed as the union of simple events), then it is called a *compoundevent*. For example, the occurrence of 6 in a throw of a die is a simple event. The occurrence of 7 in a simultaneous throw of two dice is a compound event because it can be split up in to six simple events each of which corresponds to the same compound event, namely, 6 and 1, 5 and 2, 4 and 3, 3 and 4, 2 and 5,

1 and 6.

An event is said to have occurred if any one of its elements is the outcome of the experiment.

In many cases, events can be described in terms of other events through the use of the standard constructions of set theory. The definitions of these constructions are reviewed in the following section together with the corresponding Venn diagrams that illustrate them.

Self Assessment Exercises

1. An event that cannot be decomposed further is ………………………………
2. An event that can be expressed as the union of simple events is called ……

3.3 Complements, Union and Intersection of Events

Every event is characterized by a property, which is possessed by each of the elements in the corresponding sets of the events. For instance, for the sample space of rolling a die once, let an event A be defined as

A: *An even number is observed.*

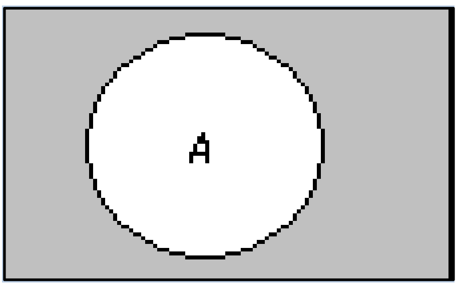
Then, we can consider another event for A where such property does not hold. The subset corresponding to this event is the set of all elements in the sample space not belonging to A. Such a set is called the *complement* of A, denoted by A'.

Definition

The complement, A', of an event A with respect to a given sample space S, is the subset which consists of all the individual outcomes of S that are not contained in A. This is given by the set

$$A' = \{x \mid x \in S \text{ and } x \notin A\}$$

The construction is illustrated by the Venn diagram below with the shaded area representing A'.



A'

Figure 1.1: The Complement of an Event A.

Example 1.01Suppose an integer is chosen from the set of first 12 positive integers. The sample space of this experiment is the set

$$S = \{1, 2, 3, \ldots, 12\}.$$

If the events below are defined on S

A: *The result of selection is an odd integer*

B : *The result of selection is an integer divisible by 4.*

We have

$$A = \{1,3,5,7,9,11\}$$
$$B = \{4,8,12\}$$

Then, the complement of A is given by

$$A' = \{2,4,6,8,10,12\}.$$

And this is the event of choosing an integer that is not odd.

Similarly, the complement of B is given by

$$B' = \{1,2,3,5,6,7,9,10,11\},$$

This is the event of choosing an integer that is not divisible by 4.

The complement of the sample space S is the *impossible* event (empty set), $\varnothing$ which is an event consisting of none of the experimental outcomes.

In application of probability theory, we are often concerned with several related events rather than with just one event. If A and B are any two events with corresponding properties e_1, e_2 (e.g., *even* and *divisible by 3*), then we can define two other natural events as follows.

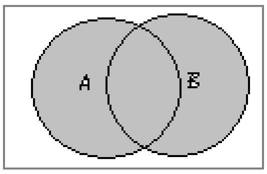
The first is the event that either e_1 or e_2 holds (e.g., a number is either *even* or *divisible by 3*) for which the corresponding set is that of all elements belonging to either A or B. Such event is called the union of the events A and B.

Definition

If A and B are any two events of a sample space S, then their union (written as $A \cup B$) is defined as the event which consists of all the individual outcomes contained in A or B. This is given by the set

$$A \cup B = \{x | x \in A \text{ or } x \in B\}$$

The Venn diagram is given below with $A \cup B$ shaded.



$A \cup B$

Figure 1.2: The union of two events A and B.

Example 1.02: Consider the above example of rolling a six-sided die for which the sample space is

$$S = \{1, 2, 3, 4, 5, 6\}$$

If A and B are two events defined on S as

A: *an even number shows on top.*

B: *a number divisible by 3 shows on top.*

Then,

$$A = \{2, 4, 6\}.$$
$$B = \{3, 6\}.$$

Therefore, the union of A and B is given by

$$A \cup B = \{2, 3, 4, 6\}.$$

This is the event of having a number showing on top, which is either even or divisible by 3.

Example 1.03: Consider an experiment of tossing a coin three times, for which the sample space is

$$S = HHH\ HHT\ HTH\ HTT\ THH\ THT\ TTH\ TTT\ .$$

If A and B are two events defined on S as

A: *First outcome is a head*

B: *Second outcome is a tail*

Then,

$$A = \{HHH, HHT, HTH, HTT\}$$

$$B = \{HTH, HTT, TTH, TTT\}$$

And the Union of A and B is given as

$$A \cup B = \{HHH, HHT, HTH, HTT, TTH, TTT\}$$

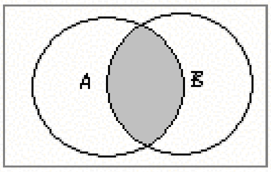
The second one is the event that both e_1 and e_2 hold (e.g., a number is *even* as well as *divisible by 3*) for which the corresponding set consists of all the elements belonging to both A and B. This event is called the intersection of the events A and B.

Definition

If A and B are any two events, then their intersection (written as $A \cap B$), is defined as the event which consists of all the outcomes contained in both A and B. Thus $A \cap B$ is the event that both A and B occur. It is given as

$$A \cap B = \{x | x \in A \text{ and } x \in B\}$$

The Venn diagram below illustrates this with the shaded portion representing $A \cap B$.



$A \cap B$

Figure 1.3: The intersection of two events A and B .

Example 1.04 Consider the experiment of example 1.03 above for which the sample space is the 8-element set $S = \{HHH\ HHT\ HTH\ HTT\ \ THH\ THT\ TTH\ TTT\}$ with the events A and B as defined. Then,

$$A \cap B = \{HTH, HTT\}.$$

Consider the experiment in which a fair die is rolled once. If A is the event that the result of the roll is an even number and B is the event that the result is a number divisible by 3, then

$$A = \{2, 4, 6\}$$

$$B = \{3, 6\}$$

The intersection of A and B is given by

$$A \cap B = \{6\}.$$

This is the event of having a number showing on top, which is even and divisible by 3.

Example 1.05 Suppose two tetrahedral dice are tossed together and the interest is on sum of the numbers that show on top for each die. The sample space of the experiment is the set

	1	2	3	4
1	2	3	4	5
2	3	4	5	6
3	4	5	6	7

If the two events below are defined on S

A: *The sum is even*

B :*A sum greater than 3*

Find

(i) $A \cup B$.

(ii) $A \cap B$.

Solution

From the sample space, we have

$$A = \{2,\ 4,\ 6,\ 8\}$$

$$B = \{4,\ 5,\ 6,\ 7,\ 8\}$$

Therefore,

(i) $A \cup B = \{2,\ 4,\ 5,\ 6,\ 7,\ 8\}$

(ii) $A \cap B = \{4,\ 6,\ 8\}$.

Self Assessment Exercises

1. If *A* is an event, the set of all elements in the sample space not belonging to A is called ...., denoted by ....
2. The complement of the sample space S is called ......represented by ....
3. An event whose outcome can occur in event A or B is written as ....
4. The intersection of any two events, A and B is denoted by ....

Self Assessment Answers

3.4 Mutually Exclusive and Independent Events

Two or more events are said to be mutually exclusive or disjoint if the occurrence of any of them makes the simultaneous occurrence of any other of them impossible. That is

$$A_i \cap A_j = \varnothing, \quad \text{for any } i \text{ and } j \neq i.$$

For instance, if a coin is tossed once, the events *head* and *tail* are mutually exclusive.

If a die is rolled once, the events

A :*five*

B :*six*

are mutually exclusive. However, the events

C :*a prime number*

D :*an even number*

are not mutually exclusive because if the roll resulted in a two, both those events would occur together.

Events are said to be independent if the occurrence or nonoccurrence of any of them has no effect at all on the probability of occurrence of the others. This implies that when two events are independent, then they can occur together simultaneously. For instance, in throwing a coin and a die together once, the events

A : *an even number on the die*

B : *head on coin*

are independent.

Self Assessment Exercises

1. Two or more mutually exclusive or disjoint events can occur simultaneously. True / False.
2. An event(s) whose occurrence is (are) of no effect at all on the probability of occurrence of the others is said to be ....

Self Assessment Answers

4.0 Conclusion

This unit, you were introduced to the basic concepts to aid you in understanding probability theory. We started with statistical experiment methods and its basic concepts, such as events, sample space, finding complements, union and intersection of various events as well as determining when events are mutually exclusive or independent.

Other events specification using the standard set theory would be discussed in the subsequent units together with the corresponding Venn diagrams that illustrate them.

5.0 Summary

In this unit, we have discussed:

i. **Statistical experiment** asany operation that is performed according to a well-defined set of rules and which when repeated generates a number of outcomes that cannot be predetermined.
ii. Simple event is an event that cannot be decomposed any further, but if it can be decomposed into a set of simple events (i.e., can be expressed as the union of simple events), then it is called a Compound event.

iii. The complement, A', of an event A with respect to a given sample space S, is the subset which consists of all the individual outcomes of S that are not contained in A.

iv. If A and B are any two events of a sample space S, then their union ($A \cup B$) is the event which consists of all the individual outcomes contained in A or B. If A and B are any two events, then their intersection ($A \cap B$), is the event which consists of all the outcomes contained in both A and B.

v. Two or more events are said to be mutually exclusive or disjoint if the occurrence of any of them makes the simultaneous occurrence of any other of them impossible. Events are said to be independent if the occurrence or nonoccurrence of any of them has no effect at all on the probability of occurrence of the others.

6.0 Tutor-Marked Assignment (TMA)

1. What do you understand by the following terms?

   (a) Sample space
   (b) Statistical experiment
   (c) Statistical event

2. How are statistical events generated?

3. U = {x: x = 1, 2, 3 . . . 10}

   A = {1, 2, 3 . . . 8}
   B = {2, 3}
   C = {1, 3, 5, 7, 9}

4. Define the following sets: (I) A’, B’, C’ (II) is B’⊂ A’? (III) Are A and C disjoint

If V = {d} W = {c, d} X = {a, b, c} Y = {a, b} Z = {a, b, d}

Determine that each of the following statement is true or false.

i. Y ⊂ X
ii. W ≠ Z
iii. W ⊉ V
iv. X ⊃ V
v. Y ⊄ X
vi. V ⊂ X
vii. X = W

7.0 References

Harry Frank and Steven C. Althoen (1995): Statistics: Concepts and Applications. Cambridge Edition.

W. J. Decoursey (2003): Statistics and Probability for Engineering Applications. Newness.

Unit 3

Concept of Probability

Content

1.0 Introduction

Events arising out of an experiment in nature or in a laboratory vary randomly in their frequencies of occurrence. Besides, these events cannot be predicted with certainty and are thought to be subject to "chance". The study of probability helps us figure out the likelihood of such events happening. Probability of an event is a numerical measure of the likelihood that the event will occur when an experiment is performed. It can be expressed as a fraction or a decimal from 0 to 1 with unlikely events having a probability near 0 and those that are likely to happen having probabilities near 1. Two approaches for computing the probability of events are considered here. These are basic interpretations, or ways of thinking about probabilities that help us to assign probabilities to uncertain outcomes.

2.0 Objectives

At the end of this unit, you should be able to:

i. Estimate the possibility of events' occurrence
ii. Proof some probability axioms to assert the chances of events' occurrence
iii. Generalize arbitrary events

3.0 Learning Outcomes

3.1 The Relative Frequency approach

This approach approximates the probability of an event E as the proportion of times that the event occurs when the experiment is performed repeatedly under similar conditions. That is

$$P(E) = \frac{\text{number of times } E \text{ occurs}}{\text{number of times experiment is performed}}.$$

We should note that this approach only gives an approximation to the true probability of E. However, if we perform our experiment more and more times, the relative frequency will eventually approach the actual probability.

Example 2.1 Suppose we roll a six-sided die 50 times and the frequency distribution of the outcomes is given as in the table below.

Score	1	2	3	4	5	6
Frequency	9	10	5	7	6	13

Let E denotes the event "the outcome is even", F the event "a number greater than or equals to 4". Then from the table of frequency distribution of tosses above, we have

$E = \{2,4,6\}$ and $F = \{4,5,6\}$ and by the relative frequency approach, we can estimate the probability of E as

$$P(E) = \frac{(10+7+13)}{50} = 0.6,$$

and that of F as

$$P(F) = \frac{(7+6+13)}{50} = 0.52.$$

The relative frequency is often used as an estimate of probability and the more trials we perform the better the estimate will be.

3.2 The Classical Approach

This approach rests on the assumption that all the possible outcomes of an experiment are equally likely. That is, these possible outcomes have equal probability of occurrence. It states that if a trial of an experiment can result in m mutually exclusive and equally likely outcomes and if exactly n of these outcomes correspond to an event E, then the probability of E is given by

$$P(E) = \frac{\text{number of outcomes favourable to } E}{\text{total possible number of outcomes}} = \frac{n}{m}.$$

For instance, in tossing a die once, the sample space

S = {1 2 3 4 5 6}.

Then by this approach, each of the elements of S has equal probability $1/6$ so that the probability of an event

A: *Score divisible by 3*

is given by

$$P(A) = \frac{n(A)}{n(S)} = \frac{2}{6} = 0.33,$$

and that of an event

B: *an even number*

is

$$P(B) = \frac{n(B)}{n(S)} = \frac{3}{6} = 0.5.$$

Self Assessment Exercises

1. The relative frequency result is better with plenty trials. True or False.
2. Events that have have equal likelihood of occurrence is denoted by …………

Self Assessment Answers

3.3 Probability Axioms

The probabilities assigned to events by a probability distribution function on a sample space S must satisfy the following properties:

1. $0 \le p(E) \le 1 \text{ for every } E \subset S.$

Proof:

For any event E the probability $P(E)$ is determined from the distribution p by

$$P(E) = \sum_{x \in E} p(x),$$

For every $E \subset S$. Since the function p is nonnegative, it follows that $P(E)$ is also nonnegative and this property proved.

(1) $0 \le p(E) \le 1 \text{ for every } E \subset S.$

2. $P(S) = 1.$

Proof:

This Property is proved by the fact that

$$0 \le p(x) \le 1 \text{ for every } x \in S,$$

and the equations

$$P(S) = \sum_{x \in S} p(x) = 1.$$

(1) $0 \le p(E) \le 1 \text{ for every } E \subset S.$

(2) $P(S) = 1.$

3. If A and B are *disjoint* subsets of S, then

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

Proof:

Suppose that A and B are disjoint subsets of S (i.e., A and B have no element in common). Then every element x of $A \cup B$ lies either in A and not in B or in B and not in A. It follows that

$$P(A \cup B) = \sum_{x \in A \cup B} p(x) = \sum_{x \in A} p(x) + \sum_{x \in B} p(x)$$
$$= P(A) + P(B),$$

and this property is proved.

Example 2.2 An experiment consists of rolling a die once. If the two events below are as defined:

A: *a five* is observed

B :*an even score* is observed.

Then A and B are disjoint subsets of S. Then

$A \cup B = \{2, 4, 5, 6\}$,

each of these elements with probability $1/6$. Therefore

$$P(A \cup B) = \sum_{x \in A \cup B} p(x) = \tfrac{4}{6} = \tfrac{2}{3}.$$

Now

$$P(five) + P(evenscore) = \frac{1}{6} + \frac{3}{6} = \frac{2}{3},$$

which verifies that

$P(A \cup B) = P(A) + P(B)$, provided A and B are disjoint.

This property can be extended to more than two events. If S is finite and $A_1, ..., A_n$ are pairwise disjoint events of S (i.e., no two of the A_i's have an element in common), then

$$P(A_1 \cup ... \cup A_n) = \sum_{i=1}^{n} P(A_i).$$

Self Assessment Question(s)

1. Specify the property a probability distribution function (pdf) must satisfy on a sample space S.
2. If two events, A and B are *disjoint* subsets of S, then the probability function is ....

Self Assessment Answer(s)

3.4 Addition Rule for Arbitrary Events

Property 3 can be generalized in another way. Suppose that A and B are any events in S that are not necessarily disjoint, then

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

Proof:

Now, the event

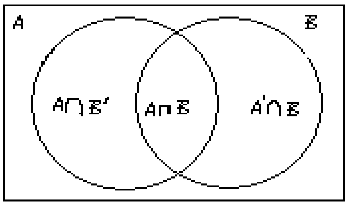
$$A \cup B$$

occurs if and only if the mutually exclusive events

$$A \cap B \text{ or } A \cap B^c \text{ or } A^c \cap B$$

occurs.

This is illustrated in the Venn diagram below



That is,

$$P(A \cup B) = P(A \cap B) + P(A \cap B^c) + P(A^c \cap B) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(1)$$

Now from this diagram and by property 3,

$$P(A) = P(A \cap B) + P(A \cap B^{'}) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(2)$$

$$P(B) = P(A \cap B) + P(A^{'} \cap B) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(3)$$

Adding (2) and (3), we obtain

$$P(A) + P(B) = 2P(A \cap B) + P(A \cap B^{'}) + P(A^{'} \cap B) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(4)$$

By comparing (4) with (1), we have

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

From which we obtain

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

Example 2.3: Let A and B be events such that $P(A \cap B) = 1/4$, $P(A^{'}) = 1/3$, and What is $P(A \cup B)$?

Solution

A and B are not disjoint since $\{A \cap B\} \neq \varnothing$.

Now

$$P(A^{'}) = 1/3 \Rightarrow P(A) = 1 - P(A^{'})$$
$$= 1 - 1/3 = 2/3.$$

by property (4) below.

Therefore

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$
$$= \tfrac{2}{3} + \tfrac{1}{2} - \tfrac{1}{4}$$
$$= \tfrac{7}{6} - \tfrac{1}{4} = \tfrac{11}{12}.$$

Example 2.4: A fair die is rolled once. What is the probability of an *even score* or *a score divisible by 3*?

Solution

The events

A: *an even score*

B: *a score divisible by 3*

are not disjoint events since

$$\{A \cap B\} \neq \varnothing.$$

Their probabilities are

$$P(A) = \frac{3}{6} = \frac{1}{2},$$
$$P(B) = \frac{2}{6} = \frac{1}{3},$$
$$P(A \cap B) = P(6) = \frac{1}{6}.$$

Then the probability of an *even score* or *a score divisible by 3* is

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

$$= \frac{1}{2} + \frac{1}{3} - \frac{1}{6} = \frac{2}{3}.$$

4. $P(A^{'}) = 1 - P(A)$ for every $A \subset S$.

Proof:

To prove this property, we consider the disjoint subsets A and $A^{'}$ whose union is given by $A \cup A^{'} = S$.

Now, by property 3 (property of disjoint additivity),

we have

$P(A) + P(A^{'}) = P(S)$

But by property 2

$P(S) = 1$

Hence

$$P(A) + P(A^{'}) = 1$$

and so

$$P(A^{'}) = 1 - P(A), \quad \text{for every } A \subset S,$$

and this property is proved.

Self Assessment Question(s)

1. A fair die is rolled once. What is the probability of an *even score* or *a score divisible by 3*

Self Assessment Answer(s)

4.0 Conclusion

In this unit, you have learnt how to estimate the possibility of events' occurrence, proof events' property axioms and generalization of arbitrary events. Baye's probability and formula was also discussed.

In the next unit, we shall further our study of probability theory by discussing a more important aspect of it, known as Permutations and Combinations involving determining the number of ways that a certain event can occur.

5.0 Summary

In this unit, we have discussed the following:

i). The Relative Frequency approach for approximating the probability of an event. If the experiment is performed manymore times, the relative frequency will yield actual probability.

ii). Probability Axioms are the properties the probabilities assigned to events by a probability distribution function on a sample space S must satisfy, such as:

(1) $0 \leq p(E) \leq 1$ for every $E \subset S$.

(2) $P(S) = 1$.

(3) If A and B are *disjoint* subsets of S, then

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

iii). Mutual exclusiveness of events i.e the occurrence of one event precludes the occurrence of the others, hence independence and mutual exclusiveness can be thought of as opposites. The conditional probability is a situation where the occurrence of one event does have some effect on the probability of occurrence of another without making that probability zero.

iv). Independent events are those events such that Event B is said to be statistically independent of Event A, if the occurrence of Event A has no effect on the probability of occurrence of B.

6.0 Tutor-Marked Assessment (TMA)

1. Let A and B be events such that $P(A \cap B) = 1/8$, $P(A^{'}) = 1/9$, and $P(B) = 1/4$. What is $P(A \cap B)$?
2. Proof that $0 \leq p(E) \leq 1$ for every $E \subset S$.

7.0 References

W. J. Decoursey (2003): Statistics and probability for engineering applications. Newness.

Harry Frank and Steven C. Althoen (1995): Statistics: Concepts and applications. Cambridge Edition.

Unit 4

Conditional Probability and Statistical Independence

Content

1.0 Introduction

Recall that mutual exclusiveness of events is a way of stating that if one of a set of events occurs, then none of the others can possibly occur; and independent events are thoseevents whose occurrences has no effect (at all) on the probabilities of occurrence of other events. In this unit, you will learn how to vary the probability of events' occurrence and observe their outcomes.

You will learn more about conditional probabilities and assigning distribution function to sample spaces when performing experiments.

2.0 Learning Outcome

At the end of this unt, you should be able to:

i. Find conditional probability of events
ii. Determine statistical Independence of events
iii. Assign distribution function to sample spaces of performed experiments
iv. Altering probabilities of events

3.0 Learning Outcomes

3.1 Conditional Probability and Statistical Independence

As stated in unit 2, mutual exclusiveness of events means that if one of them occurs, then none of the others can possibly occur and when events are independent, then the occurrence of one has no effect at all on the probabilities of occurrence of the others. Therefore, independence and mutual exclusiveness can be thought of as opposites. The conditional probability considers the in- between situations where the occurrence of one event does have some effect on the probability of occurrence of another without making that probability zero, as we shall see in this section.

Suppose an experiment is performed and we assign a distribution function to the sample space S and then learn that an event A has occured. How should we change the probabilities of the remaining events? Consider the examples below.

Example 2.5 An experiment consists of rolling a die once. Let the two events below be defined on S

B : six is observed

A:a number greater than 4 is observed.

Now S consists of 6 elemnts each with probability $1/6$. That is

$p(x)=1/6$ for $x=1,2,...,6$. Thus $P(B)=1/6.$

Now suppose that the die is rolled and we are told that the event A has occurred. Based on this information, we now want to know the new probability of the event B .

We consider A to be the new sample space and note that B is a subset of A because the only outcomes that can have occurred are those in the set A. Hence, the only way B can occur is if we have an outcome that is in both A and B, that is, an outcome in $A \cap B$. The chance of this happening is the chance of an $A \cap B$ outcome as a proportion of all the possible A outcomes which might have occurred. This is denoted by

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

The ratio $P(B \cap A)/P(A)$ when $P(A) \neq 0$ is called the conditional probability of B given A, denoted by $P(B/A)$.

We see that whenever we compute $P(B/A)$ we are essentially computing $P(B)$ with respect to the reduced sample space A, rather than with respect to the original sample space S.

Now

$$A = \{5,6\} \Rightarrow P(A) = 1/3$$

$$B = \{6\} \Rightarrow P(B) = 1/6$$

$$A \cap B = \{6\} \Rightarrow P(A \cap B) = 1/6$$

Therefore we have

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

$$= \frac{1/6}{1/3} = \frac{1}{2}.$$

Example 2.6 Consider an experiment of tossing a fair coin three times. What is the probability of obtaining exactly 2 heads, given that

(a) the first outcome was a head?

(b) the first outcome was a tail?

Solution

The sample space is

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

Now let the event 'getting exactly 2 heads' be denoted by A, the event 'first toss was a head' be denoted by B, and the event 'first toss was a tail' by C. Then we have

$$B = \{HHH \;\; HHT \;\; HTH \;\; HTT\},$$

$$A \cap B = \{HHT \; HTH\}$$

and $P(B) = 4/8 = 1/2$, $P(A \cap B) = 2/8$;

$$C = \{THH\ THT\ TTH\ TTT\},$$

$$A \cap C = \{THH\}$$

and $P(C) = 4/8 = 1/2$, $P(A \cap C) = 1/8$

Therefore

(a) $P(\text{getting exactly 2 heads / first toss was a head}) = P(A/B)$

$$= \frac{P(A \cap B)}{P(B)} = \frac{2/8}{4/8} = \frac{1}{2}.$$

(b) $P(\text{getting exactly 2 heads / first toss was a tail}) = P(A/C)$

$$= \frac{P(A \cap C)}{P(C)} = \frac{1/8}{4/8} = \frac{1}{4}.$$

Example 2.7: A random sample of 150 graduands from a Polytechnic during a session are classified below according to sex and class of diploma obtained.

Sex / Class	Male	Female
Distinction	11	3
Upper credit	22	15
Lower credit	57	42

If a person is picked at random from this group, what is the probability that

i. the person graduated with Upper credit given that the person is a male?
ii. the student did not graduate with Distinction given that the student is a female?

Solution

Let D denotes distinction, U denotes upper credit, and L denotes lower credit, M denotes male and F female. Then we have

Sex / Class	M	F	Total
D	11	3	14
U	22	15	37
L	57	42	99
Total	90	60	**150**

Therefore,

$$(i) \quad P(U/M) = \frac{P(U \cap M)}{P(M)} = \frac{22/150}{90/150} = \frac{11}{45} = 0.24.$$

$$(ii) \qquad P(D'/F) = \frac{P(D' \cap F)}{P(F)} = \frac{57/150}{60/150} = \frac{19}{20} = 0.95.$$

We can also represent the sample space of this experiment as the paths through a tree with the probabilities assigned as shown in the figure below.

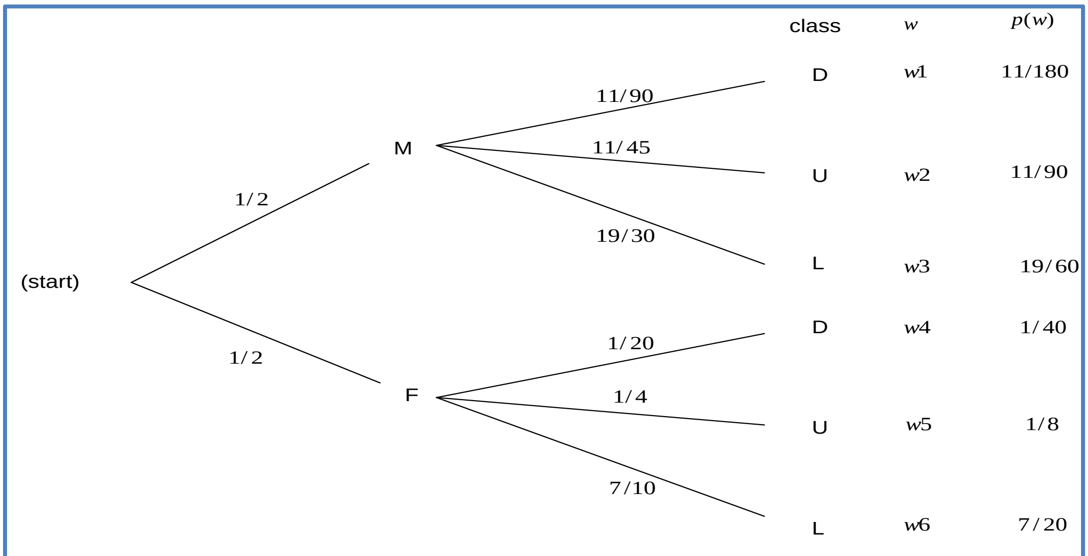


Figure 2.4: Tree diagram

Let U be the event "a student graduated with upper credit", and M the event "a male is chosen." Then the branch weight $11/45$ that is shown on one branch in the figure can now be interpreted as the conditional probability $P(U/M)$.

Self Assessment Exercises

1. A mutually exclusive events means ………………………………………………
2. An independent event means …………………………………………………….

Self Assessment Answers

3.2 Independent Events

Event B is said to be statistically independent of event A if the knowledge that A has occurred has no effect on the probability of occurrence of B. That is, if

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

In this case

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

and one would expect that the equation

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

would also be true, i.e., A is also independent of B.

For example, consider a fair coin tossed twice. Here one would not expect the knowledge of the outcome of the first toss of the coin to have any effect on the probability of the possible outcomes of the second toss. That is, the second toss is independent of the first.

Definition

Let A and B be two events. Then they are said to be independent if either (1) both events have positive probability and

$$P(B/A) = P(B) \text{ and } P(A/B) = P(A),$$

or (2) at least one of the events has probability 0.

From this definition, we see that to verify whether two events are independent, only one of these equations must be checked.

Now assume that A and B are independent.

Then $P(A/B) = P(A)$, and so

$$\begin{aligned} P(A \cap B) &= P(A/B)P(B) \\ &= P(A)P(B). \end{aligned}$$

and if the this equation is true, then it implies that

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

Also,

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

Therefore, A and B are independent.

Example 2.8 A box contains 40 bolts 16 of which are defective and 24 non-defectives. Two bolts are selected in succession with replacement. Find the probability that

(a) the first bolt is defective and the second is non-defective.

(b) the second bolt is non-defective given that the first is defective.

Solution

Let D and ND denote, respectively, the events *defective* and *non-defective* bolts. These two events are independent since the selection is with replacement and so

$(a)\ P(D \cap ND) = P(D)P(ND)$

$$= \frac{16}{40} \cdot \frac{24}{40} = 0.24$$

$$(b)\ P(ND / D) = \frac{P(ND \cap D)}{P(D)} = \frac{0.24}{0.4} = 0.6$$

Because the two events are independent, we could see from the last result that

$$P(ND / D) = P(ND) = 0.6\,.$$

If the selection was without replacement, we shall have

$(a)\ P(D \cap ND) = P(D).P(ND / D)$

$$= \frac{16}{40} \cdot \frac{24}{39} = 0.25$$

and

$$(b)\ P(ND / D) = \frac{P(ND \cap D)}{P(D)} = \frac{24}{39} \cdot \frac{16}{40} / \frac{16}{40} = 0.615\,.$$

Example 2.9 A fair coin is tossed twice. Let A be the event "first toss is a head" and B the event "the two outcomes are the same." Show that A and B are independent.

Solution

The sample space is $S = \{HH\ HT\ TH\ TT\}$,

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

$$B = \{HH\ TT\}.$$

Then $\qquad P(A) = P(B) = 1/2\,.$

Therefore

$$P(B / A) = \frac{P(B \cap A)}{P(A)} = \frac{1/4}{1/2} = \frac{1}{2} = P(B)\,,$$

and

$$P(A \cap B) = P(A / B)P(B) = \left(\tfrac{1}{2}\right)\left(\tfrac{1}{2}\right) = \left(\tfrac{1}{4}\right) = P(A)P(B).$$

Therefore, A and B are independent.

Next we consider an example involving two events that are not independent.

Example 2.10 In example 2.08, let E be the event "first toss is a head" and F the event "two heads turn up." Show that E and F are not independent.

Solution

Here we have

$$P(E) = 1/2 \text{ and } P(F) = 1/4.$$

$$E \cap F = \{HH\} \text{and } P(E \cap F) = 1/4.$$

But $P(E)P(F) = 1/8 \neq P(E \cap F)$.

Thus E and F are not independent.

Self Assessment Exercises

1. Independence and mutual exclusiveness can be considered as opposites. True or false.
2. A box contains 40 balls, 16 of which are white and 24 black. Two bolts are selected in succession with replacement. Find the probability that
   a. The first bolt is white and the second is black.
   b. The second bolt is black given that the first is white.

Self Assessment Answers

3.3 Bayes Probabilities

Suppose we have two boxes 1 and 2. Box 1 contains 3 black balls and 4 white balls. Box 2 contains 1 black ball and 2 white balls. Suppose that a ball is drawn at random and observed to be black but we do not know from which box it was drawn. Here we calculate the conditional probability that a particular box was chosen, given the color of the ball, namely $P(I/b)$. Such a probability is an *inverseprobability* called the *Bayesprobability*.Bayes probabilities are calculated using the bayes formula, which is derived as follows.

3.3.1 Bayes Formula

Let A be an arbitrary event that can occur only in conjunction with one of k mutually exclusive and exhaustive events $H_1, H_2, \ldots, H_k$ of S. Before we receive information on the occurrence of A, we have a set of probabilities $P(H_1), P(H_2), \ldots, P(H_k)$ for the events $H_1, H_2, \ldots, H_k$, called the *priorprobabilities*. We then want to find the probabilities for the events $H_1, H_2, \ldots, H_k$ given that A has occurred. That is, we want to find the conditional probabilities $P(H_i / A)$, called the *posteriorprobabilities*.

Now if A is observed, the probability that it occurs in conjunction with $H_i \, (i = 1, \ldots, k)$ is given by

$$P(H_i / A) = \frac{P(H_i \cap A)}{P(A)} \qquad (1)$$

We can calculate the numerator $P(H_i \cap A)$ of (1) from our given information by

$$P(H_i \cap A) = P(H_i)P(A/H_i). \qquad (2)$$

Now since one and only one of the events $H_1, H_2, \ldots, H_k$ can occur, we can write the probability of A as

$$P(A) = P(H_1 \cap A) + P(H_2 \cap A) + \ldots + P(H_k \cap A)$$

Using equation 2, the above expression can be seen to equal

$$P(A) = P(H_1)P(A/H_1) + P(H_2)P(A/H_2) + \ldots + P(H_k)P(A/H_k) \qquad (3)$$

Using (1), (2) and (3), we have the B*ayesformula*

$$P(H_i / A) = \frac{P(H_i)P(A/H_i)}{\sum_{j=1}^{k} P(H_j)P(A/H_j)}, \ (i = 1, 2, .., k)$$

This is the Baye's formula. Here, $P(H_i)$ is the *prior* probability of the event H_i; the probability of H_i before it is known whether A occurs, and $P(H_i / A)$ is the *posterior* probability of H_i.

Example 2.11: A doctor is trying to decide if a patient has one of the three diseases d_1, d_2, or d_3. From the national records on the distribution of diseases and test results for 10,000 people having one of these three diseases, he assigned the probability 0.3215 for disease d_1, 0.2125 for disease d_2, and 0.466 for d_3. He carries out a test that will be positve with probability 0.656 if the patient has disease d_1, 0.186 if he has disease d_2, and 0.109 if he has disease d_3. Given that the outcome of the test was positive, what probabilities should the doctor now assign to the three possible diseases?

Solution

The prior probabilities for the three diseases are $P(d_1) = 0.3215$, $P(d_2) = 0.2125$ $P(d_3) = 0.466$. Then using the Bayes formula, the posterior probabilities for the three diseases are

$$\begin{aligned} P(d_1 / t) &= \frac{P(d_1)P(t/d_1)}{P(d_1)P(t/d_1) + P(d_2)P(t/d_2) + P(d_3)P(t/d_3)} \\ &= \frac{(0.3215)(0.656)}{(0.3215)(0.656) + (0.2125)(0.186) + (0.466)(0.109)} \\ &= 0.700. \end{aligned}$$

$$P(d_2/t) = \frac{P(d_2)P(t/d_2)}{P(d_1)P(t/d_1) + P(d_2)P(t/d_2) + P(d_3)P(t/d_3)}$$
$$= \frac{(0.2125)(0.186)}{(0.3215)(0.656) + (0.2125)(0.186) + (0.466)(0.109)}$$
$$= 0.131$$

and

$$P(d_3/t) = \frac{P(d_2)P(t/d_2)}{P(d_1)P(t/d_1) + P(d_2)P(t/d_2) + P(d_3)P(t/d_3)}$$
$$= \frac{(0.466)(0.109)}{(0.3215)(0.656) + (0.2125)(0.186) + (0.466)(0.109)}$$
$$= 0.169$$

Example 2.12: In a Polytechnic, 60%, 25% and 15% of the total number of graduands in a set fall, respectively, in the class of lower credit, upper credit and distinction. Furthermore, 40% of those in the lower class, 20% of those in the upper class and 20% of those with distinction are women.

(i) If a graduand is selected at random, find the probability that the graduand is a woman.

(ii) If a graduand is selected at random and turns out to be a woman, find the probability that she belongs to the distinction class.

Solution

Let L = lower class, U = upper class, D = distinction. Then, $P(L) = 0.6,\ P(U) = 0.25,\ P(D) = 0.15.$

(i) Let W be the event that the graduand is a woman, then by total probability rule,

$$P(W) = P(L)P(W/L) + P(U)P(W/U) + P(D)P(W/D)$$
$$= 0.6x0.4 + 0.25x0.2 + 0.15x0.2$$
$$= 0.32.$$

(ii) Using Baye's theorem, the probability that the selected graduand belongs to the distinction class is

$$P(D/W) = \frac{P(W/D)P(D)}{P(W/D)P(D) + P(W/U)P(U) + P(W/L)P(L)}$$
$$= \frac{0.15x0.2}{0.15x0.2 + 0.25x0.2 + 0.6x0.4} = 0.09$$

Example 2.13: A ball is picked at random from a box containing balls that are black or white. Some balls are labeled and some are not. If half the balls are labeled and three fifths of the labeled balls and two fifths of the non- labeled balls are black, find the probability that the ball drawn is labeled, given that it is black.

Solution

Let B = black ball, W = white ball, L = labeled ball and NL = non-labeled ball. Then

$$P(L) = P(NL) = \frac{1}{2},$$

$$P(B/L) = \frac{3}{5},$$

$$P(B/NL) = \frac{2}{5}.$$

$$P(L \ / B) = \frac{P(B/L)P(L)}{P(B/L)P(L) + P(B/NL)P(NL)}$$

$$= \frac{(3/5)x(1/2)}{(3/5)x(1/2) + (2/5)x(1/2)} = \frac{3}{5}.$$

Self Assessment Exercises

1. Bayes probabilities is also known as …………………………………………
2. The probability of event H, given that A has occurred can be specified by …………

Self Assessment Answers

4.0 Conclusion

In this unit, you have learnt how to estimate the possibility of events' occurrence, proof events' property axioms and generalization of arbitrary events. Baye's probability and formula was also discussed.

In the next unit, we shall further our study of probability theory by discussing a more important aspect of it, known as Permutations and Combinations involving determining the number of ways that a certain event can occur.

5.0 Summary

In this unit, you have learnt the following:

i) Independent events are those events such that Event B is said to be statistically independent of Event A, if the occurrence of Event A has no effect on the probability of occurrence of B.

ii) Bayes Probabilitiesis an *inverseprobability*using the Bayes formula, stated as follows:

$$P(H_i/A) = \frac{P(H_i)P(A/H_i)}{\sum_{j=1}^{k} P(H_j)P(A/H_j)}, \ (i = 1, 2, .., k)$$

Where, $P(H_i)$ is the *prior* probability of the event H_i; the probability of H_i before it is known whether A occurs, and $P(H_i/A)$ is the *posterior* probability of H_i.

6.0 Tutor-Marked Assignment (TMA)

(1) In a class of ND statistics students, 25% failed MTH 212, 15% failed STA 213 and 10% failed both. A student is selected at random and noted to have failed STA 213.

i. What is the probability that the student failed MTH 212?

ii. What is the probability that the student failed MTH 212 or STA 213?

(2) An experiment consists of throwing a coin and a die simultaneously. If the events A: *head on coin* and B: *5 on die* are defined on the sample space of this experiment, investigate the dependence of these two events.

(3) 3 machines A, B and C produce, respectively, 50%, 30%, and 20% of the totalnumber of items in a factory. The percentages of defective output of thesemachines are 3%, 4% and 5% respectively.

i) If an item is selected at random, find the probability that the item is defective.

ii) If an item is selected at random and found to be defective, find the probabilitythat the item was produced by machine B.

(4) In a certain University, 4% of the men and 1% of the women are taller than 1.6.Furthermore, 56% of the students are women. If a student is selected at randomand is taller than 1.6, what is the probability that the student is a woman.

7.0 References/Further Readings

Harry Frank and Steven C. Althoen (1995): Statistics: concepts and applications. Cambridge Edition.

W. J. Decoursey (2003): Statistics and probability for engineering applications. Newness.

Sheldon M. Ross (1997): Introduction to probability models, sixth edition. Academic

Press. New York.

Mario Lefebre (2000): Applied probability and statistics. Springer.

http://www.stanford.edu/class/polisci100a/newprob2.pdf

http://en.wikipedia.org/wiki/Random_variable

http://www.stats.gla.ac.uk/steps/glossary/probability_distributions.html

http://faculty.palomar.edu/mmumford/120/notes/Chap5.pdf

Answers to SAAs

Module 1 Unit 1

1. A statistical experiment is any operation that is performed according to a well-defined set of rules and which when repeated generates a number of outcomes that cannot be predetermined.
2. A single performance of an experiment.
3. The possible result of a trial.
4. Sample space.
5. Events

6. A Simple event.

7. A compound event.

8. The Complement of A, denoted by A'.

9. The Empty set represented by, $\varnothing$ symbol.

10. $A \cup B$)

11. $A \cap B$)

12. False

1. True13. Independent event(s)

Module 1 Unit 2

2. $P(E) = \dfrac{\text{number of outcomes favourable to } E}{\text{total possible number of outcomes}} = \dfrac{n}{m}$.

3. $0 \leq p(E) \leq 1$ for every $E \subset S$.

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

Module 1 Unit 3

1. If one event occurs, then none of the others can occur.
2. The occurrence of one event has no effect at all on the probabilities of occurrence of the others.
3. Inverse probabilities
4. $P(H_i / A)$
5. True
6. a) 0.24 b) 0.6

Module 2

Combinatorics and Random Variables

Unit 1

Combinatorics

Content

1.0 Introduction

We now study the topics of *Permutations* and *Combinations*. This subject is important because many problems in probability theory involve counting the number of ways that a particular event can occur.

Before we continue, it is useful to study a general counting technique that will enable us to solve a variety of counting problems including the problem of counting the number of points in the sample space, S and in subsets of S.

2.0 Learning Outcome

At the end of this unit, you should be able to:

i. Develop a new counting technique for iterative sequences
ii. Use tree diagrams to represent events occurring in stages
iii. Arrange and select experimental objects using permutation and combination techniques of counting

3.0 Learing Outcomes

3.1 Counting Technique

Consider a task that is to be carried out in a sequence of r stages. There are n_1 ways to carry out the first stage; for each of these n_1 ways, there are n_2 ways to carry out the second stage; for each of these n_2 ways, there are n_3 ways to carry out the third stage, and so on.

Then the total number of ways in which the entire task can be accomplished is given by the product

$$N = n_1 . n_2 . \ldots . n_r .$$

It is often more convenient to represent the outcomes of this type of experiment by a *tree diagram*.

3.2 Tree Diagram

It is often useful to use a tree diagram when studying probabilities of events relating to experiments that take place in stages and for which we are given the probabilities for the outcomes at each stage.

Example 3.1 A fair coin is tossed three times. The tree diagram for this experiment is as shown in the figure below.

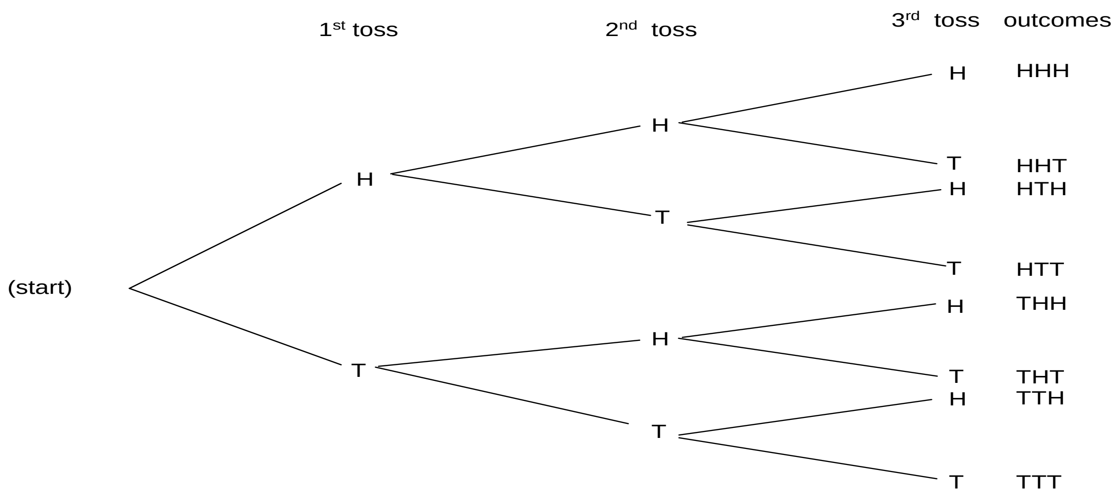


Figure 2.1: Tree diagram for three tosses of a coin.

A *path* through the tree corresponds to a possible outcome of the experiment. From the tree diagram above, we have eight paths, and assuming each outcome to be equally likely, we assign equal probability $1/8$ to each path.

Now, let E be the event "at least one head turns up". Then we have

$$E = \{HHH \;\; HHT \;\; HTH \;\; HTT \;\; THH \;\; THT \;\; TTH\}$$

and so

$$P(E) = 7/8.$$

Example 3.2 A box contains 5 red and 4 white marbles. If two marbles are drawn at random, what is the probability that

(a) i. one marble is red and the other white?

ii. both are of the same colour?

(b) If three marbles are drawn at random, what is the probability that exactly two are red?

Solution

The tree diagram is

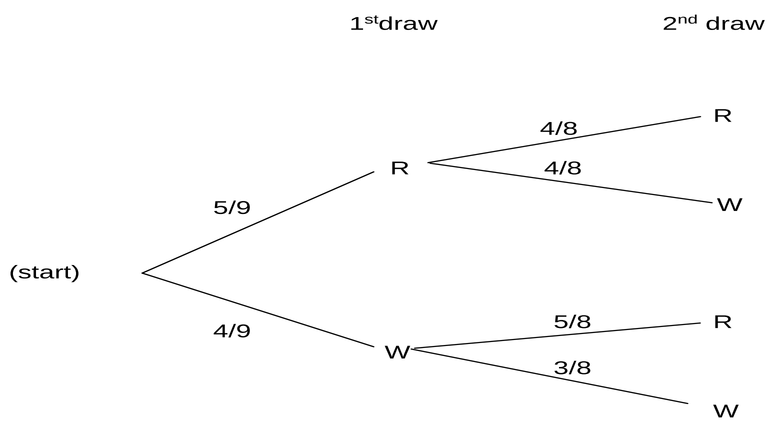


Figure 2.2: Tree diagram for drawing two marbles.

(a) i. P(one white and one red marble) = P(WR) + P(RW)

$$= \left(\frac{4}{9} x \frac{5}{8}\right) + \left(\frac{5}{9} x \frac{4}{8}\right)$$
$$= \frac{20}{72} + \frac{20}{72} = \frac{5}{9}.$$

ii. P(same colour) = P(RR) + P(WW)

$$= \left(\frac{5}{9} x \frac{4}{8}\right) + \left(\frac{4}{9} x \frac{3}{8}\right)$$
$$= \frac{20}{72} + \frac{12}{72} = \frac{4}{9}.$$

(c) If three marbles are drawn at random, the tree diagram is

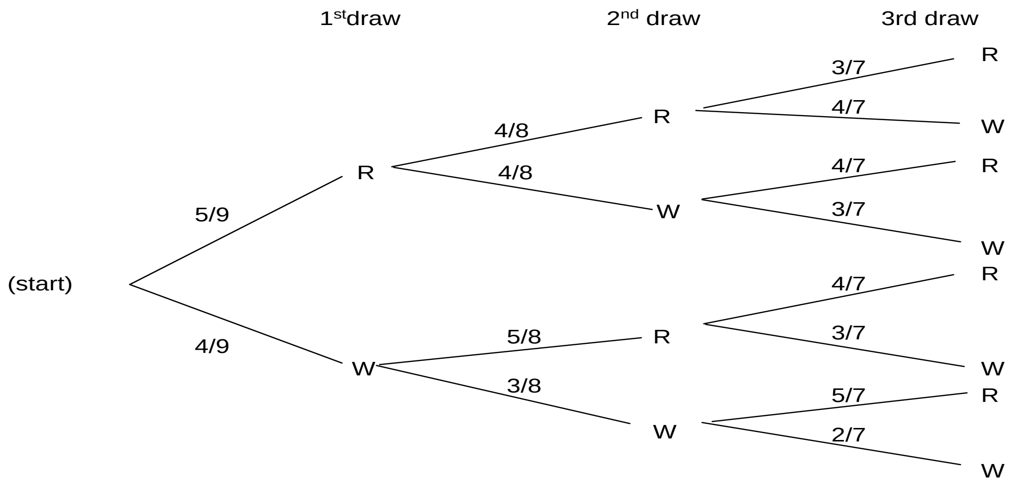


Figure 2.3: Tree diagram for drawing three marbles.

Therefore,

P(exactly 2 red balls) = P(RRW) + P(RWR) + P(WRR)

$$= \left(\frac{5}{9} x \frac{4}{8} x \frac{4}{7}\right) + \left(\frac{5}{9} x \frac{4}{8} x \frac{4}{7}\right) + \left(\frac{4}{9} x \frac{5}{8} x \frac{4}{7}\right)$$

$$= \frac{80}{504} + \frac{80}{504} + \frac{80}{504}$$

$$= \frac{10}{21} = 0.476.$$

Self Assessment Exercises

1. Multi-stage experiments such that there are n_1 ways to carry out the first stage; n_2 ways to carry out the second stage; for each of these n_2 ways, there are n_3 ways to carry out the third stage, and so on. Then the total number of ways in which the entire task can be accomplished is given by the product.
2. A path through a tree corresponds to …………………………………………..
3. If two red balls are drawn at random out of a total of six (three red, three blue) balls, what is the probability that exactly two are red?

Self Assessment Answers

3.3 Permutation

Consider the number of ways of arranging 3 of 7 items in 3 empty spaces. Here the first space can be filled in 7 ways, the second space can be filled in (7-1) = 6 ways and the third space can be filled in (7-2) = 5 ways. Therefore there are (7)(7-1)(7-2) ways of arranging the 3 items taken from 7 items. This is the number of permutations of 7 items taking 3 at a time. It is also the number of ways of filling 3 empty spaces taken 7 different items at a time.

Definition

Let A be an n- element set, and let k be an integer between 0 and n. Then a k-permutations of A is an ordered listing of a subset of the set A of size k.

That is, the total number of k-permutations of a set A of n elements is given by

$${}^{n}P_{k} = n(n-1)(n-2)...(n-k+1).$$

The total number of permutations of a set A of n distinct elements taking all at a time is given by

$$^{n}P_{n} = n(n-1)(n-2)...1.$$

This is called n *factorial* and is denoted by $n!$. By definition, $0! = 1$.

Example 3.3: Evaluate $7P_3$.

Solution

$$7P_3 = \frac{7!}{(7-3)!} = 7x6x5 = 210.$$

Example 3.4 A student must attempt five questions out of seven. How many different sequences of five might appear in his script?

Solution

This is

$$7P_5 = \frac{7!}{2!} = 7x6x5x4x3 = 2520.$$

3.4 Combination

This involves selection of objects without regard to order. It is the total number of combinations of r objects selected from n, or the combinations of n *objects taking r at a time.* Given n different objects, $r \le n$ of them can be selected in

$$\binom{n}{r} = C_r^n = \frac{n!}{r!(n-r)!} = \frac{n(n-1)...(n-r+1)}{r(r-1)...1}$$

ways, when no attention is paid to the order.

The symbol $\binom{n}{r}$, which is sometimes written as $^{n}C_{r}$ is read as *the number of combinations of* n *things taken* r *at a time*.

Example 3.5 (example 3.04 cont'd) If a student must answer 5 out of 7 questions in a semester examination, how many ways has he to select his questions?

Solution

Since the order in which the answers appear in the script is not relevant, we have

$$7C_5 = \frac{7!}{5!2!} = 21.$$

Example 3.6 In how many ways can we select 5 people from a group of ten to form a committee?

Solution

This is

$$10C_5 = \frac{10!}{5!5!} = 252.$$

Self Assessment Exercises

1. The selection of n objects from r objects without regard to order is known as ……………………………………………………………………………………..
2. In how many ways can a pair of socks be selected out of 6 pairs.

Self Assessment Answer(s)

4.0 Conclusion

In this unit, you have learnt how to determine the counting of experiments that are perfomed ins stages and how to represent them diagrammatically, using a tree diagram. Observe that a tree diagram is also useful for studying probabilities of events relating to experiments that take place in stages.

You are now familiar with Permutation i.e. the number of ways of arranging items, taking a certain number at a time; as well as Combinationwhich is the selection of objects without regard to order. These are basic techniques that are to be applied in forecasting stochastic events, as you will see later in this course.

5.0 Summary

You have learnt that:

i) There are $n_1.n_2….n_r$ possible ways in which a task involving r stages can be accomplished, for cases where each of these n_1 ways, there are n_2 ways to carry out the second stage; for each of these n_2 ways, there are n_3 ways to carry out the third stage, and so on.

ii) A tree diagram is used to study probabilities of events relating to experiments that take place in stages.

iii) Permutation deals with items arrangement by drwaing a certain number of the items at a time, denoted by $^nP_k = n(n-1)(n-2)...(n-k+1)$, called *n!* (n-factorial)

iv) Combination is the selection of objects without regard to order, denoted by ^nC_r i.e the number of combinations of n items taken r items at a time.

6.0 Tutor-Marked Assignment (TMA)

1. How many 3 letter words can be formed from the letter MANGO?
2. In how many ways can 5 students in a class sit on 1 desk?
3. Find the number of permutations that can be formed using the words:
   i. STATISTICS (ii)MISSISSIPI (iii) BANANA (iv) CASHEW
4. An urn contains 6 balls, all of different colours, find the number of ordered arrangement
   i. of size 3 with replacement.
   ii. of size 3 without replacement
   iii. of size with one more replacement
5. Consider the four letters A, B, C, D. How many unordered selection are possible taken 3 letters at a time if
   i. Replacement is not allowed.
   ii. replacement is allowed
   iii. list the samples in (a) or (b) above
6. Suppose 6 students are to represent their university in an inter – university tennis competition
   a. How many double teams can be found?
   b. If one of the students is the tennis captain and play in each competition, how many double teams are possible?
7. A welfare committee of 5 members is to be formed from 12 peoples which 7 are women and 5 are men. Find the possibility of selecting the 5 members committee if it is required that 3 must be women.
8. A committee of 3 people is to be chosen from 4 married couples.
   i. How many different committees are possible?
   ii. How many contain 2 women and 1 man?
   iii. What is the probability that the committee will be made up of more women than men?
   iv. What is the probability that the committee will contain only men?

7.0 References/Further Reading

Harry Frank and Steven C. Althoen (1995): Statistics: concepts and applications. Cambridge Edition.

W. J. Decoursey (2003): Statistics and probability for engineering applications. Newness.

Unit 2

Random Variables and Their Distributions

Content

1.0 Introduction

Observe that most of the examples of statistical experiments you have used in the previous units resulted in quantitative (numerical) data. In this unit, we shall develop models for such numerical data, with associated real components otherwise known as random variables.

In addition, you are going to study the different types of the data model (random variables) such as discrete random variables, continuous random variables, density functions, continuous distribution functions as well as their characteristic properties. With these, you will be able to develop models for all your experimental data by specifying their respective variables and state the data properties.

2.0 Learning Outcome

At the end of this unit, you should be able to:

i. Develop models for your experimental data
ii. Identify different data variables
iii. Fit data into their respective model
iv. Specify data characteristic properties

3.0 Learning Outcomes

3.1 Concept of Random Variables

We have seen that most of the examples of statistical experiments given in the previous units generated quantitative (numerical) data. This is frequently true. To develop models for such data, a real number or a vector with real components is associated with each possible outcome, s in the sample space, S, of the experiment. The function mapping the outcome to the number or vector is called a random variable.

Definition

A random variable is a real-valued function $X(s)$, defined for every outcome, s in the sample space of an experiment.

Capital letters are often used to designate the random variable itself and lower-case letters for an unspecified value of a random variable.

For example, if we toss a coin twice and denote by X the number of heads that turn up, then heads may be associated with X = 1, 2 and tails with X = 0. So, X is called a random variable and its values are 0, 1 and 2.

In practice, the sample space of the random variable is either a discrete set of values (as in throwing a die) or a continuous set (as in measuring a length).

3.2 Discrete Random Variables

These are integer-valued random variables with discrete sample space. For instance, the number of defective items in a carton of manufactured items, the number of leaves on a tree and the number of heads that comes up when a pair of fair coins is tossed are examples of random variables with discrete value space.

> ***Definition***
>
> A discrete random variable X is a random variable with finite or countably infinite set of possible outcomes.

Self Assessment Exercises

1. Numerical data can be replaced with ……………………………………………
2. Discrete random variables have …………………………..…. Values.
3. Examples of discrete random variables are ………………………………….…...

Self Assessment Answer(s)

3.3 Probability Function of Discrete Random Variables

It may be convenient to assign a probability to each point x in the sample space S of an experiment. The discrete probability function specifies the probability associated with each possible value the random variable in question can assume.

> ***Definition***
>
> Let S be a discrete sample space of an experiment. Then a probability distribution function for each of the elements, x of S is a real-valued function p whose domain is S and which satisfies
>
> 1) $0 \le p(x) \le 1, \quad \text{for all } x \in S, \text{ and}$
>
> 2) $\sum_{x \in S} p(x) = 1.$
>
> Furthermore, for any subset E of S, the probability of E is defined to be the number $P(E)$ given by
>
> $$P(E) = \sum_{x \in E} p(x).$$

Example 4.1: A fair coin is tossed twice, the sample space of this experiment is

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

Here we see that every outcome is equally likely and so each of the four outcomes will have an assigned probability of $1/4$. That is

$$p(x) = 1/4, \quad \text{for every } x \in S$$

which is nonnegative.

Now to verify condition (2), we see that

$$\sum_{x \in S} p(x) = \tfrac{1}{4} + \tfrac{1}{4} + \tfrac{1}{4} + \tfrac{1}{4} = 1.$$

Therefore, $P(x)$ is a valid discrete probability function.

Eample 4.2:

From example 3.01, let E be the event that 'atleast one head turns up'. Then, the elements of E are $E = \{HH, HT, TH\}$ and the probability of E is

$$P(E) = \sum_{x \in E} p(x) = \tfrac{1}{4} + \tfrac{1}{4} + \tfrac{1}{4} = \tfrac{3}{4}.$$

Example 4.3:

Consider an experiment of rolling a fair die once for which the sample space is $S = \{1, 2, 3, 4, 5, 6\}$. Since the die is fair, we define the probability distribution function for the elements of this sample space by

$$p(x) = \tfrac{1}{6}, \text{ for } x = 1, \ldots, 6,$$

which is nonnegative.

Also, for every $x \in S$,

we have

$$\sum_{x \in S} p(x) = 6\left(\tfrac{1}{6}\right) = 1.$$

Therefore,

is a valid discrete probability function.

Now, if E is the event that the result of the roll is an even number, we have the elements of E as

$E = \{2, 4, 6\}$,

and so

$$P(E) = \sum_{x \in E} p(x) = 3\left(\tfrac{1}{6}\right) = \tfrac{1}{2}.$$

Example 4.4

A pair of fair dice is rolled once. What is the probability of

(i) getting a sum of 5?

(ii) getting a sum of 11?

Solution

The sample space of this experiment is the set of all ordered pairs (i, j) of integers with $1 \le i \le 6$ and $1 \le j \le 6$. Therefore

$$S = \{(i, j) : 1 \le i, j \le 6\}.$$

Here, we note that there are six choices for i, and for each choice of i, there are six choices for j, leading to 6x6 =36 different outcomes as the size of the sample space S for the experiment.

Now, since the dice are fair, each of the 36 outcomes is equally likely and the probability distribution function on S is

$$p([i, j]) = \tfrac{1}{36},\ 1 \le i, j \le 6.$$

(i) Let the event “getting a sum of 5” be denoted by E, then

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

and

$$P(E) = \sum_{x \in E} p(x) = 4\left(\tfrac{1}{36}\right) = \tfrac{1}{9}.$$

That is, the probability of getting a sum of 5 is $\tfrac{1}{9}$.

(ii) Let F denote the event of getting a sum of 11. Then,

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

So,

$$P(F) = \sum_{x \in F} p(x) = 2\left(\tfrac{1}{36}\right) = \tfrac{1}{18}.$$

Therefore, the probability of getting a sum of 11 is $\tfrac{1}{18}$.

3.4 Cumulative Distribution Function for Discrete RandomVariables

For a discrete random variable X with probability function $f(x)$, for $x = x_1, x_2, ..., x_n$, the distribution function, $F(x)$, is defined by

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

That is, the corresponding cumulative probabilities of the discrete random variable X are obtained by summing all the probabilities up to a particular level.

Example 4.5: The probability distribution for a discrete random variable X is given by the table below.

x	0	1	2	3	4	5	6	
$P(X = x)$	0.04		0.2	0.3	0.1	0.16	0.08	0.12

Obtain the corresponding CDF table for X.

Solution

By definition, the cdf is given by

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

Therefore, the CDF table is

x	0	1	2	3	4	5	6	
$F(X)$		0.04	0.24	0.54	0.64	0.80	0.88	1

From this table, it is obvious that the function $F(x)$ increases from zero at the bottom of the range to unity at the top of the range.

Self AssessmentExercises

1. The probability function of a discrete random variable is called …
2. A pair of fair dice is rolled once. What is the probability of obtaining
   a. Sum of 7?
   b. Sum of 9?
3. 6. The summation of all the probabilities of a discrete random variable is called ….

Self Assessment Answer(s)

3.5 Continuous Random Variables

A random variable with sample space consisting of a continuous set of values is referred to as a continuous random variable. That is, X is a continuous random variable when it can only assume any value within a specified range or interval. A good example is the measurement of weight or height.

Definition

A random variable X is said to be a continuous type when all of its possible values are contained in a particular range or interval, say [a, b].

If X is continuous, then the probability that it assumes any one particular value is generally zero. For instance, let X be a continuous random variable denoting the height of adult males in a group. Then, if an individual is selected at random from this group, the probability that his height X is precisely 42 inches would be zero. However, there is a probability greater than zero that X is between 54.5 and 55.5 inches.

3.6 Density Functions of Continuous Random Variables

A continouos random variable represents the outcome of an experiment with a continuous sample space. In such experiments, the probabilities for the outcomes to fall in a given interval are assigned by means of the area under a suitable function called the density function.

Definition

Let X be a continuous real-valued random variable. A *density function* for X is a real-valued function f which satisfies

$$P(a \le X \le b) = \int_a^b f(x)\,dx,$$

for all $a, b \in \Re$.

It should be noted at this point that, unlike the case of discrete sample spaces, the value $f(x)$ of the density function for the outcome x is not the probability of x occurring (it has been stated earlier that this probability is always 0) and in general, $f(x)$ is not at all, a probability.

Nevertheless, the density function f does contain all the probability information about the experiment, since the probabilities of all events can be derived from it. In particular, the probability that the outcome of the experiment falls in the interval $[a,b]$ is given by

$$P([a,b]) = \int_a^b f(x)\,dx,$$

That is, by the *area* under the graph of the density function in the interval $[a,b]$.

Example 4.6A continuous random variable X has density function

$$f(x) = \begin{cases} Cx, & 2 \le x \le 10, \\ 0, & otherwise \end{cases}$$

Find

(a) The value of the constant C.

(b) $P(3 \leq X \leq 7)$.

Solution

(a) Since $f(x)$ is a density function, we must have

$$\int_2^{10} f(x)\,dx = 1.$$

That is,

$$\int_2^{10} Cx dx = C\int_2^{10} x\,dx = 1$$
$$\Rightarrow C\left[\tfrac{x^2}{2}\right]_2^{10} = 48C = 1$$
$$\Rightarrow C = \tfrac{1}{48}.$$

Hence,

$$f(x) = \tfrac{1}{48}x, \qquad 2 \leq x \leq 10,$$
$$0, \qquad otherwise.$$

(b) $P(3 \leq X \leq 7) = \int_3^7 \tfrac{x}{48}\,dx = \tfrac{1}{96}\left[x^2\right]_3^7$

$$= \tfrac{40}{96} = \tfrac{5}{12}.$$

Therefore,

$$P(3 \leq X \leq 7) = 0.417.$$

Example 4.7 A continuous random variable X has pdf

$$f(x) = \begin{cases} Cx, & 0 \leq x \leq 2, \\ C(4-x), & 2 \leq x \leq 4, \\ 0, & otherwise. \end{cases}$$

Find

(a) the value of the constant C

(b) $P(\tfrac{1}{2} \leq X \leq 2\tfrac{1}{2})$.

Solution

(a) $\int_0^2 Cx\,dx + \int_2^4 C(4-x)dx = 1$, since $f(x)$ is a density function.

That is,

$$C\left[\tfrac{x^2}{2}\right]_0^2 + C\left[4x - \tfrac{x^2}{2}\right]_2^4 = 1.$$
$$\Rightarrow 2C + C\left[(16-8)-(8-2)\right] = 1.$$
$$\Rightarrow 4C = 1.$$

So

$$C = \tfrac{1}{4}.$$

Therefore,

$$f(x) = \begin{cases} \frac{1}{4}x, & 0 \le x \le 2, \\ \frac{1}{4}(4-x), & 2 \le x \le 4, \\ 0, & otherwise \end{cases}$$

(b) $P(\frac{1}{2} \le X \le 2\frac{1}{2}) = \int_{\frac{1}{2}}^{2} \frac{x}{4}\,dx + \int_{2}^{2\frac{1}{2}} \frac{1}{4}(4-x)dx$

$$= \tfrac{1}{8}\left[x^2\right]_{\frac{1}{2}}^{2} + \tfrac{1}{4}\left[4x - \tfrac{x^2}{2}\right]_{2}^{2\frac{1}{2}}$$
$$= \tfrac{1}{8}\left[4 - \tfrac{1}{4}\right] + \tfrac{1}{4}\left[\left[10 - \tfrac{25}{8}\right] - \left[8 - 2\right]\right]$$
$$= \tfrac{15}{32} + \tfrac{7}{32} = \tfrac{11}{16}.$$

Therefore,

$$P(\tfrac{1}{2} \le X \le 2\tfrac{1}{2}) = \tfrac{11}{16}.$$

Self Assessment Exercises

1. A continuous random variable has a sample space consisting ………………. Values
2. Measuring the quantity of petrol is an example of discrete random variable. True or False.
3. The probabilities for the outcomes of continuous random variables to fall in a given interval are assigned a function called …………………………………….

Self Assessment Answers

3.7 Cumulative Distribution Functions for Continuous Random Variables

In the previous section, we considered the density functions of continuous random variables. In this section we shall consider another kind of function that is closely related to these density functions and is also of great importance. These functions are called cumulative distribution functions.

Definition

If X is a continuous random variable there exists the probability $P(X \le x)$, for X to be less than or equal to x. This probability is called the cumulative distribution function of X denoted by $F(x)$. Thus

$$F(x) = P(X \le x).$$

Therefore, if X is a continuous real-valued random variable which possesses a density function $f(x)$, then it also has a cumulative distribution function defined by

$$F(x) = \int_{-\infty}^{x} f(t)dt.$$

Furthermore, $f(x)$ and $F(x)$ are related by the fact that

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

That is, if F has a derivative, then $dF(x)/dx = f(x)$ for almost all x.

3.7.1 Properties of Distribution Functions

The following properties of a distribution function are consequences of the definition.

i. $F(-\infty) = 0; \quad F(\infty) = 1$,
   that is, the function $F(x)$ obviously increases fromzero at the bottom of the range to unity at the top of the range.
ii. $F(x_1) \le F(x_2),\ if\ x_1 < x_2$, that is, $F(x)$ is monotonically non-decreasing.
iii. The probability in an interval, say $x_1 < X \le x_2$, is given by the difference in the values of F evaluated at the endpoints of the interval. That is $F(x_2) - F(x_1) = P(x_1 < X \le x_2)$.

Example 4.8: A continuous random variable X has density function

$$f(x) = 2kx, \quad 0 \le x \le 5,$$
$$0, \quad otherwise$$

obtain (i) the value of k

(ii) the CDF

Solution

(i) to obtain the value of the constant k, we have

$$\int_0^5 2kx\,dx = 1$$

$$\Rightarrow 2k\int_0^5 x\,dx = k\left[x^2\right]_0^5 = 1.$$

That is

$$25k = 1 \;\Rightarrow\; k = \tfrac{1}{25}.$$

Hence,

$$f(x) = \frac{2x}{25}, \quad 0 \le x \le 5,$$
$$0, \quad otherwise.$$

(ii) The CDF is derived as

$$F(t) = \int_{-\infty}^{t} f(x)dx = \tfrac{2}{25}\int_0^t x\,dx$$
$$= \tfrac{2}{25}\left[\tfrac{x^2}{2}\right]_0^t = \tfrac{t^2}{25}, \quad 0 \le t \le 5.$$

Therefore, $F(x) = \frac{x^2}{25}, \quad 0 \le x \le 5.$

Example 4.9: A continuous random variable X has CDF given by

$$F(x) = \begin{cases} 0, & x \le 0, \\ \frac{x^3}{27}, & 0 \le x \le 3. \\ 1, & x \ge 3. \end{cases}$$

Obtain the density function $f(x)$.

Solution

Using the relationship

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

we have the density function derived as

$$\tfrac{d}{dx}F(x) = \tfrac{d}{dx}\left(\frac{x^3}{27}\right) = \frac{x^2}{9} = f(x).$$

Therefore the density function for X is

$$f(x) = \tfrac{x^2}{9}, \quad 0 \le x \le 3$$
$$0, \quad otherwise.$$

Self Assessment Exercises

1. The probability such that, $F(x) = P(X \le x)$ is called …
2. $F(x_1) \le F(x_2),\ if\ x_1 < x_2$, means $F(x)$ is ….

Self Assessment Answers

4.0 Conclusion

This unit has taught you how to represent experimental data values with random variables as well as distingusnig between a discrete and continuous random quantities. Also, the technique generating a data model for quantitative data was accomplished.

Further, you learnt how to obtain density functions, discrete and continuous distribution functions as well as their characteristic properties for experimental data. These are useful tools for fitting data into their respective models, subsequent units will introduce you to other data models, specification and generalization.

5.0 Summary

You have learnt the following in this unit:

i) Quantitative data generated from experiments are converted into data models. A random variable as a real-valued function $X(s)$, was introduced and defined for every outcome, s in the sample space of an experiment. Random variables are represented by upper-case lettersand lower-case letters for an unspecified value of a random variable.

ii) There are discrete random variables as well as continuous random variables. Discrete random variables are integer-valued with discrete sample space, while a continuous random variable assumes any value within a specified range or interval. Examples of discrete random variables are die tossing, number of books on a table, etc; and examples of continuous random variables are measurement of weight or height or quantity of liquids.

iii) Probability distribution function of discrete random variables which assigns a probability to each point x in the sample space S of an experiment is defined as

1) $0 \le p(x) \le 1, \quad \text{for all } x \in S,$

for S being a discrete sample space of an experiment. Then a probability distribution function for each of the elements, x of S is a real-valued function p whose domain is S. Also, for any subset E of S, the probability of E is defined to be the number $P(E)$ given by

$$P(E) = \sum_{x \in E} p(x).$$

iv) Density functions of continuous random variables X is a continuous real-valued random variable satisfying

$$P(a \le X \le b) = \int_a^b f(x)\,dx,$$

for all $a, b \in \Re$.

v) Cumulative distribution functions for continuous random variables, X is a continuous random variable where the probability $P(X \le x)$, is denoted by $F(x)$ as

$$F(x) = P(X \le x).$$

If X is a continuous random variable with a density function $f(x)$, then it has a cumulative distribution function defined by

$$F(x) = \int_{-\infty}^{x} f(t)dt.$$

vi) Properties of distribution functionsas consequences of the definition was stated as:.

i. $F(-\infty) = 0;\ \ F(\infty) = 1$,

ii. $F(x_1) \le F(x_2),\ if\ x_1 < x_2$,

iii. The probability in an interval, with $x_1 < X \le x_2$, is given by the difference in the values of F evaluated at the endpoints of the interval, i.e. $F(x_2) - F(x_1) = P(x_1 < X \le x_2)$.

6.0 Tutor-Marked Assignments (TMA)

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

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

Describe in words the events specified by the following subsets of S:

(a) $E = \{HHH, HHT, HTH, HTT\}$

(b) $F = \{HHH, TTT\}$

(c) $D = \{HHT, HTH, THH\}$

(d) $K = \{HHT, HTH, HTT, THH, THT, TTH, TTT\}$

2. State the probabilities of each of the events described in exercise 1 above.

3. A die is loaded in such a way that the probability of each face turning up is proportional to the number of dots on that face. (For example, a six is three times as probable as a two.) What is the probability of getting an even number in one throw?

4. A continuous random variable X has density function

$$f(x) = 2kx, \quad 0 \le x \le 5,$$
$$0, \quad otherwise$$

obtain (i) the value of k

(ii) the CDF

5. A continuous random variable X has CDF given by

$$F(x) = \begin{cases} 0, & x \le 0, \\ \frac{x^3}{27}, & 0 \le x \le 3. \\ 1, & x \ge 3. \end{cases}$$

Obtain the density function $f(x)$.

7.0 References/Further Reading

Harry Frank and Steven C. Althoen (1995): Statistics: concepts and applications. Cambridge Edition.

W. J. Decoursey (2003): Statistics and probability for engineering applications. Newness.

Unit 3

Expected Value and Variance

Content

1.0 Introduction

When data is gathered during a research or an experiment, the observed data or the gathered research data are not enough rather some descriptive quantities such as the mean or the median are desirable metrics of the data. This unit will step you into the measurement of such descriptive metrics and more such as the *expectedvalue* and the *variance* of numerically-valued random variables.

In addition, you will be able to measure the likely deviation from central tendencies (mean and median) of random variables such as the variance and standard deviation.

2.0 Learning Outcome

At the end of this unit, you should be able to:

i. Determine the mean and median of numerically-valued random variables
ii. Compute the deviation of random and continuous variables from central measurements
iii. State the properties of expected values and variances of random variables
iv. State the properties of expected values and variances of continuous variables

3.0 Learning Outcomes

3.1 Mathematical Expectation of Discrete Random Variables

When a research of any kind is conducted and data is gathered, we are usually interested not in the individual data items, but rather in certain descriptive quantities such as the mean or the median. In general, the same is true for the probability distribution of a numerically valued random variable. In this chapter, we shall consider two such descriptive quantities - the *expectedvalue* and the *variance* – as they apply to numerically valued random variables.

3.1.1 Expected Value

For a discrete random variable X, the mathematical expectation is obtained by considering the various values that the variable can take; multiplying those values by their corresponding probabilities and then sum the products.

> ***Definition***
>
> Let X be a discrete random variable with sample space S and probability function $p(x)$. The *expectedvalue* $E(X)$ is defined by
>
> $$E(X)=\sum_{x\in S} xp(x)$$
>
> This is often referred to as the mean and denoted by μ_x or μ.

Example 5.1: A fair coin is tossed three times. What is the expected value for the number of heads that appear?

Solution

The sample space for the experiment is

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

in which each element has the probability $\frac{1}{8}$.

Now, let the number of heads that appear be denoted by the random variable X. Then the possible values of X are 0, 1, 2 and 3 with the respective probabilities $\frac{1}{8}, \frac{3}{8}, \frac{3}{8}$ and $\frac{1}{8}$. Therefore, the expected value of X is

$$\begin{aligned} E(X) &= \sum_{i=1}^{4} x_i\, p(x_i) \\ &= 0\left(\tfrac{1}{8}\right) + 1\left(\tfrac{3}{8}\right) + 2\left(\tfrac{3}{8}\right) + 3\left(\tfrac{1}{8}\right) = \tfrac{3}{2}. \end{aligned}$$

Example 5.2: The probability distribution for a discrete random variable X is given by the table below.

x	0	1	2	3	4	5	6
$p(x)$	0.04	0.2	0.3	0.1	0.16	0.08	0.12

Obtain the expected value of X.

Solution

From the table, the expected value is given by

$$\begin{aligned} E(X) &= \sum_{i=1}^{n} x_i p_i = x_1 p_1 + x_2 p_2 + \ldots + x_n p_n \\ &= 0(0.04) + 1(0.2) + 2(0.3) + 3(0.1) + 4(0.16) + 5(0.08) + 6(0.12) \\ &= 2.86. \end{aligned}$$

Example 5.3 A fair die is rolled once. Denote by X the number that turns up. Find the expected value of X.

Solution

The sample space for the experiment is

$$S = \{1, 2, 3, 4, 5, 6\},$$

in which each element has the probability $\frac{1}{6}$.

Therefore, the expected value of X is

$$\mu = E(X) = \sum_{i=1}^{n} x_i p(x_i)$$
$$= 1\left(\tfrac{1}{6}\right) + 2\left(\tfrac{1}{6}\right) + 3\left(\tfrac{1}{6}\right) + 4\left(\tfrac{1}{6}\right) + 5\left(\tfrac{1}{6}\right) + 6\left(\tfrac{1}{6}\right)$$
$$= \tfrac{7}{2}.$$

3.2 Properties of Expected Values of Discrete Random Variables

The following are properties of the expected values of discrete random variables.

If C is any constant and X is a discrete random variable, then

i. $E(C) = C.$

This can easily be investigated since

$$E(C) = \sum_{i=1}^{n} Cf(x_i) = C\sum_{i=1}^{n} f(x_i)$$
$$= C(1) = C.$$

ii. Product Of A Constant And The Random Variable

$E(CX) = C\,E(X)$

That is, the expected value of the product of a constant and a random variable is equal to the product of the constant and the expected value of the random variable.

Proof

$$E(CX) = \sum_{i=1}^{n} cx_i P(X = x_i)$$
$$= c\sum_{i=1}^{n} x_i P(X = x_i)$$
$$= cE(X).$$

iii. Sum Of Two Random Variables.

Let X and Y be random variables with finite expected values. Then

$E(X+Y) = E(X) + E(Y)$

Proof:

We consider the random variable $X+Y$ to be the result of applying the function $\phi(x,y) = x + y$ to the joint random variable (X,Y).

$$E(X+Y) = \sum_i \sum_j (x_i + y_j) P(X = x_i, Y = y_j)$$
$$= \sum_i \sum_j x_i P(X = x_i, Y = y_j) + \sum_i \sum_j y_j P(X = x_i, Y = y_j)$$

Now, using the fact that

$$\sum_j P(X = x_i, Y = y_j) = P(X = x_i)$$

and

$$\sum_i P(X = x_i, Y = y_j) = P(Y = y_j)\,,$$

we then have

$$E(X+Y) = \sum_i x_i P(X = x_i) + \sum_j y_j P(Y = y_j)$$
$$= E(X) + E(Y).$$

It is important for us to note at this point that mutual indepenedence of the summands (X and Y) is not needed here.

iv. Sum of a constant and the random variable

$$E(X+C) = \sum_{i=1}^{n} (X+C) f(x_i)$$

$$= \sum_{i=1}^{n} Xf(x_i) + \sum_{i=1}^{n} C f(x_i)$$
$$= E(X) + C$$

v. Independence

If X and Y are two independent random variables, then

$$E(X.Y) = E(X)E(Y).$$

Proof

$$E(X.Y) = \sum_j \sum_k x_j y_k P(X = x_j, Y = y_k).$$

Now, if X and Y are independent,

$$P(X = x_j, Y = y_k) = P(X = x_j)P(Y = y_k).$$

Therefore,

$$E(X.Y) = \sum_j \sum_k x_j y_k P(X = x_j)P(Y = y_k)$$

$$= \left(\sum_j x_j P(X = x_j)\right)\left(\sum_k y_k P(Y = y_k)\right)$$
$$= E(X)E(Y).$$

Example 5.4A coin is tossed twice. $X_i = 1$ if the ith toss is head and 0 otherwise. Now, X_1 and X_2 are independent. They each have expected value ½.

Therefore,

$$E(X.Y) = E(X)E(Y) = (1/2)(1/2) = 1/4.$$

The following example shows that the expected values need not multiply if the random variables are not independent.

Example 5.5A coin is tossed once. Let the random variable X have a value 1 if heads turns up and 0 if tails turns up, and let $Y = 1 - X$. Then $E(X) = E(Y) = 1/2.$ Now, $X.Y = 0$ for either outcome. Therefore,

$$E(X.Y) = 0$$

and this implies that

$$E(X.Y) \neq E(X)E(Y).$$

Self Assessment Exercises

1. If X is a discrete random variable with sample space S and probability function $p(x)$ what is the expected value of *X*?

2. What is the symbol of mean of a discrete random variable, *X*?
3. A fair coin is tossed twice, what is the expected value for the number of tails that appear?

Self Assessment Answers

3.3 Variance and Standard Deviation of Discrete Random Variables

In the previous sections, we considered the expected value as one of the means of predicting the outcome of an experiment. The prediction becomes more and more accurate when the outcome is not likely to deviate too much from the expected value. A measure of this deviation or spread of the values (which the corresponding numerically valued random variable X can assume) is called the variance, and is introduced in this section.

Definition

If X is a real-valued random variable having probability function $f(x)$ and expected value $\mu = E(X)$, then the variance of X (denoted by $V(X)$ or $Var(X)$) is defined by

$$V(X) = E((X-\mu)^2).$$

This can further be expressed as

$$V(X) = E(X-\mu)^2 = \sum_x (X-\mu)^2 f(x)$$
$$= \sum_x (x^2 - 2\mu x + \mu^2) f(x)$$
$$= \sum_x x^2 f(x) - 2\mu \sum_x x f(x) + \mu^2 \sum_x f(x)$$
$$= E(X^2) - \mu^2$$

3.3.1 Standard Deviation

The *standarddeviation* of X denoted by σ_x, where

$$\sigma_x = \sqrt{V(x)} = \sqrt{E(x-\mu)^2} ,$$

is the positive square root of variance. It is a measure of variability or spread around the mean; a small standard deviation indicating little variability among the possible values of the random variable, a large standard deviation indicating considerable variability.

Example 5.6 Consider the experiment of die rolling in example 5.3 above, with X denoting the number that turns up. Find the variance of X.

Solution

The expected value of X is $\frac{7}{2}$ as calculated in example 5.3. Now to find the variance of X, we may form the new random variable $(X-\mu)^2$ and compute its expectation. This can easily be done if we use the table below.

x	$p(x)$	$(x-7/2)^2$
1	1/6	25/4
2	1/6	9/4
3	1/6	¼
4	1/6	¼
5	1/6	9/4
6	1/6	25/4

From this table we then find $E((X-\mu)^2)$ as

$$V(X) = E((X-\mu)^2) = \sum (x-\mu)^2 p(x)$$
$$= \frac{1}{6}\left(\frac{25}{4} + \frac{9}{4} + \frac{1}{4} + \frac{1}{4} + \frac{9}{4} + \frac{25}{4}\right)$$
$$= \frac{35}{12}.$$

Or we may use the formula

$$V(X) = E(X^2) - \mu^2,$$

as derived above. That is

$$V(X) = E(X^2) - \mu^2$$
$$= 1\left(\tfrac{1}{6}\right) + 4\left(\tfrac{1}{6}\right) + 9\left(\tfrac{1}{6}\right) + 16\left(\tfrac{1}{6}\right) + 25\left(\tfrac{1}{6}\right) + 36\left(\tfrac{1}{6}\right) - \left(\tfrac{7}{2}\right)^2$$
$$= \tfrac{91}{6} - \left(\tfrac{7}{2}\right)^2$$
$$= \frac{35}{12}$$

in agreement with the value obtained directly from the definition of $V(X)$ above.

The standard deviation is

$$\sigma_x = \sqrt{V(x)} = \sqrt{E(x-\mu)^2}$$
$$= \sqrt{35/12}$$
$$\approx 1.707.$$

Example 5.7(example 5.2 cont'd) the probability distribution for a discrete random variable X is given by the table below.

x	0	1	2	3	4	5	6
$P(X = x)$	0.04	0.2	0.3	0.1	0.16	0.08	0.12

Obtain (i) the expected value of X

(ii) the variance of X

Solution

(i) By the above definition of expected value, we have

$$E(X) = \sum_{i=1}^{n} x_i p_i = x_1 p_1 + x_2 p_2 + \ldots + x_n p_n$$
$$= 0(0.04) + 1(0.2) + 2(0.3) + 3(0.1) + 4(0.16) + 5(0.08) + 6(0.12)$$
$$= 2.86$$

(ii) Now

$$
\begin{aligned}
E(X^2) &= \sum_{i=1}^{n} x_i^2 p(x_i) \\
&= 0(0.04) + 1(0.2) + 4(0.3) + 9(0.1) + 16(0.16) + 25(0.08) + 36(0.12) \\
&= 11.18
\end{aligned}
$$

Then, the variance is given by

$$
\begin{aligned}
Var(X) &= E(X^2) - \mu^2 \\
&= 11.18 - (2.86)^2 \\
&= 3.0004
\end{aligned}
$$

Example 5.8A random variable X has the distribution

$$
p(x) = \begin{pmatrix} 0 & 1 & 2 & 4 \\ 1/3 & 1/3 & 1/6 & 1/6 \end{pmatrix}.
$$

Obtain the expected value, variance, and standard deviation of X .

Solution

(i) The expected value is given as

$$
\begin{aligned}
E(X) &= \sum_{x} xp(x) \\
&= 0(1/3) + 1(1/3) + 2(1/6) + 4(1/6) \\
&= 0 + 1/3 + 1/3 + 2/3 \\
&= 4/3.
\end{aligned}
$$

(ii) To calculate the variance, we have

$$
\begin{aligned}
E(X^2) &= \sum_{x} x^2 p(x) \\
&= 0(1/3) + 1(1/3) + 4(1/6) + 16(1/6) \\
&= 0 + 1/3 + 2/3 + 8/3 \\
&= 11/3.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
Var(X) &= E(X^2) - \mu^2 \\
&= 11/3 - (4/3)^2 \\
&= 11/3 - 16/9 \\
&= 17/9.
\end{aligned}
$$

(iii) The Standard deviation is

$$\sigma_x = \sqrt{V(x)} = \sqrt{E(x-\mu)^2}$$
$$= \sqrt{17/9}$$
$$\approx 1.374.$$

Example 5.9: A number is chosen at random from the integers $1, 2, 3, ..., n$. If X denotes the chosen number, show that

$$E(X) = (n+1)/2 \text{ and } V(X) = (n-1)(n+1)/12.$$

Hint: The following identity may be useful:

$$1^2 + 2^2 + ... + n^2 = \frac{(n)(n+1)(2n+1)}{6}.$$

Solution

The integers $1, 2, 3, ..., n$ are equally likely and so the uniform probability distribution function for X, the chosen number is given by

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

(i) The expected value for X is given as

$$E(X) = \sum_x xp(x)$$
$$= \frac{1}{n}(1+2+3+...+n)$$
$$= \frac{1}{n}\left(\frac{1}{2}\right)n(n+1)$$
$$= \frac{n+1}{2}.$$

(ii) The variance is computed as follows

$$E(X^2) = \sum_x x^2 p(x)$$
$$= \frac{1}{n}(1^2 + 2^2 + ... + n^2)$$
$$= \frac{1}{n}\frac{(n)(n+1)(2n+1)}{6}$$
$$= \frac{(n+1)(2n+1)}{6}$$

Now, the variance is

$$Var(X) = E(X^2) - \mu^2$$
$$= \frac{(n+1)(2n+1)}{6} - \left(\frac{n+1}{2}\right)^2$$
$$= \frac{(n+1)(2n+1)}{6} - \frac{(n^2 + 2n + 1)}{4}$$

$$= \frac{2n^2+3n+1}{6} - \frac{n^2+2n+1}{4}$$

$$= \frac{4n^2+6n+2-3n^2-6n-3}{12}$$

$$= \frac{n^2-1}{12} = \frac{(n-1)(n+1)}{12}.$$

3.4 Properties of Variance of Discrete Random Variables

In section 4.2, we considered the properties of expectation of random variables and we saw that the expectation is a linear function since, if c is any constant, then $E(cX) = cE(X)$ and $E(X+c) = E(X)+c$. However, the variance is not linear, as we shall see in the following properties.

If X is any random variable and c is any constant, then

(i) $V(C) = E(C^2) - (E(C))^2$

$$= C^2 - C^2$$
$$= 0.$$

That is, the variance of a constant is zero, since it doesn't vary.

(ii) $V(cX) = c^2V(X)$.

That is, the variance of the product of a constant and a random variable is the product of the square of the constant and the variance of the random variable.

Proof:

$$\begin{aligned} V(CX) &= E((cX - c\mu)^2) \\ &= E(c^2X^2 - 2c^2X\mu + c^2\mu^2) \\ &= E(c^2(X-\mu)^2) \\ &= c^2E((X-\mu)^2) \\ &= c^2V(X). \end{aligned}$$

$V(X+c) = V(X)..$

That is, the variance of the sum of a constant and a random variable equals the variance of the random variable.

Proof: to prove this assertion, we note that to compute $V(X+c)$, we would replace X by $X+c$ and μ by $\mu+c$ in the equation given above for variance. Then the c's would cancel, leaving $V(X)$, as follows

$$
\begin{aligned}
V(X+c) &= \sum_x ((X+c)-(\mu+c))^2 p(x) \\
&= \sum_x (X-\mu)^2 p(x) \\
&= V(X).
\end{aligned}
$$

(iv) Let X and Y be two *independent* random variables. Then

$V(X+Y) = V(X) + V(Y)$.

That is, in the case of independent random variables, the variance of the sum is the sum of the variance.

Proof: Let $E(X) = \mu_X$ and $E(Y) = \mu_Y$. Then

$$
\begin{aligned}
V(X+Y) &= E((X+Y)^2) - [E(X)+E(Y)]^2 \\
&= E((X+Y)^2) - (\mu_X + \mu_Y)^2 \\
&= E(X^2) + 2E(X)E(Y) + E(Y^2) - (\mu_X{}^2 + 2\mu_X \mu_Y + \mu_Y{}^2) \\
&= E(X^2) + 2E(X)E(Y) + E(Y^2) - \mu_X{}^2 - 2\mu_X \mu_Y - \mu_Y{}^2 \\
&= E(X^2) - \mu_X{}^2 + E(Y^2) - \mu_Y{}^2 \\
&= V(X) + V(Y).
\end{aligned}
$$

3.5 Expectation and Variance of Continuous Random Variables

In this section we shall consider the properties of the expected value and the vaiance of a continuous random variable. These quantities are defined just as for discrete random variables and share the same properties.

3.5.1 Expected Value

Definition

Let X be a real-valued random variable with density function $f(x)$. The *expectedvalue* $\mu = E(X)$ is defined by

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

The properties of $E(X)$ can be summarized just as for the discrete random variables. That is, if c is any constant and X and Y are real-valued random variables then

$E(cX) = cE(X)$.

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

and so on.

These properties are proved in exactly the same way that the corresponding properties for discrete random variables were proved in section 4.3.

Example 5.10 Let X be a continuous random variable with range [0, 5] and density function

$f(x) = \frac{2x}{25}$.

Obtain the expected value $E(X)$.

Solution

The expected value is given as

$$E(X) = \int_0^5 xf(x)\,dx = \int_0^5 x\,\tfrac{2x}{25}\,dx$$
$$= \frac{2}{75}\left[x^3\right]_0^5 = \frac{250}{75}$$
$$= \frac{10}{3}.$$

That is, the expected value of X is $\frac{10}{3}$.

Example 5.11 If X is a continuous random variable with density function $f(x) = \frac{3}{64}x^2,\ 0 \le x \le 4$, find $E(X)$.

Solution

$$E(X) = \int_0^4 xf(x)\,dx = \int_0^4 x\left(\tfrac{3}{64}x^2\right)dx$$
$$= \frac{3}{64}\int_0^4 x^3 dx = \frac{3}{64}\left[\frac{x^4}{4}\right]_0^4 = 3.$$

That is $E(X) = 3$.

Example 5.12A continuous random variable X is uniformly distributed on the interval [0, 1]. Obtain the expected value $E(X)$.

Solution

Since X is uniformly distributed, the density function is given by

$f(x) = \frac{1}{1-0} = 1$.

Therefore

$$E(X) = \int_0^1 xf(x)\,dx = \int_0^1 x\,dx$$
$$= \left[\frac{x^2}{2}\right]_0^1 = \frac{1}{2}.$$

3.5.2 The Variance

Let X be a continuous random variable with density function $f(x)$. The variance is defined as

$$\sigma^2 = V(X) = E((X-\mu)^2).$$

This can further be simplified as

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

The properties of variance for continuous random variables can be summarized as follows. If X is a real-valued random variable defined on S, the sample space, and c is any constant, then

$$V(cX) = c^2 V(X).$$

$$V(X+c) = V(X).$$

If X and Y are independent then

$$V(X+Y) = V(X) + V(Y).$$

These properties are all proved in exactly the same way that the corresponding properties for discrete random variables were proved in section 4.4.

Example 5.13(example 4.11 contin'd). A continuous random variable X is uniformly distributed on the interval [0, 1]. Obtain the variance.

Solution

$$V(X) = E(X^2) - \mu^2.$$

Now,

$$\begin{aligned} E(X^2) &= \int_0^1 x^2 f(x)\,dx = \int_0^1 x^2 dx \\ &= \left[\frac{x^3}{3}\right]_0^1 = \frac{1}{3}. \end{aligned}$$

Then

$$\begin{aligned} V(X) &= E(X^2) - \mu^2 \\ &= \frac{1}{3} - \left(\frac{1}{2}\right)^2 = \frac{1}{12}. \end{aligned}$$

Example 5.14 Let X be a random variable with range [-1, 1] and density function $f(x)=1/2$ Obtain the mean μ and variance σ^2 of X .

Solution

The mean is given by

$$E(X)=\int_{-1}^{1} xf(x)\,dx=\int_{-1}^{1}\tfrac{1}{2}x\,dx$$

$$=\frac{1}{2}\left[\frac{x^2}{2}\right]_{-1}^{1}=\frac{1}{2}[0]=0.$$

That is, the mean $E(X)=0$.

To find the variance, we have

$$E(X^2)=\int_{-1}^{1} x^2 f(x)\,dx=\int_{-1}^{1}\tfrac{1}{2}x^2dx$$

$$=\frac{1}{2}\left[\frac{x^3}{3}\right]_{-1}^{1}=\frac{1}{2}\left[\frac{1}{3}+\frac{1}{3}\right]=\frac{1}{3}.$$

Therefore

$$V(X)=E(X^2)-\mu^2$$

$$=\tfrac{1}{3}-(0)^2=\tfrac{1}{3}.$$

Self Assessment Exercises

1. If a worn machine tool produces 1% defective parts and the parts produced are independent, find (i) the mean number of defective parts out of 25 (ii) the variance of the number of defective parts
   n = 25, p = 0.01, q = 0.99
2. Mrs Kofo, a business tycoon, ventured into poultry business ten times. If the success and failure rate of poultry is equally likely, find the probability of success twice. Compute the mean and standard deviation.
   n = 20, p = 1/, q = 0.99

Self Assessment Answers

4.0 Conclusion

This unit has acquainted you with finding the measure of central tendency for discrete random variable as well as continuos random variables. The proof of the properties was also established.

In the next unit, you will be introduced to discrete probability distribution, density function and the normal distribution that gives you the probability of every possible outcome of a discrete experiment, trial or observation.

5.0 Summary

In this unit, you have learnt the following:

1. Mathematical expectation and variance of discrete random rariables. Expected Value $\mu_x \; or \; \mu$ of a discrete random variable X ,with sample space S and probability function $p(x)$ is defined as:

$$E(X) = \sum_{x \in S} xp(x)$$

2. Properties of expected values of discrete random variables are:

(i) If C is any constant and X is a discrete random variable then,

$$E(C) = C.$$

(ii) The expected value of the product of a constant and a random variable is equal to the product of the constant and the expected value of the random variable, thus:

$$E(CX) = C\,E(X)$$

(iii) Sum of two random variables. If X and Y are random variables with finite expected values, then

$$E(X+Y) = E(X) + E(Y)$$

(iv) Sum of a constant and the random variable

$$E(X+C) = \sum_{i=1}^{n} (X+C) f(x_i)$$

$$= \sum_{i=1}^{n} Xf(x_i) + \sum_{i=1}^{n} C f(x_i)$$
$$= E(X) + C$$

(v) Independence, if X and Y are two independent random variables, then

$$E(X.Y) = E(X)E(Y).$$

3. A measure of deviation or spread of values is the variance and standard deviation of discrete random variables such that if X is a real-valued random variable with probability function $f(x)$ and expected value $\mu = E(X)$, then the variance of X , $V(X)$ or $Var(X)$) is:

$$V(X) = E((X-\mu)^2).$$

4. For continuous random variables, the expectation and variance is defined by:

(i) $\mu = E(X) = \int_{-\infty}^{\infty} xf(x)\,dx$

Where X is a real-valued random variable with density function $f(x)$.

The properties of $E(X)$ is similar as for the discrete random variables, i.e., if c is any constant and X and Y are real-valued random variables then,

$$E(cX) = cE(X).$$

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

The properties are exactly the same for the corresponding discrete random variables.

(ii) $\sigma^2 = V(X) = E((X-\mu)^2)$

Where X is a continuous random variable with density function $f(x)$.

This can be simplified as

$$V(X) = E((X-\mu)^2) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx$$
$$= \int_{-\infty}^{\infty} x^2 f(x)\,dx - 2\mu\int_{-\infty}^{\infty} xf(x)\,dx + \mu^2\int_{-\infty}^{\infty} f(x)\,dx$$
$$= \int_{-\infty}^{\infty} x^2 f(x)\,dx - \mu^2$$
$$= E(X^2) - \mu^2$$

The properties of variance for continuous random variables can be summarized as:

$$V(cX) = c^2 V(X).$$

$$V(X+c) = V(X).$$

Whereas, if X and Y are independent then

$$V(X+Y) = V(X) + V(Y).$$

The proof of these properties is exactly the same as the corresponding properties'proof for discrete random variables.

6.0 Tutor-Marked Assignment (TMA)

1. A coin is tossed three times. Let X be the number of heads that turn up. Calculate the variance and standard deviation of X.
2. A die is loaded so that the probability of a face coming up is proportional to the number on that face. The die is rolled with outcome X. Calculate the expected value and variance of X.

7.0 References/Further Reading

Harry Frank and Steven C. Althoen (1995): Statistics: concepts and applications. Cambridge Edition.

W. J. Decoursey (2003): Statistics and probability for engineering applications. Newness.

Sheldon M. Ross (1997): Introduction to probability models, sixth edition. AcademicPress. New York.

Mario Lefebre (2000): Applied probability and statistics. Springer.

http://www.stanford.edu/class/polisci100a/newprob2.pdf

http://www.stats.gla.ac.uk/steps/glossary/probability_distributions.html

Answers to SAAs

Module 2 Unit 1

1. $N = n_1.n_2 ._{...}.n_r .$
2. A possible outcome
3. 60
4. Combination
5. 65.5 ways

Module 2 Unit 2

1. Data model or random variables
2. Integer
3. Die or coin tossing, number of Aces in a deck of cards, etc.
4. Discrete probability function
5. i) $\frac{1}{6}$, ii) $\frac{1}{18}$
6. Cummulative distribution function (CDF) for a random variable.
7. Continuous set of values.
8. False
9. Density function
10. Cumulative distribution function of X
11. Monotonically non-decreasing.

Module 2 Unit 3

1. The expected value of X is $E(X) = \sum_{x \in S} xp(x)$
2. $\mu_x \; or \; \mu$
3. 3/4
4. (i)0.25 (ii)0.2475
5. (i)5 (ii)1.58

Module 3

Discrete Probability Distribution and Normal Distribution

Unit 1

Discrete Probability Distribution

Content

1.0 Introduction

A discrete probability distribution gives the probability of every possible outcome of a discrete experiment, trial or observation. Some special types of these distributions are considered in this chapter and their properties are also examined.

2.0 Learning Outcome

At the end of thi unit, you should be able to:

i. Perform chance experiments with two outcomes
ii. Use Bernoulli distribution to compute experimental data spread: Mean and Variance
iii. Computethe probability of independent successive experiments with probability of success p on each trial
iv. Determine the probability of occurrence of random events
v. Sample with and without replacement

3.0 Learning Outcomes

3.1 The Bernoulli Distribution

A Bernoulli trial is a performance of a chance experiment consisting of only two possible outcomes generally referred to as *success* and its complement (*failure*).

The probability of success is denoted by p while that of failure is denoted by $q = 1 - p$ Examples of such experiments include the following:

i) A toss of a coin whose two possible outcomes are *Head* (success) with probability $p = \frac{1}{2}$ and *Tail* (failure) with probability $q = 1 - p = \frac{1}{2}$.
ii) Asking a person if he/she will favour a particular candidate during an election.

The two possible outcomes here are *yes* (success) and *no* (failure).

Now, let a random variable X assume only the value 0 if the outcome is a failure or 1 if the outcome is a success. Then, X is known as a *Bernoullivariable* and its distribution $f(x)$ given by the table

X	0	1

$p(X = x)$ q p

which can also be expressed as

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

is called the ***Bernoulli Distribution.***

To show that this is really a distribution, we have

$$\sum_{x=0}^{1} f(x) = \sum_{x=0}^{1} p^x q^{1-x}$$
$$= (q+p) = 1 - p + p = 1.$$

The Mean is given by

$$\mu = E(X) = \sum_{x=0}^{1} x f(x) = \sum_{x=0}^{1} x p^x q^{1-x}$$
$$= 0.q + 1.p = p$$

The Variance is computed as

$$V(X) = E(X^2) - \mu^2$$
$$= \sum_{x=0}^{1} x^2 f(x) - \mu^2$$
$$= 0^2.q + 1^2.p - p^2$$
$$= p - p^2 = pq.$$

Therefore,

$$\mu = p, \quad \sigma^2 = pq.$$

3.2 The Binomial Distribution

Suppose a Bernoulli experiment with probability of success p on each trial is performed independently n times. For instance:

i) Tossing a coin ten times with heads and tails as the two possible outcomes on each trial and $\frac{1}{2}$ as the probability of head (success) on any one toss.

ii) Asking 1000 people randomly chosen from a population of eligible voters in an opinion poll so as to know if they will favour a particular candidate during an election for which the two outcomes are yes and no. The probability p of a yes answer (i.e., a success) indicates that proportion of people in the entire population that favour this candidate, etc.

Now let the number of successes in these n trials be counted by the random variable X. Then, the distribution of X, given by

$$b(n, p, x) = \binom{n}{x} p^x q^{n-x}, \qquad x = 0, 1, 2, \ldots, n$$

is called the ***Binomial Distribution.***

This distribution gives the probability of exactly x successes in a sequence of n independent Bernoulli trials with probability of success p on each trial.

It should be noted that for the binomial distribution to be applicable to an experiment, the following four conditions must be met:

i. There must be a fixed number n (>1) of repeated trials of the experiment.
ii. Each trial must have two possible outcomes- success or failure.
iii. The probability of success p must be the same on every trial.
iv. The trials must be independent of each other.

The general properties (2.1) and (2.2) are satisfied by this pdf since $0 \le p(x) \le 1$ and

$$\sum_{x=0}^{n} p(x) = \sum_{x=0}^{n} \binom{n}{x} p^x q^{n-x}$$

$$= q^n + \binom{n}{1} pq^{n-1} + \binom{n}{2} p^2 q^{n-2} + \ldots + p^n \text{ , (by binomial expansion)}$$

$$= (q+p)^n$$

$$= 1$$

3.2.1 Expectation and Variance of Binomial Distribution

The mean is given by

$$\mu_x = E(X) = \sum_{x=0}^{n} x \binom{n}{x} p^x q^{n-x}$$

$$= \sum_{x=1}^{n} x \frac{n(n-1)!}{x(x-1)!(n-x)!} p.p^{x-1} q^{n-x}$$

$$= np \sum_{x=1}^{n} \frac{(n-1)!}{(x-1)!(n-x)!} p^{x-1} q^{n-x}$$

Let $v = x-1$, so that $x = 1+v$ and substituting, we have

$$E(X) = np \sum_{v=0}^{n-1} \frac{(n-1)!}{v!(n-1-v)!} p^v q^{n-1-v}$$

$$= np \sum_{v=0}^{n-1} \binom{n-1}{v} p^v q^{n-1-v}$$

$$= np(p+q)^{n-1}$$

$$= np$$

Next we compute the variance as

$$V(X) = E(X^2) - \mu^2 = \sum_{x=0}^{n} x^2 \binom{n}{x} p^x q^{n-x} - (np)^2$$

$$= \sum_{x=1}^{n} x^2 \frac{n(n-1)!}{x(x-1)!(n-x)!} p.p^{x-1} q^{n-x} - (np)^2$$

$$= np\sum_{x=1}^{n} x\frac{(n-1)!}{(x-1)!(n-x)!}p^{x-1}q^{n-x} - (np)^2$$

Let $v = x-1$, so that $x = 1+v$ and substituting, we have

$$V(X) = np\sum_{v=0}^{n-1}\frac{(v+1)(n-1)!}{v!(n-1-v)!}p^{v}q^{n-1-v} - (np)^2$$

$$= np\sum_{v=0}^{n-1}\frac{v(n-1)!}{v!(n-1-v)!}p^{v}q^{n-1-v} + np\sum_{v=0}^{n-1}\frac{(n-1)!}{v!(n-1-v)!}p^{v}q^{n-1-v} - (np)^2$$

$$= np\left[\sum_{v=0}^{n-1}\frac{v(n-1)!}{v(v-1)!(n-1-v)!}p^{v}q^{n-1-v} + \sum_{v=0}^{n-1}\binom{n-1}{v}p^{v}q^{n-1-v}\right] - (np)^2$$

$$= np\left[\sum_{v=0}^{n-1}\frac{(n-1)!}{(v-1)!(n-1-v)!}p^{v}q^{n-1-v} + \sum_{v=0}^{n-1}\binom{n-1}{v}p^{v}q^{n-1-v}\right] - (np)^2$$

$$= np\left[\sum_{v=0}^{n-1}\frac{(n-1)(n-2)!}{(v-1)!(n-1-v)!}p.p^{v-1}q^{n-1-v} + \sum_{v=0}^{n-1}\binom{n-1}{v}p^{v}q^{n-1-v}\right] - (np)^2$$

$$= np\left[(n-1)p\sum_{v=0}^{n-1}\frac{(n-2)!}{(v-1)!(n-1-v)!}p^{v-1}q^{n-1-v} + \sum_{v=0}^{n-1}\binom{n-1}{v}p^{v}q^{n-1-v}\right] - (np)^2$$

$$= np\left[(n-1)p\sum_{v=0}^{n-1}\binom{n-2}{v-1}p^{v-1}q^{n-1-v} + \sum_{v=0}^{n-1}\binom{n-1}{v}p^{v}q^{n-1-v}\right] - (np)^2$$

$$= np\left[(n-1)p(p+q)^{n-2} + (p+q)^{n-1}\right] - (np)^2$$

$$= np\left[(n-1)p + 1\right] - (np)^2$$

$$= np[np - p + 1] - (np)^2 = np[np + 1 - p] - (np)^2$$

$$= np[np + q] - (np)^2$$

$$= [np]^2 + npq - (np)^2$$

$$= npq.$$

Therefore, $\mu_x = np$, $\sigma_x{}^2 = npq.$

Example 6.1 If a coin is tossed 4 times and we are interested only in the number of heads showing, what is the probability of obtaining exactly 0, 1, 2, 3, 4 heads?

Solution

In this case, n = 4, $p = \frac{1}{2}$, $q = 1 - p = \frac{1}{2}$. Then let X be the random variable denoting the number of heads that turn so that the distribution of X is $P(x) = \binom{4}{x}\left(\frac{1}{2}\right)^x\left(\frac{1}{2}\right)^{4-x}$, $\quad x = 0,1,2,3,4.$

Then,

$$P(X = 0) = \binom{4}{0}\left(\tfrac{1}{2}\right)^0\left(\tfrac{1}{2}\right)^{4-0} = 1/16.$$

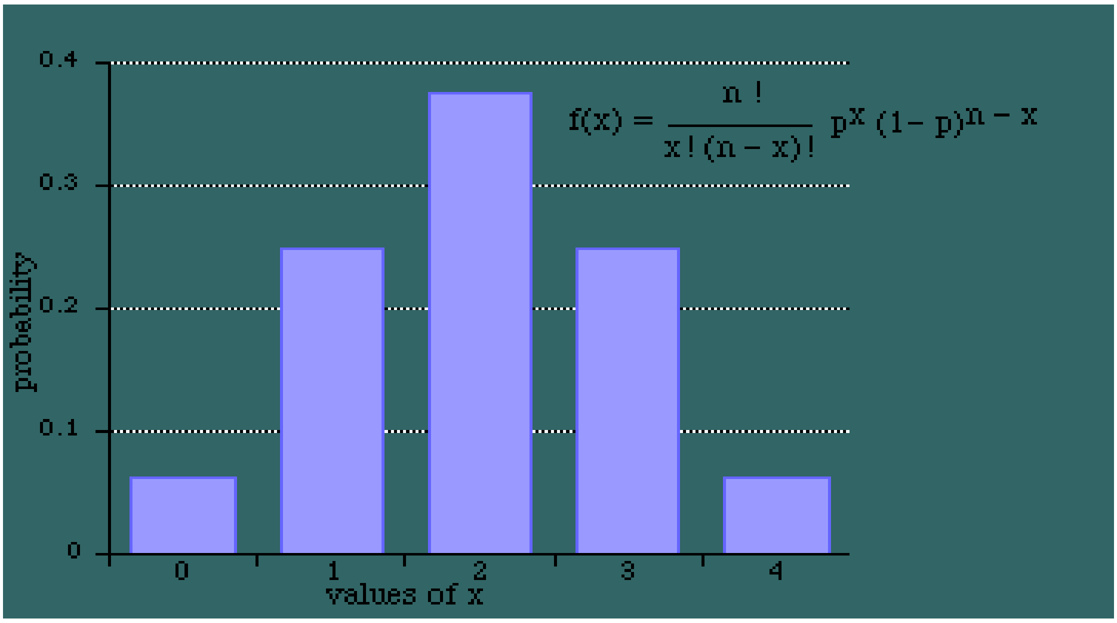
$$P(X = 1) = \binom{4}{1}\left(\tfrac{1}{2}\right)^1\left(\tfrac{1}{2}\right)^{4-1} = 4\left(\tfrac{1}{2}\right)^4 = 1/4.$$

$P(X=2) = \binom{4}{2}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^{4-2} = 6\left(\frac{1}{2}\right)^4 = 6/16$. and so on.

Thus we obtain 0, 1, 2, 3, or 4 heads with the respective probabilities

$\frac{1}{16}$, $\frac{1}{4}$, $\frac{6}{16}$, $\frac{1}{4}$, and $\frac{1}{16}$.

This distribution is represented by the bar chart below with p = ½.



Example 6.2 An ordinary die is thrown seven times. What is the probability of obtaining exactly three fives?

Solution

The probability of obtaining a five as success in one throw of the die is

$p(5) = \frac{1}{6}$, *and* $q = 1 - p(5) = \frac{5}{6}$, where q is the probability of *not obtaining a five*.

Therefore, in seven throws, we have

$p(exactly\ 3\ fives) = \binom{7}{3}\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^4$

= 0.078.

Self Assessment Exercises 1

1. A Bernoulli variable has ……………… Outcomes, namely …………………..
2. The binomial distribution is given by ……………………………………………
3. The mean of Bernoulli distribution is ……………………………………………
4. The variance of binomial distribution is …………………………………………

3.3 The Poisson Distribution

Poisson distribution is an important model for situations in which certain kinds of occurrences happen at random in a continuous time interval or region of space. A unit of time may be a minute, an hour, a day, a week etc, while a region of space may be a length, an area or a volume.

Examples of such occurrences include arrival of patients into a doctor's waiting room, the number of incoming telephone calls arising at an exchange in a particular minute, the number of errors on a page of printing, etc. Let the number of occurrences in a given time interval or region of space be counted by the random variable X, this shows that X can only take discrete values 0, 1, 2, 3, . . . The Poisson distribution gives the distribution of X.

Definition

A discrete random variable X is said to have a Poisson distribution with parameter λ if the probability function is given by

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

Where λ (>0) is the mean number of occurrences in a given time interval or region of space.

Based on certain assumptions, the Poisson distribution enables us to compute the probabilities of events such as "more than 10 phone calls occurring in a 5-minute time interval". The assumptions are that:

i. The average number (mean rate) of occurrences per minute or per unit space (denoted by λ) is a constant. This means, for instance, that in a given time interval of length 5 minutes, we would expect about 5λ occurrences.
ii. The number of occurrences in two non-overlapping (or disjoint) time intervals or regions of space are independent. For instance, the event that there are j telephone calls between 5:00 and 5:10 p.m. and k calls between 6:00 and 6:10 p.m. on the same day are independent.
iii. The probability of exactly one occurrence in a small interval of time or region of space is approximately proportional to the width of the interval or region.

It can easily be checked that the general properties (i) and (ii) are satisfied by this density function, i.e., that its values are nonnegative and sum to 1, since

$$0 \le f(x) \le 1$$

and

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

A very important relation that was used here is that

$$\sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = (1 + \lambda + \frac{\lambda^2}{2!} + \dots) = e^{\lambda}.$$

3.3.1 Expectation and Variance of Poisson Distribution

The mean of the Poisson distribution is given by

$$E(X) = \sum_{x=0}^{\infty} x \frac{\lambda^x}{x!} e^{-\lambda} = e^{-\lambda} \sum_{x=0}^{\infty} x \frac{\lambda^x}{x!}$$
$$= e^{-\lambda} \sum_{x=1}^{\infty} x \frac{\lambda^x}{x!} = e^{-\lambda} \lambda \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!}$$
$$= \lambda e^{-\lambda} \sum_{v=0}^{\infty} \frac{\lambda^v}{v!}, \quad (\text{where } v = x-1)$$
$$= \lambda e^{-\lambda} e^{\lambda} = \lambda.$$

The variance is derived as follows:

$$E[X(X-1)] = \sum_{x=0}^{\infty} x(x-1) f(x) = \sum_{x=1}^{\infty} x(x-1) \frac{\lambda^x}{x!} e^{-\lambda}$$
$$= \lambda^2 e^{-\lambda} \sum_{x=2}^{\infty} \frac{\lambda^{x-2}}{(x-2)!} = \lambda^2 e^{-\lambda} \sum_{r=0}^{\infty} \frac{\lambda^r}{r!}, \quad (\text{where } r = x-2)$$
$$= \lambda^2.$$

That is

$$E(X^2) - E(X) = \lambda^2, \text{ so that}$$
$$E(X^2) = \lambda^2 + \lambda.$$

But

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

That is, the Poisson distribution has mean and variance equal to λ.

Example 6.3

If the number of telephone calls an operator receives from time t_1 to t_2 follows a Poisson distribution with λ= 2, obtain the probability that the operator will

a. Not receive a phone call in the same time interval tomorrow.
b. Receive phone call twice.

Solution

Let the number of phone calls the operator receives in time interval t be denoted by X. Then the pdf of X is

$$f(x) = \frac{2^x e^{-2}}{x!},\ x = 0,1,\dots$$

(a) P(no phone call) = P(x = 0)

$$= \frac{2^0}{0!}e^{-2} = e^{-2} = 0.1353$$

(b) P(receives phone call twice) = P(x = 2)

$$= \frac{2^2}{2!}e^{-2} = 2e^{-2} = 0.2706.$$

Example 6.4: In a certain general hospital, arrivals of patients at a doctor's waiting room are assumed to follow Poisson distribution at an average rate of 120 per hour. What is the probability that in a particular minute interval

a. no patients arrive the room?
b. one or more patients will arrive?

Solution

The average number in a minute interval is λt. Here $\lambda = 120$ arrivals per hour, which, alternatively, equals 2 arrivals per minute and so $\lambda t = 2$, since $t = 1$ minute. Let the number of patients that arrive the doctor's room in time interval t be denoted by the random variable X. Then the pdf of X is

$$f(x) = \frac{2^x e^{-2}}{x!},\ x = 0,1,\dots$$

Therefore,

(a) P (no patients in a particular minute) $= P(X = 0) = e^{-2}\frac{2^0}{0!} = e^{-2}$

$= 0.14$.

(b) P(one or more patients in a particular minute)

= 1 – P (no patients)

=1 - 0.14 = 0.86.

Example 6.5:A typesetter makes, on the average, one mistake per 150 words. If he is setting a book with 100 words to a page, what is the probability that a seven-page pamphlet prepared by the typist contains more than two errors?

Solution

Here, $\lambda = 1$ mistake per 150 words. This, alternatively, can be given as $\lambda = 1/150$ mistakes per word. Now, let the number of words per page be denoted by w. Then, the average number of errors per page is $\lambda w = 100(1/150) = 0.67,$ since $w = 100$ words per page. So in a seven-page pamphlet, the mean number of errors is $7x0.67 = 4.7$.

Let the random variable X denotes the number of mistakes that the typesetter makes on a single page. Then the distribution of X is given as

$$P(X = x) = \frac{\lambda^x}{x!}e^{-\lambda}. = \frac{4.7^x}{x!}e^{-4.7}.$$

Now,

P (0 errors) = $\frac{4.7^0}{0!}e^{-4.7} = e^{-4.7} = 0.01$

P (1 error) = $\frac{4.7^1}{1!}e^{-4.7} = (4.7)(0.01) = 0.047.$

P (2 errors) = $\frac{4.7^2}{2!}e^{-4.7} = (11.045)(0.01) = 0.110.$

Therefore, P(pamphlet contains more than two errors)

= 1 – P(0 error) + P(1 error) + P (2 errors)

= 1 – (0.01 + 0.047 + 0.110)

= 1 - 0.167

= 0.833

The CDF of Poisson distribution is given by

$$F(x) = e^{-\lambda}\sum_{s \le x}\frac{\lambda^s}{s!}, \ x \ge 0.$$

3.3.2 Poisson as an Approximation to the Binomial Distribution

Poisson distribution is a convenient approximation of the binomial distribution in cases of a large number n of trials and a small probability p of success in a single trial. That is, the binomial distribution approaches the Poisson distribution for large n and small p. In fact, the degree of approximations improved as the number of observations increases.

Proof

Now, the mean of the Binomial distribution is derived above to be

$$\mu = np$$

from which

$$p = \frac{\mu}{n}, \text{ and so } p^x = \frac{\mu^x}{n^x}.$$

Also, $q^{n-x} = (1-p)^{n-x} = \left(1-\frac{\mu}{n}\right)^{n-x} = \left[1-\frac{\mu}{n}\right]^n \left\{1-\frac{\mu}{n}\right\}^{-x}.$

Substituting the above expressions for p^x and q^{n-x} in the binomial probability distribution function, we have

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x} = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}.$$

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

$$= \frac{n!}{x!(n-x)!} \frac{\mu^x}{n^x} \left[1-\frac{\mu}{n}\right]^n \left\{1-\frac{\mu}{n}\right\}^{-x}.$$

$$= \frac{\mu^x}{x!} \frac{n(n-1)\ldots(n-x+1)}{n^x} \left[1-\frac{\mu}{n}\right]^n \left\{1-\frac{\mu}{n}\right\}^{-x} \text{-------}(5.01)$$

Now, as $n \to \infty$, the expression

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

approaches 1, and so does the expression $\left\{1-\frac{\mu}{n}\right\}^{-x}$ for fixed x,

While $\left[1-\frac{\mu}{n}\right]^n$ can be expanded as

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

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

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

So that as $n \to \infty, \left(1-\frac{1}{n}\right) \to 1$ and

$$\left[1-\frac{\mu}{n}\right]^n \to 1-\mu+\frac{\mu^2}{2!}-\frac{\mu^3}{3!}+\ldots=e^{-\mu}.$$

Therefore, the RHS of (5.01) becomes

$$\frac{\mu^x}{x!}e^{-\mu},$$

which is the distribution of the Poisson random variable.

Example 6.6 If, in a class of national diploma in statistics students, 1% is the probability that a student will fail STA 122 in a semester examination, what is the probability that in a sample of 300 such students, exactly 5 will fail the course?

Solution

In this case, p is small and the number of trials n is large. Therefore, it is a binomial trial to be approximated by Poisson distribution and we have

$$\lambda = np = 300X\frac{1}{100} = 3.$$

$$\therefore P(x) = \frac{3^x}{x!}e^{-3}, \quad x = 0,1, \ldots$$

So, P (exactly 5) = P(x = 5) = $\frac{3^5}{5!}e^{-3} = 0.1008$

Therefore the probability that exactly 5 students will fail the course in a sample of 300 students is 0.1008.

Self Assessment Exercises 2

1. Poisson distribution is an approximation of the binomial distribution in cases where ……………………………………………..….
2. The margin of error deceases as the number of observations increases, True or False.

Self Assessment Answers 2

3.4 The Geometric Distribution

Consider a sequence of infinite number of independent Bernoulli trials with probability of success p on eah trial. For instance, a coin tossed an infinite number of times. Let the number of trials up to and including the first success be denoted by a random variable X so that

$$P(X = 1) = p,$$
$$P(X = 2) = qp,$$
$$P(X = 3) = q^2 p,$$

and in general,

$$P(X = n) = q^{n-1} p$$

That is, if $0 < p < 1$, and $q = 1 - p$, then we say that the random variable X has a *GeometricDistribution*if $P(X = x) = q^{x-1} p$ for $x = 1, 2, 3, \ldots$

The general properties are satisfied by this density function since

$0 \le f(x) \le 1$And

$$\sum_{x=1}^{\infty} f(x) = \sum_{x=1}^{\infty} q^{x-1} p = p \sum_{x=1}^{\infty} q^{x-1}$$

$$= p + pq + pq^2 + \ldots$$

The right-hand expression is just a geometric series with first term p and common ratio q, and so its limiting sum (sum to infinity) is given by

$$\lim_{n \to \infty} S_n = S_\infty = \frac{p}{(1-q)} = 1,$$ Since p = 1- q.

3.4.1 Expectation and Variance of Geometric Distribution

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

To derive the variance, we have

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

But, $\frac{d}{dq}\left(\frac{q}{(1-q)^2}\right) = \frac{1}{(1-q)^2} + \frac{2q}{(1-q)^3} = \frac{1}{p^2} + \frac{2q}{p^3}.$

Therefore,

$$\begin{aligned}
V(X) &= p\left(\frac{1}{p^2} + \frac{2q}{p^3}\right) - \left(\frac{1}{p}\right)^2 \\
&= \frac{2q}{p^2} + \frac{1}{p} - \frac{1}{p^2} \\
&= \frac{2q - 1 + p}{p^2} \quad = \frac{q}{p^2}
\end{aligned}$$

That is $\mu = \frac{1}{p},\ \sigma^2 = \frac{q}{p^2}$.

Note:

If $X \sim Geo(p)$, then

i. $p(X = r) = q^{r-1} p$

ii. $p(X \le r) = p\sum_{x=1}^{r} q^{x-1} = p\left[1 + q + q^2 + ... + q^{r-1}\right]$

$= p\dfrac{(1-q^r)}{(1-q)}$, sum of first n terms of a geometric series.

$= 1 - q^r$.

iii. $p(X > r) = 1 - p(X \le r) = q^r$.

Example 6.7 A coin is biased so that the probability of obtaining a head is 0.6. If the number of tosses up to and including the first head is a random variable denoted by X, find

a. $P(X \le 4)$.

b. $P(X > 5)$.

Solution

Given that p = 0.6, q = 0.4, then $f(x) = (0.6)(0.4)^{x-1},\ x = 1, 2, 3, ...$

a. $P(X \le 4) = 1 - (0.4)^4$, *(from (ii) above)*

$= 0.9744$.

b. From (iii), we have

$$P(X > 5) = 1 - P(X \le 5) = 1 - (1 - (0.4)^5)$$
$$= (0.4)^5$$
$$= 0.01024$$

Example 6.8 A random variable, K, has probability function given by

$$P(K = k) = \frac{4}{5}\left(\frac{1}{5}\right)^{k-1},\ k = 1, 2, 3, ...$$

$$P(K = k) = 0,\ \text{otherwise}.$$

$$\text{Given that } \sum_{n=1}^{\infty} np^n = \frac{p}{(1-p)^2},\ \text{find } E(K).$$

Solution

$$E(k)=\sum_{k=1}^{\infty}k\,f(k)=\sum_{k=1}^{\infty}k\left(\frac{4}{5}\right)\left(\frac{1}{5}\right)^{k-1}$$

$$=\left(\frac{4}{5}\right)\sum_{k=1}^{\infty}k\left(\frac{1}{5}\right)^{k-1}$$

$$=\frac{4}{5}\left(\frac{1}{(1-\frac{1}{5})^2}\right)=\frac{5}{4}$$

3.5 Hypergeometric Distribution

Suppose we have a set of N items, of which k are defective and $N-k$ are non-defective. If we draw n items from this set *withreplacement*, then we saw in section 3.2 that the probability that exactly x of the n items drawn are defective is given by the binomial distribution. Now, suppose that our sampling is without replacement and we denote by the random variable X, the number of defective items in our sample. Then, we have $\binom{N}{n}$ different samples of size n, and the total number of such samples with exactly x defective items is obtained by multiplying the number of ways of choosing x defective items from the set of k defective items and the number of ways of choosing $n-x$ non-defective items from the set of $N-k$ non-defective items. This is given as

$$\binom{k}{x}\binom{N-k}{n-x}$$

each with the probability $\dfrac{1}{\binom{N}{n}}$,

Hence, the distribution of X (the number of defective items in our sample), is given as

$$h(x;N,k,n)=\frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}},\quad x=0,1,2,\ldots$$

and is known as the Hypergeometric distribution.

Example 6.9 A box contains 20 screws, of which 5 are defective. A random sample of 5 screws is chosen to be inspected. Find the probability that the sample contains exactly one defective item.

Solution

Let N be the total number of screws in the box, k be the number of defective screws, and n be the size of the sample drawn. Let X be a random variable that gives the number of defective items in the sample. Then we have

$$P(X=1)=\frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}=\frac{\binom{5}{x}\binom{20-5}{5-x}}{\binom{20}{5}}=\frac{\binom{5}{1}\binom{15}{4}}{\binom{20}{5}}=0.44$$

Self Assessment Exercises 2

1. If we have a set of N items, of which k are defective then………… are non-defective.

Self Assessment Answers 2

4.0 Conclusion

This unit discussed how to perform chance experiments with two outcomes; success or failure with the corresponding probabilities of p and q respectively. Different types of experiments were considered, using the appropriate probability distribution techniques and worked examples for illustrations.

5.0 Summary

You have learnt the following in this unit:

i) The Bernoulli Distribution as a performance of chance experiments yielding only two possible outcomes: *success* (with prob., p) and its complement *failure* (of prob., q = 1- p). For a random variable X a *Bernoullivariable* and its distribution, $f(x)$, given by the table

X	0	1
$p(X=x)$	q	p

which can also be expressed as

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

The Mean and the variance are given by

$$\mu=p,\ \ \sigma^{2}=pq.$$

ii) The Binomial distributionis the probability of exactly x successes in a sequence of n independent Bernoulli trials with probability of success p on each trialis defined as:

$$b(n, p, x) = \binom{n}{x} p^x q^{n-x}, \quad x = 0, 1, 2, \ldots, n$$

The following **four** conditions must be met before applying the binomial distridution to sample sample data:

- There must be a fixed number n (>1) of repeated trials of the experiment.
- Each trial must have two possible outcomes- success or failure.
- The probability of success p must be the same on every trial.
- The trials must be independent of each other.

The mean and the variance are: *np* and *npq* respectively.

iii) The Poisson Distribution is used when certain the eventsoccur at random over time interval or region of space. The Poisson distribution is defined as:

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

Where λ(>0) is the mean number of occurrences in a given time interval or region of space.

The Poisson distribution has Mean and Variance equal to λ.

iv) The Geometric distribution is a sequence of infinite number of independent Bernoulli trials with probability of success p on eah trial such that, if the number of trials up to and including the first success is denoted by a random variable X then,

$$P(X = 1) = p,$$
$$P(X = 2) = qp,$$
$$P(X = 3) = q^2 p,$$

Generally,

$$P(X = n) = q^{n-1} p$$

Where, $0 < p < 1$, and $q = 1 - p$, then we say that the random variable

The Geometric Distribution Mean and Variance are:

$$\mu = \frac{1}{p}, \ \sigma^2 = \frac{q}{p^2}.$$

v) Hypergeometric distribution is defined as:

$$h(x;N,k,n)=\frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}},\quad x=0,1,2,\ldots$$

For a set of N items, of which k are defective and $N-k$ are non-defective, such that the probability that exactly x of the n items drawn are defective is given by the binomial distribution

6.0 Tutor-Marked Assignment (TMA)

1. A test is conducted and nine out of ten students in the class passed. If the same test is administered ten times, what is the probability that exactly nine passes will be recorded and one failure?
2. In an experiment, the probability of success, p, is 0.5. If the number of trials is n and the number of success is x, show that the binomial pdf for this experiment takes the simple form

$$f(x)=\tfrac{1}{2^n}\binom{n}{x}$$

Hence, find the mean, μ, and variance σ^2, of x in this case.

7.0 References/Further Reading

W. J. Decoursey (2003): Statistics and probability for engineering applications. Newness.

Harry Frank and Steven C. Althoen (1995): Statistics: Concepts and applications. Cambridge Edition.

Unit 2

Continuous Density Functions

Content

1.0 Introduction

Recall that a continuous random variable x , is one that can assume any value within some interval or intervals and that its corresponding probability distribution function $f(x)$, is valid only if it satisfies the properties.

In this unit, we are going to consider some other models such as the normal (or even) distribution, waiting–time (exponential) distribution as well as other distributions such as the Gamma and Beta distributions to fit the characteristics of other various experiments.

2.0 Learning Outcome

At the end of this unit, you should be able to:

i. Define a model for evenly distributed random variables
ii. Model waiting-time events
iii. Fit other experiments' behaviour with Gamma and Beta distributions

3.0 Learning Outcomes

3.1 The Uniform (or Rectangular distribution)

This distribution provides a model for continuous random variables that are *evenly distributed* (i.e., one that is just as likely to assume a value in one interval as it is to assume a value in any other interval of equal size).

A continuous random variable X is said to be uniformly distributed (or has rectangular distribution) in an interval $a \le X \le b$ if its density function is defined by

$$f(x) = \begin{cases} \frac{1}{b-a}, & a \le x \le b \\ 0, & otherwise \end{cases}$$

i.e., X~U (a, b)

Where the parameters a and b are real constants. This distribution has a graph that is a rectangle with base of length (b – a) and height $\frac{1}{b-a}$.

To show that this is really a distribution, we must prove that

$$\int_a^b f(x)dx = 1$$

$$\text{that is, } \frac{1}{b-a}\int_a^b dx = \frac{1}{b-a}[x]_a^b = 1.$$

The distribution function (cdf) is given by

$$F(t)=\int_a^t \frac{1}{b-a}\,dx=\frac{1}{b-a}[x]_a^t$$
$$=\frac{t-a}{b-a},\ \ a\le t\le b$$
$$\therefore\ \ F(x)=\frac{x-a}{b-a},\ \ a\le x\le b$$

$$F(b)=\frac{1}{b-a}\int_a^b dx=1,\ \text{as expected.}$$

Therefore we have :

$$\begin{cases} & x\le a & a\le x\le b & b<x \\ f(x) & 0 & 1/(b-a) & 0 \\ F(x) & 0 & (x-a)/(b-a) & 1 \end{cases}$$

3.1.1 Mean and Variance of Uniform Distribution

The mean is given by

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

We then compute the variance as

$$V(X)=E(X^2)-\mu^2=\frac{1}{b-a}\int_a^b x^2\,dx-\left(\frac{b+a}{2}\right)^2$$
$$=\frac{1}{b-a}\left[\frac{x^3}{3}\right]_a^b-\left(\frac{b+a}{2}\right)^2=\frac{1}{b-a}\left[\frac{b^3-a^3}{3}\right]_a^b-\left(\frac{b+a}{2}\right)^2$$
$$=\frac{(b^2+ab+a^2)(b-a)}{3(b-a)}-\left(\frac{b+a}{2}\right)^2$$
$$=\frac{b^2+ab+a^2}{3}-\left(\frac{b+a}{2}\right)^2=\frac{(b-a)^2}{12}.$$

That is,

$$\mu=\frac{b+a}{2},\ \ \sigma^2=\frac{(b-a)^2}{12}.$$

Example 7.1

A continuous random variable is such that X ~ U(3,6). Find

i. The pdf of X.
ii. E(X).
iii. Var(X).
iv. P(X > 5).

Solution:

i. The pdf of X is given by

$$f(x) = \frac{1}{b-a} = \frac{1}{3},\ 3 \le X \le 6$$

$$ii.\ \ E(X) = \frac{1}{3}\int_3^6 x\,dx = \frac{1}{3}\left[\frac{x^2}{2}\right]_3^6$$

$$= \frac{1}{6}[36-9] = \frac{27}{6} = 4.5$$

$$iii.\ \ Var(X) = E(X^2) - [E(X)]^2$$

$$E(X^2) = \frac{1}{3}\int_3^6 X^2\,dx = \frac{1}{3}\left[\frac{X^3}{3}\right]_3^6$$

$$= \frac{1}{9}(189) = 21.$$

$$\therefore Var(X) = 21 - (4.5)^2$$

$$= 0.75.$$

iv. $P(X > 5) = 1 - P(X \le 5).$

$$but\ P(X \le 5) = \int_3^5 \frac{1}{3}\,dx = \frac{1}{3}[X]_3^5 = \frac{2}{3}.$$

$$\therefore P(X > 5) = 1 - \tfrac{2}{3} = \frac{1}{3}.$$

Self Assessment Exercises

1. A uniformly distributedrandom variable X in an interval $a \le X \le b$ is defined by ……………………….
2. The mean and vaviance of uniform distribution are ……………………………..

3.2 The Exponential Distribution

This is also known as a waiting–time distribution. It is used to model situations where we observe a sequence of occurrences which occur at "random times". For example, we might be observing cars passing a milepost on a highway, or light bulbs burning out. In such cases, we might denote the time from one occurrence to the next one by a random variable X . Here, X is a continuous random variable whose range consists of the non-negative real numbers and which has an exponential distribution with the density function

$$f(x) = \lambda e^{-\lambda x},\ x \geq 0,$$
$$0,\ otherwise$$

Whereλ is a nonnegative real number and represents the reciprocal of the average value of X Thus if the average time between occurrences is 30 minutes, then $\lambda = \frac{1}{30}$.

To prove that this is a pdf, we have:

$$\int_0^\infty f(x)\,dx = \int_0^\infty \lambda e^{-\lambda x} dx$$
$$= \lambda \int_0^\infty e^{-\lambda x} dx = -e^{-\lambda x}\Big|_0^\infty = 1.$$

The distribution function (CDF) is given by

$$F(x) = \int_0^x \lambda e^{-\lambda t} dt$$
$$= \lambda \int_0^x e^{-\lambda t} = -e^{-\lambda t}\Big|_0^x$$
$$= 1 - e^{-\lambda x}, \qquad x \geq 0.$$

3.2.2 Mean and Variance of exponential distribution

The mean is given by

$$\mu = E(X) = \int_0^\infty x f(x) dx = \lambda \int_0^\infty x e^{-\lambda x} dx$$

Integrating by part using $\int u\,dv = uv - \int v\,du$, we have

$$E(X) = \lambda \int_0^\infty xe^{-\lambda x} dx = -xe^{-\lambda x}\Big|_0^\infty + \int_0^\infty e^{-\lambda x} dx$$

$$= 0 + \frac{e^{-\lambda x}}{-\lambda}\Big|_0^\infty = \frac{1}{\lambda}.$$

Similarly, the variance is given by

$$V(X) = E(X^2) - \mu^2 = \int_0^\infty x^2 f(x) dx - \frac{1}{\lambda^2}$$

$$= \lambda \int_0^\infty x^2 e^{-\lambda x} dx - \frac{1}{\lambda^2}$$

$$= -x^2 e^{-\lambda x}\Big|_0^\infty + 2\int_0^\infty xe^{-\lambda x} dx - \frac{1}{\lambda^2}$$

$$= -x^2 e^{-\lambda x}\Big|_0^\infty - \frac{2xe^{-\lambda x}}{\lambda}\Big|_0^\infty - \frac{2}{\lambda^2} e^{-\lambda x}\Big|_0^\infty - \frac{1}{\lambda^2}$$

$$= \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

That is, $E(X) = \frac{1}{\lambda}$, $V(X) = \frac{1}{\lambda^2}$.

Example 7.2A continuous random variable X is such that the pdf is given by

$$f(x) = 3e^{-3x}, \quad x \geq 0$$
$$0, \quad otherwise$$

Find a. $E(X)$ b. $Var(X)$ c. $P(X > 0.5)$.

Solution

$$(a)\ E(X) = \int_0^\infty Xf(x)\, dx = 3\int_0^\infty Xe^{-3x} dx$$

$$= 3\left[\left[\frac{Xe^{-3x}}{-3}\right]_0^\infty - \int_0^\infty \frac{e^{-3x}}{-3} dx\right]$$

$$= \int_0^\infty e^{-3x} dx = \frac{1}{3}$$

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

$$Now,\ E(X^2) = 3\int_0^\infty X^2 e^{-3x} dx$$

$$= 3\left[\left[-X^2 \frac{e^{-3x}}{3}\right]_0^\infty - 2\int_0^\infty X \frac{e^{-3x}}{-3} dx\right]$$

$$= 2\int_0^\infty Xe^{-3x} dx$$

$$= 2\left[X\frac{e^{-3x}}{-3} - \int_0^{\infty} \frac{e^{-3x}}{-3} dx \right] = \frac{2}{3}\int_0^{\infty} e^{-3x} dx$$

$$= \frac{2}{9}.$$

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

$$\frac{2}{9} - \frac{1}{9} = \frac{1}{9}.$$

$$(c) \;\; P(X > 0.5) = 3\int_{0.5}^{\infty} e^{-3x} dx = -\left[e^{-3x}\right]_{0.5}^{\infty}$$

$$= 0.223$$

Self Assessment Exercises

1. The exponential distribution is for modelling……………………………………..
2. What is the density function of exponential distribution ………………………..
3. The Mean and Variance of exponential distribution are ………………………..

Self Assessment Answers

3.3 The Gamma Distribution

A random variable X is said to have a gamma distribution with parameters α and β if the pdf f(x) is

$$f(x) = \frac{x^{\alpha-1}e^{-x/\beta}}{\beta^{\alpha}\,\Gamma\alpha}, \qquad 0 < x < \infty,\ \alpha > 0,\ \beta > 0$$

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

To show that the pdf is valid, we have

$$\int_0^{\infty} f(x)dx = \int_0^{\infty} \frac{1}{\beta^{\alpha}\,\Gamma\alpha} x^{\alpha-1}e^{-x/\beta} dx = \frac{1}{\beta^{\alpha}\,\Gamma\alpha}\int_0^{\infty} x^{\alpha-1}e^{-x/\beta} dx$$

$$\text{let } y = x/\beta$$

$$\Rightarrow x = \beta y \quad \text{and so, } dx = \beta dy$$

$$\therefore \quad \int_0^{\infty} f(x)dx = \frac{1}{\beta^{\alpha}\,\Gamma\alpha}\int_0^{\infty} (\beta y)^{\alpha-1} e^{-y}\beta\, dy$$

3.3.1 Mean and Variance of Gamma Distribution

To determine the mean and variance we consider $E(X^r)$, the rth moment about the origin.

$$\text{Now, } E(X^r) = \int_0^\infty X^r f(x)\,dx = \int_0^\infty X^r \frac{1}{\beta^\alpha \Gamma\alpha} X^{\alpha-1} e^{-x/\beta} dx$$

$$= \frac{1}{\beta^\alpha \Gamma\alpha} \int_0^\infty X^{\alpha+r-1} e^{-x/\beta} dx$$

$$= \frac{\beta^{\alpha+r}}{\beta^\alpha \Gamma\alpha} \int_0^\infty y^{\alpha+r-1} e^{-y} dy$$

$$= \frac{\beta^r}{\Gamma\alpha} \int_0^\infty y^{\alpha+r-1} e^{-y} dy$$

$$= \frac{\beta^r \Gamma(\alpha+r)}{\Gamma\alpha}$$

$$\text{Now, r} = 1 \Rightarrow E(X)$$

$$\therefore E(X) = \frac{\beta\Gamma(\alpha+1)}{\Gamma\alpha} = \frac{\beta\alpha\Gamma\alpha}{\Gamma\alpha}$$

$$= \alpha\beta.$$

$$\text{Also, r} = 2 \Rightarrow E(X^2)$$

$$\therefore E(X^2) = \frac{\beta^2\Gamma(\alpha+2)}{\Gamma\alpha}$$

$$= \frac{\beta^2(\alpha+1)(\alpha)\Gamma\alpha}{\Gamma\alpha}$$

$$= \beta^2\alpha(\alpha+1)$$

$$\text{But } Var(X) = E(X^2) - (E(X))^2$$

$$= \beta^2\alpha^2 + \alpha\beta^2 - \alpha^2\beta^2$$

$$= \alpha\beta^2.$$

3.4 The Beta distribution

A continuous random variable X is said to follow a Beta distribution if the pdf is defined as

$$f(x) = \frac{1}{\beta(m,n)} X^{m-1}(1-X)^{n-1}, \quad 0 < X < 1$$

$$0, \qquad \text{otherwise}$$

where m and n are the parameters of the distribution and

$$\beta(m,n) = \frac{\Gamma m \Gamma n}{\Gamma(m+n)}.$$

To verify that this pdf is valid, we have

$$\int_0^1 f(x)dx = \int_0^1 \frac{1}{\beta(m,n)} X^{m-1}(1-X)^{n-1}dx$$
$$= \frac{1}{\beta(m,n)} \int_0^1 X^{m-1}(1-X)^{n-1}dx$$
$$= \frac{\beta(m,n)}{\beta(m,n)} = 1, \text{ since the Beta function of parameter m and n is}$$
$$\beta(\mathrm{m,n}) = \int_0^1 X^{m-1}(1-X)^{n-1}dx.$$

3.4.1 Mean and Variance of Beta Distribution

To determine the mean and variance, we consider, as before, $E(X^r)$, the r^{th} moment about the origin. That is

$$E(X^r) = \frac{1}{\beta(m,n)} \int_0^1 X^r X^{m-1}(1-X)^{n-1}dx$$
$$= \frac{1}{\beta(m,n)} \int_0^1 X^{m+r-1}(1-X)^{n-1}dx$$
$$= \frac{\beta(m+r,n)}{\beta(m,n)}$$
$$= \frac{\Gamma(m+r)\Gamma n}{\Gamma(m+r+n)} \cdot \frac{\Gamma(m+n)}{\Gamma m \Gamma n}$$
$$= \frac{\Gamma(m+r)\Gamma(m+n)}{\Gamma(m+r+n)\Gamma m}$$

$$Now, r = 1 \Rightarrow E(X)$$
$$\therefore E(X) = \frac{\Gamma(m+1)\Gamma(m+n)}{\Gamma(m+n+1)\Gamma m} = \frac{m\Gamma m \Gamma(m+n)}{(m+n)\Gamma(m+n)\Gamma m}$$
$$= \frac{m}{m+n}.$$

$$\text{Also}, r = 2 \Rightarrow E(X^2).$$
$$\therefore E(X^2) = \frac{\Gamma(m+2)\Gamma(m+n)}{\Gamma(m+n+2)\Gamma m} = \frac{(m+1)m\Gamma m \Gamma(m+n)}{(m+n+1)(m+n)\Gamma(m+n)\Gamma m}$$
$$= \frac{m(m+1)}{(m+n)(m+n+1)}$$

But $Var(X) = E(X^2) - (E(X))^2$

$$= \frac{m(m+1)}{(m+n)(m+n+1)} - \left(\frac{m}{m+n}\right)^2$$

$$= \frac{mn}{(m+n+1)(m+n)^2}\ .$$

Self Assessment Exercises

1. What is the pdf of Gamma distribution?
2. What is the Mean and Variance of Gamma Distribution?
3. State the Beta distribution pdf.
4. What is the Mean and Variance of Beta Distribution?

Self Assessment Answers

4.0 Conclusion

This unit advanced further in discussing how to develop models for continuous random variables and defined its corresponding probability distribution function $f(x)$, here, you learnt about the normal (or even) distribution, waiting–time (exponential) distribution as well as the Gamma and Beta distributions to fit the features of various random variables.

5.0 Summary

In this unit, you have learnt:

i) The uniform (or rectangular distribution), which provides a model for continuous random variables that are evenly distributed, defined as:

$$f(x) = \begin{cases} \frac{1}{b-a}, & a \le x \le b \\ 0, & otherwise \end{cases}$$

For a continuous random variable, X in an interval $a \le X \le b$, where ameters *a* and *b* are real constants.

The mean and variance are given by:

$$\mu = \frac{b+a}{2},\ \ \sigma^2 = \frac{(b-a)^2}{12}.$$

ii) The exponential distributionis used to model situations where events sequence occur at randomly like the arrival of passengers at a bus park. The exponential distributionis defined by the pdf: $\begin{matrix} f(x) = \lambda e^{-\lambda x}, \; x \geq 0, \\ 0, \; otherwise \end{matrix}$ where, X is a continuous random variable whose range consists of the non-negative real numbers and λ is a nonnegative real number.

The Mean and Variance are: $E(X) = \frac{1}{\lambda}, \; V(X) = \frac{1}{\lambda^2}$

iii) The Gamma distribution of a random variable X is given as:

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

The Mean and Variance are: $E(X) = \alpha\beta$

$Var(X) = \alpha\beta^2.$

iv) A Beta distribution is one with a continuous random variable, X and the pdf defined as:

$$f(x) = \begin{cases} \dfrac{1}{\beta(m,n)} X^{m-1}(1-X)^{n-1}, & 0 < X < 1 \\ 0, & \text{otherwise} \end{cases}$$

where *m* and *n* are the parameters of the distribution.

The Mean and Variance are: $E(X) = \frac{m}{m+n}, \; Var(X) = \frac{mn}{(m+n+1)(m+n)^2}$.

6.0 Tutor-Marked Assignment (TMA)

A continuous random variable X is such that the pdf is given by

$$f(x) = \begin{cases} 3e^{-3x}, & x \geq 0 \\ 0, & otherwise \end{cases}$$

Find a. $E(X)$ b. $Var(X)$ c. $P(X > 0.5)$.

7.0 References/Further Reading

W. J. Decoursey (2003): Statistics and probability for engineering applications. Newness.

Harry Frank and Steven C. Althoen (1995): Statistics: Concepts and applications. Cambridge Edition.

Unit 3

The Normal Distribution

Content

1.0 Introduction

We are now going to turn to one of the most important of all probability distributions — the normal distribution. We shall outline its properties, discuss the standardization of normal random variable, the computation of cumulative distribution for random variates and how to use the normal distribution to estimate the the binomial distribution., especially when the number of trials is large.

2.0 Learning Outcome

At the end of this uniy, you should be able to:”

i. Specify the normal distribution
ii. State the properties of the normal distribution
iii. Standardize normal random variable
iv. Compute the value of a cumulative distribution function for random variates
v. Use the normal distribution to approximate the Binomial distribution

3.0 Learning Outcomes

3.1 The Normal Distribution

This is the most important density function. A continuous random variable, X , is said to be normally distributed if its density function is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}, \qquad -\infty < x < \infty$$
$$0, \qquad \text{otherwise.}$$

Where the parameter μ represents the “center” of the density; the parameter σ (assumed to be positive) is a measure of the “spread” of the density.

The graph of the normal distribution is given below.

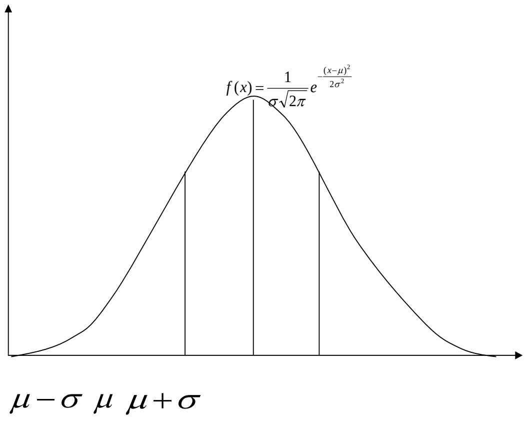


Figure 7.1: Normal distribution curve.

Properties of Normal Distribution

i. The mean and the variance are respectively, μ and σ^2.
ii. The total area under the curve and above the horizontal axis is equal to 1.
iii. The mean, median and mode all coincide in the center.
iv. The curve is symmetrical about the mean; the coefficient of skewness is zero.

3.2 The Standard Normal Distribution

Let Z be a normal random variable with parameters $\mu = 0$ and $\sigma = 1$. A normal random variable with these parameters is said to be a *standard* normal random variable. The process of changing a normal random variable to a standard normal random variable is known as *standardization*. This process involves computing the number of standard deviations the normal random variable concerned is away from the mean μ. If X has a normal distribution with paramneters μ and σ and if

$$Z = \frac{X-\mu}{\sigma},$$

Then, Z is said to be the standardized version of X.

Also,

$$E(Z) = E\left(\frac{X-\mu}{\sigma}\right) = 0, \text{ since } \mathrm{E(X)} = \mu$$

and

$$Var(Z) = Var\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma^2}Var(X-\mu) = 1,$$

as expected.

Definition

A normal random variable z is said to be a standard normal random variable if its density function is given by

$$f(z) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}, \quad -\infty < z < \infty.$$

3.3 The Cumulative Distribution Function for the Normal Random Variable

Let X be a normal random variable with parameters μ and σ. Suppose we wish to calculate the value of a cumulative distribution function for X, then this calculation can be reduced to one concerning the standard normal random variable Z as follows:

$$\begin{aligned}F_X(x) &= P(X \le x)\\ &= P\left(Z \le \frac{x-\mu}{\sigma}\right)\\ &= \phi\left(\frac{x-\mu}{\sigma}\right) = \phi(z).\end{aligned}$$

Note that this last expression can be found in a table of values of the cumulative distribution function for a standard normal random variable.

Example 8.1

A normally distributed random variable X has parameters $\mu = 10$ and $\sigma = 3$. Find the probability that X is between 4 and 16.

Solution

Here we compute the standardized version of X, which is

$$Z = \frac{X-10}{3}.$$

So, we have

$$\begin{aligned}P(4 \le X \le 16) &= P(X \le 16) - P(X \le 4)\\ &= P\left(Z \le \tfrac{16-10}{3}\right) - P\left(Z \le \tfrac{4-10}{3}\right)\\ &= \phi\left(\tfrac{16-10}{3}\right) - \phi\left(\tfrac{4-10}{3}\right)\\ &= \phi(2) - \phi(-2)\\ &= 2\phi(2) - 1\\ &= 0.9544.\end{aligned}$$

Example 8.2

In a semester examination, the scores of 400 students in a department are normally distributed with mean 56 and variance 25. What is the number of students having 60 marks and above to the nearest whole number?

Solution

Let X be the normal random variable denoting the exam scores, with parameters $\mu = 56$ and $\sigma = 5$. Then, the probability of students having 60 marks and above is

$$P(X \geq 60) = P(Z \geq \frac{60-56}{5}) = P(Z \geq 0.8)$$
$$= 1 - P(Z \leq 0.8)$$
$$= 1 - \phi(0.8)$$
$$= 1 - 0.7881$$
$$= 0.2119$$

Therefore, the number of students having 60 marks and above, to the nearest whole number, is

$$21.19\% \; of \; 400 = 85$$

Example 8.3

A manufacturer knows from experience that the resistance of the resistors he produces is normally distributed with mean $\mu = 100\,\text{ohms}$ and standard deviation $\sigma = 2$ ohms. What percentage of resistors will have resistance between 98 ohms and 102 ohms?

Solution

Let X be the normal random variable denoting the resistance of the resistors, with parameters $\mu = 100\,\text{ohms}$ and $\sigma = 2$ ohms.

Then, the probability of resistors with resistance between 98 ohms and 102 ohms is

$$P(98 \leq X \leq 102) = P(X \leq 102) - P(X \leq 98)$$
$$= \phi\left(\frac{102-100}{2}\right) - \phi\left(\frac{98-100}{2}\right)$$
$$= \phi(1) - \phi(-1)$$
$$= 2\phi(1) - 1$$
$$= 2(0.8413) - 1$$
$$= 0.6826$$
$$= 68.26\%.$$

Therefore, the percentage of the resistors with their resistance between 98 and 102 ohms is 68.26%.

Self Assessment Exercises

1. Specify the pdf of a normal distribution.
2. State the four properties of a normal distribution.
3. The standard normal distribution of a variate X with Mean, μ and variance, σ is defined as………………………………

3.4 The Normal Approximationto the Binomial Distribution

Under certain circumstances, the normal distribution can be used to give a useful approximation of the Binomial distribution when the number of trials, n, is large.

Recall that if a discrete random variable, X is such that $X \sim Bin(n, p)$, then the probability function of X is

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, \ldots, n$$

Now if $0 < p < 1$, then for large n the normal distribution can be used as an approximation to the binomial distribution with mean, $\mu = np$, and variance, $\sigma^2 = npq$. That is,

if $X \sim Bin(n, p)$, we have

$E(X) = np$

and $Var(X) = npq$

then for large n and $0 < p < 1$,

$X \sim N(np, npq)$ approximately

The density function of X is then given by

$$f(x) = \frac{1}{\sqrt{2\pi}\sqrt{npq}} e^{-z^2/2}, \qquad \text{where } z = \frac{x - np}{\sqrt{npq}}$$

The practical advantage of this approximation is that calculations are much less tedious to perform.

Example 8.4 If 10% of the screws produced in a factory are defective, what is the probability that in a random sample of 1000 screws

(a) less than 80 are defective? (b) between 90 and 115 inclusive are defective?

Solution

Let X be the random variable, ‘the number of defective screws when a random sample of 1000 screws is selected’.

Let ‘success’ be ‘obtaining a defective screw’.

Then $X \sim Bin(n, p)$ where $n = 1000$ and $p = \frac{1}{10}$, so $X \sim Bin(1000, \frac{1}{10})$.

Now n is large and p is not too small, so we use the normal approximation:

$X \sim N(np, npq)$ Where $np = (1000)(\frac{1}{10}) = 100$

$npq = (1000)(\frac{1}{10})(\frac{9}{10}) = 90$

StandardDeviation $= 3\sqrt{10}$

$\therefore X \sim N(100, 90)$.

(a) $P(X < 80) \rightarrow P(X < 79.5)$ (correction for continuity)

$$\begin{aligned} P(X < 79.5) &= P\left(\frac{X-100}{3\sqrt{10}} < \frac{79.5-100}{3\sqrt{10}}\right) \\ &= P(Z < -2.160) \\ &= 1 - \phi(2.160) \\ &= 0.0154. \end{aligned}$$

The probability of obtaining less than 80 defectives = 0.0154.

(b) $P(90 \le X \le 115) \rightarrow P(89.5 < X < 115.5)$ (correction for continuity)

$$\begin{aligned} P(89.5 < X < 115.5) &= P\left(\frac{89.5-100}{3\sqrt{10}} < \frac{X-100}{3\sqrt{10}} < \frac{115.5-100}{3\sqrt{10}}\right) \\ &= P(-1.107 < Z < 1.634) \\ &= \phi(1.634) + \phi(1.107) - 1 \\ &= 0.8145 \end{aligned}$$

The probability of obtaining between 90 and 115 inclusive = 0.8154.

Example 8.5 Find the probability of obtaining more than 110 ones in 400 tosses of an unbiased tetrahedral die with faces marked 1, 2, 3 and 4.

Solution

Let X be the random variable: 'the number of *ones* when an unbiased tetrahedral die is tossed'.

Let 'success' be 'obtaining a *one'*.

Then $X \sim Bin(n, p)$ where $n = 400$ and $p = \frac{1}{4}$, so $X \sim Bin(400, \frac{1}{4})$.

Since n is large and p is not too small, we use normal approximation:

$X \sim N(np, npq)$ where $np = 400(\frac{1}{4}) = 100$

$npq = (400)(\frac{1}{4})(\frac{3}{4}) = 75.$

standard deviation = $5\sqrt{3}$

$\therefore X \sim N(100, 75).$

To find the probability of obtaining more than 110 ones, we require

$P(X > 110) \rightarrow P(X > 110.5)$ (correction for continuity)

$$P(X > 110.5) = P\left(\frac{X - 100}{5\sqrt{3}} > \frac{110.5 - 100}{5\sqrt{3}}\right)$$
$$= P(Z > 1.212)$$
$$= 1 - \phi(1.212)$$
$$= 0.1127.$$

The probability of obtaining more than 110 ones = 0.1127.

Self Assessment Exercises

1. The normal distribution can be used to approximate the Binomial distribution when the number of trials, n, is small. True or false
2. Normal approximation of the binomial distribution is cumbersome. True or False.

Self Assessment Answers

4.0 Conclusion

This unit discussed the normal distribution, outlined its properties, as well as the standardization of normal random variable, the computation of cumulative distribution for random variates and how to use the normal distribution to estimate the the binomial distribution., especially when the number of trials is large. In addition, worked examples were used to illustrate the calculation of a cumulative distribution function for random variates.

5.0 Summary

You have learnt the following in this unit:

i) The Normal distribution as the most important density function, its pdf is defined as:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}, \qquad -\infty < x < \infty$$

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

Where μ is the "center" of the density and σ (assumed to be positive) is the "spread" of the density.

The properties of the nomal distribution are:

- The mean and the variance are respectively, μ and σ^2.
- The total area under the curve and above the horizontal axis is equal to 1.
- The mean, median and mode all coincide in the center.
- The curve is symmetrical about the mean; the coefficient of skewness is zero

ii) The standard normal distribution, Z for a normal random variable with parameters $\mu = 0$ and $\sigma = 1$ is

$$Z = \frac{X - \mu}{\sigma},$$

iii) The normal distribution can be used to obtain a helpful approximation of the Binomial distribution when the number of trials, n, is large.

6.0 Tutor-Marked Assignment (TMA)

1. If 10% of the screws produced in a factory are defective, what is the probability that in a random sample of 1000 screws
   (a) Less than 80 are defective? (b) Between 90 and 115 inclusive are defective?
2. Find the probability of obtaining more than 110 ones in 400 tosses of an unbiased tetrahedral die with faces marked 1, 2, 3 and 4.

7.0 References/Further Reading

Harry Frank and Steven C. Althoen (1995): Statistics: Concepts and applications. Cambridge Edition.

Unit 4

Joint Probability Function

Content

1.0 Introduction

Quite often an experiment is performed on pairs or vectors of random variables, whereby multiple variables are being investigated simultaneously and are defined on the same sample space. In this unit, we shall seek to deduce a model for such two functions, in what is called their *joint* distribution. Covariance shall be discussed, which measures how variables are related.

2.0 Learning Outcome

At the end of this uniy, you should be able to:”

i. Deduce a model for two joint probability functions
ii. Measure the Covariance of two joint random variables
iii. Find correlation between the joint random variables

3.0 Learning Outcomes

3.1 Joint Probability Function

It is often the case that when an experiment is performed, pairs or vectors of random variables are investigated. In this case, we consider these variables simultaneously and defined them on the same sample space. Though each member of a pair (X,Y) could be studied individually as a random variable, certain interesting and relevant relationships may exist between them, which can only be analyzed in the context of a model for the two together, in what is called their *joint* distribution.

This joint distribution is defined by a joint probability function that gives the probability of each possible pair of values:

$$f(x,y) = P(X = x,\ Y = y).$$

Definition

Let X and Y be two discrete random variables defined on the same sample space S, where X takes the values $x_1, x_2, ..., x_k$ and Y takes the values $y_1, y_2, ..., y_h$. Then the function $f(x,y)$ such that

$$f(x_i, y_j) = P(X = x_i,\ Y = y_j)$$

is called the joint probability function of X and Y if it satisfies the following conditions

(i) $f(x,y) \geq 0$ $\quad (ii)$ $\sum_x \sum_y f(x,y) = 1.$

The distributions of the individual random variables are called *MarginalDistributions* and are used to make probability statements that involve any one of the variables without regard to the value of the other variable.

From the joint distribution above, the marginal distributions of X and Y are respectively:

$$f_1(x_i) = P(X = x_i) = \sum_{j=1}^{h} f(x_i, y_i)$$

and

$$f_2(y_j) = P(Y = y_j) = \sum_{i=1}^{k} f(x_i, y_j).$$

For continuous random variables X and Y, the joint density function is obtained by analogy with the discrete case on replacing sums by integrals.

Thus, $f(x, y)$ is called the *Joint Density Function* of X and Y if it satisfies the conditions below

$$(i)\quad f(x, y) \geq 0 \qquad\qquad (ii)\quad \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)dxdy = 1.$$

The marginal density functions are

$$f_1(x) = \int_{-\infty}^{\infty} f(x, y)dy$$

and

$$f_2(y) = \int_{-\infty}^{\infty} f(x, y)dx.$$

Example 9.1 The table below is on the numbers of those who do or do not smoke and those who do or do not have cancer in a group of 60 people.

		smoke		
		No	Yes	Total
Cancer	No	40	10	50
	Yes	7	3	10
	Total	47	13	60

Obtain the joint distribution of $\{CS\}$ and the marginal distributions.

Solution

The joint distribution of $\{CS\}$ is given in the table below with $S=1$ if a person smokes and 0 if not, and $C=1$ if a person has cancer and 0 if not.

		S	
		0	1
C	0	40/60	10/60
	1	7/60	3/60

In this table for example we have $P(C=0, S=0)=40/60,\ P(C=0, S=1)=10/60,$ and so on.

The Marginal Distributions are

$$P(C)=\begin{pmatrix} 0 & 1 \\ 50/60 & 10/60 \end{pmatrix}$$

and

$$P(S)=\begin{pmatrix} 0 & 1 \\ 47/60 & 13/60 \end{pmatrix}.$$

Example 9.2 Let the random variables X and Y have the joint distribution given in the table below.

		Y		
		1	2	3
X	1	1/12	1/6	1/12
	2	1/6	0	1/6
	3	0	1/3	0

Table 7.3: Joint distribution of X and Y

(a) Find the marginal probability functions of X and Y.

(b) Find $P(X \geq 2, Y \leq 2)$.

Solution

(a) From the table, the marginal probabilities are respectively

$$P(X) = \begin{pmatrix} 1 & 2 & 3 \\ 1/3 & 1/3 & 1/3 \end{pmatrix}$$

and

$$P(X) = \begin{pmatrix} 1 & 2 & 3 \\ 1/4 & 1/2 & 1/4 \end{pmatrix}$$

(b) We see from the joint distribution table above that

$$\begin{aligned} P(X \geq 2, Y \leq 2) &= \sum_{x \geq 2} \sum_{y \leq 2} p(x, y) \\ &= 1/6 + 0 + 0 + 1/3 \\ &= 1/2. \end{aligned}$$

Example 9.3 If the joint density function of two continuous random variables X and Y is

$$f(x, y) = kxy, \quad 0 \leq x \leq 3, \ \ 1 \leq y \leq 4.$$
$$0, \quad \text{otherwise}$$

Find

(a) The value of the Constant k.

(b) $P(1 \leq X \leq 3, \ \ 0 \leq Y \leq 3)$

(c) $P(X \geq 1, \ Y \leq 2)$.

(d) The marginal density function of X.

(e) The marginal density function of Y

Solution

(a) Since $f(x, y)$ is a density function, we must have

$$\int_0^3 \int_1^4 kxy\,dx\,dy = 1.$$

That is, $k \int_0^3 \int_1^4 xy\,dx\,dy = k \int_{x=0}^{3} \left[\int_{y=1}^{4} xy\,dy \right] dx = 1.$

$$\Rightarrow \ k \int_{x=0}^{3} \left[\frac{xy^2}{2} \right]_1^4 dx = k \int_{x=0}^{3} \left[\frac{16x - x}{2} \right] dx = k \left[\frac{15x^2}{4} \right]_0^3 = \frac{135}{4} k = 1.$$

$$\therefore k = \frac{4}{135}.$$

Hence,

$$f(x,y) = \frac{4xy}{135}, \quad 0 \le x \le 3, \ \ 1 \le y \le 4.$$
$$0, \ \text{otherwise}$$

(b) $P(1 \le X \le 3, \ \ 0 \le Y \le 3) = P(1 \le X \le 3, 1 \le Y \le 3)$

$$= \int_{x=1}^{3}\int_{y=1}^{3} \frac{4}{135} xydxdy = \frac{4}{135}\int_{x=1}^{3}\left[\int_{y=1}^{3}(xy)dy\right]dx$$

$$= \frac{4}{135}\int_{x=1}^{3}\left[\frac{xy^2}{2}\right]_1^3 dx \ = \frac{4}{135}\int_{x=1}^{3} 4xdx = \frac{4}{135}\left[2x^2\right]_1^3$$

$$\therefore P(1 \le X \le 3, 1 \le Y \le 3) = \frac{64}{135}.$$

(c) $P(X \ge 1, Y \le 2) = P(1 \le X \le 3, \ 1 \le Y \le 2)$

$$= \frac{4}{135}\int_{x=1}^{3}\int_{y=1}^{2} xydxdy = \frac{4}{135}\int_{x=1}^{3}\left[\frac{xy^2}{2}\right]_1^2 dx$$

$$= \frac{4}{135}\int_{x=1}^{3} \frac{3x}{2} dx = \frac{4}{135}\left[\frac{3x^2}{4}\right]_1^3$$

$$= \frac{4}{135}\left[\frac{27-3}{4}\right] = \frac{24}{135} = \frac{8}{45}.$$

$$\therefore P(X \ge 1, \ \ Y \le 2) = \frac{8}{45}.$$

(d) The marginal density function of X is given by

$$f_1(x) = \int_{y=1}^{4} \frac{4}{135} xydy = \frac{4}{135}[xy^2]_1^4$$
$$= \frac{4}{135}\frac{(15x)}{2} = \frac{2x}{9}.$$
$$\therefore \ f_1(x) = \frac{2}{9}x, \ \text{for} \ 0 \le x \le 3$$

(e) The marginal density function of Y is given by

$$f_2(y) = \int_{x=0}^{3} \frac{4}{135} xydx = \frac{4}{135}\left[\frac{x^2 y}{2}\right]_0^3$$
$$= \frac{4}{135}\frac{(9y)}{2} = \frac{2y}{15}.$$
$$\therefore \ f_2(y) = \frac{2}{15}y, \ \text{for} 1 \le y \le 4$$

Self Assessment Exercises

1. Pairs or vectors of random variables are investigated concurrently with.........
2. A marginal distributions of continuous random variables X and Y, is defined as.........

Self Assessment Answers

3.2 Conditional Probability Function

If X and Y are two continuous random variables with joint probability density function $f(x, y)$ and the value of one of the variables becomes known, then, probability statements involving the other variable should be conditional on what is known. The conditional density function of the random variable Y given that X has the value x is given as

$$f(y/x) = \frac{f(x, y)}{f_1(x)}.$$

Similarly, that of X given Y is

$$f(x/y) = \frac{f(x, y)}{f_2(y)},$$

Where $f_1(x)$ and $f_2(y)$ are the marginal probability density functions of X and Y respectively.

For the discrete case the conditional probability of Y given X is

$$P(Y = y_i / X = x_j) = \frac{P(x_j, y_i)}{P(x_j)},$$

and that of X given Y is

$$P(X = x_j / Y = y_i) = \frac{P(x_j, y_i)}{P(y_i)}$$

Where $P(x_j)$ and $P(y_i)$ are the marginal probabilities of X and Y respectively?

3.3 Independent Random Variables

Two random variables X and Y are said to be statistically independent (or independently distributed) if their joint density function can be written as the product of the two marginal density functions:

$$f(x, y) = f_1(\mathrm{x}) f_2(\mathrm{y}).$$

It is noteworthy that when the random variables X and Y are independent and the joint density, $f(x, y)$ is known, then knowledge of the value of one of the variables furnishes no information about the other variable because in this case,

$$f(x / y) = f_1(\mathrm{x}) \text{ and } f(y / x) = f_2(\mathrm{y}).$$

Example 9.4 (example 8.1 continued). In the smoking and cancer example above, obtain the conditional probability $P(C = 1 / S = 1)$ and show that the random variables S and C are not independent.

Solution

The conditional probability is obtained as

$$P(C = 1 / S = 1) = \frac{P(C = 1, S = 1)}{P(S = 1)} = \frac{3/60}{13/60} = 0.23.$$

Now

$$P(C = 1, S = 1) = \frac{3}{60} = 0.05,$$

and from the marginal probabilities,

$$P(C = 1)P(S = 1) = \frac{10}{60} . \frac{13}{60} = 0.036.$$

Therefore S and C are not independent since $P(C = 1, S = 1) \neq P(C = 1)P(S = 1)$.

Example 9.5 (example 8.2 continued) From the joint probability function of the discrete random variables X and Y above, obtain the conditional probability $P(X = 1 / Y = 3)$ and show that these random variables are not independent.

Solution

This conditional probability is given as

$$P(X = 1 / Y = 3) = \frac{P(X = 1, Y = 3)}{P(Y = 3)} = \frac{1/6}{1/3} = 0.5.$$

Now from the marginal probability of X we have

$P(X=1)=\frac{1}{4}=0.25.$

Therefore X and Y are not independent since $P(X=1/Y=3)\neq P(X=1)$.

Example 9.6 (example 8.3 continued) From the joint density function of the two continuous random variables given above, obtain

(a) The conditional density function of Y given X.

(b) The conditional density function of X given Y

Solution

The given joint density function is

$$f(x,y)=\begin{cases} kxy, & 0\le x\le 3,\ \ 1\le y\le 4. \\ 0, & \text{otherwise} \end{cases}$$

Therefore,

(a) The conditional density function of Y given X is

$$f(y/x)=\frac{f(x,y)}{f_1(x)}=\frac{\frac{4}{135}xy}{\frac{2}{9}x}$$

$$=\frac{4\text{xy}}{30\text{x}}=\frac{2}{15}y.$$

(b) The conditional density function of X given Y is

$$f(x/y)=\frac{f(x,y)}{f_2(y)}=\frac{\frac{4}{135}xy}{\frac{2}{15}y}$$

$$=\frac{4\text{xy}}{18\text{y}}=\frac{2}{9}x.$$

3.4 Expectation and Variance for Joint Distributions

If X and Y are two continuous random variables with joint density function $f(x,y)$, the mean for each of the variables is given, respectively, by

$$\mu_X=E(X)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}xf(x,y)dxdy$$

$$=\int_{-\infty}^{\infty}xf_1(x)dx,\qquad -\infty\le X\le\infty$$

and similarly

$$\mu_Y = E(Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} yf(x,y)dxdy$$

$$= \int_{-\infty}^{\infty} yf_2(y)dy, \qquad -\infty \le Y \le \infty$$

The variance is given by

$$\sigma_X^2 = Var(X) = E(X-\mu_X)^2$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\mu_X)^2 f(x,y)dxdy$$

$$= \int_{-\infty}^{\infty}(x-\mu_X)^2 f_1(x)dx$$

Similarly,

$$\sigma_Y^2 = Var(Y) = E(Y-\mu_Y)^2$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(y-\mu_Y)^2 f(x,y)dxdy$$

$$= \int_{-\infty}^{\infty}(y-\mu_Y)^2 f_2(y)dy$$

Where $f_1(\mathrm{x})$ and $f_2(\mathrm{y})$ are the marginal densities of X and Y respectively.

Therefore, the mean and variance for each of the variables may be obtained from the marginal densities.

Self Assessment Exercises

1. State the Mean and Variance of a joint distribution function.

Self Assessment Answers

3.5 The Covariance

Another statistic, in addition to the means μ_X and μ_Y, and the variances σ_X^2 and σ_Y^2, of the marginal density functions of the random variables X and Y, is the *Covariance*, which is a measure of the "co-variability" of the two variables about their respective means and is defined as

$$\sigma_{XY} = E(X-\mu_X)(Y-\mu_Y)$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\mu_X)(y-\mu_Y)f(x,y)dxdy$$
$$= E(XY)-\mu_X\mu_Y.$$

Still, for the same random variables, we can define the *correlation* coefficient as a ratio of the covariance to the product of the standard deviations of the variables and is given as

$$\rho = \frac{\sigma_{XY}}{\sigma_X\sigma_Y}.$$

For independently distributed random variables X and Y, the covariance, and, consequently, the correlation are zero.

However, the reverse is not necessarily true; two random variables can be uncorrelated (i.e., can have correlation coefficient zero) without being independent.

Example 9.7 Let the joint pdf of X and Y be given by

$$f(x,y) = \begin{cases} 4xy, & 0 \le x \le 1, \quad 0 \le y \le 1 \\ 0, & \text{otherwise.} \end{cases}$$

Obtain(i) the mean and variance for each of the variables.

(ii) the covariance.

Solution

$$(i)\quad E(X) = \int_0^1\int_0^1 x4xydxdy$$
$$= \int_0^1\left[\int_0^1 4x^2ydy\right]dx$$
$$= \int_0^1\left[2x^2y^2\right]_0^1dx = \int_0^1 2x^2dx$$
$$= \left[\frac{2}{3}x^3\right]_0^1 = \frac{2}{3}.$$

$$E(Y)=\int_0^1\int_0^1 y4xydxdy$$

$$=\int_0^1\left[\int_0^1 4xy^2dy\right]dx$$

$$=\int_0^1\left[\frac{4}{3}xy^3\right]_0^1 dx=\frac{4}{3}\int_0^1 xdx$$

$$=\left[\frac{4}{6}x^2\right]_0^1=\frac{2}{3}.$$

$$\therefore E(X)=E(Y)=\frac{2}{3}.$$

$$(ii)\quad Cov(X,Y)=\sigma_{XY}=E(X-\mu_X)(Y-\mu_Y)$$

$$=E(XY)-\mu_X\mu_Y$$

Now,

$$E(XY)=\int_0^1\int_0^1 xy(4xy)dxdy$$

$$=\int_0^1\int_0^1 4x^2y^2dxdy=\frac{4}{3}\int_0^1 x^2[y^3]_0^1dx$$

$$=\frac{4}{9}\left[x^3\right]_0^1=\frac{4}{9}.$$

$$\therefore \sigma_{XY}=E(XY)-\mu_X\mu_Y$$

$$=\frac{4}{9}-\frac{2}{3}.\frac{2}{3}=0.$$

Now, it can easily be checked that

$$f_1(x)=2x,\quad f_2(y)=2y,$$

from which we compute

$$f_1(x)f_2(y)=4xy=f(x,y).$$

Then, since $f(x,y)=f_1(x)f_2(y),$ and $\sigma_{XY}=0,$ it follows that the two jointly distributed random variables X and Y are statistically independent.

Self Assessment Exercises

1. Correlation coefficient of independently distributed random variables X and Y, is the ratio of the covariance to the …and is given as …
2. If the random variables X and are independently distributed, what is their likely their co-variance?

Self Assessment Answers

4.0 Conclusion

This unit had introduced you to how to investigate pairs or vectors of random variables simultaneouslyon the same sample space, in order to find association or otherwise on the variable pairs. Also, another statistic, the co-variance was discussed in order to measure the covariance of the two random variables as well as determine the level of correlation among the two variables.

5.0 Summary

In this unit, you have learnt the following:

i) Joint probability function is used whenpairs or vectors of random variables are being investigatedsimultaneouslyon the same sample space. A joint probability function that gives the probability of each possible pair of values is:

$$f(x, y) = P(X = x,\ Y = y).$$

From the joint distribution above, the marginal distributions of X and Y are respectively:

$$f_1(x_i) = P(X = x_i) = \sum_{j=1}^{h} f(x_i, y_i)$$

and

$$f_2(y_j) = P(Y = y_j) = \sum_{i=1}^{k} f(x_i, y_j).$$

for continuous random variables X and Y.

ii) Conditional probability function is a situation where X and Y are two continuous random variables with joint probability density function $f(x, y)$ and the value of one of the variables is known then, the probability statements involving the other variable should be conditional on what is known.

The conditional density function of the random variable Y given that X has the value x is given as

$$f(y/x) = \frac{f(x, y)}{f_1(x)}.$$

Similarly, that of X given Y is

$$f(x/y) = \frac{f(x, y)}{f_2(y)},$$

Where $f_1(x)$ and $f_2(y)$ are the marginal probability density functions of X and Y respectively.

iii) The Covariance is an additional statistic other thanthe means and variances of the marginal density functions of the random variables X and Y, is the *Covariance*, that measures the “co-variability” of the two variables about their respective means and is defined as

$$\begin{aligned}\sigma_{XY} &= E(X-\mu_X)(Y-\mu_Y)\\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\mu_X)(y-\mu_Y)f(x,y)dxdy\\ &= E(XY)-\mu_X\mu_Y.\end{aligned}$$

The *correlation* coefficient as a ratio of the covariance to the product of the standard deviations of the variables and is given as

$$\rho = \frac{\sigma_{XY}}{\sigma_X\sigma_Y}.$$

for independently distributed random variables X and Y, the covariance, and, hence, the correlation are zero.

iv) Two random variables X and Y are said to be statistically independent (or independently distributed), if their joint density function can be written as the product of the two marginal density functions as:

$$f(x,y) = f_1(\mathrm{x})f_2(\mathrm{y}).$$

6.0 Tutor-Marked Assignment (TMA)

1. Let the joint pdf of X and Y be given by

$$f(x,y) = \begin{cases} 4xy, & 0 \le x \le 1, \quad 0 \le y \le 1 \\ 0, & \text{otherwise.}\end{cases}$$

Obtain (i) the mean and variance for each of the variables.

(ii) the covariance.

7.0 References/Further Reading

W. J. Decoursey (2003): Statistics and probability for engineering applications. Newness.

Answer To SAAs

Module 3: Unit 1

1. Two outcomes; success(1) and failure (0)
2. 2. $b(n,p,x) = \binom{n}{x} p^x q^{n-x}, \quad x = 0,1,2,\ldots,n$
3. p
4. npq
5. 5. n (Number of trials is large,) and p (probability of success) is small in a single trial.
6. True
7. N - k

Module 3: Unit 2

1. $$f(x) = \begin{cases} \frac{1}{b-a}, & a \le x \le b \\ 0, & otherwise \end{cases}$$

2. $$\mu = \frac{b+a}{2}, \ \sigma^2 = \frac{(b-a)^2}{12}.$$

3. A sequence of events which occur at “random times”.

4. $$f(x) = \lambda e^{-\lambda x}, \ x \ge 0, \quad 0, \ otherwise$$ where λ is a nonnegative real number.

5. $$E(X) = \frac{1}{\lambda}, \ V(X) = \frac{1}{\lambda^2}$$

6. $$f(x) = \begin{cases} \frac{x^{\alpha-1} e^{-x/\beta}}{\beta^\alpha \Gamma\alpha}, & 0 < x < \infty, \ \alpha > 0, \ \beta > 0 \\ 0, & \text{otherwise.} \end{cases}$$

7. $$E(X) = \alpha\beta$$ $$Var(X) = \alpha\beta^2.$$

8. $$f(x) = \begin{cases} \frac{1}{\beta(m,n)} X^{m-1}(1-X)^{n-1}, & 0 < X < 1 \\ 0, & \text{otherwise} \end{cases}$$

9. $$E(X) = \frac{m}{m+n}, Var(X) = \frac{mn}{(m+n+1)(m+n)^2}.$$

Module 3: Unit 3

1. $f(x) = \dfrac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}, \qquad -\infty < x < \infty$ where μ = "centre" of the density and

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

σ = "spread" of the density

2. i. The mean and the variance are respectively, μ and σ^2.

   ii. The total area under the curve and above the horizontal axis is equal to 1.

   iii. The mean, median and mode all coincide in the center.

   iv. The curve is symmetrical about the mean; the coefficient of skewness is zero.

3. $Z = \dfrac{X - \mu}{\sigma}$

4. False

5. False

Module 3: Unit 4

1. Joint probability function (distribution)

2. $f_1(x_i) = P(X = x_i) = \sum_{j=1}^{h} f(x_i, y_i)$ and $f_2(y_j) = P(Y = y_j) = \sum_{i=1}^{k} f(x_i, y_j)$.

3. i)

$$\mu_X = E(X) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xf(x,y)dxdy$$
$$= \int_{-\infty}^{\infty} xf_1(x)dx, \qquad -\infty \le X \le \infty$$

and

$$\mu_Y = E(Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} yf(x,y)dxdy$$
$$= \int_{-\infty}^{\infty} yf_2(y)dy, \qquad -\infty \le Y \le \infty$$

ii)

$$\sigma_X^2 = Var(X) = E(X - \mu_X)^2$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x - \mu_X)^2 f(x,y)dxdy$$
$$= \int_{-\infty}^{\infty} (x - \mu_X)^2 f_1(x)dx$$

$$\sigma_Y^2 = Var(Y) = E(Y - \mu_Y)^2$$

and $$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (y - \mu_Y)^2 f(x, y) dx dy$$

$$= \int_{-\infty}^{\infty} (y - \mu_Y)^2 f_2(y) dy$$

where $f_1(\mathrm{x})$ and $f_2(\mathrm{y})$ are the marginal densities of X and Y respectively.

4. i) Product of the standard deviations of the variables

ii) $\rho = \dfrac{\sigma_{XY}}{\sigma_X \sigma_Y}$.

for independently distributed random variables X and Y ,

5. Zero (0)

Module 4

The Law of Large Numbers and Introduction to Statistical Inference

Unit 1

Law of Large Numbers

Content

1.0 Introduction

In the previous module on probability concept, we saw that an intuitive way to view the probability of a certain outcome of an experiment is as the frequency with which that outcome occurs in the long run, when the experiment is repetated a large number of times. The law of large numbers is a fundamental theorem proved about the mathematical model of probability and shows that this model is consistent with the frequency interpretation of probability.

To discuss this law, we would start with an important inequality called the *chebyshevInequality*.

2.0 Learning Outcome

At the end of this unit, you should be able to:”

i. State Chebyshev’s Inequality law
ii. Define the law of large numbers
iii. Estimate boundary values

3.0 Learning Outcomes

3.1 Chebyshev’s Inequality

This states as follows:

Let X be a discrete random variable with mean μ and variance σ^2. Then if $\varepsilon > 0$ is any positive real number, we have

$$P(|X - \mu| \geq \varepsilon) \leq \frac{V(X)}{\varepsilon^2}.$$

Proof. Let the distribution function of X be denoted by $p(x)$. Then the probability that X differs from μ by at least ε is given by

$$P(|X - \mu| \geq \varepsilon) = \sum_{|x-\mu|\geq\varepsilon} p(x).$$

Now we know that the variance of X is given by

$$V(X) = \sum_x (x-\mu)^2 p(x),$$

and this is at least as large as

$$\sum_{|x-\mu|\geq\varepsilon} (x-\mu)^2 p(x),$$

Since all the summands are positive and the range of summation in the second sum have been restricted. Now, since $|x-\mu| > \varepsilon$, we have $(x-\mu)^2 > \varepsilon^2$ and so the above last sum is

$$\sum_{|x-\mu|\geq\varepsilon} \varepsilon^2 p(x) = \varepsilon^2 \sum_{|x-\mu|\geq\varepsilon} p(x)$$

$$= \varepsilon^2 P(|X-\mu| \geq \varepsilon).$$

So,

$$P(|X-\mu| \geq \varepsilon) \leq \frac{V(X)}{\varepsilon^2}.$$

Therefore, if X is any random variable with mean μ and variance σ^2, and $\varepsilon = k\sigma$, then this inequality states that

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

That is, for any random variable, the probability of a deviation from the mean of more than k standard deviations is $\leq 1/k^2$.

Example10.1 Let X be a continuous random variable with values uniformly distributed over the interval [0, 20].

i. Find the mean and variance of X .
ii. Using chebyshev's Inequality, find an upper bound for $P(|X-\mu| \geq 9)$.
iii. Calculate $P(|X-\mu| \geq 9)$ exactly and compare with the bound in (b). How good is Chebyshev's Inequality in this case?

Solution

Since $X \sim U[0,20]$, we have

$$f(x) = \frac{1}{20}, \quad 0 \leq x \leq 20.$$

(a) $\mu = E(X) = \int_0^{20} xf(x)dx$

$$= \frac{1}{20}\int_0^{20} xdx = \frac{1}{20}\left[\frac{x^2}{2}\right]_0^{20} = 10.$$

$$\sigma^2 = E(X^2) - \mu^2 = \int_0^{20} x^2 f(x)dx - \mu^2$$

$$= \frac{1}{20}\int_0^{20} x^2 dx - \mu^2 = \frac{1}{20}\left[\frac{x^3}{3}\right]_0^{20} - \mu^2$$

$$= \frac{400}{3} - 100 = \frac{100}{3}.$$

Therefore,

$$\mu = 10, \quad \sigma^2 = 100/3.$$

(b) Chebyshev's inequality states that, given mean and variance,

$$P(|X-\mu| \geq \varepsilon) \leq \frac{V(X)}{\varepsilon^2}.$$

Thus,

$$P(|X-10| \geq 9) \leq \frac{100}{3(9^2)} = 0.412.$$

(c) Since $X \sim U[0,20]$, the exact probability is calculated as follows

$$P(|X-\mu| < 9) = P(|X-10| < 9)$$

$$= P(1 < X < 19) = \int_1^{19} f(x)dx$$

$$= \int_1^{19} \frac{1}{20}dx = \frac{1}{20}[x]_1^{19} = 0.9.$$

But $P(|X-10| \geq 9) \leq 1 - P(|X-10| < 9)$.

Thus

$$P(|X-10| \geq 9) = 1 - 0.9 = 0.1.$$

Comparing this exact probability with the one in (b) we see that the Chebyshev's estimate, here, is *not* very accurate.

Example 10.2 Let X be a continuous random variable with values exponentially distributed over $[0,\infty]$ with parameter $\lambda = 0.1$.

(a) Find the mean and variance of X .

(b) Using chebyshev's Inequality, find an upper bound for $P(|X-\mu| \geq 20)$.

(c) Calculate the probability $P(|X-\mu| \geq 20)$ exactly and compare with the Chebyshev's upper bound in (b) above. How good is Chebyshev's Inequality in this case?

Solution

Since $X \sim \exp(\lambda)$, we have

$$f(x) = \lambda e^{-\lambda x}, \qquad x \geq 0..$$

Now $\lambda = 0.1$, therefore as derived in section 6.3, we have

(a) $$\mu = E(X) = \frac{1}{\lambda} = 10, \ \sigma^2 = V(X) = \frac{1}{\lambda^2} = 100.$$

(b) Chebyshev's inequality states that, given mean and variance,

$$P(|X - \mu| \geq \varepsilon) \leq \frac{V(X)}{\varepsilon^2}.$$

Thus since $\mu = 10 \text{ and } \sigma^2 = 100$,

$$P(|X - 10| \geq 20) \leq \frac{100}{400} = 0.25.$$

(c) Since $X \sim \exp(\lambda)$, the exact probability is calculated as follows

$$\begin{aligned} P(|X - \mu| < 20) &= P(|X - 10| < 20) \\ &= P(-10 < X < 30) = \int_0^{30} f(x)dx \\ &= \int_0^{30} \lambda e^{-\lambda x} dx = -e^{-\lambda x}\Big|_0^{30} = -e^{-0.1x}\Big|_0^{30} \\ &= 1 - e^{-3}. \end{aligned}$$

But $P(|X - 10| \geq 20) \leq 1 - P(|X - 10| < 20)$.

Thus

$$\begin{aligned} P(|X - 10| \geq 20) &= 1 - (1 - e^{-3}) \\ &= e^{-3} = 0.0497. \end{aligned}$$

Again, comparing this exact probability with the one in (b) we see that the Chebyshev's estimates are in general *not* very accurate.

Example 10.3 For the binomial variable with $n = 50$ and $p = \frac{1}{2}$, use Chebyshev's inequality to obtain an upper bound for the probability that the r. v. X deviates from its mean by two standard deviations or more.

Solution

$$n = 50, \ p = \tfrac{1}{2}. \quad \therefore \mu = np = 25$$

$$\sigma^2 = np(1 - p) = 12.5$$

$$\sigma = \sqrt{12.5}$$

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

Here, $k = 2$,

$$\therefore P(|X-25| \geq 2\sigma) \leq \frac{1}{4} = 0.25.$$

Example 10.4 A random variable X has $\mu = 10$, $\sigma^2 = 4$, using Chebyshev's inequality, find $P(|X-10| \geq 3)$.

Solution

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

$\sigma^2 = 4$, $\sigma = 2$, $k\sigma = 3 \Rightarrow 2k = 3$ and $k = \frac{3}{2}$.

$$P(|X-10| \geq 3) \leq \frac{4}{9}.$$

Example 10.5 A random variable X has $\mu = 12$, $\sigma^2 = 9$. Using Chebyshev's inequality, find $P(6 < X < 18)$.

Solution

Since $\mu = 12$, we have

$$P(6 < X < 18) = P(6-12 < X-12 < 18-12)$$
$$= P\left(|X-12| < 6\right) = 1 - P\left(|X-12| \geq 6\right).$$

By Chebyshev's inequality,

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

Now $\sigma^2 = 9 \Rightarrow \sigma = 3$.

$$\therefore P(|X-12| \geq 3k) \leq \frac{1}{k^2}.$$

Now, 3k = 6 $\Rightarrow$ k = 2

So,

$$P(|X-12| \geq 6) \leq \frac{1}{4},$$

and

$$P(|X-12| < 6) \geq 1 - \frac{1}{4} = \frac{3}{4}.$$

Therefore,

$$P(6 < X < 18) = \frac{3}{4}.$$

Now we discuss the law of large numbers.

Self Assessment Exercises

1. State Chebyshev's Inequalitylaw.
2. Chebyshev's estimates are always very accurate. True or false.

Self Assessment Answers

3.2 Law of Large Numbers

Let $X_1,..., X_n$ be a set of n independent and identically distributed random variables with finite expected value $\mu = E(X)$ and finite variance $\sigma^2 = V(X_j)$.

Let $S_n = X_1 + ... + X_n$

be the sum, and

$$A_n = \frac{S_n}{n}$$

be the average.

Then for any $\varepsilon > 0$, we have

$$P\left(\left|\frac{S_n}{n} - \mu\right| \geq \varepsilon\right) \to 0$$

as $n \to \infty$. Equivalently,

$$P\left(\left|\frac{S_n}{n} - \mu\right| < \varepsilon\right) \to 1$$

as $n \to \infty$.

That is, in an independent trials process where the individual summands have finite expected value and variance, for large n, the value of the arithmetic mean, A_n, is usually very close to its expected value, which equals μ.

Proof

Since $X_1, X_2, ..., X_n$ are independent and identically distributed, we have

$$E(S_n) = E(X_1) + ... + E(X_n)$$
$$= \mu_1 + \mu_2 + ... + \mu_n = n\mu.$$

$$V(S_n) = V(X_1) + ... + V(X_n)$$
$$= \sigma_1{}^2 + \sigma_2{}^2 + ... + \sigma_n{}^2 = n\sigma^2.$$

$$E(A_n) = E\left(\frac{S_n}{n}\right) = \frac{1}{n}E(S_n) = \mu ,$$

$$V(A_n) = V\left(\frac{S_n}{n}\right) = \frac{1}{n^2}V(S_n) = \frac{\sigma^2}{n}.$$

Then by Chebyshev's Inequality, for any $\varepsilon > 0$, we have

$$P\left(\left|\frac{S_n}{n} - \mu\right| \geq \varepsilon\right) \leq \frac{\sigma^2}{n\varepsilon^2} ,$$

and, for fixed ε, taking the limit as $n \to \infty$, we obtain

$$\lim_{n\to\infty} P\left(\left|\frac{S_n}{n} - \mu\right| \geq \varepsilon\right) = 0.$$

Or equivalently,

$$\lim_{n\to\infty} P\left(\left|\frac{S_n}{n} - \mu\right| < \varepsilon\right) = 1.$$

Example 10.6 A die is rolled n times. Let X_j denotes the outcome of the *j*th roll. Then $S_n = X_1 + X_2 + ... + X_n$ is the sum of the outcomes of the first n rolls, and $A_n = \frac{S_n}{n}$ is the average of the outcomes. Then, each X_j has distribution

$$p(x) = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \end{pmatrix}$$

and

$$E(X_j) = 7/2 ,$$

$$V(X_j) = 35/12 .$$

Then,

$$E(A_n) = \frac{1}{n} E(S_n) = \frac{7}{2},$$

$$V(A_n) = \frac{1}{n^2} V(S_n) = \frac{35}{12n}.$$

By the law of large numbers, for any $\varepsilon > 0$, we have

$$P\left(\left| \frac{S_n}{n} - \frac{7}{2} \right| \geq \varepsilon \right) \leq \frac{35}{12n\varepsilon^2}.$$

Thus, for fixed ε, we have

$$P\left(\left| \frac{S_n}{n} - \frac{7}{2} \right| \geq \varepsilon \right) \to 0$$

as $n \to \infty$, or equivalently,

$$P\left(\left| \frac{S_n}{n} - \frac{7}{2} \right| < \varepsilon \right) \to 1$$

as $n \to \infty$.

That is, for large n, we can expect the average, A_n, of the outcomes to be very near the expected value.

Example 10.7 (a) Suppose 20 real numbers are chosen independently from the interval [0, 20] with uniform distribution. Find a lower bound for the probability that their average lies between 8 and 12.

(b) Suppose 100 real numbers are chosen independently from the interval in (a). Find a lower bound for the probability that their average lies between 8 and 12.

Solution

In this experiment, if X_i describes the *i*th choice, we have

$$\mu = E(X_i) = \frac{1}{20}\int_0^{20} x\,dx = 10,$$

$$\sigma^2 = V(X_i) = \frac{1}{20}\int_0^{20} x^2 dx - \mu^2$$

$$\frac{400}{3} - 100 = \frac{100}{3}.$$

Hence,

(a) $$E\left(\frac{S_{20}}{20}\right) = \frac{1}{20} E(S_{20}) = 10,$$

$$V\left(\frac{S_{20}}{20}\right)=\frac{1}{400}V(S_{20})=\frac{5}{3}.$$

Now,

$$P\left(8\le\frac{S_{20}}{20}\le 12\right)=P\left(\left|\frac{S_{20}}{20}-10\right|<2\right)=1-P\left(\left|\frac{S_{20}}{20}-10\right|\ge 2\right).$$

By the law of large numbers,

$$P\left(\left|\frac{S_{20}}{20}-10\right|\ge k\sigma\right)\le\frac{1}{k^2}.$$

Now,

$$\sigma^2\left(\frac{S_{20}}{20}\right)=\frac{5}{3}\Rightarrow\sigma\left(\frac{S_{20}}{20}\right)=\sqrt{\frac{5}{3}},$$

$$k\sigma=2\Rightarrow k\sqrt{\frac{5}{3}}=2\Rightarrow k=\frac{2\sqrt{3}}{\sqrt{5}}.$$

Therefore,

$$P\left(\left|\frac{S_{20}}{20}-10\right|\ge 2\right)\le 1/\left(2\sqrt{3/5}\right)^2=5/12.$$

Hence,

$$P\left(\left|\frac{S_{20}}{20}-10\right|<2\right)=1-\frac{5}{12}=\frac{7}{12}.$$

Thus,

$$P\left(8\le\frac{S_{20}}{20}\le 12\right)=\frac{7}{12}.$$

(b) $$E\left(\frac{S_{100}}{100}\right)=\frac{1}{100}E(S_{100})=10,$$

$$V\left(\frac{S_{100}}{100}\right)=\frac{1}{(100)^2}V(S_{100})=\frac{1}{3}.$$

Now,

$$P\left(8\le\frac{S_{100}}{100}\le 12\right)=P\left(\left|\frac{S_{100}}{100}-10\right|<2\right)=1-P\left(\left|\frac{S_{100}}{100}-10\right|\ge 2\right).$$

By the law of large numbers,

$$P\left(\left|\frac{S_{100}}{100}-10\right|\geq k\sigma\right)\leq\frac{1}{k^2}.$$

Now,

$$\sigma^2\left(\frac{S_{100}}{100}\right)=\frac{1}{3}\Rightarrow\sigma\left(\frac{S_{100}}{100}\right)=\sqrt{\frac{1}{3}},$$

$$k\sigma=2\Rightarrow k\sqrt{\frac{1}{3}}=2\Rightarrow k=2\sqrt{3}.$$

Therefore,

$$P\left(\left|\frac{S_{100}}{100}-10\right|\geq 2\right)\leq\frac{1}{\left(2\sqrt{3}\right)^2}=\frac{1}{12}.$$

Hence,

$$P\left(\left|\frac{S_{100}}{100}-10\right|<2\right)=1-\frac{1}{12}=\frac{11}{12}.$$

Thus,

$$P\left(8\leq\frac{S_{100}}{100}\leq 12\right)=\frac{11}{12}.$$

From this example, we can see that as we increase n, the chances that $\left|\frac{S_n}{n}-10\right|$ is less than 2 become closer and closer to 1.

Self Assessment Exercises

1. The law of large number states that for large number of trials, n we can expect the average, A_n, of the outcomes to be very near …..
2. As $n\rightarrow\infty$, $\left|\frac{S_n}{n}-10\right|\rightarrow$ ….

Self Assessment Answers

4.0 Conclusion

The law of large numbers was discussed in this unit. Note that this is a fundamental theorem about the mathematical model of probability and it was shown that this model is consistent with the frequency interpretation of probability.

You were introduced to an important inequality law called the chebyshev Inequality and the law of large numbers. The two laws were contrasted.

5.0 Summary

In this unit, you have learnt the following:

i) Chebyshev's inequalitystates that if X is a discrete random variable with mean μ and variance σ^2. Then if $\varepsilon > 0$ is any positive real number, we have

$$P(|X - \mu| \geq \varepsilon) \leq \frac{V(X)}{\varepsilon^2}.$$

ii) If $X_1, \ldots, X_n$ is a set of n independent and identically distributed random variables with finite expected value $\mu = E(X)$ and finite variance $\sigma^2 = V(X_j)$. Then, if

$$S_n = X_1 + \ldots + X_n$$

is the sum, and

$$A_n = \frac{S_n}{n} \text{ is the average}$$

Then, for any $\varepsilon > 0$, we have

$$P\left(\left|\frac{S_n}{n} - \mu\right| \geq \varepsilon\right) \to 0$$

as $n \to \infty$. Equivalently,

$$P\left(\left|\frac{S_n}{n} - \mu\right| < \varepsilon\right) \to 1, \text{ as } n \to \infty.$$

6.0 Tutor-Marked Assignment (TMA)

1. Let X be a random variable with $E(X) = 0$ and $V(X) = 1$. What integer value k will assure us that $P(|X| \geq k) \leq 0.01$?
2. In a final semester examination, a student's score on a particular statistics course is a random variable with values of [0, 100], mean 70, and variance 25.
   (a) Find a lower bound for the probability that the student's score will fall between 65 and 75.

(b) If 100 students take the final, find a lower bound for the probability that the class average will fall between 65 and 75.

7.0 References/Further Reading

Sheldon M. Ross (1997): Introduction to probability models, sixth edition. Academic Press. New York.

Mario Lefebre (2000): Applied probability and statistics. Springer.

Unit 2

Introduction to the Central Limit Theorem

Content

1.0 Introduction

The central limit theorem (CLT) is one of the remarkable results of the theory of probability. It occupies a basic position not only in theory but also in application.

2.0 Learning Outcome

At the end of this unit, you should be able to:

i. State and apply the Central limit theorem
ii. Undestand the underlying assumptions of the theorem
iii. Proof the Central limit theorem

3.0 Learning Outcomes

3.1 The Central Limit Theorem

In its simplest form, the theorem states that if several samples of size n are taken from a population with mean μ and finite variance σ^2, then for large n, the distribution of sample means, $\bar{X}$, approaches normality with mean μ and variance σ^2/n. That is

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right).$$

Moreover, the approximation steadily improves as n increases without bound.

We now show that the mean and variance of the sample mean are μ and σ^2/n respectively.

Suppose that $X_1, X_2, ..., X_n$ are independent and identically distributed observations in a random sample of size n from a population with mean μ and finite variance σ^2. If $\bar{X}$ is the sample mean, then

$$\begin{aligned} E(\bar{X}) &= E\left[\frac{1}{n}(X_1 + X_2 + ... + X_n)\right] \\ &= \frac{1}{n}E[X_1 + X_2 + ... + X_n] \\ &= \frac{1}{n}[E(X_1) + E(X_2) + ... + E(X_n)] \\ &= \frac{1}{n}[n\mu] = \mu. \end{aligned}$$

And

$$V(\bar{X}) = V\left[\frac{1}{n}(X_1 + X_2 + ... + X_n)\right]$$

$$= \left(\frac{1}{n}\right)^2 V[X_1 + X_2 + ... + X_n]$$

$$= \left(\frac{1}{n}\right)^2 [V(X_1) + V(X_2) + ... + V(X_n)]$$

$$= \left(\frac{1}{n}\right)^2 [\sigma_1{}^2 + \sigma_2{}^2 + ... + \sigma_n{}^2]$$

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

The mathematical statement and proof of the **CLT** now follows:

Let $\bar{X}_n$ denotes the mean of the observations in a random sample of size n from a population with mean μ and finite variance σ^2.

Define the random variable

$$U_n = \frac{\bar{X}_n - \mu}{\sigma / \sqrt{n}}$$

Then as n increases indefinitely, the distribution function of U_n converges to the standard normal distribution function.

Proof

Define a standardized random variable Z_i by

$$Z_i = \frac{X_i - \mu}{\sigma},$$

where $E(Z_i) = 0$ and $Var(Z_i) = 1$.

Then, the moment generating function of Z can be written as

$$M_Z(t) = 1 + tE(Z_i) + \tfrac{t^2}{2!}E(Z_i{}^2) + \tfrac{t^3}{3!}E(Z_i{}^3) + ...$$

$$= 1 + \tfrac{t^2}{2!} + \tfrac{t^3}{3!}E(Z_i{}^3) + ..., \qquad \text{(by Maclaurin series expansion).}$$

Now,

$$U_n = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} = \sqrt{n}\left(\frac{\bar{X}-\mu}{\sigma}\right)$$

$$= \sqrt{n}\left(\frac{\frac{1}{n}\left(\sum_{i=1}^{n} X_i - n\mu\right)}{\sigma}\right) = \frac{\sqrt{n}}{n}\left(\frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma}\right)$$

$$= \frac{1}{\sqrt{n}}\left(\frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma}\right) = \frac{1}{\sqrt{n}} \frac{\sum_{i=1}^{n}(X_i - \mu)}{\sigma}$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_i \quad -------- \quad (12.1)$$

Because the random variables X_i are independent, so are the random variables Z_i. We now take the moment generating function (mgf) of the standardized variable, U_n, as

$M_n(t) = \left[M_Z\left(\frac{t}{\sqrt{n}}\right)\right]^n$ *(mgf of sum of independent random variables* Z_i*).*

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

$= \left[1 + \frac{t^2}{2n} + \frac{t^3}{3! n^{3/2}} E(Z_i^{\ 3}) + \ldots\right]^n$, *(since* $E(Z_i) = 0, V(Z_i) = E(Z_i^{\ 2}) = 1.$*)*

By taking logarithm to base e of $M_n(t)$, we obtain

$$\ln(M_n(t)) = \ln\left[M_n\left(\frac{t}{\sqrt{n}}\right)\right]^n$$

$$= n\ln\left(1 + \left(\frac{t^2}{2n} + \frac{t^3}{3! n^{3/2}} E(Z_i^{\ 3}) + \ldots\right)\right) \quad -------- (12.2)$$

Recall that the Taylor Series expansion of ln(1+x) is

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \ldots$$

Now, if we let $x = \left[\frac{t^2}{2n} + \frac{t^3}{3! n^{3/2}} E(Z_i^{\ 3}) + \ldots\right]$ in (12.2) above, we obtain

$$\ln(M_n(t)) = n\ln(1+x) = n\left(x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \ldots\right)$$

Rewriting this expression by substituting for x, we have

$$\ln(M_n(t)) =$$

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

Now, if we multiply through by the initial n, we see that all terms except the first have some positive power of n in the denominator. Consequently, if we take the limit of $\ln(M_n(t))\ as\ n \to \infty,$ all terms but the first go to zero, leaving

$$\lim_{n \to \infty} \ln(M_n(t)) = \frac{t^2}{2}$$

and so

$$\lim_{n \to \infty} (M_n(t)) = e^{t^2/2},$$

Which is the moment generating function for a standard normal random variable.

Probabilities for $\bar{X}_n$ can then be calculated, approximately, using the standard normal table as

$$P(\bar{X}_n \le x) = \phi\left(\frac{x-\mu}{\sigma/\sqrt{n}}\right).$$

Example 11.1 The lifetimes of a particular fluorescent bulb produced by a company are known to come from a distribution with mean 35 months and variance 64 months. If a random sample of 45 such bulbs is taken from this population, derive an approximate probability that the mean of this sample is greater than 38 months.

Solution

Given that n = 45, $\mu = 35,\ \sigma^2 = 64$.

Then, $\sigma_{\bar{X}} = \sigma/\sqrt{n} = 8/\sqrt{45} = 1.19$

$$\therefore \quad P(\bar{X} > 38) = P\left(\frac{\bar{X}-\mu}{\sigma_{\bar{X}}} \le \frac{38-35}{1.19}\right)$$

$$= P(z > 2.52) = 1 - P(z \le 2.52)$$

$$= 1 - \phi(2.52)$$

$$= 1 - 0.0058$$

$$= 0.9942.$$

Example 11.2 If a random sample of size 30 is taken from Poisson distribution with parameter $\lambda = 2,$ find $P(3/2 \le \overline{X} \le 7/4)$.

Solution

$X \sim P_o(2),\ so\ \mu_X = \sigma_X{}^2 = 2.$

The sample size is large and so by the CLT,

$\overline{X} \sim N(2,\ 2/30)$ approximately.

$\therefore \sigma_{\overline{X}} = \sqrt{\tfrac{2}{30}} = \sqrt{0.07},$

$$\begin{aligned}
P(3/2 \le \overline{X} \le 7/4) &= P\left(\frac{1.5-2}{\sqrt{0.07}} \le \frac{\overline{X}-2}{\sqrt{0.07}} \le \frac{1.75-2}{\sqrt{0.07}}\right) \\
&= P(-1.88 \le z \le -0.945) \\
&= \phi(-0.945) - \phi(-1.88) \\
&= \phi(1.88) - \phi(0.945) \\
&= 0.1435.
\end{aligned}$$

Example 11.3

If a random sample of size 10 is taken from the binomial distribution with probability of success $p = \frac{1}{2}$, find $P(3 \le y \le 6)$.

Solution

Here $y \sim Bin(10, \tfrac{1}{2})\ and\ so\ \mu = np = 5, \sigma^2 = npq = 2.5.$

$\therefore \sigma = \sqrt{npq} = \sqrt{2.5},\ so,$

$P(3 \le y \le 6) = P(2.5 \le y \le 6.5)$ *(Continuity correction)*

$$\begin{aligned}
&= P\left(\frac{2.5-5}{\sqrt{2.5}} < \frac{y-5}{\sqrt{2.5}} < \frac{6.5-5}{\sqrt{2.5}}\right) \\
&= P(-1.58 < z < 0.95) \\
&= \phi(0.95) - \phi(-1.58) \\
&= \phi(0.95) + \phi(1.58) - 1 \\
&= 0.7719.
\end{aligned}$$

Now if

$$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \text{ then } n\overline{X} \sim N\left(n\mu, n^2\frac{\sigma^2}{n}\right)$$

But

$$n\overline{X} = X_1 + X_2 + ... + X_n = S_n.$$

Therefore

$$S_n \sim N\left(n\mu, n\sigma^2\right).$$

That is, for large n, the distribution of S_n is approximately normal with mean $n\mu$ and variance $n\sigma^2$. The standardized sum of S_n is given by

$$S_n^* = \frac{S_n - n\mu}{\sqrt{n\sigma^2}}.$$

This standardizes S_n to have expected value 0 and variance 1.

Example 11.4

A coin is tossed 100 times. Estimate the probability that the number of heads that turn up

(a) Lies between 40 and 60.

(b) Is 45 or less.

Solution

The expected number of heads in 100 tosses is $np = (1/2)(100) = 50$, and the standard deviation is $\sigma = \sqrt{100.\frac{1}{2}..\frac{1}{2}} = 5$. Let S_{100} be the number of heads that turn up in 100 tosses of the coin. Then we have

$$(a) \qquad P\left(40 \le S_{100} \le 60\right) \approx P\left(\frac{39.5 - 50}{5} \le S_{100}^* \le \frac{60.5 - 50}{5}\right)$$
$$= P(-2.1 \le S_{100}^* \le 2.1)$$
$$= 2\phi(2.1) - 1$$
$$= 0.9642.$$

$$(b) \qquad P\left(S_{100} \le 45\right) \approx P\left(S_{100}^* < \frac{45.5 - 50}{5}\right)$$
$$= P(S_{100}^* < 0.9)$$
$$= 1 - P\left(S_{100}^* < 0.9\right)$$
$$= 1 - \phi(0.9)$$
$$= 0.1841.$$

Self Assessment Exercises

1. For samples of size n taken from a population with mean μ and finite variance σ^2, then for large n, the distribution of sample means, $\overline{X}$, is …….

4.0 Conclusion

This unit discussed the central limit theorem (CLT) as one of the remarkable results of the theory of probability. It is useful theoretically and in application. The theorem was stated and proved, with mean and variance as well as illustration with worked examples.

5.0 Summary

This unit discussed the following:

i) The Central Limit Theorem which states that if several samples of size n are taken from a population with mean μ and finite variance σ^2, then for large n, the distribution of sample means, $\overline{X}$, approaches normality with mean μ and variance σ^2/n. i.e.

$$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right).$$

ii) As n increases indefinitely, the distribution function of U_n converges to the standard normal distribution function.

iii) In the proof of CLT, if a standardized random variable Z_i is defined by

$$Z_i = \frac{X_i - \mu}{\sigma},$$

where $E(Z_i) = 0$ and $Var(Z_i) = 1$, then, the moment generating function of Z can be written as

$$M_Z(t) = 1 + tE(Z_i) + \tfrac{t^2}{2!}E(Z_i^{\,2}) + \tfrac{t^3}{3!}E(Z_i^{\,3}) + \ldots$$

$$= 1 + \tfrac{t^2}{2!} + \tfrac{t^3}{3!}E(Z_i^{\,3}) + \ldots, \qquad \text{(by Maclaurin series expansion).}$$

6.0 Tutor-Marked Assignment (TMA)

1. A population of students in a Polytechnic has an average weight of 100 and standard deviation of 25. A student selected at random from the population will have a weight that is random, with the distribution of the population. If a random sample of weights of 55 students is taken from this population, what is an approximate probability that the mean weight, $\bar{x}$ of this sample is greater that 105?

2. What is an approximate probability that the mean of a random sample of size 25 taken from a normally distributed population with mean 30 and variance 16 is

   (a) less than or equal to 28

   (b) between 15 and 25

7.0 References/Further Reading

Sheldon M. Ross (1997): Introduction to probability models, sixth edition. Academic Press. New York.

Mario Lefebre (2000): Applied probability and statistics. Springer.

Unit 3

Introduction to Statistical Inference

Content

1.0 Introduction

In statistics,we are often concerned with making inferences from samples about the populations from which they have been drawn. In other words, if we find a difference between two samples, we would like to know, is this a"real" difference (i.e., is it present in the population) or just a"chance" difference (i.e. it could just be the result of random sampling error). The procedure by which we achieve this is termed *statisticalinference*.

Statistical inference refers to extending your knowledge obtained from a random sample drawn from the entire population to the whole population. Any inferred conclusion from a sample data to the population from which the sample is drawn must be expressed in a probabilistic term. Inference in statistics is of two types:

i. Estimation
ii. Testing of Hypothesis

2.0 Learning Outcome

At the end of this unit, you should be able to:

i. Estimate population parameters
ii. Identify Good Estimator properties
iii. Postulate and test Hypothesis

3.0 Learning Outcomes

3.1 Estimation of Population Parameters

Estimation involves the determination, with a possible error due to sampling, of the unknown value of a population characteristic, such as the proportion having a specific attribute or the average value of some numerical measurement.To express the accuracy of the estimates of population characteristics, one must also compute the standard errors of the estimates.

In statistics, one is often faced with the problem of predicting the values of population parameters of interest using the information obtained from the sample of reasonable size drawn from that population. The process of achieving this is termed *estimation*. It provides the most likely location of a population parameter. It involves selecting a random sample of size n from the population, collecting data on the sample members and then computing a sample statistic such as the mean, denoted by $\bar{x}$, the variance s^2, etc. These statistics are then used as estimators to predict or estimate the corresponding population parameters like the mean μ, the variance σ^2, etc.

> ***Definition***
>
> An estimator $\hat{\theta}$ of a population parameter θ is a function of the sample observations $(x_1, x_2, ..., x_n)$ which is closest to the true value of the parameter in some sense.

An estimate is the numerical value taken by the estimator in a particular instance.

There are two methods of estimating a population parameter from the sample statistic. These are the *point* and *interval* estimation.

3.2 Point Estimation

This gives an estimate of the population parameter as a single number using the sample statistic. For instance, the sample mean $\bar{x}$ can be used to predict a specific value for the population mean μ; the sample standard deviation s can also be used to predict a value for the population standard deviation σ. In such cases $\bar{x}$ and s are point estimators, respectively, of the population mean μ and population standard deviation σ.

3.3 Properties of a Good Estimator

As stated above, the process of estimation involves finding a sample statistic (an estimator) that comes closer to the true value of the parameter of the population from which the sample was drawn. But various sample statistics might be tried as estimators, some of which are better than others. Our task therefore is to find a good estimator of the parameter in question; an estimator whose closeness to the parameter we can say something definite about. There are a number of desirable properties that serve as measures of closeness of an estimator to the true value of its corresponding parameter. Some of these properties are discussed in this section.

Consider a random sample $X_1, X_2, ..., X_n$ independently drawn from a population F, the parameter of which is θ. Let $\hat{\theta} = h(x_1, x_2, ..., x_n)$ be a statistic that gives the estimate of the parameter θ. Then $\hat{\theta}$ would be regarded as a good estimator of θ if it possesses the properties below

Unbiasedness: $\hat{\theta}$ is an unbiased estimator of θ if its expected value equals θ.

That is if

$$E(\theta) = \theta.$$

Example 12.1 Given that $X_1, X_2, ..., X_n$ is a random sample of size n from an exponential distribution with parameter θ such that

$$f(x) = \theta e^{-\theta x}, \ \theta \geq 0, \ x > 0$$

Show that $\bar{x}$ is an unbiased estimator of $\frac{1}{\theta}$.

Solution

By definition, $\bar{x} = \frac{1}{n}\sum_i X_i$ and for all i, $E(X_i) = \frac{1}{\theta}$.

Therefore,

$$E(\bar{X}) = E\left(\tfrac{1}{n}\sum_i X_i\right) = \tfrac{1}{n}\sum_i E(X_i)$$
$$= \frac{1}{n}\sum_i\left(\frac{1}{\theta}\right) = \frac{1}{n}\left(\frac{n}{\theta}\right) = \frac{1}{\theta}.$$

$\therefore \bar{x}$ is an unbiased estimator of θ^{-1}.

Example 12.2 Let $X_1, X_2, ..., X_n$ be a random sample of size n from geometric distribution with mean $\frac{1}{p}$. Show that $\bar{x}$ is an unbiased estimator of $\frac{1}{p}$.

Solution

$\bar{x} = \frac{1}{n}\sum_i X_i$ and for all i, $E(X_i) = \frac{1}{p}$.

Therefore,

$$E(\bar{X}) = E\left(\tfrac{1}{n}\sum_i X_i\right) = \tfrac{1}{n}\sum_i E(X_i)$$
$$= \frac{1}{n}\sum_i\left(\frac{1}{p}\right) = \frac{1}{n}\left(\frac{n}{p}\right) = \frac{1}{p}.$$

$\therefore \bar{x}$ is an unbiased estimator of $\frac{1}{p}$.

Example 12.3 Let $X_1, X_2, ..., X_n$ be a random sample of size n from a normal distribution with mean μ and variance σ^2. Show that

(i) The sample mean $\bar{x}$ is an unbiased estimator of μ.

(ii) The sample variance $s^2 = \dfrac{\sum (X - \bar{x})^2}{n-1}$ is an unbiased estimator of σ^2.

Solution

(i) $\bar{x} = \frac{1}{n}\sum_i X_i$,

$$E(\bar{X}) = E\left(\tfrac{1}{n}\sum_i X_i\right) = \tfrac{1}{n}\sum_i E(X_i)$$

But for all i, $E(X_i) = \mu$.

$$\therefore E(\overline{X}) = \tfrac{1}{n}\sum_i E(X_i) = \tfrac{1}{n}\sum_i \mu = \tfrac{1}{n}(n\mu) = \mu.$$

(ii) s^2 is an unbiased estimator of σ^2 if $E(s^2) = \sigma^2$.

Now,

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

Then

$$\begin{aligned}(n-1)s^2 &= \sum\left(X^2 - 2\overline{X}X + \overline{X}^2\right)\\ &= \sum X^2 - 2\overline{X}\sum X + n\overline{X}^2\\ &= \sum X^2 - 2\tfrac{(\sum X)^2}{n} + \tfrac{(\sum X)^2}{n}\\ &= \sum X^2 - \frac{(\sum X)^2}{n} = \sum X^2 - n\overline{X}^2\end{aligned}$$

and

$$\begin{aligned}E(n-1)s^2 &= E\sum X^2 - E\left(n\overline{X}^2\right)\\ &= \sum E(X^2) - nE(\overline{X}^2)\end{aligned}$$

But $E(X^2) = \sigma^2 + \mu^2$ and $E(\overline{X}^2) = \dfrac{\sigma^2}{n} + \mu^2$.

So,

$$\begin{aligned}E(n-1)s^2 &= \sum(\sigma^2 + \mu^2) - n\left(\frac{\sigma^2}{n} + \mu^2\right)\\ &= n\sigma^2 - \sigma^2 = (n-1)\sigma^2\end{aligned}$$

Therefore,

$$E(s^2) = \frac{(n-1)\sigma^2}{n-1} = \sigma^2 .$$

(ii) **Efficiency**: An efficient estimator is the one such that among several asymptotically unbiased estimators, no other estimator has smaller asymptotiv variance. That is, if $\hat{\theta}_1$ is an unbiased estimator of a parameter θ of a given distribution and the variance of $\hat{\theta}_1$ is minimum compared to any other unbiased estimators of θ, then $\hat{\theta}_1$ is said to be the most efficient estimator of θ and is also known as the minimum variance unbiased estimator.

(iii) **Consistency**: An estimator $\hat{\theta}_n$ (based on a random sample of size n) of a parameter θ is said to be a consistent estimator if it converges in probability to the true value of the parameter as the sample size, n increases.

That is,

$\hat{\theta}_n$ is consistent for θ if $\hat{\theta}_n \to \theta$ as $n \to \infty$.

Therefore, consistency is not valid with reference to a particular sample size but the condition is that the larger the sample size, the closer the estimator to the true value of the parameter being estimated.

So we have

$$P\left(\left|\hat{\theta}_n - \theta\right| \to 0\right) \to 1 \text{ as } n \to \infty.$$

Alternatively, $\hat{\theta}$ is consistent for θ if $V(\hat{\theta}) \to 0$ as $n \to \infty$.

Example 12.4 A biased estimator of σ^2 based on a random sample of size n from a normal population is

$$s^2 = \frac{\sum (X - \bar{x})^2}{n} \text{ with } E(s^2) = \frac{(n-1)\sigma^2}{n} = \sigma^2\left(1 - \frac{1}{n}\right).$$

We see that $E(s^2) \to \sigma^2$ as $n \to \infty$.

Hence, s^2, though biased, is consistent for σ^2.

(iv) **Minimum mean square error**: Among the properties of a good estimator is that its mean square error

$E(\hat{\theta} - \theta)^2$ be small.

The amount of bias is the difference between the mean of an estimator and the value of the parameter one wish to estimate. It can be expressed as

$$b(\hat{\theta}) = E(\hat{\theta}) - \theta,$$

so that $E(\hat{\theta}) = \theta$ if $\hat{\theta}$ is unbiased and then $b(\hat{\theta}) = 0$.

For an unbiased estimator $\hat{\theta}$, the mean square error can be expressed as

$$mse(\hat{\theta}) = E(\hat{\theta} - \theta)^2 = E(\hat{\theta} - E(\hat{\theta}))^2 = V(\hat{\theta}).$$

The mean square error combines variability of the estimator with its bias because

$$mse(\hat{\theta}) = E(\hat{\theta} - \theta)^2 = V(\hat{\theta}) + [b(\hat{\theta})]^2.$$

Example 12.5 The mean square error in estimating the population mean μ by the sample mean $\bar{x}$ is

$$mse(\overline{X}) = E(\overline{X} - \mu)^2 = E[\overline{X} - E(\overline{X})]^2$$
$$= E(\overline{X}^2) - [E(\overline{X})]^2$$
$$= E(\overline{X}^2) - \mu^2, \quad \text{since } E(\overline{X}) = \mu$$
$$= \frac{\sigma^2}{n} = V(\overline{X}).$$

Self Assessment Exercises

1. The two methods of estimating a population parameter from the sample statistic are ………………………………………….....
2. State the four properties of a good estimator.

Self Assessment Answers

3.4 Methods of Estimation

Some methods of estimation that offer general technique for finding estimators for population parameters are as follows.

(i) Method of Moments

With this method, we make use of the moments of the distributions together with the sample moments to get estimates of the parameters. We equate the observed values of the sample moments with the corresponding population moments expressed as functions of the parameters and then solve the resulting equations for estimates of the parameters.

Let $x_1, x_2,..., x_n$ be independent and identically distributed observations from a distribution involving unknown parameters $\theta_1, \theta_2,..., \theta_k$. Let the first k raw moments of the distribution exist as functions $\alpha_r(\theta_1, \theta_2,..., \theta_k)$, $r = 1,..., k$ of the parameters. The *r*th raw moment of the distribution is denoted by $E(X^r)$, $r = 1, 2, 3,...$

Now if we denote the sample moment functions by

$$a_r = \sum_i X_i^r / n,$$

then the method of moments consist of equating the observed value of a_r to $E(X^r)$ and solving for $\theta_1, \theta_2,..., \theta_k$.

Example 12.6 Let $X_1, X_2,..., X_n$ be a random sample from an exponential distribution with pdf

$f(x) = \theta e^{-\theta x}, \quad x \geq 0, \; \theta > 0.$

Obtain the moment estimate of the unknown parameter θ.

Solution

$$E(X) = \int_0^\infty x\theta e^{-\theta x} dx = \theta \int_0^\infty x e^{-\theta x} dx = \frac{1}{\theta}.$$

The sample mean is $a_1 = \sum_i X_i \,/\, n = \bar{x}.$

Therefore equating these corresponding moments and solve for θ we have

$$\bar{x} = 1/\hat{\theta}$$

and so

$$\hat{\theta} = 1/\bar{x}\,.$$

Example 12.7 Let $X_1, X_2, ..., X_n$ be a random sample from Poisson distribution with parameter λ.

Obtain the moment estimate for the population mean μ.

Solution

$$\mu_1^{'} = E(X) = \sum_{x=0}^{\infty} x \frac{\lambda^x}{x!} e^{-\lambda} = \lambda \text{ and } a_1 = \frac{\sum_i X_i}{n} = \bar{x}\,.$$

Equating $\mu_1^{'}$ and a_1, we have

$$\hat{\lambda} = \bar{x}\,.$$

(ii) Method of Maximum Likelihood

This is a widely used method of estimation that has attractive large- sample properties.

Definition

Let $X_1, X_2, ..., X_n$ be a random sample from a discrete distribution with probabilities $f(x;\theta)$, in which θ may be a vector. Then, the likelihood function gives the probability of getting the particular sample values and is defined as the product of n probability mass functions, one evaluated at each sample value. That is

$$L(x_1, x_2, ..., x_n; \theta) = P(x_1;\theta)P(x_2;\theta)...P(x_n;\theta)$$
$$= \prod_{i=1}^{n} P(x_i;\theta)$$

If the sample is from a continuous distribution, the likelihood function is defined to be the product of the densities evaluated at the sample points. That is

$$L(x_1, x_2, ..., x_n; \theta) = f(x_1;\theta) f(x_2;\theta)...f(x_n;\theta)$$
$$= \prod_{i=1}^{n} f(x_i;\theta)$$

We then choose as estimates, those values of the parameters that maximize this probability. That is $(\hat{\theta}_1, ..., \hat{\theta}_k)$ is said to be the maximum likelihood estimate of the parameter $(\theta_1, ..., \theta_k)$ if it maximizes the likelihood function.

Example 12.8 Let $X_1, X_2, ..., X_n$ be a random sample from an exponential distribution with density $f(x) = \lambda e^{-\lambda x}, \quad x \geq 0, \; \lambda > 0.$

Obtain the maximum likelihood estimate of λ.

Solution

$$f(x) = \lambda e^{-\lambda x}, \quad x \geq 0, \; \lambda > 0.$$

$$L(x_1, x_2, ..., x_n; \lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda x} = \lambda^n e^{-\lambda \sum_{i=1}^{n} X_i}$$

$$\ln L(x;\lambda) = n \ln \lambda - \lambda \sum_{i=1}^{n} X_i \; .$$

$$\frac{D \ln L}{D\lambda} = \frac{n}{\lambda} - \sum_i X_i, \text{ and } \frac{D \ln L}{D\lambda} = 0 \Rightarrow \frac{n}{\hat{\lambda}} = \sum_i X_i$$

Therefore,

$$\hat{\lambda} = \frac{n}{\sum_i X_i} = \frac{1}{\bar{x}}.$$

Example 12.9 Let $X_1, X_2, ..., X_n$ be a random sample from an normal distribution with mean μ and variance σ^2.

Obtain the maximum likelihood estimate of μ and σ^2.

Solution

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(X-\mu)^2}{2\sigma^2}}$$

Therefore,

$$L(x_1, x_2,..., x_n; \mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(X-\mu)^2}{2\sigma^2}} = (2\pi\sigma^2)^{-n/2} e^{-\sum_{i=1}^{n}\frac{(X_i-\mu)^2}{2\sigma^2}}$$

$$\ln L(x; \mu, \sigma^2) = -\frac{n}{2}\ln\sigma^2 - \frac{n}{2}\ln(2\pi) - \sum_{i=1}^{n}\frac{(X_i-\mu)^2}{2\sigma^2}.$$

$$\frac{D\ln L}{D\lambda} = \sum_{i=1}^{n}\frac{(X_i-\mu)}{\sigma^2} = 0.$$

That is,

$$\sum_{i=1}^{n} X_i - n\mu = 0$$

hence,

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{x}.$$

Now,

$$\frac{D\ln L}{D\sigma^2} = \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{2\sigma^4} - \frac{n}{2\sigma^2} = 0$$

i.e.,
$$\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{2\sigma^4} = \frac{n}{2\sigma^2},$$

and,

$$\sum_{i=1}^{n}(X_i-\mu)^2 = n\sigma^2.$$

Hence

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{n}.$$

(iii) Method of Least Squares

This is most suitable for estimating parameters involve in linear models of the form

$$y = \alpha + \beta x + e$$

where y is the response variable, α and β are unknown parameters to be estimated x is a known independent constant and e is a random error component. The least squares estimators $\hat{\alpha}$ and $\hat{\beta}$ and the values of α and β which minimize the sum of squares of the error term

$$SSe = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \alpha - \beta x_i)^2 \qquad ....(1)$$

The minimum will occur for values of α and β that satisfy

$\frac{\partial SSe}{\partial \alpha} = 0$ and $\frac{\partial SSe}{\partial \beta} = 0$.

From the above SSe, we have

$$\frac{\partial SSe}{\partial \alpha} = -2\sum_{i=0}^{n} (y_i - \alpha - \beta x_i) = 0 \ \Rightarrow \sum_{i=0}^{n} \left(y_i - \hat{\alpha} - \hat{\beta} x_i\right) = 0 \ ....(2)$$

and

$$\frac{\partial SSe}{\partial \beta} = -2\sum_{i=1}^{n} x_i (y_i - \alpha - \beta x_i) = 0 \ \Rightarrow \sum_{i=0}^{n} \left(x_i y_i - \hat{\alpha} x_i - \hat{\beta} x_i^2\right) = 0. \ ....(3)$$

So that

$$\sum y_i = n\hat{\alpha} + \hat{\beta} \sum x_i \ ....(6)$$

$$\sum x_i y_i = \hat{\alpha} \sum x_i + \hat{\beta} \sum x_i^2 \ ....(7)$$

By solving the equations (6) and (7) simulataneously, we obtain the least squares estimator $\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$,

$$\boxed{\hat{\beta} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - \left(\sum x_i\right)^2}}$$

3.5 Interval Estimation

Point estimators provide us with single value that helps locate the parameter. But this estimate is always inexact and may contain some error. Besides, point estimation provides no information about the precision of the estimate. Interval estimation provides us with the opportunity of determining the size of the error we are likely to make when we estimate. With this method, an interval is given within which a population parameter is likely to lie, together with the chance that the interval includes the parameter.

Supposea random sample of size n is taken from a normal population with unknown mean μ and known standard deviation σ . Then if the mean of this sample is $\bar{x}$, a $100(1-\alpha)\%$ confidence interval for μ is given by

$$\left(\bar{X} - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right).$$

This can be written as

$$\bar{X} \pm Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

A $100(1-\alpha)\%$ confidence interval for a parameter is constructed so that $100(1-\alpha)\%$ of a large number of intervals calculated in the same way will capture or include the parameter.

Example 12.10 A random sample of weights of 10 students chosen from a class has a mean weight of 29.0 kg. Ifit is known that the standard deviation of the weights of all students in the class is 13.586, obtain a 95% confidence interval for the mean weight of all students in the class.

Solution

$$n = 10,\ \bar{X} = 29,\ \sigma = 13.586.$$
$$1-\alpha = 95\% = 0.95 \Rightarrow \alpha = 1 - 0.95 = 0.05,$$
$$\alpha/2 = 0.025 \Rightarrow z_{\alpha/2} = 1.96.$$

$$sem = \frac{\sigma}{\sqrt{n}} = \frac{13.586}{\sqrt{10}} = 4.30\,.$$

Therefore, a 95% confidence interval for μ is given by

$$\begin{aligned}\bar{X} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} &= 29.0 \pm (1.96)(4.30)\\ &= 29.0 \pm 8.4\\ &= (20.6,\ 37.4).\end{aligned}$$

Therefore, we are 95% confident that the above interval of 20.6 to 37.4 includes the parameter (μ).

3.5.1 Test of Hypotheses

In this section, we discuss test of hypothesis as the second basic class of statistical inference.Hypothesis testing involves the definitions of a hypothesis (null) as one set of possible population values and an alternative, a different set. There are many statistical procedures for determining, on the basis of a sample, whether the true population characteristic belongs to the set of values in the hypothesis or the alternative. In attempting to reach a decision about population on the basis of sample information, it is useful to make assumptions or guesses about the probability distribution of the population involved. Such assumptions are called *statisticalhypotheses*.A hypothesis under test is referred to as the *null hypothesis*, denoted by H_0. It states the absence of an effect.

A test of hypothesis H_0 is a rule that specifies for each possible set of values of the observations, whether to accept or reject H_0, should these particular values be observed. It consists in dividing the sample space S into two regions.The first region is that subset of S for each element of which the null hypothesis, H_0, will be rejected and is referred to as the *rejection*or *critical region*,denoted by w. The second is that subset for which H_0 will be accepted, referred to as the *acceptanceregion* and denoted by $S-w$.

Hence the null hypothesis is rejected, an alternative hypothesis, usually denoted H_1, is accepted in its place.Your decision to accept H_0, or to reject H_0 in favour of your alternative H_1 is based on the observed value $T=t_{obs}$ of a suitably chosen test statistic T.

If a hypothesis completely specifies the distribution, it is called *simple*. For instance, the hypothesis $H_0:\mu=\mu_0$, where μ_0 is a numerical value.But if the distribution is not completely specified, it is *composite*, e.g., the hypothesis $H_0:\mu\neq\mu_0$ or $\mu>\mu_0$.

Self AssessmentExercises

1. List three methods of estimation that offer general technique for finding estimators for population parameters.
2. What are the defects of point estimate in locating a parameter?
3. Hypothesis testing involves the testing of…………. and…………..hypothesis, denoted by ………….. and……..…. respectively.

Self Assessment Answers

3.6 One-Tailed and Two-Tailed Tests

There are two types of tests which could be performed, depending on the alternative hypothesis being made. Consider the hypothesis $H_0:\mu=\mu_0$ and the three different sets of alternatives: $H_1:\mu<\mu_0$, $H_1:\mu>\mu_0$ or $H_1:\mu\neq\mu_0$.

Alternatives of the first two types and the associated tests are called *one-sided*or *one-tailed*tests while those corresponding to the third type are called *two-sided* or *two-tailed* tests.

3.7 Type of Errors

Whatever procedure may be employed for testing H_0, there are two types of errors involved. The first is the error of rejecting H_0 when it is true, called the *Type I* error.

The second one is the error of not rejecting (accepting) H_0 when it is false, called *Type II* error. These are presented in the table below.

Decision	H_0 True	H_0 False
Accept H_0	Correct Decision	Type II Error
Reject H_0	Type I Error	Correct Decision

Let w denotes the subset of S consisting of the values for which H_0 will be rejected(called the critical region), then the probability of Type I error (called the *significance level*) is given by α where

$$\alpha = P(\text{reject} H_0 / H_0 \text{ is true}) = \text{P}(w / H_0)$$

$$\therefore \; 1-\alpha = P(\text{accept } H_0 / H_0 \text{ is true})$$

The probability of Type II error is given by β where

$$\beta = P(\text{accept} H_0 / H_0 \text{ is false}) = P(S - w / H_1)$$

$\because \; 1-\beta = P(\text{reject} H_0 / H_0 \text{ is false})$ called power of the test.

These probabilities are presented in the table below.

Decision	H_0 True	H_0 False
Accept H_0	$1-\alpha$	β
Reject H_0	α	$1-\beta$

Therefore, the steps for testing a given hypothesis include:

(i) State the null and alternative hypotheses.

(ii) Choose a suitable test statistic *T*(e.g., Z_c, t_c, χ_c^2, etc.).

(iii) Determine the critical (rejection) region w.

(iv) Based on sample information, compute the value t_c of your test statistic chosen in (ii) above.

(v) Make decision. If the observed value of the test statistic falls in the rejection region, i.e.,if $t_c \in w$, we reject H_0 at a significant level α and say that the test is significant, otherwise we accept H_0.

Example 12.11 A fluorescent bulbs manufacturer claims that his light bulbs have a mean life of 1000 hrs and standard deviation of 40 hrs. A random sample of 64 such bulbs is tested and the mean life was found to be 988 hrs. Test the manufacturer's claim at 5% significant level.

Solution

The hypotheses are

$$H_0 : \mu = 1000$$

$$H_1 : \mu < 1000$$

Since σ is known and the sample size, n, is large ($n \geq 30$), we use the test statistic

$$z_{obs} = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$$

We will reject H_0 if $z_{obs} < z_\alpha$, where

$$z_{obs} = \frac{988 - 1000}{40 / \sqrt{64}} = -2.4$$

and

$$z_\alpha = z_{.05} = 1.645$$

Therefore, since $z_{obs} < z_\alpha$, we reject H_0 at 5% significant level.

Now if n is large and the population variance is unknown, we use an estimator $\hat{\sigma}^2$ for it where $\hat{\sigma}^2 = \frac{nS^2}{n-1}$, and S^2 is the sample variance. Then the test statistic above becomes

$$z_{obs} = \frac{\overline{X} - \mu}{\hat{\sigma} / \sqrt{n}} \sim N(0,1).$$

In case of small sample size ($n < 30$), where the population variance is unknown, we use the test statistic

$$T_{obs} = \frac{\overline{X} - \mu}{S / \sqrt{n-1}} \sim t(n-1).$$

3.8 The Likelihood Ratio Test

This is a general method of test construction that frequently leads to satisfactory results.

Definition

Let $X_1, X_2,..., X_n$ be a random sample from a distribution with pdf $f(x;\theta)$, and required to test a simple hypothesis $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1 > \theta_0$ at significance level α. The likelihood ratio of the required test is defined by

$$\lambda = \frac{L(x/H_0)}{L(x/H_1)}$$

If the numerator is sufficiently smaller than the denominator, this indicates that x has a much higher probability under one of the alternatives than under H_0, and it then seems reasonable to reject H_0 when x is observed.

3.9 The Most Powerful Test

As already noted, the probability

$$P(\text{reject} H_0 / H_0 \text{ is false}) = 1 - \beta ,$$

where β is the probability of Type II error, is called the *power* of the test of H_0 against the alternative H_1. Therefore, a good test is the one for which α, the size of typeI error is small, and similarly, for which β, the size of Type II error, is also very small (ideally zero). That is, for which the power $1 - \beta$, is large (ideally unity). Hence, a good test could be achieved by fixing α and minimizing β (maximizing the power of the test).

Definition

A test of $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1$ is defined to be a most powerful test of size α if it has the size of its Type I error equal to α and if among all other tests of size α or less, it has the largest power or smallest size of Type II error.

The fundamental result underlying construction of a test of H_0 with maximum power against H_1 is the Neyman – Pearson lemma, discussed in the nextsection.

Self Assessment Exercise

1. There are……………………………types of error, namely………………………
2. A good test is the one for which ………………………………………………..
3. A way of maximizing the power of the test is by …………………………………

3.10 Neyman-Pearson Lemma

Let $X_1, X_2, ..., X_n$ be a random sample from a distribution with pdf $f(x;\theta)$. For testing a simple hypothesis $H_0 : \theta = \theta_0$ against a particular alternative $H_1 : \theta = \theta_1$, the critical region, w, of the most powerful test is determined by the likelihood ratio

$$\frac{L(x/H_0)}{L(x/H_1)} \leq k, \quad \text{for } k > 0$$

with k chosen to give the desired significance level.

Example 12.12 Let $X_1, X_2, ..., X_n$ be a random sample of size n from a normal distribution with an unknown mean μ and a known variance σ^2. Find a most powerful (MP) test for testing a simple hypothesis $H_0 : \mu = \mu_0$ against $H_1 : \mu = \mu_1 > \mu_0$.

Solution

The pdf is

$$f(x;\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Therefore,

$$L\ f_{\mu,\sigma^2}(x) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Now,

$$L(x,\mu_0,\sigma^2)) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu_0)^2}{2\sigma^2}} = \left(2\pi\sigma^2\right)^{-n/2} e^{-\frac{\sum(x_i-\mu_0)^2}{2\sigma^2}}$$

and

$$L(x,\mu_1,\sigma^2)) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu_1)^2}{2\sigma^2}} = \left(2\pi\sigma^2\right)^{-n/2} e^{-\frac{\sum(x_i-\mu_1)^2}{2\sigma^2}}$$

By N-P lemma, we have

$$\frac{L(\mu_0)}{L(\mu_1)} = \frac{\left(2\pi\sigma^2\right)^{-n/2} e^{-\frac{\sum(X_i-\mu_0)^2}{2\sigma^2}}}{\left(2\pi\sigma^2\right)^{-n/2} e^{-\frac{\sum(X_i-\mu_1)^2}{2\sigma^2}}} \leq k.$$

$$= e^{-\left[\frac{\sum (X_i-\mu_0)^2}{2\sigma^2} - \frac{\sum (X_i-\mu_1)^2}{2\sigma^2}\right]} \le k = e^{-\frac{1}{2\sigma^2}\left[-2\sum X_i(\mu_0-\mu_1)+n(\mu_0^2-\mu_1^2)\right]} \le k$$

Taking logarithms, we have

$$2\sum x_i(\mu_0-\mu_1) - n(\mu_0^2-\mu_1^2) \le 2\sigma^2 \ln k$$

i.e., $$\frac{\sum x_i}{n}(\mu_0-\mu_1) \le \frac{(\mu_0^2-\mu_1^2)}{2} + \frac{\sigma^2}{n}\ln k$$

$$\Rightarrow \quad \frac{\sum x_i}{n} \le \frac{(\mu_0^2-\mu_1^2)}{2(\mu_0-\mu_1)} + \frac{\sigma^2}{n(\mu_0-\mu_1)}\ln k$$

That is

$$\bar{x} \le \frac{\mu_0+\mu_1}{2} + \frac{\sigma^2}{n(\mu_0-\mu_1)}\ln k = c$$

$$\Rightarrow \bar{x} \le c$$

Therefore, the best critical region of size α for testing $H_0 : \mu = \mu_0$ against $H_1 : \mu = \mu_1 > \mu_0$ is

$$c = \{(x_1,..., x_n) : \bar{x} \le c\},$$

where c is selected such that $P(\bar{x} \le c / H_0 : \mu = \mu_0) = \alpha$.

Example 12.13 Let $X_1, X_2,..., X_n$ be a random sample of size n from a Poisson distribution with mean λ. Find a most powerful (MP) test for testing a simple hypothesis $H_0 : \lambda = \lambda_0$ against $H_1 : \lambda = \lambda_1$.

Solution

The pdf is $f(x;\lambda) = \frac{\lambda^x e^{-\lambda}}{x!} \Rightarrow Lf(x;\lambda) = \prod_{i=1}^{n} \frac{\lambda^x e^{-\lambda}}{x!}$

Therefore,

$$L(\lambda_0) = \prod_{i=1}^{n} \frac{\lambda_0^x e^{-\lambda_0}}{x!} = \frac{\lambda_0^{\sum x_i} e^{-\lambda_0 n}}{x!}$$

$$L(\lambda_1) = \prod_{i=1}^{n} \frac{\lambda_1^x e^{-\lambda_{11}}}{x!} = \frac{\lambda_1^{\sum x_i} e^{-\lambda_1 n}}{x!}$$

By N-P lemma, we have

$$\frac{L(\mu_0)}{L(\mu_1)} = \frac{\left(\lambda_0^{\sum x_i} e^{-\lambda_0 n}\right)/ x!}{\left(\lambda_1^{\sum x_i} e^{-\lambda_1 n}\right)/ x!} \le k \quad \Rightarrow \quad \frac{\lambda_0^{\sum x_i} e^{-\lambda_0 n}}{\lambda_1^{\sum x_i} e^{-\lambda_1 n}} \le k.$$

i.e., $\left(\frac{\lambda_0}{\lambda_1}\right)^{\sum x_i} e^{-(\lambda_0-\lambda_1)n} \le k \;\Rightarrow \sum x_i \ln\left(\frac{\lambda_0}{\lambda_1}\right) - (\lambda_0 - \lambda_1)n \le \ln k$

$$\Rightarrow \sum x_i \ln\left(\frac{\lambda_0}{\lambda_1}\right) + n(\lambda_1 - \lambda_0) \le \ln k$$

Therefore,

$$\sum x_i \le \frac{\ln k - n(\lambda_1 - \lambda_0)}{\ln(\lambda_0/\lambda_1)} = c\,.$$

Therefore, the best critical region of size α for testing $H_0 : \lambda = \lambda_0$ against $H_1 : \lambda = \lambda_1$ is

$$c = \{(x_1,..., x_n) : \sum x_i \le c\}\,.$$

Self Assessment Exercises

1. The error of not rejecting (accepting) H_0 when it is false, called …………….... error.
2. The error of rejecting H_0 when it is true, called the …………………………..… error.
3. What are the steps involved in testing for a given hypothesis?

Self Assessment Answers

4.0 Conclusion

Theis unit discussed theuse of statitical inference from samples about the populations from which they have been drawn in order to determine the sample characteristic, i.e. to know, if there is a"real" difference (i.e., is it present in the population) or just a"chance" difference (i.e. it could just be the result of random sampling error).

Statistical inference is an extension of your knowledge of random sampling, already discussed in the last unit.You were introduced to how to make estimates of population parameters; the error types as well as how to postulate and test hypothesis.

5.0 Summary

In this unit, we have discussed the following:

i) Estimation of population parameterswhich involves the determination, with a possible error due to sampling, of the unknown value of a population characteristic features. To express the accuracy of the estimates of population characteristics, one must also compute the standard errors of the estimates.

By definition, an estimator, $\hat{\theta}$ of a population parameter θ is a function of the sample observations $(x_1, x_2, ..., x_n)$ which is closest to the true value of the parameter in some sense.

There are two methods of estimating a population parameter from the sample statistic, viz: the **point** and **interval** estimation.

ii) Properties of a good Estimator include:

a) **Unbiasedness**: $\hat{\theta}$ is an unbiased estimator of θ if its expected value equals θ, i.e. if $E(\theta) = \theta$.

b) **Efficiency**: such that of all asymptotically unbiased estimators, no other estimator has smaller asymptotiv variance.

c) **Consistency**: An estimator $\hat{\theta}_n$ (having a random sample of size n) of a parameter θ is said to be a consistent estimator if it converges in probability to the true value of the parameter as the sample size n increases, i.e. $\hat{\theta}_n$ is consistent for θ if $\hat{\theta}_n \to \theta$ as $n \to \infty$.

d) **Minimum mean square error**, i.e. $E(\hat{\theta} - \theta)^2$ must be small.

iii) Methods of estimationfor finding estimators for population parameters are:

a) Method of moments

Using the moments of the distributions and the sample moments to obtain estimates of the parameters. Equate the observed values of the sample moments with the corresponding population moments expressed as functions of the parameters and solve the resulting equations for estimates of the parameters.

b) Method of Maximum Likelihood

A widely used method of estimation with attractive large-sample properties, defined as the product of n probability mass functions, one evaluated at each sample value, i.e.

$$L(x_1, x_2, ..., x_n; \theta) = P(x_1; \theta) P(x_2; \theta) ... P(x_n; \theta)$$
$$= \prod_{i=1}^{n} P(x_i; \theta)$$

For $X_1, X_2, ..., X_n$ as a random sample from a discrete distribution with probabilities $f(x; \theta)$, in which θ may is a vector, wheraasif the sample is from a continuous distribution,the likelihood function is defined as:

$$L(x_1, x_2, ..., x_n; \theta) = f(x_1;\theta) f(x_2;\theta) ... f(x_n;\theta)$$
$$= \prod_{i=1}^{n} f(x_i;\theta)$$

Then choose as estimates, those values of the parameters that maximize this probability.

c) Method of Least Squares

Most suitable for estimating parameters of linear models e.g. $y = \alpha + \beta x + e$ where y is the response variable, α and β are unknown parameters to be estimated x is a known independent constant and e is a random error component. The least squares estimators $\hat{\alpha}$ and $\hat{\beta}$ and the values of α and β which minimize the sum of squares of the error term.

Therefore,

$$\sum y_i = n\hat{\alpha} + \hat{\beta}\sum x_i \text{ ....(1)}$$

$$\sum x_i y_i = \hat{\alpha}\sum x_i + \hat{\beta}\sum x_i^2 \text{ ....(2)}$$

By solving the equations (1) and (2) simulataneously, we obtain the least squares estimators

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} \text{ ,}$$

$$\hat{\beta} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - \left(\sum x_i\right)^2} \text{ .}$$

iii) Interval estimation enables one to determine the size of the error during the estimate. For a random sample of size n] taken from a normal population with unknown mean μ and known standard deviation σ. Then if the mean of this sample is $\bar{x}$, a $100(1-\alpha)\%$ confidence interval for μ is given by

$$\left(\bar{X} - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X} + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right).$$

or

$$\bar{X} \pm Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

iv) Test of hypothesesinvolves the definitions of a hypothesis (null) as one set of possible population values and an alternative, a different set. Assumptions or guesses about the probability distribution of the population are made. Such assumptions are called *statisticalhypotheses*. A hypothesis under test is referred to as the *null hypothesis*, denoted by H_0. An alternative hypothesis, is denoted H_1.

The decision to accept H_0, or to reject H_0 in favour of your alternative H_1 is based on the observed value $T = t_{obs}$ of a suitably chosen test statistic T.

If a hypothesis completely specifies the distribution, it is called *simple*. For instance, the hypothesis $H_0 : \mu = \mu_0$, where μ_0 is a numerical value. But if the distribution is not completely specified, it is *composite*, e.g., the hypothesis H_0: $\mu \neq \mu_0$ or $\mu > \mu_0$.

v) Type of Errors: There are two types of errors, viz: a) *Type I* error — the error of rejecting H_0 whenit is true, and b) *Type II* error — the error of not rejecting (accepting) H_0 when it is false.

vi) Tests that lead to satisfactory results e.g. The Likelihood Ratio Test, The Most Powerful Test andNeyman-Pearson Lemma were considered.

6.0 Tutor-Marked Assignment (TMA)

1. Let $X_1, X_2, ..., X_n$ be a random sample of size n from a normal distribution with variance 1 and an unknown mean μ. Find a most powerful (MP) test for testing a simple hypothesis $H_0 : \mu = \mu_0$ against $H_1 : \mu = \mu_1 > \mu_0$.
2. Let $X_1, X_2, ..., X_n$ be a random sample of size n from a distribution with pdf $\frac{1}{\theta} e^{-x/\theta}$; $0 < x < \infty$. Obtain a best critical region for testing $H_0 : \mu = \mu_0$ against $H_1 : \mu = \mu_1 > \mu_0$.

7.0 References/Further Readings

Sheldon M. Ross (1997): Introduction to probability models, sixth edition. Academic Press. New York.

Mario Lefebre (2000): Applied probability and statistics. Springer.

Unit 4

Moments and Moment Generating Functions

Content

1.0 Introduction

This unit will take you through the measurements of the characteristics of a distribution, as it provides information about the distribution function. Different types of moments would be considered, then we turn Generating function, a simpler method of computing moments. A generating function is a transform of density and probability functions that simplifies analyses for certain sums of random variables.

2.0 Learning Outcome

At the end of this unit, you should be able to:

i. Measure the characteristics of a distribution
ii. Use a generating function to calculate moments

3.0 Learning Outcomes

3.1 Moments

Moments are one of the devices for measuring the characteristics of a distribution. They are of interest because of the information that they provide about the distribution function. In fact, in many circumstances, knowledge of the moments of all orders uniquely determines the distribution function.

The types of moments considered here are as follows.

(i) The moment about the origin (the n^th^ raw moment) of a random variable X , or of the corresponding distribution $f(x)$, is defined as

$$\mu_n' = E(X^n), \quad n = 1, 2, \ldots$$

where

$$E(X^n) = \begin{cases} \sum_j X_j^n P(x_j) & (discrete\ case) \\ \int_{-\infty}^{\infty} X^n f(x)\,dx & (continuous\,\mathrm{case}) \end{cases}$$

when n = 1, we have

$$\mu_1' = E(X^1) = \mu.$$

and n = 2 gives

$$\mu_2' = E(X^2) \text{ etc.}$$

That is, the first raw moment is the mean.

(ii) The n^th^ moment about the mean (i.e., the central moment) is defined as

$$\mu_n = E((X - \mu)^n), n = 1, 2, \ldots$$

where

$$E((X-\mu)^n) = \begin{cases} \sum_j (X_j - \mu)^n P(X_j) & (discrete case) \\ \int_{-\infty}^{\infty} (X-\mu)^n f(x)dx & (continuous case) \end{cases}$$

when n = 1, we have

$$\mu_1 = E(X-\mu)^1 = E(X) - \mu = 0, \text{ since } E(X) = \mu.$$

That is, the first central moment is zero.

When n = 2, we have

$$\begin{aligned}\mu_2 = E(X-\mu)^2 &= E(X^2 - 2\mu X + \mu^2) \\ &= E(X^2) - 2\mu E(X) + \mu^2 \\ &= E(X^2) - \mu^2, \quad \text{since } E(X) = \mu.\end{aligned}$$

This shows that the second central moment is the variance of the random variable. The mean μ, and the variance μ_2, generally represented by σ^2, play an important role in statistical data analysis. Purely as descriptive measures of the distribution, the mean represents a central value of the random variable and the variance represents the scatter around the central value.

The central moments can be represented in terms of the raw moments as

$$\mu_0 = 1, \ \ \mu_1 = 0$$

$$\begin{aligned}\mu_2 &= E(X-\mu_1')^2 \\ &= (X^2) - 2\mu_1' E(X) + \mu_1'^2 \\ &= \mu_2' - 2\mu_1'^2 + \mu_1'^2 \\ &= \mu_2' - \mu_1'^2\end{aligned}$$

$$\begin{aligned}\mu_3 &= E(X-\mu_1')^3 \\ &= (X^3) - 3\mu_1' E(X^2) + 2\mu_1'^3 \\ &= \mu_3' - 3\mu_1'\mu_2' + 2\mu_1'^3\end{aligned}$$

And in general,

$$\boxed{\mu_n = \mu_n' - \binom{n}{1}\mu_1'\mu_{n-1}' + \binom{n}{2}\mu_1'^2\mu_{n-2}' - \ldots (-1)^{n-1}(n-1)\,\mu_1'^n.}$$

Example 13.1

The probability distribution for a discrete random variable X is given by the table below

x	1	2	3	4
$P(x)$	$\frac{1}{16}$	$\frac{3}{16}$	$\frac{5}{16}$	$\frac{7}{16}$

Find the first three moments about the origin and the corresponding moments about the mean.

Solution

The first three moments about the origin are calculated as follows:

$$\mu_1' = E(X) = \sum XP(x)$$

$$= 1x\tfrac{1}{16} + 2x\tfrac{3}{16} + 3x\tfrac{5}{16} + 4x\tfrac{7}{16}$$

$$= \frac{50}{16} = 3.125$$

$$\mu_2' = E(X^2) = \sum X^2 P(x)$$

$$= 1x\tfrac{1}{16} + 4x\tfrac{3}{16} + 9x\tfrac{5}{16} + 16x\tfrac{7}{16}$$

$$= \frac{170}{16} = 10.625$$

$$\mu_3' = E(X^3) = \sum X^3 P(x)$$

$$= 1x\tfrac{1}{16} + 8x\tfrac{3}{16} + 27x\tfrac{5}{16} + 64x\tfrac{7}{16}$$

$$= \frac{608}{16} = 38.$$

For the corresponding moments about the mean, we use the relationship between the two types of moments discussed above.

$$\mu_1 = E(X - \mu_1')^1 = 0, \quad \text{since}\, E(X) = \mu_1'.$$

$$\begin{aligned}\mu_2 &= E(X - \mu_1')^2 \\ &= \mu_2' - \mu_1'^2 \\ &= 10.625 - (3.125)^2 \\ &= 0.86\end{aligned}$$

$$\begin{aligned}\mu_3 &= E(X - \mu_1')^3 \\ &= \mu_3' - 3\mu_1'\mu_2' + 2\mu_1'^3 \\ &= 38 - 3(3.125)(10.625) + 2(3.125)^3 \\ &= -0.57.\end{aligned}$$

Example 13.2

A continuous random variable X has density function given by

$$f(x) = 4x^3, \; 0 < x < 1$$
$$\quad 0, \; otherwise$$

Find the first three raw moments and hence obtain the 3rd central moment.

Solution

The first 3 raw moments are

$$\mu_1' = E(X) = \int_0^1 x f(x)dx$$
$$= \int_0^1 x 4x^3 dx$$
$$= \frac{4}{5}\left[x^5\right]_0^1$$
$$= 0.8$$

$$\mu_2' = E(X^2) = \int_0^1 x^2 f(x)dx$$
$$= 4\int_0^1 x^5 dx$$
$$= \frac{4}{6}\left[x^6\right]_0^1$$
$$= 0.67.$$

$$\mu_3' = E(X^3) = \int_0^1 x^3 f(x)dx$$
$$= 4\int_0^1 x^6 dx$$
$$= \frac{4}{7}\left[x^7\right]_0^1$$
$$= 0.57$$

The 3rd central moment is given by

$$\mu_3 = E(X - \mu_1')^3$$
$$= \mu_3' - 3\mu_1'\mu_2' + 2\mu_1'^3$$
$$= 0.57 - 3(0.8)(0.67) + 2(0.8)^3$$
$$= -0.014$$

Self Assessment Exercises

1. What is a moment?
2. What is the purpose of a moment?

Self Assessment Answers

3.2 Generating Functions

Instead of computing moments from their definitions, it is often simpler to make use of a generating function. A generating function is a transform of density and probability functions that simplifies analyses for certain sums of random variables.

3.2.1 The Moment Generating Functions

This is defined as the expected value of the random variable e^{tx}, where t is a real variable.

Definition

Given a random variable X of the discrete type with probability function $P(x)$, if there exists an $h>0$ such that

$$E(e^{tx}) = \sum_x e^{tx} P(x)$$

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

$$M_X(t) = E(e^{tx})$$

is called the moment generating function of X.

This function exists only for random variables that have moments of all orders and when expanded in powers of t, it yields the moments of the distribution as the coefficients of those powers. MGF is formally defined as

$$M_X(t) = E(e^{tx}) = \begin{cases} \sum_j e^{tx_j} P(x_j) & \text{(discretecase)} \\ \int_{-\infty}^{\infty} e^{tx} f(x)dx & \text{(continuous case)} \end{cases}$$

It can be shown that the existence of $M(t)$ for $|t| < h$ implies that the derivatives of all orders exist for it at $t=0$.

That is if $M_X(t) = E(e^{tx})$ exists, then for $i = 1, 2, ...,$ the r^th^ moment can be found as the r^th^ derivative of $M_X(t)$ evaluated at $t = 0$, and $\mu_r' = \frac{d^r M(t)}{dt^r}\Big|_{t=0}$. Therefore,

$$M'(t) = E(xe^{tx}) = \begin{cases} \sum_j x_j e^{tx_j} f(x_j) \\ \int_{-\infty}^{\infty} xe^{tx} f(x)dx \end{cases}$$

$$M''(t) = E(x^2 e^{tx}) = \begin{cases} \sum_j x_j^2 e^{tx_j} P(x_j) \\ \int_{-\infty}^{\infty} x^2 e^{tx} f(x)dx \end{cases}$$

and for any positive integer r,

$$M^{(r)}(t) = E(x^r e^{tx}) = \begin{cases} \sum_j x_j^r e^{tx_j} P(x_j) \\ \int_{-\infty}^{\infty} x^r e^{tx} f(x)dx \end{cases}$$

Setting $t = 0$, we see that the exponential function equals 1 and the expression on the right equals the r^{th} raw moment. To generate those moments, we replace the exponential function in the integrand of $M_X(t)$ by its power series expansion. That is

$$\begin{aligned} M_X(t) &= \int_{-\infty}^{\infty} e^{tx} f(x)dx \\ &= \int_{-\infty}^{\infty} \sum_0^{\infty} \frac{(tx)^r}{r!} f(x)dx \\ &= \int_{-\infty}^{\infty} (1 + tx + \frac{1}{2!}t^2x^2 + \ldots) f(x)dx \\ &= 1 + t\mu_1^{'} + \frac{1}{2!}t^2\mu_2^{'} + \ldots \\ &= \sum_{r=0}^{\infty} \frac{\mu_r^{'}}{r!} t^r. \end{aligned}$$

where $\mu_r^{'}$ is the r^{th} raw moment. This is also equivalent to the r^{th} derivative of $M_X(t)$ evaluated at $t = 0$. That is

$$\mu_r^{'} = \frac{d^r}{dt^r} M_X(t)\Big|_{t=0} = M^{(r)}(0).$$

Thus

$$\mu_1^{'} = M'(0)$$
$$\mu_2^{'} = M''(0)$$

and so on.

Therefore, if the MGF exists, we have

$$\mu = M'(0$$
$$\sigma^2 = M''(0) - [M'(0)]^2$$

Example 13.3 A discrete random variable X has Bernoulli distribution with probability function

$$f(x) = \begin{cases} P^x(1-P)^{1-x}, & x = 0,1. \\ 0, & \text{otherwise}. \end{cases}$$

i) The MGF is

$$M_X(t) = E(e^{tx})$$
$$= \sum_{x=0}^{1} e^{tx} P^x (1-P)^{1-x}$$
$$= \sum_{x=0}^{1} (Pe^t)^x (1-P)^{1-x}$$
$$= 1 - P + Pe^t$$

ii) Mean and Variance are computed as follows

$$M'(t) = Pe^t$$
$$\therefore \mu = M'(0) = P$$

$$M''(t) = Pe^t$$
$$\therefore \mu_2^{'} = M''(0) = P.$$

$$Var(X) = M''(0) - (M'(0))^2$$
$$= P - P^2$$
$$= Pq.$$

Example 13.4

A discrete r.v. X has binomial distribution with pdf

$$f(x) = \begin{cases} \binom{n}{x} P^x (1-P)^{n-x}, & x = 0,1,2,\cdots,n \\ 0 & ,\text{otherwise.} \end{cases}$$

i) The MGF of X is

$$M_X(t) = E(e^{tx})$$
$$= \sum_{x=0}^{n} e^{tx} \binom{n}{x} P^x (1-P)^{n-x}$$
$$= \sum_{x=0}^{n} \binom{n}{x} (Pe^t)^x (1-P)^{n-x}$$
$$= [(1-P) + Pe^t]^n$$

ii) The mean and variance are

$$M(t) = [(1-p) + pe^t]^n$$
$$M'(t) = npe^t[(1-p) + pe^t]^{n-1}$$
$$\therefore E(X) = M'(0) = np[1 - p + p]^{n-1}$$
$$= np.$$

Differentiating again we have

$$M''(t) = n(n-1)[(1-p)+pe^t]^{n-2}(pe^t)^2 + npe^t[(1-p)+pe^t]^{n-1}$$

$$\therefore M''(0) = n(n-1)[1-p+p]^{n-2}p^2 + np[1-p+p]^{n-1}$$

$$= n(n-1)p^2 + np$$

therefore,the variance is

$$V(X) = M''(0) - [M'(0)]^2$$

$$= n(n-1)p^2 + np - (np)^2$$

$$= (np)^2 - np^2 + np - (np)^2$$

$$= np - np^2$$

$$= np(1-p)$$

$$= npq.$$

Example 13.5

Let a discrete random variable X have a Poisson distribution with pdf

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

$$0, \; otherwise$$

Obtain (a) the MGF (b) mean and variance.

Solution

(a) The moment generating function is

$$M_X(t) = E(e^{tx}) = \sum_{x=0}^{\infty} e^{tx}\frac{\lambda^x}{x!}e^{-\lambda}$$

$$= e^{-\lambda}\sum_{x=0}^{\infty}\frac{(\lambda e^t)^x}{x!}$$

$$= e^{-\lambda}e^{\lambda e^t}$$

$$= e^{\lambda(e^t-1)}$$

(b) The mean and variance are derived as follows

$M_X(t) = e^{\lambda(e^t-1)}$, we have

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

and

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

so,

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

and

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

Therefore,

$$V(X) = M''(0) - [M'(0)]^2$$
$$= \lambda + \lambda^2 - \lambda^2$$
$$= \lambda$$

That is, $\mu = \sigma^2 = \lambda > 0.$

Example 13.6

A continuous random variable X has an exponential distribution with pdf

$$f(x) = \lambda e^{-\lambda x}, \quad x \geq 0,\ \lambda > 0$$
$$0, \quad \text{otherwise.}$$

Find (i) the MGF (ii) the mean and variance

Solution

(i) The mgf is

$$M_X(t) = E(e^{tx}) = \int_0^\infty e^{tx} f(x)dx$$
$$= \int_0^\infty e^{tx}\lambda e^{-\lambda x}dx = \lambda \int_0^\infty e^{-(\lambda-t)x}dx$$
$$= \lambda\left[\frac{e^{-(\lambda-t)x}}{-(\lambda-t)}\right]_0^\infty = -\frac{\lambda}{(\lambda-t)}[0-1]$$
$$= \frac{\lambda}{\lambda-t} = \left[1-\frac{t}{\lambda}\right]^{-1}$$

(ii) The mean and variance:

$$M(t) = \left[1-\frac{t}{\lambda}\right]^{-1}$$

So,

$$M'(t) = \frac{1}{\lambda}\left[1 - \frac{t}{\lambda}\right]^{-2}$$

$\therefore$ the mean is

$$\mu = M'(0) = \frac{1}{\lambda}.$$

$$M''(t) = -\frac{2}{\lambda}\left[1 - \frac{t}{\lambda}\right]^{-3}\frac{1}{\lambda}$$

$$= \frac{2}{\lambda^2}\left[1 - \frac{t}{\lambda}\right]^{-3}$$

$$\therefore M''(0) = \frac{2}{\lambda^2}$$

So,

$$V(X) = M''(0) - [M'(0)]^2$$

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

$$= \frac{1}{\lambda^2}$$

Example 13.7

A continuous random variable X has normal distribution with pdf

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty$$

$$0, \quad \text{otherwise.}$$

The moment –generating function is derived as follows:

$$M_X(t) = E(e^{tx}) = \int_{-\infty}^{\infty} e^{tx}\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

$$= \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}(x-\mu)^2 + tx} dx$$

By introducing $(t(x-\mu) + t\mu) = tx,$ we have

$$M_X(t) = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}(x-\mu)^2 + t(x-\mu) + t\mu} dx$$

$$= \frac{e^{\mu t}}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}(x-\mu)^2 + t(x-\mu)} dx$$

$$= \frac{e^{\mu t}}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}[(x-\mu)^2 - 2\sigma^2 t(x-\mu)]} dx \qquad -------(1)$$

We now complete the square in the exponent of (1) w.r.t. $(x-\mu)$. That is,

$$(x-\mu)^2 - 2\sigma^2 t(x-\mu) = (x-\mu)^2 - 2\sigma^2 t(x-\mu) + \sigma^4 t^2 - \sigma^4 t^2$$
$$= (x-\mu-\sigma^2 t)^2 - \sigma^4 t^2 ----------(2)$$

By substituting (2) in (1), we have

$$M_X(t) = \frac{e^{\mu t}}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}[(x-\mu-\sigma^2 t)^2 - \sigma^4 t^2]} dx$$
$$= \frac{e^{\mu t}}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}(x-\mu-\sigma^2 t)^2 + \frac{\sigma^4 t^2}{2\sigma^2}} dx$$
$$= \frac{e^{\mu t} e^{\frac{\sigma^2 t^2}{2}}}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}(x-\mu-\sigma^2 t)^2} dx$$

We can see that the last integral is the area under the normal curve with mean $\mu + \sigma^2 t$ and variance σ^2, which is equal to 1. Therefore,

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

Now to derive the mean and variance of this distribution using the MGF method, we have

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

So,

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

and

$$M''(t) = \mu(\mu + \sigma^2 \text{t})\exp(\mu t + \frac{\sigma^2 t^2}{2}) + 0 + \sigma^2 t(\mu + \sigma^2 t)\exp(\mu t + \frac{\sigma^2 t^2}{2}) + \sigma^2 \exp(\mu t + \frac{\sigma^2 t^2}{2}).$$

So, $M'(0) = \mu$,

$$M''(0) = \mu^2 + \sigma^2$$

Thus

$$\text{variance} = M''(0) - (M'(0))^2$$
$$= \mu^2 + \sigma^2 - \mu^2$$
$$= \sigma^2.$$

Example 13.8

Let a continuous random variable X have gamma distribution with pdf

$$f(x) = \frac{X^{\alpha-1}e^{-x/\beta}}{\beta^{\alpha}\Gamma(\alpha)}, \quad \alpha > 0,\ \beta > 0$$

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

where α and β are the parameters and

$$\Gamma(\alpha) = \int_0^\infty X^{\alpha-1}e^{-x}dx$$
$$= (\alpha-1)\Gamma(\alpha-1)$$
$$= (\alpha-1)!$$

The moment generating function is

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

now let $\mathrm{y} = \mathrm{x}(\frac{1}{\beta} - \mathrm{t}) = \mathrm{x}\left(\frac{1-\beta\mathrm{t}}{\beta}\right), \quad \mathrm{t} < \frac{1}{\beta}$

then $\mathrm{x} = \frac{\beta\mathrm{y}}{(1-\beta\mathrm{t})},\ dx = \frac{\beta}{1-\beta t}dy.$

$$\therefore M(t) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)}\int_0^\infty \left(\frac{\beta y}{1-\beta t}\right)^{\alpha-1} e^{-y}\left(\frac{\beta}{1-\beta t}\right)dy$$
$$= \frac{1}{\beta^{\alpha}\Gamma(\alpha)}\int_0^\infty \left(\frac{\beta}{1-\beta t}\right)^{\alpha} y^{\alpha-1}e^{-y}dy$$
$$= \frac{1}{(1-\beta t)^{\alpha}\Gamma(\alpha)}\int_0^\infty y^{\alpha-1}e^{-y}dy$$
$$= \left(\frac{1}{1-\beta t}\right)^{\alpha}, \ \text{provided } \mathrm{t} < \tfrac{1}{\beta}.$$

The mean and variance are derived from the MGF as follows:

$$M(t) = \left(\frac{1}{1-\beta t}\right)^{\alpha} = (1-\beta t)^{-\alpha}$$
$$\therefore M'(t) = -\alpha(1-\beta t)^{-\alpha-1}(-\beta)$$
$$= \alpha\beta(1-\beta t)^{-\alpha-1}$$

So,

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

Also,

$$M''(t) = -\alpha - 1(\alpha\beta)(1-\beta t)^{-\alpha-2}(-\beta)$$
$$= -(\alpha+1)(\alpha\beta)(1-\beta t)^{-\alpha-2}(-\beta)$$
$$= (\alpha+1)\alpha\beta^2(1-\beta t)^{-\alpha-2}$$
$$\therefore M''(0) = (\alpha+1)\alpha\beta^2$$

So, the variance is

$$\sigma^2 = V(X) = M''(0) - [M'(0)]^2$$
$$= (\alpha+1)\alpha\beta^2 - (\alpha\beta)^2$$
$$= (\alpha\beta)^2 + \alpha\beta^2 - (\alpha\beta)^2$$
$$= \alpha\beta^2$$

While we have seen that the moment generating function is a useful device for computing moments of random variables and their distributions, its main significance arises from the *uniqueness theorem*, which states that a moment- generating function uniquely determines it's distribution function. And it is often easier to find the moment generating function of a random variable than its distribution function. For instance, consider the random variable $Z = X + Y$, where X and Y are two independently distributed random variables. It then follows from the definition of a moment – generating function that the mgf of Z is the product of the moment- generating functions of X and Y. This is shown in the next section. No such simple relationship exist between the distribution function of Z and those of X and Y.

However, once the mgf of Z is known, it is theoretically possible to determine its distribution functions.

3.2.1.1 MGF of Sums of Independent Random Variables

If X and Y are two independently distributed random variables having moment generating functions $M_X(t)$ and $M_Y(t)$ respectively, then

$$M_{X+Y}(t) = M_X(t).M_Y(t)$$

That is, the moment generating function of a sum of two independent random variables is the product of the moment generating function of the summands.

Proof

The joint pdf of X, Y is $f(x, y)$.

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

Extension of this property to any finite number of independent random variables (the summands) is easily made by induction. That is, if $X_1, X_2, ..., X_n$ are independent, then $M_{X_1+X_2+...+X_n}(t) = \prod_{r=1}^{n} M_{X_r}(t)$.

Moreover, if the random variables have identical distributions, say, with common mgf, $M_X(t)$ then

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

Example 13.9

Let $X_1, X_2, ..., X_n$ be independent sequence of random variables from a Bernoulli population with parameter p, each with the distribution

$$P(X = x) = p^x q^{1-x}, \ \ x = 0,1.$$

The mgf of this distribution is

$$\begin{aligned}
M_X(t) = E(e^{tx}) &= \sum_0^1 e^{tx} P(x) \\
&= e^{t(0)} q + e^{t(1)} p \\
&= q + pe^t.
\end{aligned}$$

So, the moment generating function of the sum $X_1 + X_2 +, ..., + X_n$, is

$$\begin{aligned}
M_{X_1+X_2+...+X_n}(t) &= [M_X(t)]^n \\
&= [q + pe^t]^n
\end{aligned}$$

3.2.2 Factorial Moment Generating Functions

This is another useful generating function in determining the moments of a random variable X. For a discrete random variable X, taking values 0, 1, 2, . . . Let $p_i = P(X = i)$. Then this function is defined as

$$p(t) = E(t^X) = \sum_{x=0}^{\infty} t^X p_k .$$

This generates the factorial moments, α_k of the random variable X as the k^th derivative of $p(t)$ evaluated at $t = 1$.

Such derivatives are

$$p'(t) = E[Xt^{X-1}],$$

$$p''(t) = E[X(X-1)t^{X-2}],$$

$$p'''(t) = E[X(X-1)(X-2)t^{X-3}]$$

and so on, with values at $t = 1$ given by

$$p'(1) = E[X],$$

$$p''(1) = E[X(X-1)],$$

$$p'''(1) = E[X(X-1)(X-2)].$$

And in general, the k^th factorial moments are given by the formula

$$p^k(t) = \frac{d^k p(t)}{dt^k},$$

so that

$$\begin{aligned} p^k(1) &= \frac{d^k p(t)}{dt^k}\Big|_{t=1} \\ &= E[X(X-1)(X-2)...(X-k+1)] \\ &= \alpha_k . \end{aligned}$$

Clearly, knowledge of the factorial moments is equivalent to knowledge of the raw moments; the factorial moments are connected with the raw moments, $\mu_1^{'}$ by the relations

$$\begin{aligned} \alpha_1 &= E(X) = \mu_1^{'} \\ \alpha_2 &= E[X(X-1)] = \mu_2^{'} - \mu_1^{'} \\ \alpha_3 &= E[(X(X-1)(X-2)] = \mu_3^{'} - 3\mu_2^{'} + 2\mu_1^{'} \\ \alpha_4 &= E[X(X-1)(X-2)(X-3)] = \mu_4^{'} - 6\mu_3^{'} + 11\mu_2^{'} - 6\mu_1^{'} \end{aligned}$$

and so on.

or conversely,

$$\begin{aligned}
\mu_1' &= \alpha_1 \\
\mu_2' &= \alpha_2 + \alpha_1 \\
\mu_3' &= \alpha_3 + 3\alpha_2 + \alpha_1 \\
\mu_4' &= \alpha_4 + 6\alpha_3 + 7\alpha_2 + \alpha_1.
\end{aligned}$$

$$etc.$$

Example 13.10

A random variable X has Binomial distribution. The factorial moment generating function is derived as

$$\begin{aligned}
P_X(t) = E(t^X) &= \sum_0^n \binom{n}{x} t^x p^x (1-p)^{n-x} \\
&= \sum_0^n \binom{n}{x} (pt)^x (1-p)^{n-x} \\
&= (q+pt)^n.
\end{aligned}$$

So,

$$\begin{aligned}
P'(t) &= np[(1-p)+pt]^{n-1} \\
p''(t) &= np^2(n-1)[(1-p)+pt]^{n-2} \\
p'''(t) &= np^3(n-1)(n-2)[(1-p)+pt]^{n-3}
\end{aligned}$$

from which, by substituting $t=1$, we obtain

$$\begin{aligned}
P'(1) &= \alpha_1 = np \\
p''(1) &= \alpha_2 = np^2(n-1) \\
p'''(1) &= np^3(n-1)(n-2)
\end{aligned}$$

and so on.

So that

$$\begin{aligned}
\mu &= E(X) = np, \\
\sigma_X^2 &= E(X^2) - E(X)^2 \\
&= E[(X(X-1)] + E(X) - E(X)^2 \\
&= np^2(n-1) + np - (np)^2 \\
&= np - np^2 \\
&= npq.
\end{aligned}$$

Example 13.11

Let X have a Poisson distribution with mean value λ. The f.m.g.f. is given by

$$P_X(t) = E(t^X) = \sum_0^{\infty} t^X \frac{\lambda^x}{x!} e^{-\lambda}$$
$$= e^{-\lambda} \sum_0^{\infty} \frac{(\lambda t)^x}{x!}$$
$$= e^{-\lambda} e^{\lambda t} = e^{\lambda(t-1)}.$$

The derivatives are

$$P^{'}(t) = \lambda e^{\lambda(t-1)}$$
$$p^{''}(t) = \lambda^2 e^{\lambda(t-1)}$$
$$p^{'''}(t) = \lambda^3 e^{\lambda(t-1)}$$

and,

$$p^k(t) = \lambda^k e^{\lambda(t-1)}$$

so that

$$p^k(1) = \lambda^k.$$

In particular,

$$E(X) = \mu = p^{'}(1) = \lambda,$$
$$E[(X(X-1)] = p^{''}(1) = \lambda^2$$

so that

$$\sigma_X^2 = E(X^2) - E(X)^2$$
$$= E[(X(X-1)] + E(X) - E(X)^2$$
$$= \lambda^2 + \lambda - \lambda^2$$
$$= \lambda.$$

In conclusion, for some random variables (usually discrete), factorial moments are easier to calculate than raw moments and they provide very concise formulae for the moments of certain discontinuous distributions of the binomial type.

3.2.3 Cumulant Generating Functions (c.g.f)

Let $M_X(t)$ be the moment generating function of a random variable X, then, the cumulant generating function is defined as

$$K(t) = \ln M_X(t).$$

That is, the logarithm of the moment generating function is called the cumulant generating function.

If $E(X^r)$ exists, then $K(t)$ admits the Taylor expansion

$$K(t) = \ln M(t) = tk_1 + \frac{t^2}{2!}k_2 + \frac{t^3}{3!}k_3 + ... = \sum_{r=0}^{\infty} \frac{t^r}{r!}k_r$$

$$(\text{since} \ln M(0) = \ln(1) = 0).$$

where

$$K^{'}(t) = \ln M_X(t) = \frac{1}{M(t)} M^{'}(t)$$

from which

$$K^{'}(0) = \frac{1}{M(0)} M^{'}(0) = M^{'}(0), \quad (\text{since} M(0) = 1)$$

$$= k_1 = \mu_1^{'}$$

and

$$K^{''}(t) = \frac{1}{[M(t)]^2}[M(t)M^{''}(t) - [M^{'}(t)]^2$$

from which

$$K^{''}(0) = \frac{1}{[M(0)]^2}[M(0)M^{''}(0) - [M^{'}(0)]^2$$

$$= k_2 = \mu_2^{'} - \mu_1^{'2}$$

and so on.

That is,

$$k_1 = \mu_1^{'}$$

$$k_2 = \mu_2^{'} - \mu_1^{'2}$$

$$k_3 = \mu_3^{'} - 3\mu_1^{'}\mu_2^{'} + 2\mu_1^{'3}$$

and so on.

And in terms of the central moments

$$k_1 = \mu$$

$$k_2 = \mu_2$$

$$k_3 = \mu_3$$

$$k_4 = \mu_4 - 3\mu_2^{\ 2}$$

etc.

Example 13.12

Let X have Poisson distribution with mgf

$$M(t) = e^{\lambda(e^t - 1)}.$$

The cumulant generating function is

$$\begin{aligned} K(t) = \ln M(t) &= \lambda(e^t - 1) \\ &= \lambda\left(\sum_{r=0}^{\infty} \frac{t^r}{r!} - 1\right) \\ &= \lambda\left(\sum_{r=1}^{\infty} \frac{t^r}{r!}\right) \end{aligned}$$

So

$$k_r = \lambda, \quad \text{for all } r.$$

Thus the cumulants of this distribution are equal to λ.

Example 13.13

Let X have a Binomial distribution with moment generating function

$$M(t) = \left(q + pe^t\right)^n.$$

The cumulant generating function is

$$\begin{aligned} K(t) &= \ln M(t) = \ln[q + pe^t]^n \\ &= t(np) + \frac{1}{2!}t^2[np(1 - p + np) - n^2p^2] + \ldots \end{aligned}$$

So,

$$\begin{aligned} k_1 &= \mu_1' = np \\ k_2 &= np(1 - p + np) - n^2p^2 = np(1 - p) \\ k_3 &= np(1 - p)(1 - 2p) \\ & \qquad etc. \end{aligned}$$

That first cumulant is the np and subsequent cumulants are given by the recurrence relation

$$k_{r+1} = pq\frac{d\,k_r}{d\,p}.$$

In conclusion, cumulant generating function provides an easiest way of computing μ and σ^2 for a given distribution.

3.2.4 Characteristic Function

The characteristic function is defined as the expected value of the random variable e^{itx}, where t is real and $i = \sqrt{-1}$.

Definition

The characteristics function of a random variable X with probability density function, $f(x)$ is defined as

$$\phi(t) = E(e^{itx}) = \begin{cases} \sum_{-\infty}^{\infty} e^{itx} p(x) & \text{(discretecase)} \\ \int_{-\infty}^{\infty} e^{itx} f(x)dx & \text{(continuous case)} \end{cases}$$

where i denotes the imaginary unit and t is an arbitrary real.

Every distribution has a unique characteristic function and to each characteristic function, there corresponds a unique distribution of probability.

If X and Y are independent random variables with characteristic functions $\phi_X(t)$ and $\phi_Y(t)$ respectively, then the characteristic function of $(X+Y)$ is given by

$$\phi_{X+Y}(t) = \phi_X(t).\phi_Y(t).$$

The characteristic function has the advantage that it always exists since

$$\left|e^{itx}\right| = \left|Cos\,tx + iSin\,tx\right| = 1,$$

continuous and bounded for all x.

If $E(X^r)$ exists, then, it is possible to express $\phi(t)$ as a Maclaurin series. That is

$$\begin{aligned}\phi(t) &= \phi(0) + t\phi'(0) + \frac{1}{2!}t^2\phi''(0) + \frac{1}{3!}t^3\phi'''(0) + ... \\ &= 1 + it\mu_1' + \frac{1}{2!}(it)^2\mu_2' + \frac{1}{3!}(it)^3\mu_3' + ... \\ &= 1 + it\mu_1' - \frac{1}{2}t^2\mu_2' - \frac{1}{3!}it^3\mu_3' + ... \\ &= \sum_{k=0}^{\infty} \frac{(it)^k}{k!}\mu_k' \ .\end{aligned}$$

where μ_n' is the nth moment about origin and $\mu_0' = 1$. This indicates that the characteristic function generates moments in the same way as the moment generating function, except that each differentiation introduces factor of i. Thus the rth moment can be found as the rth derivative of $\phi(t)$ evaluated at $t = 0$, and

$$\mu_r' = (i)^{-r} d^r\phi(t)/dt^r\Big|_{t=0} = (i)^{-r}\phi^r(0).$$

Example 13.14

A continuous random variable X has pdf given by

$f(x) = e^{-x}, \; x > 0.$

The characteristic function is

$$\phi(t) = E(e^{itx}) = \int_0^\infty e^{itx} e^{-x} dx$$
$$= \int_0^\infty e^{-(1-it)x} dx$$
$$= \frac{-1}{(1-it)} e^{-(1-it)x} \Big|_0^\infty$$
$$= \frac{1}{1-it}.$$

Example 13.15

Let a continuous random variable X have Gamma distribution with pdf given by

$f(x) = \dfrac{X^{\alpha-1} e^{-x/\beta}}{\beta^\alpha \Gamma(\alpha)}, \; x > 0, \, \alpha > 0, \; \beta > 0.$

The characteristic function is

$$\phi(t) = E(e^{itx}) = \frac{1}{\beta^\alpha \Gamma(\alpha)} \int_0^\infty X^{\alpha-1} e^{-x/\beta} e^{itx} dx$$
$$= \frac{1}{\beta^\alpha \Gamma \alpha} \int_0^\infty X^{\alpha-1} e^{-(\frac{1}{\beta}-it)x} dx$$

set $y = x(\frac{1}{\beta} - it) = x(\frac{1-\beta it}{\beta})$

then, $x = \dfrac{\beta y}{1-\beta it}, \; dx = \left(\dfrac{\beta}{1-\beta it}\right) dy.$

$$\therefore \phi(t) = \frac{1}{\beta^\alpha \Gamma \alpha} \int_0^\infty \left(\frac{\beta y}{1-\beta it}\right)^{\alpha-1} e^{-y} \left(\frac{\beta}{1-\beta it}\right) dy$$

$$= \frac{1}{\beta^\alpha \Gamma \alpha} \int_0^\infty \left(\frac{\beta}{1-\beta it}\right)^{\alpha} y^{\alpha-1} e^{-y} dy$$
$$= \left(\frac{1}{1-\beta it}\right)^{\alpha} \frac{1}{\Gamma \alpha} . \Gamma \alpha$$
$$= (1-\beta it)^{-\alpha}, \; t < \tfrac{1}{\beta}.$$
$$\therefore \phi(t) = (1-\beta it)^{-\alpha}.$$

Example 13.16

Given that $f(x) = \lambda e^{-\lambda x}, \ \lambda x > 0.$

The characteristic function is

$$\phi(t) = E(e^{itx}) = \lambda \int_0^\infty e^{itx} e^{-\lambda x} dx$$

$$= \lambda \int_0^\infty e^{-(\lambda - it)x} dx$$

$$= \frac{-\lambda}{(\lambda - it)} e^{-(\lambda - it)x} \Big|_0^\infty$$

$$= \frac{\lambda}{\lambda - it} = \left(1 - \frac{it}{\lambda}\right)^{-1}.$$

Self Assessment Exercises

1. A continuous random variable X has pdf given by $f(x) = e^{-x}, \ x > 0$. Find the characteristic function

Self Assessment Answers

4.0 Conclusion

This unit discussed the measurements of the characteristics of a distribution, called the moment, which provides information about the distribution function. Different types of moments were considered. You also learnt about Generating function, a simpler method of computing moments.

5.0 Summary

In this unit, you have learnt the following:

I) Moments as a way of measuring the characteristics of a distribution. We discussed about the moment about the origin (the n^th raw moment) of a random variable X , or of the corresponding distribution $f(x)$, defined as $\mu_n^{'} = E(X^n), \quad n = 1, 2, ...$ as well as the n^th moment about the mean (i.e., the central moment) is defined as $\mu_n = E((X - \mu)^n), n = 1, 2, \ ...$

2) Generating Functions and its variants as methods of computing moments from a generating function. A generating function is a transform of density and probability functions that simplifies analyses for certain sums of random variables.

6.0 Tutor-Marked Assignment (TMA)

1. Let a continuous random variable X have Gamma distribution with pdf given by

$f(x) = \dfrac{X^{\alpha-1} e^{-x/\beta}}{\beta^{\alpha}\Gamma(\alpha)}, \; x > 0, \alpha > 0, \; \beta > 0$. Find the characteristic function.

7.0 References/Further Reading

W. J. Decoursey (2003): Statistics and probability for engineering applications. Newness.

Sheldon M. Ross (1997): Introduction to probability models, sixth edition. AcademicPress. New York.

Answer to SAAs

Module 4: Unit 1

1. Let X be a discrete random variable with mean μ and variance σ^2. Then if $\varepsilon > 0$ is any positive real number, we have

$$P(|X-\mu| \geq \varepsilon) \leq \frac{V(X)}{\varepsilon^2}.$$

2. False

3. The expected value, or the mean.

4. 1 (Unity)

Module 4: Unit 2

1. $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$.

Module 4: Unit 3

1. The point and interval estimation.
2. i) Unbiasedness ii) Efficiency, iii) Consistency iv) Small (minimum) mean square error.
3. i) Method of moments

   ii) Method of Maximum Likelihood

   iii) Method of Least Squares
4. i) Always inexact

   ii) May contain errors

   iii) No information about the precision of the estimate.
5. Null and Alternative hypothesis, denoted by H_0 and H_1 respectively.
6. Two; Type I and Type II errors.
7. α, the size of type I error is small; and for which β, the size of Type II error, is also very small (ideally zero).
8. Fixing α and minimizing β.
9. Type II error
10. Type I error
11. i) State the null and alternative hypotheses.

    (i) Choose a suitable test statistic *T*(e.g., Z_c, t_c, χ^2_c, etc.).

(ii) Determine the critical (rejection) region w.

(iii) Based on sample information, compute the value t_c of your test statistic chosen in (ii) above.

(iv) Make decision. If the observed value of the test statistic falls in the rejection region, i.e., if $t_c \in w$, we reject H_0 at a significant level α and say that the test is significant, otherwise we accept H_0.

Module 4: Unit 4

1. Moment is a device for measuring the characteristics of a distribution.
2. It provides informationabout the distribution function