Probability and Distribution Theory

Probability and Distribution Theory

A. Adedayo Adepoju

IBADAN UNIVERSITY PRESS
2014

Ibadan University Press
Publishing House
University of Ibadan
Ibadan, Nigeria.

First Published 2014

ISBN: 978 – 978 – 8456 – 68 – 1

Contents

Foreword

This book on probability and distribution theory is written in a language that is intelligible and has examples that have ready appeal. The worked examples serve to illustrate the procedures and reinforce definitions and rules.

In some textbooks on this subject, certain parts of probability suffer from lack of coherence because the usual grouping and treatment of problems depend largely on accidents of its historical development.

A salient feature of this book however is to distinguish "definition" of probability from the "method" of its calculation. The presentation style of the book is lucid; the topics are arranged in increasing level of difficulty and they cover essential constituents of probability and distribution theory. The subject matter is not new but a major attraction of the book is enough exercises of all types and these exercises, which have been specially constructed to illustrate the theory are so designed that after doing them, the student should not only have a better grasp of the theory, but should also know the motivation for the various steps. Undergraduate students in statistics will find the book useful. Statisticians and non-major students in statistics will find the book a reliable guide.

Professor G.N. Amahia
Statistics Department
University of Ibadan
Ibadan

General Introduction and Course Objectives

Students must have encountered a number of problems relating to probability distributions that are sufficiently important to warrant identification by name and a more systematic study to ascertain their properties and applications. Their significance lies in their utility as workable models for many important practical applications. It provides an introduction to probability and a detailed illustrations of many distributions used in statistics. It is important, that one be familiar with the concepts of distribution theory since the knowledge of the behaviour of these probability distributions is very useful in statistical inference. The subject matter of statistics is very broad extending from the planning and design of experiments, surveys and other studies which generate data to the collection, analysis, presentation and interpretation of the data. Hence, numerical data constitute the raw material of this subject.

The first chapter discusses useful concepts of set theory, combination, and permutations, independent and conditional events. Probability theory provides a rational basis for inference and decision making about population or a larger group from which samples are taken. Chapter three introduces functions and random variables (both discrete and continuous) and some descriptions of the probability distributions for random variables. The rest of the chapters address problems regarding probability distributions ranging from one dimensional case to two dimensional situations.

Objectives

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

(1) Discuss several methods of measuring probabilities.
(2) Discuss the theory of discrete and continuous random variables with their corresponding properties.
(3) Discuss the concept of distribution functions and their properties.
(4) Appreciate the theory of frequently occurring probability distributions, and obtain expressions for the expected values and variances.

(5) Discuss the importance of generating functions and its usefulness in connection with sums of independent random variables.
(6) Discuss the significance of some basic limit theorems in finding approximations to distributions of statistics and/ or statistics themselves.
(7) Handle cases involving two-dimensional random variables since in almost all applications random variables do not occur singly and
(8) Appreciate the usefulness of the t and the F distributions and the extensive use in tests concerning normal populations.

1
Basic Concepts of Probability

Introduction
It will be useful to begin this book by defining some basic concepts of probability. One of the fundamental tools of statistics is probability, which had its formal origin in the games of chance in the seventeenth century. Probability theory is a fascinating subject which can be studied at a variety of intellectual and mathematical levels. Probability lies at the foundation of statistical theory and application.

Objectives
At the end of this chapter, you should be able to:
(1) Explain some basic terms used in probability theory;
(2) Differentiate between the different approaches to the definition of probability; and
(3) Solve problems involving combinations and permutations.

Pre-Test
(1) What do you understand by the term 'probability'?
(2) Define a statistical experiment with simple relevant examples.
(3) What do you understand by the word permutation?
(4) Define the word combination?
(5) What is a conditional probability?

Content
Basic Set Theory
***Random Experiment*:** A random experiment is an experiment in which,
(a) All the outcomes of the experiment are known in advance;
(b) Any performance of the experiment results in an outcome that is not known in advance; and
(c) The experiment can be repeated in an identical condition.

Simply put, any experiment that can have more than one possible outcomes/results that are known in advance is called a random experiment.

Examples

(i) A coin is tossed once; the result of the experiment is either head or tail.
(ii) A die is rolled once, the possible outcomes are 1, 2, 3, 4, 5, or 6.

***Set*:** A set can be defined as any well-defined collection of objects. Individual objects that belong to a set are called members or elements of the set. Examples are:

A set of vowels is $\{a, e, i, o, u\}$

A set of animals $\{dog,\ tiger,\ elephant\}$

A *SET* is denoted by upper case letter and an element, a lower case.

(i) A unit set is a set composed of only one element.
(ii) A set that contains no elements is called the empty set, or null set, and is designated by the symbol ϕ.

***Sample Space*:** A set S which consists of all possible outcomes of a random experiment is called a sample space.

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

***Events*:** An "event" is a subset of the sample space.

(i) An elementary event is a single possible outcome of an experimental trial. It is thus an event which cannot be further subdivided into a combination of other events.

e.g. $A = \{1\}$ or $B = \{5\}$

(ii) Compound or composite event is an event that can be subdivided into small events.

e.g. $A = \{1,\ 3,\ 5\}$ or $A = \{2,\ 4,\ 6\}$

***Subset*:** A collection made up of some of the objects in a set is called a subset.

e.g. If $A = \{2, 4, 6, 8\}$ is a set, then a subset of A can be

$B = \{2\}$, or $\{2, 6, 8\}$, or $\{4, 6\}$, etc

Basic Concepts of Probability

Since probability originated from games of chance, actions such as the following are familiar in the theory of probability; tossing a coin, throwing a die, spinning a roulette wheel, drawing a card etc. Here, the outcome of a trial is uncertain, however, it is recognized that even though, the outcome of a trial is uncertain; there is a predictable long-term outcome (relative frequency). It is known, for example, that in many throws of an ideal (balanced symmetrical) coin about half of the trials will result in heads. This concept of probability may be defined and interpreted in several different ways, the chief ones arising from the following:

Classical (A Priori) Approach to Probability

In this approach, the total number of all possible outcomes is fixed and known prior to the performance of any experiment. Similarly, the number of outcomes that have the particular characteristic associated with the event in question is fixed and known beforehand. All the outcomes are mutually exclusive and equally likely.

In this situation, the probability that an event occurs is defined as the ratio of:

The number of outcomes (results) in an experiment that have the characteristic associated with the event to the total number of possible outcomes of the experiment.

Therefore, the probability of a given event can be determined without necessarily performing the experiment.

That is, for event E in the sample space S.

$$P(E) = \frac{\text{No. of results in E}}{\text{No. of results in S}} = \frac{n(E)}{n(S)}$$

That is, if there are E possible outcomes favourable to the occurrence of an event A and F possible outcomes unfavorable to the occurrence of this event, then the probability that A will occur is

$$P(A) = \frac{E}{F+E} = \frac{\text{No. of outcomes favorable to A}}{\text{Total No. of possible outcomes}}$$

Example 1.1

(1) If a fair 6-sided die is rolled, the probability that 1 will be observed is equal to $1/6$ and is the same for other five sides.

(2) If a card is picked at random from a well shuffled deck of ordinary playing cards, the probability of picking a heart is $^{13}/_{52}$.

Relative Frequency (A Posteriori) Approach to Probability

This is when the probability ratio associated with a given event is not known before experiments are performed. Rather, the probability ratio is determined only after a relatively large number of trials of the experiments under identical conditions. In this type of situation, the probability that an event may occur is defined as the ratio of:

The number of trials in which the specified event occurred to the total number of trials performed.

The two basic assumptions underlying this definition are that;

(a) A relatively large number of trials is performed under identical conditions.

(b) As the total number of trials is increased, the probability ratio approaches the true value.

Consider an experiment in which there are independently repeated trials.

The number of outcomes "f" of an event A in which we are interested is recorded in n trials of the experiment. Then, the relative frequency of occurrence of A is

$$P(A) = {}^{f}/_{n}.$$

Example 1.2
In an experiment, a box contains N balls, where N is a large but unknown number. Some of the balls may be red. The experimenter defines the event E as "obtain a red ball" and wishes to assign a probability p to this event.

Note: Since the number of results is not known and the number of favourable results is unknown, the experimenter cannot apply the classical definition of probability. He therefore decides to draw a sample with replacement of balls from the box, making sure that every ball has an equal chance of being selected.

$\Rightarrow$ He draws a total of 10 balls from the box and 4 are red, so

$$f_1/n = 4/10 \cong P = 0.4$$

$\Rightarrow$ He draws 100 balls, 38 are red,

$$f_2/n = 38/100 \approx P = 0.38$$

$\Rightarrow$ He draws 1000 balls and 388 are red,

$$f_3/n = 388/1000 = 0.388$$

$\Rightarrow$ 10,000, 3840 are red,

$$f_4/n = 3840/10000 = 0.3840$$

$\Rightarrow$ Draws 100,000 and 38500 are red,

$$f_5/n = 38500/100000 = 0.385 = P$$

In many cases, it is found that these relative frequencies differ from one another by small amounts, provided n is large. Thus in such cases, there is a tendency in the relative frequencies to accumulate in the neighbourhood of some fixed value. This limiting value of f/n as $n \rightarrow \infty$ is regarded as the probability of E in the experiment.

That is,

$$P(E) = p = \underset{n \to \infty}{\text{limit}} \, f/n$$

where f is the observed frequency and n is the sample size

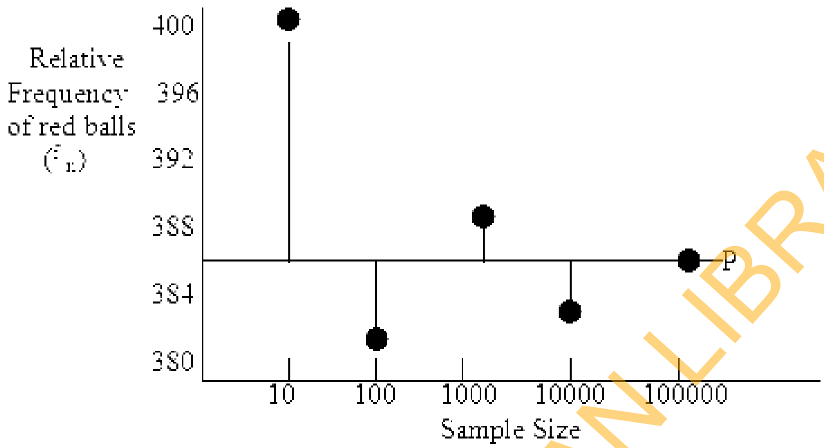


Fig. 1.1: Relative frequency of red balls in Example 1.2

Mathematical (Axiomatic) Definition of Probability

The basis of this approach is embodied in three properties from which the whole system of probability theory is constructed through the use of mathematical logic. The mathematical definition of probability is given as follows:

Let S be the sample space of an experiment. Then the probability is a function P that assigns real numbers to events in such a way that:

(i) $0 \le P(E) \le 1$ for any event E

(ii) $P(E_1 \cup E_2 \cup \cdots) = P(E_1) + P(E_2) + \cdots$ for any collection of mutually exclusive events $\{E_1, E_2 \ldots\}$

(iii) $P(S) = 1$

Counting Techniques – Permutations and Combinations

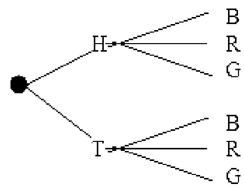
These techniques are helpful in computing the probability of an event when the total number of possible events is large.

Multiplication Principle

If one operation can be performed in n_1 ways and a second operation can be performed in n_2 ways, then there are $n_1 \cdot n_2$ ways in which both operations can be carried out.

Example 1.3

Suppose a coin is tossed once and then a marble is selected at random from a box containing one black, one red, and one green marble. The possible outcomes are HB, HR, HG, TB, TR, and TG. For each of the two possible outcomes of the coin, there are three marbles that may be selected, giving a total of $2 \times 3 = 6$ possible outcomes.



Note that the multiplication principle can be extended to more than two operations. In particular, if the ith of r successive operations can be performed in n_i ways, the total number of ways to carry out all r operations is the product.

$$\prod_{i=1}^{r} n_i = n_1 n_2 \ldots\ldots n_r$$

Permutations: A permutation is an ordered arrangement of objects. The number of permutations of n distinct objects taken r at a time is

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

Proof: In order to fill r positions from n objects, the first position may be filled in n ways using n objects, the second position may be filled in $n-1$ ways, and so on until there are $n-(r-1)$ objects left to fill in the rth position. Thus, the total number of ways of carrying out this operation is given by,

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

The possible permutations of 4 objects taken 4 at a time is = ${}_4P_4$

Ways to fill 1st Position | Ways to fill 2nd Position | Ways to fill 3rd Position | Ways to fill 4th Position | Counting the number of arrangements

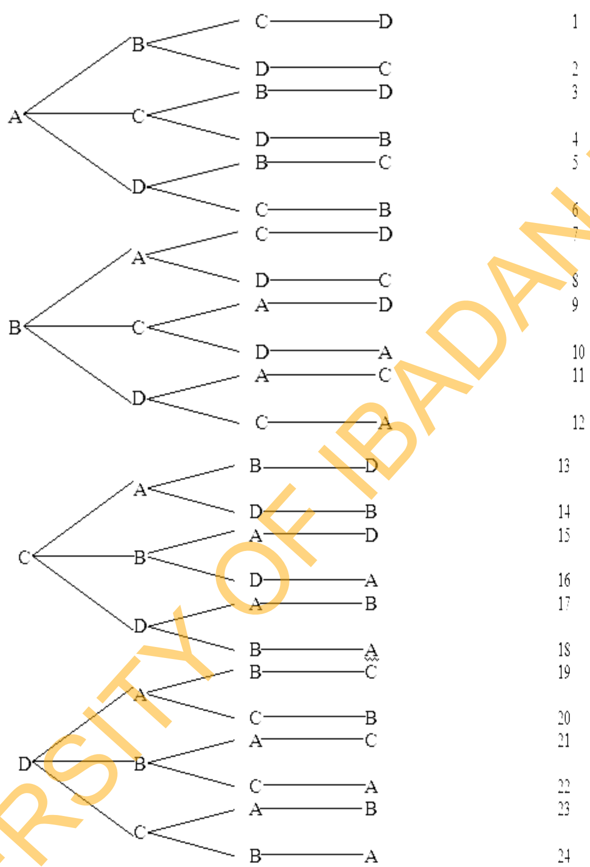


Fig. 1.2: A tree diagram showing the permutations of 4 objects

Suppose we have only two positions available on the shelf. In how many ways can we fill these two positions using four objects?

First determine the number of possible permutations of 4 objects taken two at a time; let the 4 objects be A, B, C, D.

There are four objects with which to fill the 1st position. Once that has been filled, we have 3 objects only, and using the tree diagram (fig. 1.3), we have,

Ways to fill the 1st Position	Ways to fill the 2nd Positions	Counting the number of permutations

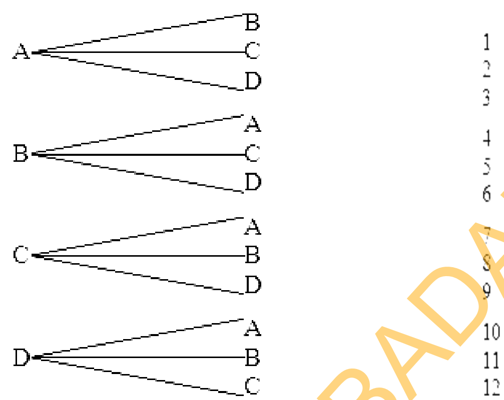


Fig. 1.3: A tree diagram of permutations of 4 objects taken 2 at a time

From figure 1.3, there are 4 x 3 = 12 possible permutations of 4 taken 2 at a time.

Designate the number of distinct objects by n from which the ordered arrangement is to be derived and by r the number of objects in the arrangement.

The number of possible such ordered arrangements is referred to as the number of permutations of n items taken r at a time, and is written as ${}_nP_r$.

In general,

$$n\Pr = n(n-1)(n-2)\cdots(n-r+1)$$

${}_nP_r$ can be evaluated by means of a fraction involving factorials as follows:

$$n\,\mathrm{Pr} = \frac{n!}{(n-r)!}$$

Example 1.4
In a County health department, there are five adjacent offices to be occupied by five nurses, A, B, C, D and E. In how many ways can the five nurses be assigned to the offices?

$${}_5P_5 = \frac{5!}{(5-5)!} = 5 \text{ x } 4 \text{ x } 3 \text{ x } 2 \text{ x } 1 = 120.$$

Suppose that there were six nurses to whom adjacent offices are to be assigned, out of only four offices available.

We need to determine the number of permutations of six items taken four at a time which is

$$6P_4 = \frac{6!}{(6-4)!} = \frac{6 \text{ x } 5 \text{ x } 4 \text{ x } 3 \text{ x } 2}{2!} = 360$$

Permutation: Indistinguishable objects; objects that are not all different.

$$n! = \left(nP_{n_1}, n_2, \ldots\ldots, n_k\right) n_1!n_2!\ldots\ldots n_k!$$

Combinations: A combination is an arrangement of objects without regard to order. The number of combinations of n distinct objects chosen r at a time is

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

Proof: ${}_nP_r$ may be interpreted as the number of ways of choosing r objects from n objects and then permuting the r objects r! ways.

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

Permutations of 4 objects taken two at a time consist of

AB	AC	AD	BC	BD	CD
BA	CA	DA	CB	DB	DC

whereas, there are only six combinations i.e. there are 2 permutations of each combination.

In certain cases we may not want to make distinction between arrangements AB and BA for example, we may want to consider them as the same subset, we say that order does not count, and refer to the arrangements as combinations.

In general, we have $r!$ permutations for each combination of n objects taken r at a time.

$$nP_r = r!\binom{n}{r}$$

$$\binom{n}{r} = \frac{nP_r}{r!}$$

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

$$\binom{4}{2} = \frac{4!}{2!\ 2!}$$

$$= \frac{4 \text{ x } 3 \text{ x } 2!}{2!\ 2!} = 6$$

Example: 1.5
Suppose a group therapy leader in a mental health clinic has 10 patients from which to form a group of six. How many combinations of patients are possible?

$$\binom{10}{6} = \frac{10!}{6!\ 4!} = 210$$

Example 1.6
(Combination): A steering committee of 7 is to be chosen at random from a club with 40 members. How many committees can be formed?

$$40C_7 = \frac{40!}{33!\ 7!} = 18{,}643{,}560$$

Properties of Probability

(1) If A is an event and A′ is its complement, then

$$P(A) = 1 - P(A')$$

Proof: $S = A \cup A'$ since $A \cap A' = \phi$, A and A' are mutually exclusive

So,

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

(2) For any event A, $P(A) \le 1$.

Proof: $P(A) = 1 - P(A')$, we know that $P(A') \ge 0$

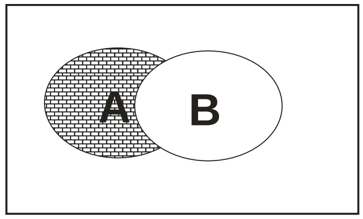
$$\therefore\ P(A) \le 1.$$

(3) For any two events A and B

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

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

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



i.e. $(A \cup B) \cap (B' \cup B)$

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

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

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

Then it follows that $(A \cap B')$ and B are mutually exclusive

Since $(A \cap B') \cap B = (A \cap B) \cap (B' \cap B) = \phi$

Since, $A \cap B = \phi$

$\therefore B' \cap B = \phi$

Then, $P(A \cup B) = P(A \cap B') + P(B)$

Similarly, $A \cap B$ and $A \cap B'$ are mutually exclusive,

So, $P(A) = P(A \cap B) + P(A \cap B')$

$\therefore \quad P(A \cap B') = P(A) - P(A \cap B)$

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

For any three events, A, B and C

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

Conditional Probability

In this section, we introduce one of the most important concepts in all of probability theory—that of conditional probability. Its importance is twofold. In the first place, we are often interested in calculating probabilities when some partial information concerning the result of the experiment is available, or in recalculating them in light of additional information. In such situations, the desired probabilities are conditional ones. Second, as a kind of a bonus, it often turns out that the easiest way to compute the probability of an event is to first "condition" it on the occurrence or nonoccurrence of a secondary event.

In such a case, we use "Conditional Probability of A given B" and write as $P(A/B)$.

Definition: The Conditional Probability of an event A, given the event B denoted as $P(A/B)$, is defined by

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

If $P(B) > 0$

This definition can be understood in a special case in which all outcomes of a random experiment are equally likely. If there are n total outcomes,

$$P(A) = (number\ of\ outcomes\ in\ A)/n$$

Also,

$$P(A \cap B) = (number\ of\ outcomes\ in\ A \cap B)/n$$

Consequently,

$$\frac{P(A \cap B)}{P(A)} = \frac{number\ of\ outcomes\ in\ A \cap B}{number\ of\ outcomes\ in\ A}$$

Example 1.7
A common test for AIDS is called the ELISA test. Among 1,000,000 people who are given ELISA test, results similar to those given in the table below can be expected.

	B_1 **Carry AIDS Virus**	B_2 Do Not Carry AIDS Virus	
A_1:Test positive	4,885	73,630	**78,515**
A_2: Test Negative	115	921,370	**921,485**
	5,000	**995,000**	**1,000,000**

If one of these 1,000,000 people is selected randomly, find the following probabilities:

(i) $P(B_1)$

(ii) $P(A_1)$

(iii) $P(A_1|B_2)$

(iv) $P(B_1|A_1)$

Solution

(i) $P(B_1) = \dfrac{5,000}{1,000,000} = 0.005$

(ii) $P(A_1) = \dfrac{78,515}{1,000,000} = 0.07852$

(iii) $P(A_1|B_2) = \dfrac{P(A_1 \cap B_2)}{P(B_2)} = \dfrac{73,630}{995,000} = 0.074$

(iv) $P(B_1|A_1) = \dfrac{P(A_1 \cap B_1)}{P(A_1)} = \dfrac{4,885}{78,515} = 0.0622$

Example 1.8
A day's production of 850 manufactured parts contains 50 parts that do not meet customer requirements. Two parts are selected randomly without replacement from the batch. What is the probability that the second part is defective given that the first part is defective?

Solution
Let A denote the event that the first part selected is defective, and let B denote the event that the second part selected is defective. The probability needed can be expressed as $P(B/A)$. If the first part is defective, prior to selecting the second part, the batch contains 849 parts, of which 49 are defective, therefore

$$P(B/A) = \frac{49}{849}$$

Example 1.9
Continuing the previous example, if three parts are selected at random, what is the probability that the first two are defective and the third is not defective?

Solution
This event can be described in shorthand notation as simply $P(ddn)$.

$$P(ddn) = \frac{50}{850} \cdot \frac{49}{849} \cdot \frac{800}{848} = 0.0032$$

Note that the third term is obtained as follows: After the first two parts are selected, there are 848 remaining. Of the remaining parts, 800 are not defective. In this example, it is easy to obtain the solution with a conditional probability for each selection.

Independent Events
The calculation of certain probabilities is greatly facilitated by the knowledge of any relationships, or lack thereof, between the events under consideration. In this section we want to examine the latter case, that is, the case in which the occurrence of one event has no influence on the probability of the other's occurrence. Such events are said to be independent of each other, and we want to see how this is reflected in the probabilities.

In some situations, knowledge that an event A which has occurred will not affect the probability that an event B will occur is $P(B/A) = P(B)$

That is $P(A \cap B) = P(A)P(B/A) = P(A)P(B)$

Definition: (Independence of Two Events). Two events A and B are called independent events if

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

Otherwise, A and B are called dependent events.

Dependent events occur in connection with repeated sampling without replacement from a finite collection.

Example 1.10

Suppose a day's production of 850 manufactured parts contains 50 parts that do not meet customer requirements. Suppose two parts are selected from the batch, but the first part is replaced before the second part is selected. What is the probability that the second part is defective (denoted as *B*) given that the first part is defective (denoted as *A*)?

Solution

The probability needed can be expressed as $P(B/A)$.

Because the first part is replaced prior to selecting the second part, the batch still contains 850 parts, of which 50 are defective. Therefore, the probability of *B* does not depend on whether or not the first part was defective. That is,

$$P(B/A) = \frac{50}{850}$$

Also, the probability that both parts are defective is

$$P(A \cap B) = P(B|A)P(A) = \left(\frac{50}{850}\right)\cdot\left(\frac{50}{850}\right) = 0.0035$$

Example 1.11
Let A *and* B be independent events, with $P(A) = 1/4$ and $P(B) = 2/3$. Compute:

(i) $P(A \cap B)$
(ii) $P(A \cap B')$
(iii) $P(A' \cap B')$
(iv) $P[(A \cup B)']$
(v) $P(A' \cap B)$

Solution

(i) $P(A \cap B) = P(A)\cdot P(B) = \frac{1}{4}\cdot\frac{2}{3} = \frac{1}{6}$

(ii) $P(A \cap B') = P(A)\cdot P(B') = \frac{1}{4}\cdot\frac{1}{3} = \frac{1}{12}$; since $P(B') = 1 - P(B)$

(iii) $P(A' \cap B') = P(A')\cdot P(B') = \frac{3}{4}\cdot\frac{1}{3} = \frac{1}{4}$; since $P(A') = 1 - P(A)$

(iv) $P[(A \cup B)'] = P(A') + P(B') - P(A \cap B)' = \frac{3}{4} + \frac{1}{3} - \frac{5}{6} = \frac{1}{4}$

(v) $P(A' \cap B) = P(A')\cdot P(B) = \frac{3}{4}\cdot\frac{2}{3} = \frac{1}{2}$

Summary
This chapter discussed definitions of terms, basic concepts of set theory and some probabilistic concepts. Recall that it discussed the following:

(1) Random experiment, set, sample space, events etc.
(2) Different approaches to the definition of probability.
(3) Counting techniques – Permutations and Combinations.
(4) Conditional and independent events.

Post-Test

(1) A box contains 4 white balls, two red, and two green balls. Use the classical definition of probability to find the probability of
 (i) drawing a red ball on one draw from the box
 (ii) drawing a black ball on one draw from the box

(2) If two fair dice are rolled once, what is the probability that the probability that the total number of spots shown is
 (i) Equal to 5?
 (ii) Divisible by 3?

(3) Twenty balls numbered from 1 to 20 are mixed in an urn and two balls are drawn successively and without replacement. If x_1 and x_2 are the numbers written on the first and second ball drawn, respectively, what is the probability that:
 (i) $x_1 + x_2 = 8$?
 (ii) $x_1 + x_2 \leq 5$?

(4) If $P(A/B) > P(A)$, then show that $P(B/A) > P(B)$.

(5) Show that:
 (i) $P(A'/B) = 1 - P(A/B)$
 (ii) $P(A \cup B/C) = P(A/C) + P(B/C) - P(A \cap B/C)$

(6) Define the following terms with relevant examples:
 (i) An experiment
 (ii) A sample space
 (iii) An event

2
Conditional Probability and Independence

Introduction

The concepts of conditional probability and independent events have been introduced in the last chapter. In this chapter, we shall attempt to discuss these concepts in detail.

Objectives

After a careful study of this chapter you should be able to do the following:

1. Interpret and calculate conditional probabilities of events;
2. Determine the independence of events and use independence to calculate probabilities;
3. Use Multiplication rule to express total probabilities;
4. Use Bayes' theorem to calculate conditional probabilities;
5. Use a tree diagram to organize and compute probabilities,

Pre-Test

(1) Distinguish between independent and dependent events.

(2) Define the term "conditional probability.

(3) Establish a relationship between conditional probability and independent events.

(4) State the condition for which "the conditional probability of an event A given event B" can exist.

Content

Conditional Probability

Conditional probability provides us with a way to reason about the outcome of an experiment, based on partial information. Here are some examples of situations we have in mind:

(a) In an experiment involving two successive rolls of a die, you are told that the sum of the two rolls is 9. How likely is it that the first roll was a 6?

(b) In a word guessing game, the first letter of the word is a "t". What is the likelihood that the second letter is an "h"?

(c) How likely is it that a person has a disease given that a medical test was negative?

(d) A spot shows up on a radar screen. How likely is it that it corresponds to an aircraft?

In more precise terms, given an experiment, a corresponding sample space and a probability law, suppose we know that the outcome is within some given event *B*. We wish to quantify the likelihood that the outcome also belongs to some other given event *A*. We thus seek to construct a new probability law, which takes into account this knowledge and which, for any event *A*, gives us the conditional probability of *A* given *B*, denoted by $P(A/B)$.

We would like the conditional probabilities $P(A/B)$ of different events *A* to constitute a legitimate probability law that satisfies the probability axioms. The laws should also be consistent with our intuition in important special cases, e.g., when all possible outcomes of the experiment are equally likely. For example, suppose all six possible outcomes of a fair die roll are equally likely. If we are told that the outcome is even, we are left with only three possible outcomes, namely, 2, 4, and 6. These three outcomes were *equally likely* to start with, and so they should remain equally likely given the additional knowledge that the outcome was even.

A major objective of probability modelling is to determine how likely it is that an event *A* will occur when a certain experiment is performed. However, there are numerous cases in which the probability assigned to *A* will be affected by the knowledge of the occurrence or non-occurrence of another event B. In such a case, we use "Conditional Probability of *A* given B" and write as $P(A/B)$.

Definition 2.1: The Conditional Probability of an event A, given the event B, is defined by

$$P(A/B) = \frac{P(A \cap B)}{P(B)}, \textbf{\textit{ if }} P(B) \neq 0$$

Example 2.1

We toss a fair coin three successive times. We wish to find the conditional probability $P(A/B)$ when A and B are the events

A= {more heads than tails come up}, B= {1st toss is a head}

The sample space consists of eight sequences,

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

which we assume to be equally likely. The event *B* consists of the four elements *HHH*, *HHT*, *HTH*, *HTT*, so its probability is

$$P(B) = \frac{4}{8}$$

The event $A \cap B$ consists of the three elements outcomes *HHH*, *HHT*, *HTH*, so its probability is

$$P(A \cap B) = \frac{3}{8}$$

Thus, the conditional probability $P(A/B)$ is

$$P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{3/8}{4/8} = \frac{3}{4}$$

Because all possible outcomes are *equally likely* here, we can also compute $P(A/B)$ using a shortcut. We can bypass the calculation of $P(B)$ and $P(A \cap B)$, and simply divide the number of elements shared by *A* and *B* (which is 3) with the number of elements of *B* (which is 4), to obtain the same result 3/4.

Example 2.2
Two cards are drawn without replacement from a deck of cards.

Let A_1 denote the event of getting "an ace on the first draw"

and

A_2 denote the event of getting "an ace on the second draw"

The number of ways in which different outcomes can occur can be enumerated as follows. (Multiplicative principle is used).

	A_1	A_1'	
A_2	4 x 3	48 x 4	4 x 51
A_2'	4 x 48	48 x 47	48 x 51
	4 x 51	48 x 51	52 x 51

(i) The probability of getting "an ace on the first draw and an ace on the second draw" is given by

$$P(A_1 \cap A_2) = \frac{4 \text{ x } 3}{52 \text{ x } 51}$$

(ii) What is the probability that an ace is drawn on the second draw given that an ace was obtained on the first draw?

$$P(A_2/A_1) = \frac{P(A_1 \cap A_2)}{P(A_1)} = \frac{(4 \text{ x } 3)/(52 \text{ x } 51)}{(4 \text{ x } 51)/(52 \text{ x } 51)} = \frac{3}{51}$$

Conditional Probabilities Specify a Probability Law
For a fixed event *B*, it can be verified that the conditional probabilities $P(A/B)$ form a legitimate probability law that satisfies the three axioms. Indeed, nonnegativity is clear. Furthermore,

$$P(S/B) = \frac{P(S \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$$

and the normalization axiom is also satisfied. In fact, since we have $P(B/B) = P(B)/P(B) = 1$, all of the conditional probability is concentrated on *B*. Thus, we might as well discard all possible outcomes outside *B* and treat the conditional probabilities as a probability law defined on the new sample space (universe) *B*.

To verify the additivity axiom, we write for any two events A_1 and A_2,

$$
\begin{aligned}
P(A_1 \cup A_2/B) &= \frac{P((A_1 \cup A_2) \cap B)}{P(B)} \\
&= \frac{P((A_1 \cap B) \cup (A_2 \cap B))}{P(B)} \\
&= \frac{P((A_1 \cap B) + (A_2 \cap B))}{P(B)} \\
&= \frac{P(A_1 \cap B)}{P(B)} + \frac{(A_2 \cap B)}{P(B)} \\
&= P(A_1/B) + P(A_2/B)
\end{aligned}
$$

where for the second equality, we used the fact that $A_1 \cap B$ and $A_2 \cap B$ are disjoint sets, and for the third equality we used the additivity axiom for the (unconditional) probability law. The argument for a countable collection of disjoint sets is similar. Since conditional probabilities constitute a legitimate probability law, all general properties of probability laws remain valid. For example, a fact such as

$P(A \cup C) \leq P(A) + P(C)$ translates to the new fact

$P(A \cup C|B) \leq P(A|B) + P(C|B)$.

Let us summarize the conclusions reached so far.

Properties of Conditional Probability

- The conditional probability of an event *A*, given an event *B* with $P(B) > 0$, is defined by

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

  and specifies a new (conditional) probability law on the same sample space *S*. In particular, all known properties of probability laws remain valid for conditional probability laws.

- Conditional probabilities can also be viewed as a probability law on a new universe *B*, because all of the conditional probability is concentrated on *B*.

- In the case where the possible outcomes are finitely many and equally likely, we have

$$P(A/B) = \frac{\text{number of elements of } A \cap B}{\text{number of elements of } B}$$

Independence

We have introduced the conditional probability $P(A/B)$ to capture the partial information that event *B* provides about event *A*. An interesting and important special case arises when the occurrence of *B* provides no information and does not alter the probability that *A* has occurred, i.e.,

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

Hence, knowledge that an event A has occurred will not affect the probability that an event B will occur is $P(B/A) = P(B)$

That is, $\quad P(A \cap B) = P(A)P(B/A) = P(A)P(B)$

In general, when this happens the two events are said to be independent or stochastically independent.

Definition 2.2: Two events A and B are called independent events if

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

Otherwise, A and B are called dependent events

We adopt this latter relation as the definition of independence because it can be used even if $P(B) = 0$, in which case $P(A/B)$ is undefined. The symmetry of this relation also implies that independence is a symmetric property; that is, if *A* is independent of *B*, then *B* is independent of *A*, and we can unambiguously say that *A* and *B* are **independent events**.

Independence is often easy to grasp intuitively. For example, if the occurrence of two events is governed by distinct and noninteracting physical processes, such events will turn out to be independent. On the other hand, independence is not easily visualized in terms of the sample space. A common first thought is that two events are independent if they are disjoint, but in fact the opposite is true: two disjoint events *A* and *B* with $P(A) > 0$ and $P(B) > 0$ are never independent, since their intersection $A \cap B$ is empty and has probability 0.

Example 2.3

Consider an experiment involving two successive rolls of a 4-sided die in which all 16 possible outcomes are equally likely and have probability 1/16.

(a) Are the events

A_i = {1st roll results in i}, B_j = {2nd roll results in j},

independent? We have,

$$P(A \cap B) = P(\text{the results of the two rolls is } (i, j)) = \frac{1}{16},$$

$$P(A_i) = \frac{\text{number of elements of } A_i}{\text{total number of possible outcomes}} = \frac{4}{16},$$

$$P(B_i) = \frac{\text{number of elements of } B_j}{\text{total number of possible outcomes}} = \frac{4}{16}.$$

We observe that $P(A_i \cap B_j) = P(A_i)P(B_j)$, and the independence of A_i and B_j is verified.

(b) Are the events

A = {1st roll is a 1}, B= {sum of the two rolls is a 5},

independent? The answer here is not quite obvious. We have,

$$P(A \cap B) = P(\textit{the results of the two rolls is } (1,4)) = \frac{1}{16}$$

and also,

$$P(A) = \frac{\textit{number of elements of } A}{\textit{total number of possible outcomes}} = \frac{4}{16}$$

The event B consists of the outcomes (1,4), (2,3), (3,2), and (4,1), and

$$P(B) = \frac{\textit{number of elements of } B}{\textit{total number of possible outcomes}} = \frac{4}{16}.$$

Thus, we see that $P(A \cap B) = P(A)P(B)$, and the events A and B are independent

(c) Are the events

A = {maximum of the two rolls is 2}, B= {minimum of the two rolls is 2},

independent? Intuitively, the answer is "no" because the minimum of the two rolls tells us something about the maximum. For example, if the minimum is 2, the maximum cannot be 1.

More precisely, to verify that A and B are not independent, we calculate

$$P(A \cap B) = P(\textit{the results of the two rolls is } (2,2)) = \frac{1}{16}$$

and also

$$P(A)=\frac{\text{number of elements of } A}{\text{total number of possible outcomes}}=\frac{3}{16}$$

$$P(B)=\frac{\text{number of elements of } B}{\text{total number of possible outcomes}}=\frac{5}{16}.$$

We have $P(A)P(B)=15/(16)^2$, so that $P(A\cap B)\neq P(A)P(B)$, and A and B are not independent.

Independence of Several Events

The discussion that has just been given for two events can be extended to any number of events. If k events $A_1, A_2, \ldots, A_k$ are independent in the sense that they are physically unrelated to each other, then it is natural to assume that the probability $P(A_1\cap A_2 \cap\ldots\cap A_k)$ that all k events will occur is the product $P(A_1)\cdot P(A_2)\cdots P(A_k)$. Furthemore, since the events $A_1, A_2, \ldots, A_k$ are unrelated, this product rule should hold not only for the intersection of all k events, but also for the intersection of any two of them, any three of them or any other number of them. These consideration lead to the following definition:

Definition 2.3: the k events $A_1, A_2, \ldots, A_k$ are *independent* if, for every subset $A_{i_1}, A_{i_2}, \ldots, A_{i_j}$ of j of these events $(j=2,3,\ldots,k)$.

$$P\left(A_{i_1}\cap A_{i_2} \cap\ldots\cap A_{i_j}\right)=P\left(A_{i_1}\right)P\left(A_{i_2}\right)\ldots P\left(A_{i_j}\right)$$

In particular, in order for three events A, B, and C to be independent, the following four relations must be satisfied:

$$\begin{aligned} P(AB)&=P(A)P(B),\\ P(AC)&=P(A)P(C),\\ P(BC)&=P(B)P(C), \end{aligned} \qquad (2.1)$$

and

$$P(ABC)=P(A)P(B)P(C) \qquad (2.2)$$

Note that it is possible that Eq. (2.2) will be satisfied but one or more of the three relations in Eq. (2.1) will not be satisfied. On the other hand, it is also possible that each of the three relations In (2.1) will be satisfied but Eq. (2.2) will not be satisfied.

Independence

- Two events A and B are said to independent if
  $P(A \cap B) = P(A)P(B)$.
- If in addition, $P(B) > 0$, independence is equivalent to the condition
  $P(A/B) = P(A)$.
- If A and B are independent, so are A and B^c.

Example 2.4
If two cards are drawn in succession from a deck, what is the probability of ace on the 1st draw and ace on the 2nd draw?

$P(A_2) = 4/52 \qquad P(A_2/A_1) = 3/51$

the events are dependent.

But if the sampling is with replacement; then the draws are independent trials.

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

i.e. $P(A_2) = 4/52 \quad P(A_2/A_1) = \dfrac{P(A_2 \cap A_1)}{P(A_1)} = 4/52$

Example 2.5
In a certain high school class, consisting of 60 girls and 40 boys, it is observed that 24 girls and 16 boys wear eye glasses. If a student is picked at random from the class, what is the probability that the student wears eye glasses?

$P(E) = 40/100 = 0.4$

What is the probability that a student picked at random wears eye glasses, given that the student is a boy?

$$P(E/B) = \frac{P(E \cap B)}{P(B)} = \frac{16/100}{40/100}$$

i.e $$\frac{16/40 \text{ x } 40/100}{40/100} = 0.4$$

Thus, the additional information that the student is a boy does not alter the probability that the student wears eyeglasses.

$$\therefore \; P(E) = P(E|B)$$

The events being a boy and wearing glasses for this example are independent. Show that the events of wearing eyeglasses, E_1 and not being a boy $\bar{B}$ are also independent.

$$P(E/\bar{B}) = \frac{P(E \cap \bar{B})}{P(B)} = \frac{24/100}{60/100} = \frac{24}{60} = 0.4$$

For mutually exclusive events,

$$P(E_1 \text{ and } E_2) = P(E_1 \cap E_2) = \phi$$

and $$P(E_1 or E_2) = P(E_1) + P(E_2)$$

Multiplication Rule

Assuming that all of the conditioning events have positive probability, we have

$$P\left(\bigcap_{i=1}^{n} A_i\right) = P(A_1)P(A_2 / A_1)P(A_3 / A_1 \cap A_2)\cdots P\left(A_n \middle| \bigcap_{i=1}^{n} A_i\right)$$

The multiplication rule can be verified by writing

$$P\left(\bigcap_{i=1}^{n} A_i\right) = P(A_1)\frac{P(A_2 \cap A_1)}{P(A_1)}\frac{P(A_1 \cap A_2 \cap A_3)}{P(A_1 \cap A_2)}\cdots\frac{P\left(\bigcap_{i=1}^{n} A_i\right)}{P\left(\bigcap_{i=1}^{n-1} A_i\right)}$$

and using the definition of conditional probability to rewrite the right-hand side of the equation above as

$$P(A_1)P(A_2/A_1)P(A_3/A_1 \cap A_2)\cdots P\left(A_n \middle| \bigcap_{i=1}^{n-1} A_i\right)$$

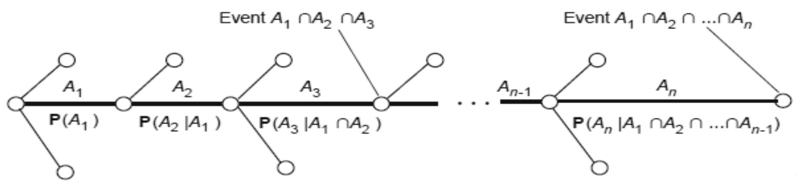


Fig. 2.1: Visualization of the total probability theorem.

The intersection event $A = A_1 \cap A_2 \cap \cdots \cap A_n$ is associated with a path on the tree of a sequential description of the experiment. We associate the branches of this path with the events $A_1, A_2 \cdots A_n$, and we record next to the branches the corresponding conditional probabilities.

The final node of the path corresponds to the intersection event *A*, and its probability is obtained by multiplying the conditional probabilities recorded along the branches of the path

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

Note that any intermediate node along the path also corresponds to some intersection event and its probability is obtained by multiplying the corresponding conditional probabilities up to that node. For example, the event $A_1 \cap A_2 \cap A_3$ corresponds to the node shown in the figure, and its probability is

$$P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2)$$

For the case of just two events, A_1 and A_2, the multiplication rule is simply the definition of conditional probability.

Example 2.6

If an aircraft is present in a certain area, radar correctly registers its presence with probability 0.99. If it is not present, the radar falsely registers an aircraft presence with probability 0.10. We assume that an aircraft is present with probability 0.05. What is the probability of false alarm (a false indication of aircraft presence), and the probability of missed detection (nothing registers, even though an aircraft is present)?

A sequential representation of the sample space is appropriate here, as shown in figure 2.2. Let A and B be the events

A = {an aircraft is present},
B = {the radar registers an aircraft's presence},

and consider also their complements

A^c = {an aircraft is not present},
B^c = {the radar does not register an aircraft's presence}.

The given probabilities are recorded along the corresponding branches of the tree describing the sample space, as shown in figure 2.2. Each event of interest corresponds to a leaf of the tree and its probability is equal to the product of the probabilities associated with the branches in a path from the root to the corresponding leaf. The desired probabilities of false alarm and missed detection are

$$P(\textit{false alarm}) = P(A^c \cap B) = P(A^c)P(B|A^c) = 0.95 \cdot 0.10 = 0.095$$

$$P(\textit{missed detection}) = P(A \cap B^c) = P(A)P(B^c|A) = 0.05 \cdot 0.01 = 0.0005$$

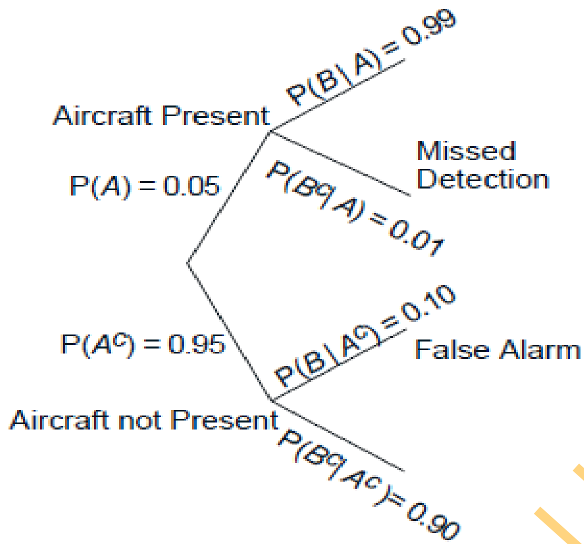


Fig. 2.2: Sequential description of the sample space for the radar detection problem in Example 2.6

Extending the preceding example, we have a general rule for calculating various probabilities in conjunction with a tree-based sequential description of an experiment. In particular:

(a) We set up the tree so that an event of interest is associated with a leaf. We view the occurrence of the event as a sequence of steps, namely, the traversals of the branches along the path from the root to the leaf.

(b) We record the conditional probabilities associated with the branches of the tree.

(c) We obtain the probability of a leaf by multiplying the probabilities recorded along the corresponding path of the tree.

Example 2.7
Three cards are drawn from an ordinary 52-card deck without replacement (drawn cards are not placed back in the deck). We wish to find the probability that none of the three cards is a heart. We assume that at each step, each one of the remaining cards is equally likely to be picked. By symmetry, this implies that every

triplet of cards is equally likely to be drawn. A cumbersome approach, that we will not use, is to count the number of all card triplets that do not include a heart, and divide it with the number of all possible card triplets.

Instead, we will use a sequential description of the sample space in conjunction with the multiplication rule (fig. 2.3).

Define the events

A_i = {the i^{th} card is not a heart}, i= 1, 2, 3.

We will calculate $P(A_1 \cap A_2 \cap A_3)$, the probability that none of the three cards is a heart, using the multiplication rule,

$$P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2 | A_1)P(A_3 | A_1 \cap A_2)$$

We have,

$$P(A_1) = \frac{39}{52}$$

since there are 39 cards that are not hearts in the 52-card deck. Given that the first card is not a heart, we are left with 51 cards, 38 of which are not hearts, and

$$P(A_2 | A_1) = \frac{38}{51}$$

Finally, given that the first two cards drawn are not hearts, there are 37 cards which are not hearts in the remaining 50-card deck, and

$$P(A_3 | A_1 \cap A_1) = \frac{37}{50}$$

These probabilities are recorded along the corresponding branches of the tree describing the sample space, as shown in figure 2.3. The desired probability is now obtained by multiplying the probabilities recorded along the corresponding path of the tree:

$$P(A_1 \cap A_1 \cap A_3) = \frac{39}{52} \cdot \frac{38}{51} \cdot \frac{37}{50}$$

Note that once the probabilities are recorded along the tree, the probability of several other events can be similarly calculated. For example,

$$P(1st\ is\ not\ a\ heart\ and\ 2nd\ is\ a\ heart) = \frac{39}{52} \cdot \frac{13}{51}$$

$$P(1st\ two\ are\ not\ hearts\ and\ 3rd\ is\ a\ heart) = \frac{39}{52} \cdot \frac{38}{51} \cdot \frac{13}{50}$$

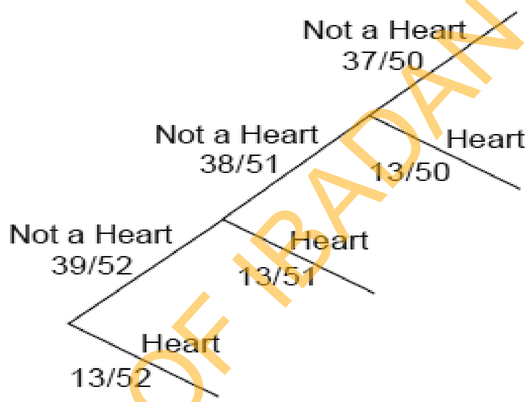


Fig. 2.3: Sequential description of the sample space of the 3-card selection problem in Example 2.7.

Example 2.8
A class consisting of 4 graduate and 12 undergraduate students is randomly divided into 4 groups of 4. What is the probability that each group includes a graduate student? We interpret randomly to mean that given the assignment of some students to certain slots, any of the remaining students is equally likely to be assigned to any of the remaining slots. We then calculate the desired

probability using the multiplication rule, based on the sequential description shown in figure 2.4. Let us denote the four graduate students by 1, 2, 3, 4, and consider the events

A_1 = {students 1 and 2 are in different groups},
A_2 = {students 1, 2, and 3 are in different groups},
A_3 = {students 1, 2, 3, and 4 are in different groups}.

We will calculate $P(A_3)$ using the multiplication rule:

$$P(A_3) = P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2)$$

we have,

$$P(A_1) = \frac{12}{15}$$

since there are 12 student slots in groups other than the one of student 1, and there are 15 student slots overall, excluding student 1. Similarly,

$$P(A_2|A_1) = \frac{8}{14}$$

since there are 8 student slots in groups other than the one of students 1 and 2, and there are 14 student slots, excluding students 1 and 2. Also,

$$P(A_3|A_1 \cap A_2) = \frac{4}{13},$$

since there are 4 student slots in groups other than the one of students 1, 2, and 3, and there are 13 student slots, excluding students 1, 2, and 3. Thus, the desired probability is

$$\frac{12}{15} \cdot \frac{8}{14} \cdot \frac{4}{13},$$

and is obtained by multiplying the conditional probabilities along the corresponding path of the tree of figure 2.4.

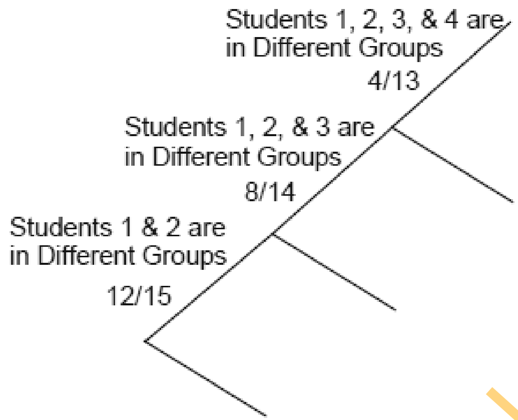


Fig. 2.4: Sequential description of the sample space of the student problem in Example 2.8.

Total Probability Theorem and Bayes' Rule

In this section, we explore some applications of conditional probability. We start with the following theorem, which is often useful for computing the probabilities of various events, using a "divide-and-conquer" approach.

Total Probability Theorem

Let $A_1, \ldots, A_n$ be disjoint events that form a partition of the sample space (each possible outcome is included in one and only one of the events $A_1, \ldots, A_n$) and assume that $P(A_i) > 0$, for all $i = 1, \ldots, n$. Then, for any event *B*, we have

$$\begin{aligned} P(B) &= P(A_1 \cap B) + \cdots + P(A_n \cap B) \\ &= P(A_1)P(B|A_1) + \cdots + P(A_n)P(B|A_n) \end{aligned}$$

The theorem is visualized and proved in figure 2.5. Intuitively, we are partitioning the sample space into a number of scenarios (events) A_i. Then, the probability that *B* occurs is a weighted average of its conditional probability under each scenario, where each scenario is weighted according to its (unconditional) probability. One of the uses of the theorem is to compute the probability of various events B for which the conditional probabilities $P(B|A_i)$ are known or easy to derive. The key is to choose appropriately the partition $A_1, \ldots, A_n$, and this choice is often suggested by the problem structure.

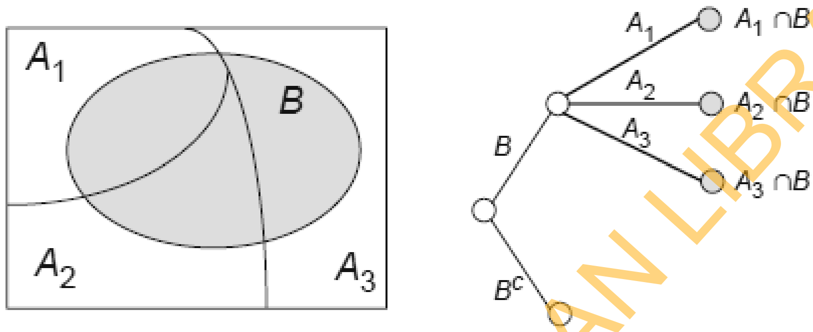


Fig. 2.5: Visualization and verification of the total probability theorem.

Example 2.9
We roll a fair four-sided die. If the result is 1 or 2, we roll once more but otherwise, we stop. What is the probability that the sum total of our rolls is at least 4?

Let A_i be the event that the result of first roll is i, and note that $P(A_i) = 1/4$ for each i. Let B be the event that the sum total is at least 4. Given the event A_1, the sum total will be at least 4 if the second roll results in 3 or 4, which happens with probability 1/2. Similarly, given the event A_2, the sum total will be at least 4 if the second roll results in 2, 3, or 4, which happens with probability 3/4. Also, given the event A_3, we stop and the sum total remains below 4. Therefore,

$$P(B|A_1) = \frac{1}{2}, \quad P(B|A_2) = \frac{3}{4}, \quad P(B|A_3) = 0, \; P(B|A_4) = 1$$

Using the total probability theorem,

$$P(B) = \frac{1}{4}\cdot\frac{1}{2} + \frac{1}{4}\cdot\frac{3}{4} + \frac{1}{4}\cdot 0 + \frac{1}{4}\cdot 1 = \frac{9}{16}$$

The total probability theorem can be applied repeatedly to calculate probabilities in experiments that have a sequential character. The total probability theorem is often used in conjunction with the following celebrated theorem, which relates conditional probabilities of the form $P(A|B)$ with conditional probabilities of the form $P(B|A)$, in which the order of the conditioning is reversed.

Bayes' Rule

Let $A_1, A_2, \ldots, A_n$ be disjoint events that form a partition of the sample space, and assume that $P(A_i) > 0$, for all *i*. Then, for any event *B* such that $P(B) > 0$, we have

$$\begin{aligned} P(A_i|B) &= \frac{P(A_i)P(B|A_i)}{P(B)} \\ &= \frac{P(A_i)P(B|A_i)}{P(A_1)P(B|A_1) + \cdots + P(A_n)P(B|A_n)} \end{aligned}$$

To verify Bayes' rule, note that $P(A_i)P(B|A_i)$ and $P(A_i|B)P(B)$ are equal, because they are both equal to $P(A_i \cap B)$. This yields the first equality. The second equality follows from the first by using the total probability theorem to rewrite $P(B)$.

Bayes' rule is often used for inference. There are a number of "causes" that may result in a certain "effect." We observe the effect, and we wish to infer the cause. The events $A_1, \ldots, A_n$ are associated with the causes and the event *B* represents the effect. The probability $P(B|A_i)$ that the effect will be observed when the cause A_i is present amounts to a probabilistic model of the cause-effect relationship (cf. figure 2.6). Given that the effect B has been observed, we wish to evaluate the (conditional) probability $P(A_i|B)$ that the cause A_i is present.

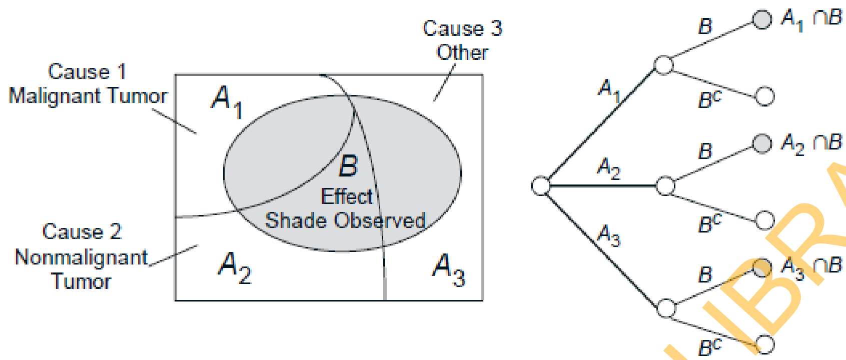


Fig. 2.6: An example of the inference context that is implicit in Bayes' rule.

An example of the inference context that is implicit in Bayes' rule. If we observe a shade in a person's X-ray (this is event B, the "effect") and we want to estimate the likelihood of three mutually exclusive and collectively exhaustive potential causes: cause 1 (event A_1) is that there is a malignant tumor, cause 2 (event A_2) is that there is a non-malignant tumor and cause 3 (event A_3) corresponds to reasons other than a tumor. We assume that we know the probabilities $P(A_i)$ and $P(B|A_i)$, $i=1,2,3$. Given that we see a shade (event B occurs); Bayes' rule gives the conditional probabilities of the various causes as

$$P(A_i|B)=\frac{P(A_i)P(B|A_i)}{P(A_1)P(B|A_1)+P(A_2)P(B|A_2)+P(A_3)P(B|A_3)},\quad i=1,2,3.$$

For an alternative view, consider an equivalent sequential model, as shown on the right (fig. 2.6). The probability $P(A_1|B)$ of a malignant tumor is the probability of the first highlighted leaf, which is $P(A_1\cap B)$, divided by the total probability of the highlighted leaves, which is $P(B)$.

Example 2.10
Let us return to the radar detection problem of Example 2.6 and figure 2.2. Let

A = {an aircraft is present},
B = {the radar registers an aircraft's presence}.

We were given that,

$$P(A) = 0.05\,,\; P(B|A) = 0.99\,,\; P(B|A^c) = 0.1$$

Applying Bayes' rule, with $A_1 = A$ and $A_2 = A^c$, we obtain

$$P(\textit{aircraft present}|\textit{radar registers}) = P(A|B)$$

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

$$= \frac{P(A)P(B|A)}{P(A)P(B|A) + P(A^c)P(B|A^c)}$$

$$= \frac{0.05 \cdot 0.99}{0.05 \cdot 0.99 + 0.95 \cdot 0.1}$$

$$\approx 0.3426$$

Summary

This chapter discussed in detail the concepts of Conditional Probability and Independent events that were introduced in the previous chapter with relevant examples for easy comprehension. Other topics treated were:

(1) Independence of two and several events.
(2) Properties of conditional probability.
(3) Demonstrating the multiplication rule as the rule for definning the conditional probability.
(4) Stating and proving the Total probability theorem.
(5) Using Bayes' theorem to calculate conditional probabilities.

Post-Test

(1) A box contains three cards, one is red on both sides, one card is green on both sides, and one card is red on one side and green on the other. One card is selected from the box at random, and the color on one side is observed. If this is green, what is the probability that the other side of the card is also green?

(2) Two students *A* and *B* are both registered for a certain course. Student *A* attends class 80 percent of the time and student *B* attends class 60 percent of the time, and the absences of the two students are independent. If at least one of the two students is in class on a given day, what is the probability that *A* is in class that day?

(3) If A and B are disjoint events and $P(B)>0$, what is the value of $P(A|B)$?

(4) Two boxes contain long bolts and short bolts. Suppose that one box contains 60 long bolts and 40 short bolts and that the other box contains 10 long bolts and 20 short bolts. Suppose also that one box is selected at random and a bolt is then selected at random from that box. Determine the probability that this bolt is long.

(5) In a game of cards, what is the probability of drawing, without replacement, two aces in succession?

(6) A simple binary communication channel carries messages by using only two signals, 0 and 1. For a given binary channel, 40% of the time a 1 is transmitted; the probability that a transmitted 0 is correctly received is 0.90, and the probability that a transmitted 1 is correctly received is 0.95. Determine

(i) the probability of a 1 being received, and
(ii) given a 1 is received, the probability that 1 was transmitted.

(7) Consider the following table;

Table 2.1: Hypothetical probabilities for getting a flu shot and getting the flu

Get a shot	**Get the flu**		
	Yes	**No**	
Yes	0.25	0.20	0. 45
No	0.28	0.27	0.55
	0.53	0 .47	1.00

(i) What is the probability of developing heart disease given that your cholesterol level is 250?
(ii) What is the probability that someone does not get the flu, given that they get a flu shot?
(iii) What is the probability of getting the flu, given that the person gets a shot?

3
Functions and Random Variables

Introduction

If as we observe a characteristic, we find that it takes on different values in people, places or things, we label the characteristic a VARIABLE e.g. heights of adults, weights of children and ages of patients seen in a dental clinic. Whenever we determine the height, weight or age of an individual, the result is frequently referred to as a value of the respective variable. When the values obtained arise by chance the variable is called a random variable.

Objectives

After careful study of this chapter you should be able to do the following:

(1) Define random variable;
(2) Distinguish between discrete and continuous random variables; and
(3) Discuss the properties of random variables, both discrete and continuous.

Pre-Test

(1) Have you heard of the term 'function' before?
(2) Describe the word "variable".
(3) Explain and give examples of sample space.

Content

Functions and Random Variables

A function is simply a rule by which every member of one set is assigned to or paired with one member of another set.

Let X and Y be sets. If f is a rule that assigns to every element x in the set X a unique element y in the set Y, then f is said to be a function that maps X into Y.

$$X \qquad Y$$

$$x_1 \to y_1$$

$$x_2 \to y_2$$

$$\left.\begin{matrix} x_3 \\ x_4 \\ x_5 \end{matrix}\right\} \to y_3$$

For example: supposing X is the set of 5 students in a seminar and that Y is the set of term paper topics. Let a rule be defined as "choose a topic for a term paper". The rule "***Choose a topic …..***" is called a ***function***. A random variable is therefore a function which assigns numerical values to the different outcomes defined by the sample space.

Definition 3.1: Given a random experiment with a sample space S. A function X that assigns to each element $c \in S$ one and only one real number $X(c) = x$ is called a Random Variable. The space of X is the set of real numbers

$$A = \{x : x = X(c), c \in S\}$$

In a group of 10 people, a, b, c, d… j, a person is selected at random from this group (random selection can have 10 possible results). Suppose that the height of each person has been measured to the nearest inch, and the weight measured to the nearest 10 pounds.

S	X	S	Y
a $\longrightarrow$	64	a $\longrightarrow$	130
b $\longrightarrow$	67	b $\longrightarrow$	150
c $\longrightarrow$	63	c $\longrightarrow$	110
d $\longrightarrow$	68	d $\longrightarrow$	160
e $\longrightarrow$	72	e $\longrightarrow$	170
f $\longrightarrow$	62	f $\longrightarrow$	180
g $\longrightarrow$	68	g $\longrightarrow$	210
h $\longrightarrow$	66	h $\longrightarrow$	160
i $\longrightarrow$	68	i $\longrightarrow$	130
j $\longrightarrow$	65	j $\longrightarrow$	150

The arrow that connects a value in X to every individual in S represents the operation "measure the height of each person to the nearest inch". A rule that assigns a numerical value to every result in the sample space of a random experiment is called a random variable, and the numbers that are assigned by this rule are called the value set of the random variable.

Draw an individual at random; what are the chances that the height of the individual drawn is 68 inches? Ans. $^3/_{10}$

Example 3.1

Suppose a coin is tossed twice so that the sample space is S. Let X represents the number of heads which can come up.

Elements of the sample space S	**No. of Heads (Random Variable)** X	**Probability** $f(x)$
HH	2	¼
HT	1	¼
TH	1	¼
TT	0	¼

The probability distribution (function) can be described as a function which assigns probabilities to these numerical values.

If X stands for the random variable "number of heads" then x is the value that the random variable can assume. The probability that random variable X takes on the value x is written as:

$P(X=x)$ or $f(x)$ is called the probability density function, p.d.f.

$$P(X=0)=f(0)=\tfrac{1}{4}$$
$$P(X=1)=f(1)=\tfrac{1}{2}$$
$$P(X=2)=f(2)=\tfrac{1}{4}$$

Random variables can be classified as discrete or continuous.

Discrete Random Variable

Let X denote a random variable with space $\mathbb{R}$, suppose we can compute $P(X \in A)$; $A \subset \mathbb{R}$. Let X denote a random variable with one-dimensional space $\mathbb{R}$, a subset of the real numbers. Suppose that the space $\mathbb{R}$ contains a countable number of points; that is, $\mathbb{R}$ contains either a *finite number of points* or the points of $\mathbb{R}$ can be put into *a one-to-one correspondence with the positive integers*. Such a set $\mathbb{R}$ is called a set of discrete points or simply a discrete sample space. The random variable X is therefore called a random variable of the discrete type and X is said to have a distribution of the discrete type.

For a random variable X of the discrete type, the probability function $P(X=x)$ is frequently denoted by $f(x)$ and this function $f(x)$ is called the probability density function. Some authors refer to $f(x)$ as the *probability function*, the *frequency function*, or the *probability mass function*.

Definition 3.2: If the set of all possible values of a random variable X, is a countable set, $x_1, x_2, \ldots$ then X is called a discrete random variable. The function that assigns the probability to each possible value x will be called the discrete probability density function (or probability mass function, *p.m.f*).

$$f(x) = P[X = x], \; x = x_1, x_2, \ldots$$

A function $f(x)$ is a discrete *p.d.f* if and only if it satisfies the following properties for at most a countably infinite set of real numbers $x_1, x_2, \ldots$

(i) $f(x) \geq 0 \quad \forall x \quad$ (Non-negative)

(ii) $\sum_{all\ x} f(x) = 1$

(iii) $P(X \in A) = \sum_{x \in A} f(x)$ where $A \subset \mathbb{R}$

Example 3.2

Let random variable X with pdf $\begin{cases} f(x) = x/6, & x = 1, 2, 3 \\ 0, \text{ otherwise} \end{cases}$

Find;

(i) $P[X = 1 \text{ or } 2]$

(ii) $P[X \geq 2]$

Solution: $P[X = 1 \text{ or } 2] = P[X = 1] \text{ or } P[X = 2]$

$$\sum_{x=1}^{2} f(x) = \sum_{1}^{2} x/6$$

$$= \quad 1/6 + 2/6$$

$$= \quad 3/6 \qquad \text{if } x = 1, \; f(1) = 1/6$$

$$= \quad 1/2 \qquad x = 2, \; f(2) = 2/6$$

$$P[X \ge 2] = \sum_{x=2}^{3} f(x) = \sum_{2}^{3} x/6$$
$$= 2/6 + 3/6$$
$$= 5/6$$

or

$$P[X \ge 2] = 1 - P[X < 2]$$
$$= 1 - f(1)$$
$$= 1 - 1/6 = 5/6$$

Example 3.3
For each of the following, determine the constant c so that $f(x)$ satisfies the conditions of being a *p.d.f* for a random variable X.

(i) $f(x) = x/c, \quad x = 1, 2, 3, 4$

(ii) $f(x) = cx, \quad x = 1, 2, 3, \ldots, 10$

(iii) $f(x) = c(x+1)^2, \quad x = 0, 1, 2, 3$

(iv) $f(x) = x/c, \ x = 1, 2, \cdots, n$

Solution:

(i) $f(x) = x/c, \quad x = 1, 2, 3, 4$

$$\sum_{x=1}^{4} f(x) = 1 \Rightarrow \sum_{x=1}^{4} x/c = 1$$

$$1/c \sum_{x=1}^{4} x = 1$$

$$1/c\,[1+2+3+4] = 1$$

$$10/c = 1 \qquad \therefore c = 10$$

(ii) $f(x) = cx, \qquad x = 1,\ 2,\ 3,\ \ldots,\ 10$

$$\sum cx = 1$$

$$c\sum x = 1$$

$$c\left[1+2+3+\cdots+10\right] = 1$$

$$c = \tfrac{1}{55}$$

(iii) $f(x) = c\left(x+1\right)^2, \qquad x = 0,\ 1,\ 2,\ 3$

$$\sum c\left(x+1\right)^2 = 1$$

$$c\sum\left(x+1\right)^2 = 1$$

$$c\left[1^2+2^2+3^2+4^2\right] = 1$$

$$c = 1/30$$

(iv) $f(x) = x/c,\ x = 1,\ 2,\ \cdots,\ n$

$$\sum_{x=1}^{n} x/c = 1$$

$$\frac{1}{c}\sum_{x=1}^{n} x = 1$$

$$\frac{1}{c}\left[\frac{n(n+1)}{2}\right] = 1$$

$$\therefore c = n(n+1)/2$$

Note the following:

(1) $\sum_{i=1}^{n} i = \dfrac{n(n+1)}{2}$

(2) $\sum_{i=1}^{n} i^2 = \dfrac{n(n+1)(2n+1)}{6}$

(3) $$\sum_{i=1}^{n} i^3 = \left[\frac{n(n+1)}{2}\right]^2$$

(4) $$\sum_{i=1}^{n} i^4 = \frac{n(n+1)(2n+1)(3n^2+3n+1)}{30}$$

(5) $$(1+t)^n = \sum_{k=0}^{n} \binom{n}{k} t^k$$

(6) $$\sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k} = (\mathrm{a} + \mathrm{b})^{\mathrm{n}}$$

(7) $$(1-t)^n = \sum \binom{n}{k} (-t)^k 1^k$$

(8) $$2^n = \sum_{k=0}^{n} \binom{n}{k}$$

(9) $$\sum_{k=0}^{n} \binom{a}{k}\binom{b}{n-k} = \binom{a+b}{n}$$

Continuous Random Variable

Random variable whose spaces are not composed of a countable number of points but are intervals or a union of intervals are said to be of the continuous type.

Definition 3.3: The probability density function, *pdf*, of a random variable X of the continuous type, with space R that is an interval or union of intervals, is an integrable function $f(x)$ satisfying the following conditions.

(i) $f(x) > 0,\ x \in \mathbb{R}$

(ii) $\int_R f(x)dx = 1$

(iii) The probability of the event $X \in A$ is
$$P(X \in A) = \int_A f(x)dx$$

Example 3.4

A machine produced copper wire, and occasionally there was a flaw at some point along the wire. The length of wire (in meters) produced between successive flaws is a continuous random variable X with p.d.f of the form

$$f(x) = \begin{cases} c(1+x)^{-3} & x > 0 \\ 0 & x \le 0 \end{cases}$$

Obtain the value of c, where c is a constant.

Solution:

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

$$\int_{0}^{\infty} c(1+x)^{-3}\, dx = 1$$

Let $U = (1+x)$ and $du = dx$ apply power rule for integral

$$c\int_{0}^{\infty} U^{-3} du = 1$$

Substituting for U gives

$$c\left[\left(\tfrac{(1+x)^{-2}}{-2}\right) - \left(\tfrac{1+0}{-2}\right)^{-2}\right]$$

$$c\left[0 + \tfrac{1}{2}\right] = \tfrac{2}{c} = 1$$

$$\therefore c = 2$$

$$c\left[\frac{U^{-2}}{-2}\right]_0^{\infty} = 1$$

$$c\left[\frac{1}{2}\right] = 1$$

$$c = 2$$

Example 3.5

For each of the following functions, find the constant c so that $f(x)$ is a p.d.f of a random variable X.

(i) $f(x) = 4x^c,\ 0 \le x \le 1$

(ii) $f(x) = c\sqrt{x},\ 0 \le x \le 4$

(iii) $f(x) = c/x^{3/4},\ 0 < x < 1$

Solution:

(i) $f(x) = 4x^c$

$$\int_0^1 f(x)dx = 1$$

$$\int_0^1 4x^c dx = 1$$

$$4\int_0^1 x^c dx = 1$$

$$4\left[\frac{x^{c+1}}{c+1}\right]_0^1 = 1$$

$$4\left[\frac{1}{c+1}\right] = 1$$

$$c = 3$$

(ii) $f(x) = c\sqrt{x}$

$$c\int_0^4 x^{1/2} dx = 1$$

$$c\left[\frac{x^{1/2+1}}{1/2+1}\right]_0^4 = 1$$

$$c\left[\frac{4^{3/2}}{3/2}\right] = 1$$

$$c = 3/16$$

(iii) $f(x) = c/x^{3/4}$

$$c\int_0^1 x^{-3/4} = 1$$

$$c\left[\frac{x^{-3/4+1}}{-3/4+1}\right]_0^1 = 1$$

$$c = 1/4$$

Summary

In this chapter, we discussed the following:

(1) The definition of a function and random variable.
(2) Discrete random variable and its properties.
(3) Continuous random variable and its properties.

Post-Test

(1) Define the following terms;
 (i) discrete random variable
 (ii) continuous random variable

(2) Let $f(x) = \frac{x}{15}$, $x = 1,2,3,4,5$, zero elsewhere be the p.d.f. of X.
Find;
 (i) $\Pr[1 \text{ or } 2]$
 (ii) $\Pr[1 \leq X \leq 3]$
 (iii) $\Pr\left[\frac{1}{2} \leq X \leq \frac{5}{2}\right]$

(3) For each of the following functions, find the constant c so that $f(x)$, satisfies the conditions of being a *p.d.f.* of a random variable X.
 (i) $f(x) = cx^3$, $o < x < 1$
 (ii) $f(x) = \frac{c}{x^4}$, $0 < x < \infty$

4

Distribution Functions of a Random Variable

Introduction

Frequently, we are interested in the probability that a random variable is equal to or less than some specified value or greater than a given value. The cumulative distribution function is particularly useful in this regard. The probability function for a discrete random variable X gives the probability of occurrence of the elements in the range of X. It can then be used to compute the probability of occurrence for any event defined by the observed value of X. We shall consider distribution functions (also frequently called the cumulative distribution function or cdf) for a random variable X. It is simply an alternative function that can be used to evaluate probabilities of events defined by the observed value of random variable.

Objectives

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

(1) Give a straightforward method for describing continuous random variables;
(2) Learn and understand the properties of cdf; and
(3) Evaluate distribution functions of random variables.

Pre-Test

(1) Define random variable.
(2) Define probability density function.
(3) What do you understand by cumulative distribution function?
(4) What is the relationship between cdf and p.d.f.?

Contents

Distribution Functions of Random Variable

Definition 4.1: Given a random variable X, the value of the cumulative distribution function at x, denoted by $F(x)$, is the probability that X takes on values less than or equal to x. Hence,

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

In the case of a discrete random variable, it is clear that

$$F(x) = \sum_{x \leq c} f(x)$$

The symbol, $\sum_{x \leq c} f(x)$ means "sum the values of $f(x)$ for all values of x less than or equal to c".

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

Theorem: Let X be a discrete random variable with p.d.f. $f(x)$ and cdf $F(x)$. If the possible values of X are indexed in increasing order, $x_1 < x_2 < x_3 < \ldots$, then $f(x_1) = F(x_1)$, for any $i > 1$,

$$f(x_i) = F(x_i) - F(x_{i-1})$$

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

$$F(x) = \sum_{x_i \leq x} (fx_i)$$

where, the summation is taken over all indices i such that $x_i < x$.

Definition 4.2: The distribution function of a random variable X of the continuous type defined in terms of the p.d.f. of X is given by

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

Note that $F'(x) = f(x) = \dfrac{dF(x)}{dx}$

Properties of a Distribution Function

The following are the properties of a distribution function $F(x)$ as a consequence of the fact that probability must be a value between 0 and 1, inclusive.

(i) $0 \le F(x) \le 1$ since $F(x)$ is a probability.

(ii) $F(x)$ is a non decreasing function of x.

(iii) $F(w) = 1$, where w is any value greater than or equal to the largest value in $\mathbb{R}$; and $F(z) = 0$, where z is any value less than the smallest value in $\mathbb{R}$.

(iv) If X is a random variable of the discrete type, then $F(x)$is a step function, and the height of a step at x, $x \in \mathbb{R}$, equal the probability $P(X = x)$.

Example 4.1

Let the random variable X of the discrete type have the p.d.f $f(x) = x/7$, $x = 1,2,4$. Find the distribution function of X.

Note that $P(X \le 0) = f(0) = 0$

But $P(X \le 1) = f(1) = \dfrac{1}{7}$,

$$P(X \le 2) = f(1) + f(2) = \frac{1}{7} + \frac{2}{7} = \frac{3}{7}$$

$$P(X \le 4) = f(1) + f(2) + f(4) = \frac{1}{7} + \frac{2}{7} + \frac{4}{7} = 1$$

So let $F(x) = P(X \le x)$ be defined for each real number x. Then

$$F(x) = \begin{cases} 0, -\infty < x < 1 \\ \frac{1}{7}, 1 \le x < 2 \\ \frac{3}{7}, 2 \le x < 4 \\ 1, 4 \le x < \infty \end{cases}$$

Note that $F(x)$ cumulates all the probability from points that are less than or equal to x.

Example 4.2

Let the random variable X be the distance in feet between bad records on a used computer tape. Suppose that a reasonable probability model of X is given by the p.d.f.

$$f(x) = \begin{cases} 0 & -\infty < x < 0 \\ \dfrac{1}{40} e^{-x/40} & , 0 \le x < \infty \end{cases}$$

the distribution function of X is

$$F(x) = 0 \text{ for } x \le 0 \text{ and for } x > 0$$

$$F(x) = \int_{-\infty}^{x} f(t)dt = \int_{0}^{x} \frac{1}{40} e^{-t/40} dt$$

$$= \quad -e^{-t/40}\Big|_0^x = 1 - e^{-x/40}$$

Note that,

$$F'(x) = \begin{cases} 0 & -\infty < x < 0 \\ \dfrac{1}{40} e^{-x/40} & , 0 \le x < \infty \end{cases}$$

Example 4.3

Let the random variable X have the p.d.f

$$f(x) = 2(1-x), \quad 0 \le x \le 1, \text{ zero otherwise.}$$

Determine the distribution function of X.

Solution:

$$F(x) = \int_{-\infty}^{x} f(t)dt = \int_{0}^{x} 2(1-t)dt$$

$$= \quad 2t\Big|_0^x - t^2\Big|_0^x$$

$$= \quad 2x - x^2 = x(2-x)$$

$$F(x) = \begin{cases} 0 & x < 0 \\ x(2-x) & 0 \le x < 1 \\ 1 & 1 \le x \end{cases}$$

Example 4.4
For each of the following functions.

(i) Find the constant c so that $f(x)$ is a p.d.f of a random variable X.

(ii) Find the distribution function, $F(x) = P(X \le x)$

(a) $f(x) = 4x^c, \quad 0 \le x \le 1 \quad c = 3$

(b) $f(x) = c\sqrt{x}, \quad 0 \le x \le 4 \quad c = 3/16$

(c) $f(x) = c/x^{3/4}, \quad 0 < x < 1 \quad c = 1/4$

Solution: $f(x) \; = \; 4x^3$

$$F(x) = \int_{-\infty}^{x} f(t)dt = \int_0^x 4t^3 dt = t^4\Big|_0^x = x^4$$

$$F(x) = \begin{cases} 0 & x < 0 \\ x^4 & 0 \le x < 1 \\ 1 & 1 \le x \end{cases}$$

Summary

We discussed the following among other things in this chapter;

(1) The description of cumulative distribution function.
(2) The properties of distribution functions.
(3) The relationship between $F(x)$ and $f(x)$ for a discrete distribution.
(4) How to determine the distribution function of a random variable X, of the discrete and continuous type.

Post-Test

(1) Let $f(x)$ be the p.d.f. of a random variable X. Find the distribution function $F(x)$ of X.

(i) $f(x) = 1, \quad x = 3$

(ii) $f(x) = \frac{1}{3}, \quad x = 1,2,3.$

(iii) $f(x) = \frac{x}{15}, \quad x = 1,2,3,4,5.$

(iv) $f(x) = \left(\frac{1}{4}\right)\left(\frac{3}{4}\right)^x, \quad x = 0,1,2,....$

(2) For each of the followings, find the distribution function $F(x) = P(X \le x)$.

(i) $f(x) = \left(\frac{3}{16}\right)x^{1/2}$

(ii) $f(x) = \dfrac{\left(\frac{1}{4}\right)}{x^{3/4}}$

5
Mathematical Expectation

Introduction

The probability distribution for a random variable can be defined by either its distribution function, $F(x)$, or its density function, $f(x)$, for the continuous or discrete type. Once the probability distribution of X is known, the probabilities of occurrence for any event of interest can thus be computed. However, in many applications, we may be interested in describing various aspects of different probability distributions, and ways of describing certain properties of probability distributions. For instance, what is a "typical" value the random variable can assume? "Typical" here may be defined in various ways. How much variability is exhibited by the probability distribution or how spread out the possible observed values for a random variable are, can be our concern. In this chapter, we will discuss some common measures of certain aspects of probability distributions such as a value that describes the "middle" or the "spread" of the probability distribution.

Objectives

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

(1) Obtain means of random variables;
(2) Obtain variances of random variables; and
(3) Prove the properties of mathematical expectations.

Pre-Test

(1) Explain the concept of mathematical expectation.
(2) Define the expected value of a random variable X of the discrete type.
(3) Define the expected value of a random variable X of the continuous type.

Content

Mathematical Expectation

An extremely important concept in summarizing important characteristics of distributions of probability is mathematical expectation, which is introduced by using example 5.1.

Example 5.1

A young man who needs a little extra money devises a game of chance in which some of his friends might wish to participate. The game that he proposes is to let the participants cast an unbiased die and then receive a payment according to the following schedule.

If the event $A = \{1, 2, 3\}$ occurs, he receives ₦1; if $B = \{4, 5\}$ occurs, he receives ₦5; and if $C = \{6\}$ occurs, he receives ₦35.

The probabilities of the respective events are assumed to be $3/6$, $2/6$, $1/6$.

The problem that now faces the young man is the determination of the amount that should be charged for the opportunity of playing the game. He reasons that if the game is played a large number of times, about $3/6$ of the trials will require a payment of ₦1; about $2/6$ of them will require a payment of ₦5 and about $1/6$ of them will require a payment of ₦35. Thus, the approximate average payment is

$$(1)\left(\frac{3}{6}\right) + (5)\left(\frac{2}{6}\right) + (35)\left(\frac{1}{6}\right) = 8$$

He expects to pay ₦8 "on the average". He never pays ₦8, the payment is either ₦1, ₦5 or ₦35. The weighted average of 1, 5 and 35 in which the weights are the respective probabilities $3/6$, $2/6$ and $1/6$, equals eight. Such a weighted average is called the ***Mathematical Expectation*** of payment.

Definition 5.1: If $f(x)$ is the pdf of the random variable X of the discrete type with space $\mathbb{R}$ and if the summation.

$$\sum_{R} U(x) f(x) = \sum_{x \in R} U(x) f(x)$$

exists, then the sum is called the Mathematical expectation or the expected value of the function $U(X)$ and it is denoted by $E[U(X)]$. That is

$$E[U(X)] = \sum_{R} U(x) f(x)$$

The expected value $E[U(X)]$ is thought of as a weighted mean of $U(X)$, $x \in \mathbb{R}$ where the weights are the probabilities $f(x) = P(X = x)$, $x \in \mathbb{R}$.

Example 5.2
Let the random variable X have the pdf

$$f(x) = 1/3, \qquad x \in R$$

Where $\mathbb{R} = \{-1, 0, 1\}$. Let $U(X) = X^2$

then,

$$E[U(X)] = E(X^2) = \sum U(x) f(x) = \sum x^2 f(x)$$

$$= (-1)^2 \left(\frac{1}{3}\right) + (0)^2 \left(\frac{1}{3}\right) + (1)^2 \left(\frac{1}{3}\right)$$

$$= \frac{2}{3}$$

For continuous type random variable, the definitions associated with mathematical expectation are the same as those in the discrete case except that integrals $\left(\int\right)$ replace summation $\left(\sum\right)$ symbols.

Definition 5.2: If X is a continuous random variable with p.d.f $f\left(x\right)$, then the expected value of X is defined by

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

Properties: When it exists, mathematical expectation E satisfies the following properties.

(a) If c is a constant,

$$E\left(c\right)=c$$

(b) If c is a constant, and U is a function

$$E\left[cU\left(X\right)\right]=cE\left[U\left(X\right)\right]$$

(c) If c_1 and c_2 are constants and U_1 and U_2 are functions, then

$$E\left[c_1U_1\left(X\right)+c_2U_2\left(X\right)\right]=c_1E\left[U_1\left(X\right)\right]+c_2E\left[U_2\left(X\right)\right]$$

Proof:

(a) $E(c)=\sum_R cf(x)=c\sum_R f(x)=c.$ Since $\sum_R f(x)=1$

In the continuous case,

$$E(c)=\int_{-\infty}^{\infty} cf\left(x\right)dx=c\int_{-\infty}^{\infty} f\left(x\right)dx=c. \text{ Since } \int_{-\infty}^{\infty} f\left(x\right)dx=1$$

(b) $E[cU(X)]=\sum_R cU(x)f(x)$

$$= c\sum_{R} U(x) f(x)$$

$$= cE[U(X)] \quad \text{Since} \quad E[U(X)] = \sum U(x) f(x)$$

For the continuous case,

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

$$= c\int_{-\infty}^{\infty} U(x)f(x)dx$$

$$= cE[U(X)] \qquad \text{Since } E[U(X)] = \int_{-\infty}^{\infty} U(x)f(x)dx$$

(c) $E[c_1U_1(X)+c_2U_2(X)] = \sum_R [c_1U_1(X)+c_2U_2(X)]f(x)$

$$= \sum_R c_1U_1(x)f(x) + \sum_R c_2U_2(x)f(x)$$

$$= c_1\sum_R U_1(x)f(x) + c_2\sum_R U_2(x)f(x)$$

$$= c_1E[U_1(X)] + c_2E[U_2(X)]$$

The Mean, Variance and Standard Deviation

If X is a random variable with p.d.f. $f(x)$ of the discrete type and space $\mathbb{R} = \{b_1, b_2, b_3, \ldots\}$ then

$$E(X) = \sum_R xf(x)$$

$$= b_1 f(x) + b_2 f(x) + b_3 f(x) + \ldots$$

is the weighted average of the numbers belonging to $\mathbb{R}$, where the weights are given by the p.d.f. $f(x)$.

If X is a continuous random variable having p.d.f $f(x)$, then

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

$E(X)$ is called the mathematical expectation or mean value or just mean of X (or the mean of the distribution) and denoted by μ. That is, $\mu = E(X)$.

Example 5.3: Let X have the p.d.f

$$f(x)=\begin{cases}\frac{1}{8}, x=0,3\\ \frac{3}{8}, x=1,2\end{cases}$$

The mean of X is

$$\mu=E(X)=0\left(\frac{1}{8}\right)+1\left(\frac{3}{8}\right)+2\left(\frac{3}{8}\right)+3\left(\frac{1}{8}\right)=\frac{3}{2}$$

Note that the mean $\mu=E(X)$ is the centroid of a system of weights or a measure of the central location of the probability distribution of X. A measure of the dispersion or spread of a distribution is defined as follows.

If $U(X)=(X-\mu)^2$ and $E\left[(X-\mu)^2\right]$ exists, the variance denoted by σ^2 or $V(X)$ of a random variable X of the discrete type is defined by

$$\sigma^2=E[(X-\mu)^2]=\sum_R (x-\mu)^2 f(x)$$

The positive square-root of the variance is called the standard deviation of X and denoted by

$$\sigma=\sqrt{Var(x)}$$

Example 5.4

(a) Let the *pdfs* of X be given by $f(0)=3/10$, $f(1)=3/10$, $f(2)=1/10$ and $f(3)=3/10$, compute the mean, variance and standard deviation of X.

(b) Find the mean and variance for the following discrete distributions:

(i) $f(x) = \frac{1}{5}, \quad x = 5, 10, 15, 20, 25$

(ii) $f(x) = 1, \quad x = 5$

(iii) $f(x) = \frac{4-x}{6}, \quad x = 1, 2, 3$

Solution: (a) $E(x) = \sum xf(x) = 0\left(\frac{3}{10}\right) + 1\left(\frac{3}{10}\right) + 2\left(\frac{1}{10}\right) + 3\left(\frac{3}{10}\right)$

$$= \frac{3}{10} + \frac{2}{10} + \frac{9}{10}$$

$$= \frac{14}{10} = 1.4$$

$$\sigma^2 = E\left[(X-\mu)^2\right] = V(X) = \sum (x-\mu)^2 f(x)$$

$$= (0-1.4)^2\left(\frac{3}{10}\right) + (1-1.4)^2\left(\frac{3}{10}\right) + (2-1.4)^2\left(\frac{1}{10}\right) + (3-1.4)^2\left(\frac{3}{10}\right)$$

$$= \frac{5.88}{10} + \frac{0.48}{10} + \frac{0.36}{10} + \frac{7.68}{10}$$

$$= \frac{14.4}{10} = 1.44$$

$$\sigma = \sqrt{V(X)} = 1.2$$

or $V(X) = E(X^2) - E^2(X)$

$$E(X^2) = \sum x^2 f(x)$$

$$= 0^2\left(\frac{3}{10}\right) + 1^2\left(\frac{3}{10}\right) + 2^2\left(\frac{1}{10}\right) + 3^2\left(\frac{3}{10}\right)$$

$$= \frac{3}{10} + \frac{4}{10} + \frac{27}{10}$$

$$= \frac{34}{10}$$

$$V(X)=\frac{34}{10}-\left(\frac{14}{10}\right)^2$$

$$= \frac{34}{10}-\frac{196}{100}$$

$$= 3.4 - 1.96 = 1.44$$

(b) i. $E(X)=\sum xf(x)=\sum x\left(\frac{1}{5}\right)$

$$= \frac{5}{5}+\frac{10}{5}+\frac{15}{5}+\frac{20}{5}+\frac{25}{5}$$

$$= 1 + 2 + 3 + 4 + 5$$

$$= 15$$

Example 5.5

Given $E(X+4)=10$ and $E\left[(X+4)^2\right]=116$. Determine

(i) $Var(X+4)$

(ii) μ

(iii) σ^2

Solution

(i) $Var(X+4)=E\left[(X+4)^2\right]-E^2(X+4)$

$$= 116 - 10^2$$

$$= 116 - 100$$

$$= 16$$

(ii) $\mu=E(X+4)=10$

$E(X)+4=10$

$\mu=E(X)=6$

(iii) $\sigma^2=Var(X)=E(X^2)-E^2(X)$

$$\text{But } E\left[(X+4)^2\right] = E\left(X^2+8X+16\right) = 116$$
$$E\left(X^2\right)+8E(X)+16 = 116$$
$$E\left(X^2\right)+48 = 100$$
$$E\left(X^2\right) = 52$$

$$Var(X) = E\left(X^2\right) - E^2(X)$$
$$= 52-6^2$$
$$= 16$$

As a measure of variability or spread in a continuous distribution, we will again consider the variance, $Var(X) = E\left[(X-\mu)^2\right]$, and the standard deviation is denoted by $\sigma = \sigma_X = \sqrt{Var(X)}$

The relationship $Var(X) = E(X)^2 - \mu^2$ holds, where $E\left(X^2\right) = \int_{-\infty}^{\infty} x^2 f(x)dx$

Example 5.6
Let Y be a continuous random variable with p.d.f.

$$f(y) = 2y, \quad 0<y<1$$

Obtain $E(Y)$ and $Var(Y)$

Solution:

$$\mu = E(Y) = \int_0^1 yf(y)dy$$
$$= \int_0^1 y(2y)dy$$
$$= \int_0^1 2y^2 dy$$

$$= \left[\frac{2y^3}{3}\right]_0^1 = \frac{2}{3}$$

$$\sigma^2 = Var(Y) = E(Y^2) - \mu^2$$

$$E(Y^2) = \int_0^1 y^2 f(y)dy$$

$$= \int_0^1 2y^3 dy$$

$$= \left[\frac{2y^4}{4}\right]_0^1$$

$$= \frac{2}{4}$$

Hence, var (Y) = $\frac{2}{4} - \frac{4}{9} = \frac{1}{2} - \frac{4}{9}$

$$= \frac{1}{18}$$

Summary

Here we learnt the following,

(1) The expected value of a random variable X of the discrete type, if it exists, is given by $E(x) = \sum_x xf(x)$.

(2) The expected value of a random variable X of the continuous type, if it exists, is given by $E(x) = \int_{-\infty}^{\infty} xf(x)dx$.

(3) From the very definition of $E[X]$, the following properties are immediate:

(i) $E[c] = c$.

(ii) $E[cX] = cE[X]$

(iii) $E[cX + d] = cE[X] + d$

(iv) $E[X + Y] = E[X] + E[Y]$

Post-Test

(1) Solve the question in example 5.4b (ii) and (iii).

(2) Show that $E[cU(X)+d]=cE[U(X)]+d$, and, in particular, $E(cX+d)=cE(X)+d$ where c and d are constants.

(3) Suppose $f(x)=\frac{1}{5}$, $x=1,2,3,4,5$, zero elsewhere, is the p.d.f. of the discrete type of random variable X. Compute $E(X)$ and $E(X^2)$. Hence or otherwise, find $Var(X)$ and $E\left[(X+2)^2\right]$.

6
Probability Distributions of Discrete Random Variables

Introduction

In many situations, it is useful to present the probability distribution of a random variable by a general algebraic expression. Probability calculations can then be made conveniently by substituting appropriate values into the algebraic model. The mathematical expression is a compact form of summarizing the nature of the process that has generated the probability distribution. In this chapter, three probability distributions of the discrete type will be discussed: the Bernoulli, binomial and Poisson distributions. The procedure for obtaining means and variances of these distributions shall also be explained. Other examples of the discrete distributions are; Multinomial distribution, Hyper geometric Distribution, Negative Binomial, etc.

Objectives

After a careful study of this chapter you should be able to do the following:

(1) Discuss the characteristics of probability distributions;
(2) Study some of the frequently occurring probability distributions; and
(3) Examine the assumed chance mechanisms that lead to the usage of these distributions.
(4) Derive expressions for obtaining means and variances of the discrete distributions; and
(5) Discuss the procedures for evaluating expected value of a discrete random variable.

Pre-Test

(1) Discuss the two types into which random variables are classified.
(2) Define the term ‘probability function’.

(3) List the properties of a probability density function.
(4) List the properties of a binomial distribution.
(5) Give the mean and variance of a Poisson distribution

Contents

Probability Distribution

A probability distribution is a mathematical idealization, or model of the relative frequency distribution of outcomes of a random experiment. If a random variable can assign only a countable number of values to the result of a random experiment, it is said to be a discrete random variable. However, a random variable that can assume any real value is called a continuous random variable.

Bernoulli Distribution

A trial with only two possible outcomes is used very frequently as a building block of a random experiment and it is called a *Bernoulli trial***.** It is usually assumed that the trials that constitute the random experiment are independent. This implies that the outcome from one trial has no effect on the outcome to be obtained from any other trial. Furthermore, it is often reasonable to assume that the probability of a success in each trial is constant.

The principal use of the binomial coefficients will occur in the study of one of the important chance processes called *Bernoulli trials*.

A *Bernoulli process* or *Bernoulli trial* is developed from a very specific set of assumptions involving the concept of a series of experimental trials.

Let us envision a process or experiment characterized by repeated trials. The trials take place under the following set of assumptions:

(1) There are two mutually exclusive possible outcomes on each trial, which are referred to as "success" and "failure." In whatever different language, the sample space of possible outcomes on each trial is $S = \{failure, success\}$. Other examples of experiments with exactly two possible results (outcomes) are; guilty and not guilty, defective and non defective, male or female etc.

(2) The probability p of success on each experiment is the same for each experiment, and this probability is not affected by any knowledge of previous outcomes. The probability q of failure is given by $q = 1 - p$.

The distribution is derived from a process known as a Bernoulli trial. When a single trial of an experiment results in only one of two mutually exclusive outcomes e.g. dead or alive, sick or well, the trial is called a Bernoulli trial.

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

The p.d.f. of a Bernoulli random variable is given as

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

Where, p is the probability of success and it remains constant from trial to trial, the corresponding probability of failure is denoted by *q* which is equal to *1-p*. The trials are independent. That is, the outcome of any given trial or sequence of trials do not affect the outcomes on subsequent trials. The outcome of any specific trial is determined by chance. Such processes are referred to as "random process" or "stochastic process." Bernoulli trials are one example of such processes.

Properties

Mean	$\mu = p$
Variance	$\sigma^2 = pq$
Standard Deviation	$\sigma = \sqrt{pq}$

Example 6.1
An urn contains 5 red and 15 green balls. Draw one ball at random from the urn. Let $X = 1$ if the ball drawn is red, and $X = 0$ if a green ball is drawn. Obtain:
(i) the *p.d.f.* of X,
(ii) mean of X and
(iii) variance of X.

Solution:

The p.d.f. of a Bernoulli distribution is $f(x) = p^x q^{1-x}, \; x = 0,1$

where $p = {}^{5}\!/_{20}$ and $q = {}^{15}\!/_{20}$

$$f(x) = \left(\frac{5}{20}\right)^x \left(\frac{15}{20}\right)^{1-x}, x = 0,1$$

Mean of $X = E(X) = \sum_{x=0}^{1} x\left(\tfrac{5}{20}\right)^x \left(\tfrac{15}{20}\right)^{1-x} = (0)\left(\tfrac{5}{20}\right)^0 \left(\tfrac{15}{20}\right)^1 + (1)\left(\tfrac{5}{20}\right)^1 \left(\tfrac{15}{20}\right)^0 = \left(\tfrac{5}{20}\right)$

Variance of $X = V(X) = \sum_{0}^{1} x^2 f(x) - [E(X)]^2 = (1)\left(\tfrac{5}{20}\right)^1 \left(\tfrac{15}{20}\right)^0 - \left(\tfrac{5}{20}\right)^2$

$$= \left(\tfrac{5}{20}\right) - \left(\tfrac{5}{20}\right)^2 = \left(\tfrac{3}{16}\right)$$

Binomial Distribution

An important distribution arising from counting the number of successes in a fixed number of independent *Bernoulli trials* is the Binomial distribution. A binomial experiment is any experiment that can be regarded as a sequence of n *Bernoulli trials* and meeting the following conditions.

(i) The underlying experiment consists of n repeated trials (n is defined before the experiment begins).

(ii) The result of every trial can be classified into one of two mutually exclusive categories.

(iii) The probability of success p does not change from trial to trial.

(iv) The result of any trial is independent of the results of all other trials.

The shape of the distribution depends on the two parameters p and n.

(i) When $p < 0.5$ and n is small, the distribution will be skewed to the right.

(ii) When $p > 0.5$ and n is small, the distribution will be skewed to the left

(iii) When $p = 0.5$ the distribution will be symmetric.

(iv) In all cases, as n gets larger the distribution gets closer to being a symmetric, bell-shaped distribution.

Properties

Mean $\mu = np$

Variance $\sigma^2 = npq$

Standard Deviation $\sigma = \sqrt{npq}$

If X is a random variable with probability of a success p, then the probability of obtaining x success in n trials is

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

That is the probability of x number of successes in n number of trials. This is the *p.d.f* of a Binomial distribution.

Example 6.2

If 20% of the bolts produced by a machine are bad. Determine the probability that out of 4 bolts chosen at random.

(i) one is defective
(ii) none is defective
(iii) at most 2 bolts will be defective.

Solution: $n = 4$, $p = 0.2$, $q = 0.8$

(i) $P[X = 1] = f(1) = \binom{4}{1} 0.2^1 0.8^3 = 0.4096$

(ii) $P[X = 0] = f(0) = \binom{4}{0} 0.2^0 0.8^4 = 0.4096$

(iii) $P[X \le 2] = P[X = 0, 1, 2] = P[X = 0] + P[X = 1] + P[X = 2]$

$= 0.4096 + 0.4096 + 0.1536$

$= 0.9728$

or $\quad 1-P[X>2] = 1-P(X=3)-P(X=4)$

or $\quad 1-P[X\geq 3] = 1-\binom{4}{3}0.2^{3}0.8^{1}-\binom{4}{4}0.2^{4}0.8^{0}$

$= 1-0.0256-0.0016$

$= 0.9728$

Example 6.3

(a) Suppose it is known that 30% of a certain population is immune to some disease. If a random sample of 10 is selected from this population. What is the probability that it will contain exactly 4 immune persons?

$n = 10, \quad p = 0.3, \quad x = 4$

$$f(4)=\binom{10}{4}(0.3)^{4}(0.7)^{6}$$

$$= 0.2$$

(b) In a certain population, 10% of the population is color-blind. If a random sample of 25 people is drawn from this population (use standard statistical tables). Find the probability that

(i) $P(X \geq 5) = 1-P(X<5) = 0.0980$

(ii) $P(X \leq 4) = 0.902$

or $1-P(X \geq 5) = 1-0.0980 = 0.902$

(iii) $P(6 \leq X \leq 10) = P(6) + P(7) + P(8) \text{-----} + P(10)$
$= 0.0333$
$= 0.0334 \qquad$ i.e. $P(X \geq 6)$

Example 6.4

From the experiment "toss four coins and count the number of tails", what is the variance of *X*?

$n = 4, \quad p = ½, \quad q = ½$

$$V(X)=npq=4\times ½\times ½=1$$

Example 6.5
Roll a fair 6-sided die 20 times and count the number of times that 6 shows up. What is the standard deviation of your random variable?

$$n = 20, \quad p = \tfrac{1}{6} \quad q = \tfrac{5}{6}$$

$$Var(X) = npq$$

$$= 20 \times \tfrac{1}{6} \times \tfrac{5}{6} = \tfrac{100}{36}$$

$$\sigma = \sqrt{V(X)} = \sqrt{\tfrac{100}{36}} = \tfrac{10}{6}$$

The Poisson Distribution
This is concerned with occurrences that can be described by a discrete random variable. The random variable can take on values, $x = 0, 1, 2$ ---- (i.e. non-negative integers)—countably infinite distribution, e.g.

- number of telephone calls per minute at a switchboard;
- number of mistakes per page in a large document; and
- numbers of traffic arrivals such as trucks at terminals, air planes at airports, ships at docks, etc.

All these have something in common, the given occurrences can be described in terms of a discrete random variable which takes values 0, 1, 2, ----.

The Poisson distribution can be used to find the probability that a certain number of events will occur in a given period of time provided that the following criteria are satisfied:

(1) The time interval used can be divided into many sub-intervals, so small that the probability of the event occurring in any one sub-interval is almost zero.
(2) The probability of more than one occurrence in any sub-interval is negligible.
(3) The occurrences of the events are independent. The occurrence of an event in an interval of space or time has no effect on the probability of a second occurrence of the event in the same or any other interval.

(4) The probability of the single occurrence of the event in a given interval is proportional to the length of the interval.
(5) The probability of an occurrence in any of the sub-intervals (or the mean rate of occurrence) remains constant throughout the entire time under consideration.
(6) The mean and the variance are equal.

When the above mentioned criteria are satisfied, the probability of X occurrences per unit of time is given by

$$f(x) = \frac{e^{-\lambda}\lambda^x}{x!} \qquad x = 0, 1, 2, ----\infty, \quad \lambda > 0$$

λ is the average number of occurrences of the random event in the interval and e is the constant, 2.7183.

The Poisson distribution gives a very good approximation to the binomial distribution. The Poisson distribution can be used when the sample size is very large, and the probability of an event occurring is very small.

Properties

Mean $\quad E(X) = \mu = \lambda$

Variance $\quad Var(X) = \sigma^2 = \lambda$

Standard deviation $\quad \sigma = \sqrt{\lambda}$

Example 6.6
Suppose that an urn contains 100,000 marbles and 120 are red. If a random sample of 1000 is drawn, what are the probabilities that 0, 1, 2, 3, and 4 respectively will be red?

$$n = 1000, \qquad p = \frac{120}{100000} = 0.0012, \ q = 0.9988$$

Solution:

Using **Binomial:** $\quad f(x) = \binom{1000}{x} 0.0012^2 . 0.9988^{1000-x}$

For x=3; $$f(3)=\binom{1000}{3}0.0012^3 \text{ x } 0.9988^{997}=0.0867$$

$$=166167000\times1.728^{-09}\times0.30206$$

$$= 0.0867$$

Using the **Poisson** method,

$$\lambda=np=1000\times0.0012=1.2$$
$$e^{-1.2}=0.3012$$
$$f(3)=\frac{e^{-1.2}1.2^3}{3!}=0.0867$$

Note that the result obtained by the two methods is the same but the computations involving the binomial distribution is quite tedious when n is large, it is however, less stressful when a simple method of approximation (i.e. Poisson) is used. Hence, the Poisson distribution is particularly suitable as an approximation when n is large and p is small.

Example 6.7
Let X have a Poisson distribution with a mean of $\lambda=5$. Find

(i) $P(X\le6)$
(ii) $P(X>5)$
(iii) $P(X=6)$
(iv) $P(X\ge4)$ P

Solution:

(i) $P(X\le6)=\sum_{x=0}^{6}\frac{5^x e^{-5}}{x!}=0.762$
(ii) $P(X>5)=1-P(X\le5)=1-0.616=0.384$
(iii) $P(X=6)=P(X\le6)-P(X\le5)=0.762 - 0.616 = 0.146$
(iv) $P(X\ge4)=1-P(X<4)$

Example 6.8

A hospital administrator, who has been studying daily emergency admissions over a period of several years, has come to the conclusion that they are distributed according to the Poisson law. Hospital records reveal that emergency admissions have averaged three per day during this period. If the administrator is correct in assuming a Poisson distribution. Find the probability that,

(1) Exactly two emergency admissions will occur on a given day;
(2) No emergency admission will occur on a particular day; and
(3) Either 3 or 4 emergency cases will be admitted on a particular day.

Solution:

(1) $$P(X=2)=\frac{e^{-\lambda}\lambda^{x}}{x!} \qquad \lambda=3$$

$$= \frac{e^{-3}3^{2}}{2!}=\frac{0.05(9)}{2}=0.225$$

(2) $$P(X=0)=\frac{e^{-3}3^{0}}{0!}=0.05$$

(3) $$P(X=3)+P(X=4)=\frac{e^{-3}3^{3}}{3!}+\frac{e^{-3}3^{4}}{4!}$$

$$= e^{-3}\left[\frac{27}{6}+\frac{81}{24}\right]$$

$$= 0.05\left[\frac{9}{2}+\frac{27}{8}\right]$$

$$= 0.05\,(7.875)$$

$$= 0.394$$

Derivation of Means and Variances of Some Discrete Distributions

Binomial Distribution

Given that X has a Binomial distribution with parameters n and p. Obtain the mean and variance of X

Solution: The pdf of a Binomial distribution is

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

$$E(X) = \sum_{x=0}^{n} x f(x)$$

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

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

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

Let x – 1 = y and n– 1 = m

$$= np \sum \frac{m!}{y!(m-y)!} p^y q^{m-y}$$

$$= np \sum_{y=0}^{n} \binom{m}{y} p^y q^{m-y}$$

$$= np$$

$$E(X^2) = \sum x^2 f(x)$$

$$= \sum x^2 \binom{n}{x} p^x q^{n-x}$$

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

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

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

Let $x-1=y$, $n-1=m$

$$= np \sum_{y=0}^{m} y+1 \frac{m!}{y!(m-y)!} p^y q^{m-y}$$

$$= np\left[\sum_{y=0}^{m} y\binom{m}{y} p^y q^{m-y} + \sum_{y=0}^{m}\binom{m}{y} p^y q^{m-y}\right]$$

$$= np[mp+1]$$

$$= np[(n-1)p+1]$$

$$= np[np-p+1]$$

$$= n^2p^2 - np^2 + np$$

$$Var(X) = E(X^2) - (E(X))^2$$

$$= n^2p^2 - np^2 + np - n^2p^2$$

$$= np - np^2$$

$$= np(1-p)$$

$$= npq$$

Poisson Distribution

Given that X has a Poisson distribution with parameter λ. Obtain the mean and variance of X.

Solution: The p.d.f. of X is given by

$$f(x) = \frac{e^{-\lambda}\lambda^x}{x!} \qquad x = 0, 1, 2, ----\infty, \quad \lambda > 0$$

$$E(X) = \sum_{x=0}^{\infty} x\frac{\lambda^x \ell^{-\lambda}}{x!} = \ell^{-\lambda}\sum_{x=0}^{\infty} x\frac{\lambda^x}{x(x-1)!} = \ell^{-\lambda}\sum_{x=1}^{\infty}\frac{\lambda^x}{(x-1)!}$$

Let $k = x-1$, then

$$E(X) = \ell^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^{k+1}}{k!} = \lambda\ell^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!} = \lambda\ell^{-\lambda}\ell^{\lambda} = \lambda\,,$$

Since from Maclaurin's series expansion $\sum_{x=0}^{\infty}\frac{\lambda^x}{x!} = \ell^{\lambda}$

$$\therefore E(X) = \lambda$$

$$E(X^2) = \sum_{x=0}^{\infty} x^2\frac{\lambda^x \ell^{-\lambda}}{x!} = \ell^{-\lambda}\sum_{x=0}^{\infty} x^2\frac{\lambda^x}{x(x-1)!} = \ell^{-\lambda}\sum_{x=1}^{\infty} x\frac{\lambda^x}{(x-1)!}$$

Let $k = x-1$, then

$$= \ell^{-\lambda}\sum_{k=1}^{\infty} k+1\frac{\lambda^{k+1}}{k!} = \lambda\ell^{-\lambda}\left[\sum_{k=0}^{\infty} k\frac{\lambda^k}{k!} + \sum_{k=0}^{\infty}\frac{\lambda^k}{k!}\right] = \lambda\left[\sum_{k=0}^{\infty} k\frac{\lambda^k\ell^{-\lambda}}{k!} + \sum_{k=0}^{\infty}\frac{\lambda^k\ell^{-\lambda}}{k!}\right]$$

$$= \lambda[\lambda+1]$$

Note that $\sum_{k=0}^{\infty}\frac{\lambda^k\ell^{-\lambda}}{k!} = 1$

$$= \lambda^2 + \lambda$$

$$\text{var}(X) = E(X^2) - E^2(X)$$
$$= \lambda^2 + \lambda - \lambda^2$$

$= \lambda$

Summary

In this chapter, we learnt the following:

(1) The Bernoulli distribution.
(2) Properties of the Bernoulli distribution which are:
  (i) $E(X) = \mu = p$
  (ii) $Var(X) = \sigma^2 = pq$
  (iii) Standard Deviation $= \sigma = \sqrt{pq}$
(3) Binomial Distribution with its properties given as follows;
  (i) $E(X) = \mu = np$
  (ii) $Var(X) = \sigma^2 = npq$
  (iii) $Standard\ Deviation = \sigma = \sqrt{npq}$
(4) The concept of the Poisson distribution.
(5) Properties of the Poisson distribution
  (i) $E(X) = \mu = \lambda$
  (ii) $Var(X) = \sigma^2 = \lambda$
  (iii) Standard Deviation $= \sigma = \sqrt{\lambda}$
(6) How to derive the mean and variance of a binomial distribution.
(7) To obtain the mean and variance of a Poisson distribution.

Post-Test

(1) Suppose that 24% of a certain population have blood group **B**. For a sample size of 20 drawn from this population, find the probability that
  (a) Exactly 3 persons with blood group **B** will be found.
  (b) Three or more persons with the characteristics of interest will be found
  (c) Fewer than three will be found.

(d) Exactly five will be found.

(2) In a large population, 16% of the members are left-handed. In a random sample of size 10, find

(a) The probability that exactly 2 will be left-handed $P(X = 2)$.

(b) $P(X \geq 2)$

(c) $P(X < 2)$

(d) $P(1 \leq X \leq 4)$

(3) Suppose mortality rate of a certain disease is 0.1, suppose 10 people in a community contract the disease, what is the probability that

(a) None will survive

(b) 50% will be

(c) At least 3 will die

(d) Exactly 3 will die

(4) Suppose it is known that the probability of recovery from a certain disease is 0.4. If 15 people are stricken with the disease, what is the probability that

(a) 3 or more will recover?

(b) 4 or more will recover?

(c) at least 5 will recover?

(d) fewer than three recover?

(5) In the study of a certain aquatic organism, a large number of samples were taken from a pond and the number of organisms in each sample was counted. The average number of organisms per sample was found to be two,, assuming the number of organisms Poisson distributed. Find the probability that:

(i) The next sample taken will contain one or more organisms,

(ii) The next sample taken will contain exactly three organisms,

(iii) The next sample taken will contain fewer than five organisms.

(6) It has been observed that the number of particles emitted by a radioactive substance, which reach a given portion of space during time *t*, follows closely the Poisson distribution with parameter $\lambda = 100$. Calculate the probability that:

(i) No particles will reach the portion of space under consideration during time t;

(ii) Exactly 120 particles do so;

(iii) At least 50 particles do so.

(7) The phone calls arriving at a given telephone exchange within one minute follow the Poisson distribution with parameter value equal to ten. What is the probability that in a given minute:

(i) No calls arrive?

(ii) Exactly 10 calls arrive?

(iii) At least 10 calls arrive?

7

Probability Distributions of Continuous Random Variables

Introduction

In the last chapter, we discussed the distributions of the discrete random variable. In this chapter, we shall look at some distributions of the continuous type. We will now consider the notion of a continuous random variable. In chapter three, we defined a continuous random variable stating clearly its properties. We also demonstrated how the mean and variance of a continuous random variable can be obtained.

Objectives

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

(1) Describe the theory of continuous distributions;
(2) Explain the properties of continuous distributions; and
(3) Evaluate the means and variances of any given function.
(4) Discuss the procedures for evaluating expected value of a continuous random variable.

Pre-Test

(1) Enumerate the relationship between Bernuolli and Binomial distributions.
(2) Define a continuous random variable X.
(3) Differentiate between discrete and continuous random variables.
(4) Write the *p.d.f.* of a normal distribution.

Contents

Uniform Distribution

Suppose that a continuous random variable X can assume values in a bounded interval only, say the open interval (a, b), and suppose the *p.d.f.* of X is given as

$$f(x;a,b) = f(x) = \begin{cases} \dfrac{1}{b-a}, & a < x < b \\ 0, & otherwise \end{cases}$$

This distribution is referred to as the Uniform or Rectangular Distribution on the interval *(a, b)* and is simply written as $X \sim U(a,b)$, where '*a*' and '*b*' are the parameters of the distribution. It provides a probability model for selecting a point at random from the interval *(a, b)*.

Properties

Mean $\mu = \dfrac{a+b}{2}$

Variance $\sigma^2 = \dfrac{(b-a)^2}{12}$

Standard Deviation $\sigma = \sqrt{\dfrac{(b-a)^2}{12}}$

Example 7.1

The hardness of a certain alloy (measured on Rockwell scale) is a random variable *X*. Assume that $X \sim U[50,75]$.

a. Find $P[60 < X < 70]$

b. Find $E(X)$

c. Find $Var(X)$

Solution:

(v) $P[60 < X < 70] = \int_{60}^{70} \frac{1}{b-a} dx$

$$= \frac{1}{75-50}[x]_{60}^{70}$$

$$= \frac{2}{5}$$

(vi) E(X) = $\frac{1}{b-a}\int_{50}^{75} xdx = \frac{125}{2}$

Or

E(X) = $\frac{b+a}{2} = \frac{75+50}{2} = \frac{125}{2}$

(vii) $Var(X) = E(X^2) - E^2(X) = \frac{1}{25}\int_{50}^{75} x^2 dx - \left(\frac{125}{2}\right)^2 = \frac{625}{12}$

Or

$Var(X) = \frac{(b-a)^2}{12} = \frac{625}{12}$

Exponential Distribution

An exponential distribution is a continuous distribution related to the Poisson distribution. In the Poisson process, the number of changes occurring in a given interval is counted which results in discrete distribution. But not only is the number of changes a random variable; the waiting times between successive changes are also random variables which are of the continuous type. The latter results into a distribution called the exponential distribution.

A continuous random variable X has the exponential distribution with parameter $\theta > 0$, if it has a p.d.f. of the form

$$f(x;\theta) = f(x) = \begin{cases} \frac{1}{\theta}\ell^{-\frac{x}{\theta}}, & x > 0 \\ 0, & otherwise \end{cases}$$

The exponential distribution, which is an important probability distribution for lifetimes, is characterized by the following properties.

Properties

Mean $E(X) = \theta$

Variance $Var(X) = \theta^2$

Standard Deviation $\sigma = \theta$

So if λ is the mean of changes in the unit interval, then $\theta = \frac{1}{\lambda}$ is the mean waiting time for the first change.

Example 7.2

Let the *p.d.f.* of *X* be $f(x) = \left(\frac{1}{2}\right)\ell^{-x/2}, 0 \le x < \infty$.

(i) What are the mean and variance of *X*?

(ii) Calculate $P(X > 3)$

(iii) Calculate $P(X > 5 | X > 2)$

(iv) Calculate $P(X < 2)$

Solution

(i) $E(X) = \theta = 2$ and $Var(X) = \theta^2 = 4$

(ii) $P(X > 3) = \frac{1}{2}\int_3^\infty \ell^{-\frac{x}{2}} dx = \ell^{-\frac{3}{2}} = 0.2231$

(iii) $P(X > 5 | X > 2) = \dfrac{\frac{1}{2}\int_5^\infty \ell^{-\frac{x}{2}}}{\frac{1}{2}\int_2^\infty \ell^{-\frac{x}{2}}} = \dfrac{\ell^{-\frac{5}{2}}}{\ell^{-2/2}} = \ell^{-\frac{3}{2}} = 0.2231$

(iv) $P(X < 2) = \frac{1}{2}\int_0^2 \ell^{-\frac{x}{2}} = 1 - \ell^{-\frac{2}{2}} = 1 - \ell^{-1} = 0.6321$

The Normal Distribution

The normal distribution plays a central role in statistical theory and practice, particularly in the area of statistical inference. The normal distribution is perhaps the most important distribution in statistical applications since many measurements have (approximate) normal

distributions. The main reason for this is its role in the Central Limit Theorem (CLT).
The random variable X has a normal distribution if its *p.d.f.* is defined by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right], -\infty < x < \infty$$

In this equation, the mean and standard deviation, which determine the location and spread of the distribution, are denoted by μ and σ, respectively. These are said to be the two parameters of the normal distribution satisfying $-\infty < \mu < \infty, 0 < \sigma < \infty$. Briefly, we say that X is $N(\mu, \sigma^2)$.

Theorem: If the random variable X is $N(\mu, \sigma^2), \sigma^2 > 0$ then the random variable $Z = (X - \mu)/\sigma$ is $N(0,1)$.

Proof: The distribution function of Z is

$$P(Z \le z) = P\left(\frac{X-\mu}{\sigma} \le z\right) = P(X \le z\sigma + \mu)$$

$$= \int_{-\infty}^{z\sigma+\mu} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[\frac{(x-\mu)^2}{2\sigma^2}\right] dx$$

Changing the variable of integration by writing $w = (x-\mu)/\sigma$, $x = w\sigma + \mu$. We then obtain;

$$P(Z \le z) = \int_{=\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \ell^{-w^2/2}\, dw$$

This is the expression for $\Phi(z)$, the cumulative distribution function of a standardized normal random variable. Hence, Z is $N(0,1)$.

This fact considerably simplifies the calculations of probabilities concerning normally distributed variables, as seen in the following illustration:

Suppose, that X is $N(\mu, \sigma^2)$, let $c_1 < c_2$, and since $P(X = c_1) = 0$, then

$$P(c_1 < X < c_2) = P(X < c_2) - P(X < c_1)$$

$$= P\left(\frac{X-\mu}{\sigma} < \frac{c_2-\mu}{\sigma}\right) - P\left(\frac{X-\mu}{\sigma} < \frac{c_1-\mu}{\sigma}\right)$$

$$= \Phi\left(\frac{c_2-\mu}{\sigma}\right) - \Phi\left(\frac{c_1-\mu}{\sigma}\right)$$

Note that $\Phi(-x) = 1 - \Phi(x)$.

The normal distribution possesses the following properties.

Properties

Mean	$E(X) = \mu$
Variance	$Var(X) = \sigma^2$
Standard Deviation	$= \sigma$

Example 7.3

(1) If Z is $N(0,1)$, find;
   (i) $P(0.53 < Z < 2.06)$
   (ii) $P(Z > 2.89)$

(2) If X is $N(75,100)$, find $P(X < 60)$.

(3) If X is normally distributed with a mean of 6 and a variance 25, find $P(6 \le X \le 12)$

Solution

(1)

(i) $P(0.53 < Z < 2.06) = \Phi(2.06) - \Phi(0.53) = 0.9803 - 0.7019 = 0.2784$

(ii) $P(Z > 2.89) = 1 - \Phi(2.89) = 1 - 0.9981 = 0.0019$

(2) $P(X < 60) = P\left(\frac{X - 75}{10} < \frac{60 - 75}{10}\right) = P(Z < -1.5) = 0.0668$

(3) $P(6 \le X \le 12) = P\left(\frac{6-6}{5} \le Z \le \frac{12-6}{5}\right) = P(0 \le Z \le 1.2)$

$$= \Phi(1.2) - \Phi(0) = 0.8849 - 0.5000 = 0.3849$$

Derivation of Means and Variances of Continuous Distribution

Uniform Distribution

Let X have a uniform distribution $U(a,b)$ with p.d.f

$$f(x) = \frac{1}{b-a} \qquad a < x < b$$

Obtain the mean and variance of X.

Solution: $E(X) = \int_a^b xf(x)dx = \int_a^b x\frac{1}{b-a}dx$

$$= \frac{1}{b-a}\int_a^b xdx = \frac{1}{b-a}\left[\frac{x^2}{2}\right]_a^b$$

$$= \frac{1}{b-a}\left[\frac{b^2 - a^2}{2}\right]$$

$$= \frac{1}{b-a}\left[\frac{(b+a)(b-a)}{2}\right]$$

$$= \frac{b+a}{2}$$

$$E(X^2) = \int_a^b x^2 f(x)dx$$

$$= \frac{1}{b-a}\int_a^b x^2 dx = \frac{1}{b-a}\left[\frac{x^3}{3}\right]_a^b = \frac{1}{b-a}\left[\frac{b^3 - a^3}{3}\right]$$

$$= \frac{1}{b-a}\left[\frac{(b-a)(b^2 + ab + a^2)}{3}\right]$$

$$= \frac{b^2 + ab + a^2}{3}$$

$$V(X) = \frac{b^2 + ab + a^2}{3} - \frac{\left(b^2 + 2ab + a^2\right)}{4}$$

$$= \frac{4b^2 + 4ab + 4a^2 - 3b^2 - 6ab - 3a^2}{12}$$

$$= \frac{b^2 - 2ab + a^2}{12} = \frac{(b-a)^2}{12}$$

Summary

In this chapter, we considered the following:

(1) The concept of uniform and exponential distributions.

(2) How the Poisson distribution is related to the exponential distribution.

(3) The uniform distribution has the following properties;

(i) Mean $\mu = \frac{a+b}{2}$

(ii) Variance $\sigma^2 = \frac{(b-a)^2}{12}$

(iii) Standard Deviation σ = $\sqrt{\frac{(b-a)^2}{12}}$

(4) The exponential distribution on the other hand has the following properties;

(i) Mean $E(X) = \theta$

(ii) Variance $Var(X) = \theta^2$

(iii) Standard Deviation $\sigma = \theta$

(5) The importance of the normal distribution.

(6) The p.d.f. of a normal distribution is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right], -\infty < x < \infty$$

(7) A theorem to show that the random variable $Z = (X-\mu)/\sigma$ is $N(0,1)$.

Post-Test

(1) Let X have an exponential distribution with a mean of $\theta = 20$. Compute

(i) $P(10 < X < 30)$

(ii) $P(0 < X < 30)$

(iii) $P(X > 30)$

(iv) $P(X > 40 | X > 10)$

(2) Telephone calls enter a college switchboard according to a Poisson process on the average of two every 3 minutes. Let X denote the waiting time until the first call that arrives after 10 A.M.

(i) What is the *p.d.f.* of X?

(ii) Find $P(X > 2)$

(3) Customers arrive randomly at a bank teller's window. Given that one customer arrived during a particular 10-minute period, let X equal the time within the 10 minutes that the customer arrived. If X is $U(0,10)$, find

(i) The *p.d.f.* of X.

(ii) $P(X \geq 8)$

(iii) $P(2 \leq X < 8)$

(iv) $E(X)$

(v) $Var(X)$

(4) Explain the relationship that exists between the poisson and the exponential distributions.

(5) If X is $N(75,100)$, find $P(X < 35)$ and $P(70 < X < 100)$.

(6) If Z is $N(0,1)$, find values of c such that

(i) $P(Z \geq c) = 0.025$

(ii) $P(|Z| \leq c) = 0.95$

(iii) $P(Z > c) = 0.05$

(7) Let X be $N(\mu, \sigma^2)$, so that $P(X < 89) = 0.90$ and $P(X < 94) = 0.95$. Find μ and σ^2.

(8) Show that the random variable $Z = (X - \mu)/\sigma$ is distributed $N(0,1)$.

(9) Suppose that $Z \sim N(0,1)$. Find the following probabilities:

(i) $P(Z \leq 1.53)$

(ii) $P(Z > -0.48)$

(iii) $P(0.35 < Z < 2.01)$

(iv) $P(|Z| > 1.28)$

Find the value of 'a' and 'b' such that

(v) $P(Z \leq a) = 0.648$

(vi) $P(|Z| \leq b) = 0.95$

(10) The *p.d.f* of X is $f(x) = {d}/{x^3}$, $1 < x < \infty$, zero elsewhere.

(i) Calculate the value of d so that $f(x)$ is a *p.d.f*

(ii) Find $E(X)$

(iii) Show that $Var(X)$ does not exist.

(11) Find the mean and variance of the following distributions.

(i) $f(x) = \left({3}/{2}\right)x^2, \quad -1 < x < 1$

(ii) $f(x) = \dfrac{1}{2}, \quad -1 < x < 1$

(iii) $f(x) = \begin{cases} x+1 & -1 < x < 0 \\ 1-x & 0 \leq x < 1 \end{cases}$

(12) Obtain the mean and variance of the exponential distribution.

(13) Derive the mean and variance of the normal distribution.

8
Generating Functions

Introduction

In the preceding chapters we saw the importance and derivation of the mean, standard deviation and variance of a random variable *X*. For some distributions, it can be fairly difficult to obtain directly $E(X)$ and $E(X^2)$, the first and second moments. We shall discuss here a function of a real variable t that can be used to find $E(X)$ and $E(X^2)$ as well as higher moments of *X*. ***Moments*** are merely the averages of powers of the variable values. In this chapter, we will treat the notions of the location, spread, symmetry, and peakedness of a histogram as measures of the characteristics of shape. It is important to note that once location and spread have been determined, it is more informative to look at standardized variables, or standardized moments, to measure the remaining shape characteristics.

Although, the moment-generating function (m.g.f.), if it exists, is a useful tool for determining moments, its major importance is in the fact that it uniquely determines the distribution.

The moments of any given random variable can be computed using the moment-generating function approach. For many random variables the cumulant generating function (c.g.f) proves easier to use in evaluating the mean and variance. The reason for this simplicity is that the first two derivatives of *c.g.f.* of *X* written as $C_x(t)$ evaluated at $t = 0$ directly give the mean and the variance of *X*. For many of the standard random variables that we will discuss, these two derivatives are very easy to compute. Either the cumulant generating function or the moment generating function can be used to evaluate means and variances (and other moments) of a random variable.

Objectives

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

(1) Obtain the means and variances of distributions;
(2) Show the usefulness of *m.g.f.* over the direct computation of expectations; and
(3) Make extensions for determining higher moments.
(4) Discuss the concept of cumulant generating function;
(5) Distinguish between the moment-generating function and the cumulant generating function; and
(6) Compute moments using *c.g.f.*

Pre-Test

(1) Define the term 'moment' of a random variable X.
(2) What do you understand by the term 'first moment about the origin?'
(3) What do you understand by the term 'r^{th} moment about the mean?'
(4) Explain the term 'factorial moments.'
(5) Obtain the mean and variance of a geometric distribution.
(6) Write down expressions for the first, second and r^{th} moments about the origin.
(7) Differentiate between 'moments about the origin' and 'moments about the mean.

Contents

Moments of a Distribution: Mean, Variance, Skewness and Kurtosis

When a set of values has a sufficiently strong central tendency, that is, a tendency to cluster around some particular value, then it may be useful to characterize the set by a few numbers that are related to its *moments*, the sums of integer powers of the values.

Best known is the *mean* of the values $x_1, x_2, \ldots, x_N$,

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

which estimates the value around which central clustering occurs. This does not imply that the *mean* is the only available estimator of this quantity, nor is it necessarily the best one. For values drawn from a probability distribution with very broad "tails," the mean may converge poorly, or not at all, as the number of sampled points is increased.

Having characterized a distribution's central value, one conventionally then characterizes its "width" or "variability" around that value. Here again, more than one measure is available. Most common is the *variance*,

$$Var(x_1, x_2, \ldots, x_N) = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2 \qquad 8.1$$

or its square root, the *standard deviation*,

$$\sigma(x_1, x_2, \ldots, x_N) = \sqrt{Var(x_1, x_2, \ldots, x_N)} \qquad 8.2$$

Equation (8.1) estimates the mean squared deviation of x from its mean value.

As the mean depends on the first moment of the data, so do the variance and standard deviation depend on the second moment. It is not uncommon, in real life, to be dealing with a distribution whose second moment does not exist (i.e., is infinite). In this case, the variance or standard deviation is useless as a measure of the data's width around its central value: The values obtained from equations (8.1) or (8.2) will not converge with increased numbers of points, nor show any consistency from data set to data set drawn from the same distribution. This can occur even when the width of the peak looks, by eye, perfectly finite. A more robust estimator of the width is the *average deviation* or *mean absolute deviation*, defined by

$$ADev(x_1, x_2, \ldots, x_N) = \frac{1}{N}\sum_{i=1}^{N}|x_i - \bar{x}| \qquad 8.3$$

Higher moments, or statistics involving higher powers of the input data, are almost always less robust than lower moments or statistics that involve only linear sums or (the lowest moment of all) counting.

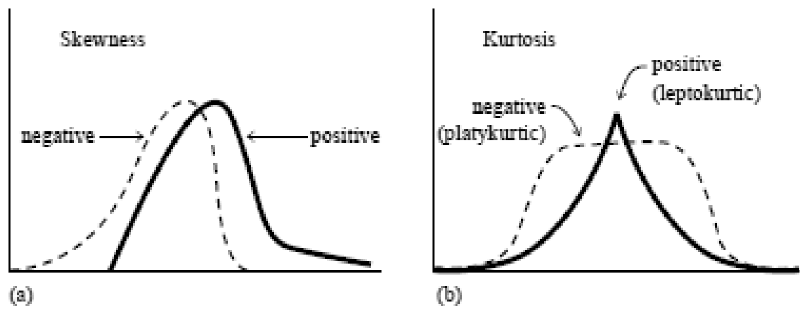


Fig. 8.1: Distributions whose third and fourth moments are significantly different from a normal (Gaussian) distribution. (a) Skewness or third moment. (b) Kurtosis or fourth moment.

The skewness characterizes the degree of asymmetry of a distribution around its mean. While the mean, standard deviation, and average deviation are *dimensional* quantities, that is, have the same units as the measured quantities x_i, the skewness is conventionally defined in such a way as to make it *non-dimensional*. It is a pure number that characterizes only the shape of the distribution. The usual definition is

$$Skew(x_1, x_2, \ldots, x_N) = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{x_i - \bar{x}}{\sigma}\right]^3 \qquad 8.4$$

where $\sigma = \sigma(x_1, x_2, \ldots, x_N)$ is the distribution's standard deviation (8.2). A positive value of skewness signifies a distribution with an asymmetric tail extending out towards more positive *x*; a negative value signifies a distribution whose tail extends out towards more negative *x* (fig. 8.1a).

Of course, any set of *N* measured values is likely to give a nonzero value for (8.4), even if the underlying distribution is in fact symmetrical (has zero skewness). For (8.4) to be meaningful, we need to have some idea of *its* standard deviation as an estimator of the skewness of the underlying distribution. Unfortunately, that depends on the shape of the underlying distribution, and rather critically on its tails! For the idealized case of a normal (Gaussian) distribution, the standard deviation of (8.4) is approximately $\sqrt{15/N}$. In real life it is good practice to believe in skewnesses only when they are several or many times as large as this.

The kurtosis is also a nondimensional quantity. It measures the relative peakedness or flatness of a distribution (relative to a normal distribution). A distribution with positive kurtosis is termed *leptokurtic*. A distribution with negative kurtosis is termed *platykurtic* (fig. 8.1b). And an in-between distribution is termed *mesokurtic*. The conventional definition of the kurtosis is

$$Kurt\left(x_1, x_2, \ldots, x_N\right) = \left\{\frac{1}{N}\sum_{i=1}^{N}\left[\frac{x_i - \bar{x}}{\sigma}\right]^4\right\} - 3 \qquad 8.5$$

where the −3 term makes the value zero for a normal distribution.

The standard deviation of (8.5) as an estimator of the kurtosis of an underlying normal distribution is $\sqrt{96/N}$. However, the kurtosis depends on such a high moment that there are many real-life distributions for which the standard deviation of (8.5) as an estimator is effectively infinite.

Moments

The k^{th} moment about the origin of a random variable X is

$$\mu_k' = E(X^k)$$

and the k^{th} moment about the mean is

$$\begin{aligned}\mu_k &= E\left[X - E(X)\right]^k \\ &= E\left[X - \mu\right]^k\end{aligned}$$

Definition 8.1: Moment Generating Functions

Let X be a random variable of the discrete type with p.d.f $f\left(x\right)$ and space $\mathbb{R}$. If there is a positive number h such that

$$E(e^{tx}) = \sum_{x \in R} e^{tx} f(x)$$

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

$$M(t) = E\left(e^{tx}\right)$$

is called the Moment – generating function (MGF) of *X*.
The derivatives of $M(t)$ of all orders exist at $t = 0$. Thus

$$M'(t) = \sum_{x \in R} xe^{tx} f(x)$$
$$M''(t) = \sum_{x \in R} x^2 e^{tx} f(x)$$

and, for each positive integer *r*,

$$M^{(r)}(t) = \sum_{x \in R} x^r e^{tx} f(x)$$

Setting $t = 0$

$$M'(0) = \sum xf(x) = E(X)$$
$$M''(0) = \sum x^2 f(x) = E(X^2)$$

and in general

$$M^{(r)}(0) = \sum_{x \in R} x^r e^{tx} f(x) = E(X^r)$$

In particular, if the *MGF* exists,

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

For continuous random variable *X*

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

The moment-generating functions have the following properties:

(1) $M_x(0) = 1$.

(2) $|M_x(t)| \leq 1$

(3) M_x is uniformly continuous.

(4) $M_{x+d}(t) = \ell^{td} M_x(t)$, where d is a constant.

(5) $M_{cx}(t) = M_x(ct)$, where c is a constant.

(6) $M_{cx+d}(t) = \ell^{td} M_x(ct)$

$$\frac{d^n}{dt^n} M_x(t)\bigg|_{t=0} = E(X^n), \qquad n = 1,2,\dots, \text{if } E|X^n| < \infty$$

The proofs of selected properties are as follows;

Property 1: $\mathrm{M_x}(\mathrm{t}) = E(\ell^{tx})$, for $t = 0$, $\mathrm{M_x}(\mathrm{t}) = E(\ell^0) = E(1) = 1$

Property 4: $M_{x+d}(t) = E(\ell^{t(x+d)}) = E(\ell^{tx}\ell^{td}) = \ell^{td} E(\ell^{tx})$

$= \ell^{td} M_x(t)$

Property 5: $M_{cx}(t) = E(\ell^{ctx}) = E(\ell^{(ct)x}) = M_x(ct)$

Example 8.1: Binomial Distribution

Let X have a binomial distribution $b(n, p)$ with *p.d.f*

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}, \qquad x = 0, 1, 2, \text{----}, n$$

The *MGF* of X is

$$M(t) = E[e^{tX}] = \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x (1-p)^{n-x}$$

$$= \sum_{x=0}^{n} \binom{n}{x} (pe^t)^x (1-p)^{n-x}$$

Using the formula for the binomial expansion with $a = 1 - p$ and $b = pe^t$

$$= \sum \binom{n}{x} b^x a^{n-x}$$

$$= (a+b)^n$$

$$\therefore \; M(t) = \left[(1-p) + pe^t\right]^n \;\; \forall \text{ real values of } t$$

The mean and variance are,

$$M'(t) = n\left[(1-p) + pe^t\right]^{n-1}\left(pe^t\right)$$

and $M''(t) = n(n-1)\left[(1-p) + pe^t\right]^{n-2}\left(pe^t\right)^2 + n\left[(1-p) + pe^t\right]^{n-1}\left(pe^t\right)$

Thus $\mu = E(X) = M'(0) = np$

and $\sigma^2 = E\left(X^2\right) - E^2(X) = M''(0) - \left[M'(0)\right]^2$

$$= n(n-1)p^2 + np - (np)^2$$
$$= n^2p^2 - np^2 + np - n^2p^2$$
$$= np(1-p)$$

Example 8.2: Poisson Distribution
Let X has a Poisson distribution with *p.d.f*

$$f(x) = \frac{\lambda^x e^{-\lambda}}{x!} \qquad x = 0, 1, 2, \text{----}$$

The MGF of X can be obtained as follows;

$$M(t) = E(e^{tX}) = \sum_{x=0}^{\infty} e^{tx} f(x)$$

$$= \frac{\sum e^{tx} \lambda^x e^{-\lambda}}{x!}$$

$$= e^{-\lambda} \sum \frac{\left(\lambda e^t\right)^x}{x!}$$

Note that, $\qquad e^X = \sum_{X=0}^{\infty} \frac{x^n}{n!}$

$$M(t) = e^{-\lambda} e^{\lambda e^t}$$

$$= e^{\lambda(e^t - 1)} \qquad \forall \text{ real values of } t$$

The first and second moments are;

$$E(X) = M'(0)$$

$$M'(t) = \lambda \ell^t e^{\lambda\left(\ell^t - 1\right)} = M'(0) = \lambda$$

$$M''(t) = \left(\lambda \ell^t\right) e^{\lambda\left(\ell^t - 1\right)} + \left(\lambda \ell^t\right)^2 e^{\lambda\left(\ell^t - 1\right)} = M''(0) = \lambda + \lambda^2$$

Then $Var(X) = M''(0) - [M'(0)]^2$

$$= \lambda + \lambda^2 - \lambda^2$$

$$= \lambda$$

Example 8.3: Normal Distribution
Let X have a normal distribution with *p.d.f.*

$$f(x) = \begin{cases} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right], -\infty < x < \infty \\ 0, \ elsewhere \end{cases}$$

The moment-generating function of X can be calculated as follows;

$$M(t)=E\left(\ell^{tx}\right)=\int_{-\infty}^{\infty}\ell^{tx}\frac{1}{\sigma\sqrt{2\pi}}\ell^{-\frac{1}{2\sigma^2}(x-\mu)^2}dx$$

$$=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty}\ell^{tx-\frac{1}{2\sigma^2}\left(x^2-2\mu x+\mu^2\right)}dx$$

$$=\frac{1}{\sigma\sqrt{2\pi}}\ell^{-\frac{\mu^2}{2\sigma^2}}\int_{-\infty}^{\infty}\ell^{-\frac{1}{2\sigma^2}\left[x-\left(\mu+\sigma^2t\right)\right]^2-\left[\left(\mu+\sigma^2t\right)^2\right]}dx$$

$$=\ell^{-\frac{1}{2\sigma^2}\left[\mu^2-\left(\mu+\sigma^2t\right)^2\right]}\int_{-\infty}^{\infty}\frac{1}{\sigma\sqrt{2\pi}}\ell^{-\frac{1}{2\sigma^2}\left[x-\left(\mu+\sigma^2t\right)\right]^2}dx$$

$$=\ell^{-\frac{1}{2\sigma^2}\left[\mu^2-\mu^2-2\mu\sigma^2t-\left(\sigma^2\right)^2t^2\right]}$$

Since $\displaystyle\int_{-\infty}^{\infty}\frac{1}{\sigma\sqrt{2\pi}}\ell^{-\frac{1}{2\sigma^2}\left[x-\left(\mu+\sigma^2t\right)\right]^2}dx=1$

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

The first and second moments are given as follows:

$$M'(t)=\left(\mu+\sigma^2t\right)\ell^{\mu t+\frac{\sigma^2t^2}{2}}$$

$$M''(t)=\sigma^2\ell^{\mu t+\frac{\sigma^2t^2}{2}}+\left[\left(\mu+\sigma^2t\right)\left(\mu+\sigma^2t\right)\ell^{\mu t+\frac{\sigma^2t^2}{2}}\right]$$

The mean is therefore

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

To get the variance, we proceed as follows:

$$M''(0) = \sigma^2 + \mu^2$$

$$Var(X) = M''(0) - [M'(0)]^2 = \sigma^2 + \mu^2 - \mu^2 = \sigma^2$$

Example 8.4: Uniform Distribution
Let X have a uniform distribution with *p.d.f*

$$f(x) = \frac{1}{b-a} \qquad\qquad a < x < b$$

Obtain the moment-generating function of X. Hence find its mean and variance.

$$M(t) = \int_a^b e^{tx} \frac{1}{b-a} dx$$

$$= \qquad \frac{1}{b-a}\int_a^b e^{tx} dx = \frac{1}{b-a}\left[\frac{e^{tx}}{t}\right]_a^b$$

$$= \qquad \frac{1}{b-a}\left[\frac{e^{tb} - e^{ta}}{t}\right]$$

$$= \qquad \frac{e^{tb} - e^{ta}}{t(b-a)}$$

$$= \qquad \frac{1}{t(b-a)}\left[e^{tb} - e^{ta}\right]$$

Note that $$e^X = 1 + \frac{X}{1!} + \frac{X^2}{2!} + \frac{X^3}{3!} ------$$

$$e^{tb} = 1 + \frac{tb}{1!} + \frac{t^2b^2}{2!} + \frac{t^3b^3}{3!} ------$$

$$e^{ta} = 1 + \frac{ta}{1!} + \frac{t^2a^2}{2!} + \frac{t^3a^3}{3!} ------$$

$$M(t) = \frac{1}{t(b-a)}\left[1 + tb + \frac{t^2b^2}{2} + \frac{t^3b^3}{6} - 1 - ta - \frac{t^2a^2}{2} - \frac{t^3a^3}{6}\right]$$

$$= \frac{t}{t(b-a)}\left[(b-a) + \frac{tb^2}{2} - \frac{ta^2}{2} + \frac{t^2b^3}{6} - \frac{t^2a^3}{6}\right]$$

$$M'(t) = \frac{1}{b-a}\left[\frac{b^2}{2} - \frac{a^2}{2} + \frac{2tb^3}{6} - \frac{2ta^3}{6}\right]$$

The mean is thus,

$$M'(0) = \frac{1}{b-a}\left[\frac{b^2 - a^2}{2}\right]$$

And the variance is

$$M''(0) = \frac{1}{b-a}\left[\frac{2b^3}{6} - \frac{2a^3}{6}\right]$$

$$= \frac{1}{b-a}\left[\frac{b^3 - a^3}{3}\right]$$

$$Var(X) = M''(0) - [M'(0)]^2 = \left[\frac{b^3 - a^3}{3(b-a)}\right] - \left[\frac{b^2 - a^2}{2(b-a)}\right]^2$$

$$= \frac{(b-a)^2}{12}$$

Definition 8.2: Cumulant Generating Functions
The cumulant generating function is defined to be the natural log of the moment generating function (assuming it exists). That is, $M(t)$ is the moment generating function of *X*, then the cumulant generating function for *X* is

$$C_x(t) = \ln M(t)$$

i.e. $$M(t) = \ell^{C_x(t)}$$

So if $C_x(t)$ were known, it is easy to find $M_x(t)$. Then

$$\frac{d}{dt}C(t) = \frac{M'(t)}{M(t)}$$

$$\frac{d^2}{dt^2}C(t) = \frac{M''(t)M(t) - (M'(t))^2}{(M(t))^2}$$

where

$$\frac{d}{dt}M(t) = M'(t), \frac{d^2}{dt^2}M(t) = M''(t)$$

Since $M(0)=1$, i.e. $M(t) = E\left(e^{tx}\right)$, then $M(0) = E\left(e^0\right) = 1$

$$\left.\frac{d}{dt}C(t)\right|_{t=0} = \frac{M'(0)}{M(0)} = \frac{M_1}{1} = \mu$$

$$\left.\frac{d^2}{dt^2}C(t)\right|_{t=0} = \frac{M''(0)M(0) - (M'(0))^2}{M^2(0)}$$

$$= \frac{M_2 - (M_1)^2}{1} = \sigma^2$$

The first two derivatives of $C_x(t)$ evaluated at $t=0$ directly give the mean and variance of *X*.

Example 8.5

Let X be a random variable with $M(t) = \frac{1}{4}(1+e^t)^2$. Find the cumulant generating function.

Solution: $C(t) = \ln M(t) = \ln \frac{1}{4} + 2\ln(1+e^t)$

$$\frac{d}{dt}C(t) = \frac{2e^t}{1+e^t}$$

$$\frac{d^2}{dt^2}C(t) = \frac{2e^t}{(1+e^t)^2}$$

$$\left.\frac{d}{dt}C(t)\right|_{t=0} = \frac{2}{2} = 1 = \mu$$

$$\left.\frac{d^2}{dt^2}C(t)\right|_{t=0} = \frac{2}{2^2} = \frac{1}{2} = \sigma^2$$

Example 8.6

A discrete random variable X has *mgf*

$$M(t) = \exp\left[2\left(e^t - 1\right)\right]$$

Find the *c.g.f* of X and use it to evaluate μ and σ^2.

Solution: $C(t) = \ln M(t) = \ln \exp\left[2\left(e^t - 1\right)\right] = 2\left[e^t - 1\right]$

$$\frac{d}{dt}C(t) = 2e^t$$

$$\frac{d^2}{dt^2}C(t) = 2e^t$$

$$\left.\frac{d}{dt}C(t)\right|_{t=0} = 2 = \mu$$

$$\left.\frac{d^2}{dt^2}C(t)\right|_{t=0} = 2 = \sigma^2$$

Example 8.7

(1) The m.g.f. of a normal random variable X is

$$M(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$$

Find the c.g.f and hence it's mean and variance

(2) (i) The mgf of a binomial random variable X is

$$M(t) = (q + pe^t)^n$$

(ii) For a Poisson random variable X is

$$M(t) = e^{\lambda(e^t - 1)}$$

Find the means and variances of X.

Solution:

(1) $M(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$

$$C(t) = InM(t) = \mu t + \frac{1}{2}\sigma^2 t^2$$

$$\frac{d}{dt}C(t) = \mu + \sigma^2 t$$

$$\frac{d^2}{dt^2}C(t) = \sigma^2$$

$$\frac{d}{dt}C(t)\Big/_{t=0} = \mu$$

(2i) $M(t) = (q + pe^t)^n$

$$C(t) = \text{In } M(t) = n \text{ In } (q + pe^t)$$

$$\frac{d}{dt}C(t) = \frac{npe^t}{q + pe^t}$$

$$\frac{d^2}{dt^2}C(t) = \frac{(q+pe^t)npe^t - npe^t(pe^t)}{(q+pe^t)^2}$$

$$\left.\frac{d}{dt}C(t)\right|_{t=0} = np$$

$$\left.\frac{d^2}{dt^2}C(t)\right|_{t=0} = \frac{npqe^t - np^2e^{2t} - np^2e^{2t}}{(q+pe^t)^2} = npq$$

(2ii) $\quad C(t) \quad = \quad \text{In } M(t) \quad = \quad \lambda(e^t - 1)$

$$\frac{d}{dt}C(t) = \lambda e^t$$

$$\frac{d^2}{dt^2}C(t) = \lambda e^t$$

$$\left.\frac{d}{dt}C(t)\right|_{t=0} = \lambda$$

$$\left.\frac{d^2}{dt^2}C(t)\right|_{t=0} = \lambda$$

Example 8.8
Let X have an exponential distribution with *p.d.f*

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

Obtain the MGF of X , hence or otherwise find μ and σ^2.

Solution:

$$M(t) = \int_0^\infty e^{tx}\left(\frac{1}{\theta}\right)e^{-x/\theta}dx$$

$$= \lim_{b\to\infty}\int_0^b\left(\frac{1}{\theta}\right)e^{-(1-\theta t)x/\theta}dx$$

$$= \lim_{b\to\infty}\left[-\frac{e^{-(1-\theta t)x/\theta}}{1-\theta t}\right]_0^b$$

$$= \frac{1}{1-\theta t}, t < \frac{1}{\theta}$$

$$M'(t) = \frac{\theta}{(1-\theta t)^2}$$

$$M''(t) = \frac{2\theta^2}{(1-\theta t)^2}$$

$$\mu = M'(0) = \theta$$

$$\sigma^2 = M''(0) - [M'(0)]^2$$

$$= \theta^2$$

Example 8.9
Let X have the *pdf*

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

$$M(t) = \int_0^\infty e^{tx}xe^{-x}dx$$

$$= \lim_{b\to\infty}\int_0^b xe^{-(1-t)x}dx$$

$$= \lim_{b\to\infty}\left[-\frac{xe^{-(1-t)x}}{1-t}-\frac{e^{-(1-t)x}}{(1-t)^2}\right]_0^b$$

$$= \lim_{b\to\infty}\left[-\frac{be^{-(1-t)b}}{1-t}-\frac{e^{-(1-t)b}}{(1-t)^2}\right]+\frac{1}{(1-t)^2}$$

$$M(t)=\frac{1}{(1-t)^2}\text{, provided } \quad t<1$$

Summary
In this chapter, we have learnt the following:

(1) The importance of moment- generating function.
(2) How to obtain the means and variances of distributions from m.g.f.
(3) The properties of m.g.f.
(4) The notion of cumulant generating function.
(5) How moment-generating function relates to the cumulant generating function.
(6) How to derive cumulant generating function from the moment-generating function.
(7) Evaluating the means and variances of random variables using c.g.f. from a given m.g.f.

Post-Test

(1) Define the r^{th} moment about the mean.
(2) Obtain the 3rd and 4th moments and state their usefulness.
(3) Define the moment-generating function of a discrete random variable *X*.
(4) Give the proof of the remaining properties given above.
(5) Obtain the *m.g.f.* of the exponential distribution and hence or otherwise find the mean and variance.

(6) Find the *m.g.f.* when the *p.d.f.* of X is defined by

(i) $f(x)=\frac{1}{5}, \quad x=1,2,3$

(ii) $f(x)=1, \quad x=5$

(iii) $f(x)=\frac{5-x}{10}, \quad x=1,2,3,4.$

(7) A random variable X has moment-generating function

$$M_x(t)=\left(0.25+0.75\ell^t\right)^{12},$$

Find the cumulant generating function of X and use it to obtain the mean and variance of X.

(8) Find the moment-generating function, mean and variance of X if the p.d.f of X is

$$f(x)=\left(\frac{1}{2}\right)\left(\frac{2}{3}\right)^x, \quad x=1,2,3,4,...$$

From the m.g.f. obtained, find the *c.g.f* of X.

(9) Given the following *m.g.fs* of X, find the *c.g.fs* and hence or otherwise obtain μ and σ^2.

(i) $M(t)=\ell^{4\left(\ell^t-1\right)}$

(ii) $M(t)=\frac{0.3\ell^t}{1-0.7\ell^t}, \quad t<-\ln(0.7)$

(iii) $M(t)=0.3\ell^t+0.4\ell^{2t}+0.2\ell^{3t}+0.1\ell^{4t}$

9
Chebyshev's Inequality

Introduction

Limit theorems are basically concerned with finding approximations to statistics and/or distributions of statistics. Finding limits may not be attained without looking at some important inequalities. One of the most important inequalities used in probability is the *Chebyshev's inequality.* Loosely speaking, it asserts that if the variance of a random variable is small, then the probability that it takes a value far from its mean is also small. Note that the Chebyshev inequality does not require the random variable to be nonnegative.

A small standard deviation for a set of values indicates that these values are located close to the mean. Conversely, a large standard deviation reveals that the observations are widely scattered about the mean. The Russian mathematician P.L. Chebyshev (1821 – 1894) developed a theorem that allows us to determine the minimum proportion of the values that lie within a specified number of standard deviations of the mean. For example, according to Chebyshev's theorem, at least three of four values, or 75 percent, must lie between the mean plus two standard deviations and the mean minus two standard deviations $(\mu \pm 2\sigma)$. This relationship applies regardless of the shape of the distribution. Further, at least eight of nine values, or 99.9 percent, will lie between plus three standard deviations and minus three standard deviations of the mean $(\mu \pm 3\sigma)$.

Objectives

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

(1) Find bounds on probabilities based on moments;
(2) Find upper and lower bounds for certain probabilities; and
(3) Show that the inequality is valid for all distributions for which the standard deviation exists.

Pre-Test

(1) Have you ever come across the word 'inequality'?
(2) Mention the inequalities you have come across in the past.
(3) Define Markov's inequality.
(4) Define the word 'statistics'.
(5) Differentiate between statistics and parameter.

Contents

Chebyshev's Inequality

Many important inequalities exist which relate expectations and probabilities. A lot of these are variations on the basic inequality called Markov's inequality. Chebyshev's inequality will be used to show that the sample mean, $\bar{x}$, is a good statistic to estimate a population mean, μ; the relative frequency of success in n Bernoulli trials, x/n, is a good statistic for estimating p; the empirical distribution function, $F_n(x)$, can be used to estimate the theoretical distribution function $F(x)$. The effect of the sample size *n* on these estimates is discussed.

First, show that Chebyshev's inequality gives added significance to the standard deviation in terms of bounding certain probabilities. The inequality is valid for all distributions for which the standard deviation exists.

In general, the theorem states that for real numbers k, $k > 1$, at least $1 - 1/k^2$ of the values lie within k standard deviations of the mean (or average).

Table 9.1: Chebyshev's Theorem for Some Values of $k > 1$

At least this proportion of data:	**Lies within this interval:**	
	Population	**Sample**
$1 - 1/2^2 = 3/4$	$\mu \pm 2\sigma$	$\bar{y} \pm 2s$
$1 - 1/3^2 = 8/9$	$\mu \pm 3\sigma$	$\bar{y} \pm 3s$
$1 - 1/4^2 = 15/16$	$\mu \pm 4\sigma$	$\bar{y} \pm 4s$
$1 - 1/k^2$	$\mu \pm k\sigma$	$\bar{y} \pm ks$

Table 9.2: The Empirical Rule

Approximately this proportion of the data:	**Lies within this interval:**	
	Population	**Large Sample**
0.682	$\mu \pm 1\sigma$	$\bar{y} \pm 1s$
0.954	$\mu \pm 2\sigma$	$\bar{y} \pm 2s$
0.997	$\mu \pm 3\sigma$	$\bar{y} \pm 3s$

Note that the theorem is true for any population or sample. Although this theory gives only a lower bound for the proportion of the data within certain intervals, it is applicable to all data sets regardless of the shape of their distribution and regardless of their size. If a population or a large sample is symmetrical and mound shaped, an estimate is possible for the proportion of the data within certain intervals.

Theorem 9.1: (***Markov's Inequality***) If X is a random variable and $U(X)$ is a non-negative real-valued function, then for any positive constant $c > 0$,

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

Proof: If $A = \{x | U(x) \geq c\}$, then for a continuous random variable,

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

$$= \int_A U(x) f(x) dx + \int_{A^c} U(x) f(x) dx$$

$$\geq \int_A U(x) f(x) dx$$

$$\geq \int_A c f(x) dx$$

$$= cP[X \in A]$$

$$= cP[U(X) \geq c]$$

Theorem 9.2: (Chebyshev's Inequality) If the random variable X has a mean μ and variance σ^2, then for every $k \geq 1$

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

Proof: Let $f(x)$ denote the pdf of X, then

$$\sigma^2 = E[(X-\mu)^2] = \sum_{x\in R}(x-\mu)^2 f(x)$$
$$= \sum_{x\in A}(x-\mu)^2 f(x) + \sum_{x\in A'}(x-\mu)^2 f(x)$$

where, $A = \{x; |x-\mu| \geq k\sigma\}$

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

However, in A, $|x-\mu| \geq k\sigma$; so

$$\sigma^2 \geq \sum_{x\in A}(k\sigma)^2 f(x) = k^2\sigma^2 \sum_{x\in A} f(x)$$

But $\sum_{x\in A} f(x) = P(X \in A)$ and thus

$$\sigma^2 \geq k^2\sigma^2 P(X \in A) = k^2\sigma^2 P(|X-\mu| \geq k\sigma)$$

That is $P(|X-\mu| \geq k\sigma) \leq \dfrac{\sigma^2}{k^2\sigma^2} = \dfrac{1}{k^2}$

Corollary If $\varepsilon = k\sigma$, then

$$P(|X-\mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}$$

In words, Chebyshev's inequality states that the probability that X differs from its mean by at least k standard deviations is less than or equal to $1/k^2$. It follows that the probability that X differs from its mean by less than k standard deviations is at least $1-1/k^2$. That is

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

From the corollary, it also follows that

$$P(|X-\mu|<\varepsilon)\geq 1-\frac{\sigma^2}{\varepsilon^2}$$

Thus, Chebyshev's inequality can be used as a bound for certain probabilities. However, in many instances, the bound is not very close to the true probability.

Example 9.1
If it is known that X has a mean of 25 and a variance of 16, then, since $\sigma=4$, a lower bound for $P(17<X<33)$ is given by

$$P(17<X<33)=P(|X-25|<8)$$
$$=P(|X-\mu|<2\sigma)\geq 1-\frac{1}{4}=0.75$$

and an upper bound for $P(|X-25|\geq 12)$ is found to be

$$P(|X-25|\geq 12)=P(|X-\mu|\geq 3\sigma)\leq\frac{1}{9}$$

Example 9.2
If X is a random variable with mean 33 and variance 16, use Chebyshev's inequality to find

(a) A lower bound for $P(23<X<43)$
(b) An upper bound for $P(|X-33|\geq 14)$

Solution:

(a) $P(|X-33|<10) = P(|X-\mu|<2.5\sigma) \geq 1-\dfrac{1}{2.5^2}$

$= \quad 1-0.16$

$= \quad 0.84$

(b) $P(|X-33|\geq 14) = P(|X-\mu|\geq 3.5\sigma) \leq \dfrac{1}{3.5^2} = \dfrac{1}{12.25}$

$= \quad 0.082$

Example 9.3

Let X denote the outcome when rolling a fair die. Then $\mu = 7/2$ and $\sigma^2 = 35/12$. (Note that, the maximum deviation of X from μ equals $5/2$) Express this deviation in terms of number of standard deviations; that is find k where $k\sigma = 5/2$. Determine a lower bound for $P(|X-3.5|\geq 2.5)$.

Solution: $\quad k\sigma = 5/2$

$$k = \frac{2.5}{\sigma}$$

but, $\quad \sigma = \sqrt{35/12}$

$$\therefore k = \frac{2.5}{\sqrt{35/12}} = 1.464$$

$$P(|X-3.5|<2.5) = P(|X-\mu|<1.4645\sigma) \geq 1-\frac{1}{1.464^2}$$

$= \quad 1.0.467$

$= \quad 0.5328 \simeq 0.533$

If Y is the number of successes in n Bernoulli trials with probability p of success on each trial, then Y is $b(n,p)$. Furthermore,

Y/n gives the relative frequency of success, and when p is unknown, Y/n can be used as an estimate of p. To gain some insight into the closeness of Y/n to p, we shall use Chebyshev's inequality with $\varepsilon > 0$.

$$p\left(\left|\frac{Y}{n} - p\right| \geq \varepsilon\right) = p\left(|Y - np| \geq n\varepsilon\right)$$

$$= \quad p\left(|Y - np| \geq \frac{\sqrt{n}\varepsilon}{\sqrt{pq}}\sqrt{npq}\right)$$

However, $\mu = np$ and $\sigma = \sqrt{npq}$ are the mean and the standard deviation of Y so that, with $k = \sqrt{n}\varepsilon \big/ \sqrt{pq}$, we have

$$p\left(\left|\frac{Y}{n} - p\right| \geq \varepsilon\right) = p\left(|Y - \mu| \geq k\sigma\right) \leq \frac{1}{k^2} = \frac{pq}{n\varepsilon^2} \text{ -------- } (9.1)$$

or, equivalently,

$$p\left(\left|\frac{Y}{n} - p\right| < \varepsilon\right) \geq 1 - \frac{pq}{n\varepsilon^2}$$

when p is completely unknown, $pq = p(1-p)$ is a maximum when $p = \frac{1}{2}$ in order to find a lower bound for the probability in equation (9.1). That is

$$1 - \frac{pq}{n\varepsilon^2} \geq 1 - \frac{\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)}{n\varepsilon^2}$$

For example, if $\varepsilon = 0.05$ and $n = 200$

$$p\left(\left|\frac{Y}{400} - p\right| < 0.05\right) \geq 1 - \frac{\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)}{200(0.0025)} = 0.75$$

Note that Chebyshev's inequality is applicable to all distributions with a finite variance, thus the bound is not always a tight one; i.e, the bound is not necessarily close to the true probability.

In general, it should be noted that with fixed $\varepsilon > 0$ and $0 < p < 1$,

$$\lim_{n\to\infty} p\left(\left|\frac{Y}{n} - p\right| < \varepsilon\right) \geq \lim_{n\to\infty}\left(1 - \frac{pq}{n\varepsilon^2}\right) = 1$$

Since, the probability of every event is less than or equal to 1, then

$$\lim_{n\to\infty} p\left(\left|\frac{Y}{n} - p\right| < \varepsilon\right) = 1$$

That is, the probability that the relative frequency Y/n is within ε of p is close to 1 when n is large enough. Thus, this is one form of the law of large numbers.

Summary
In this chapter, we have been able to discuss:

(1) The upper and lower bounds for probabilities.
(2) The theory of Chebyshev's inequality.
(3) A theorem on Chebyshev's inequality and provided the proof.

Post-Test

(1) If $E(X) = 17$ and $E(X^2) = 298$. Use Chebyshev's inequality to determine
   (i) A lower bound for $P(10 < X < 24)$.
   (ii) An upper bound for $P(|X - 17| \geq 16)$.

(2) State and prove the Chebyshev's inequality.

(3) If X is a random variable such that $E(X)=3$ and $E(X^2)=13$, use Chebyshev's inequality to determine a lower bound for the probability $P(-2 < X < 8)$.

(4) If Y is $b(n,0.5)$, give a lower bound for $P\left(\left|\frac{Y}{n}-0.5\right| < 0.08\right)$ when

(i) $n=100$
(ii) $n=500$
(iii) $n=1000$

10
Central Limit Theorem

Introduction

The relationship between the shapes of the population distribution and the sampling distribution of the mean can be summarized in what is often referred to as the most important theorem in statistics, namely, the central limit theorem. The central limit theorem is concerned with the probability distribution of sums of random variables as n, the number of terms in the sum, increases without bound. The central limit theorem is frequently relied on to justify the assumption of a normal probability distribution for any random variable whose value can be thought of as the accumulation of a large number of independent quantities.

Objective

At the end of this chapter, you should be able to provide an approximate distribution in cases where the exact distribution is unknown or intractable.

Pre-Test

(1) What is sampling distribution of means?
(2) What are limiting distributions?
(3) What do you understand by the central limit theorem?

Contents

Central Limit Theorem

Theorem: If $\bar{X}$ is the mean of a random sample $X_1, X_2, \ldots, X_n$ of size n from a distribution with a finite mean μ and a finite positive variance σ^2 then the distribution of

$$W = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma}$$

is $N(0,1)$ in the limit as $n \to \infty$

Proof: Let $M_x(t) = M(t)$

$$M_s(t) = (M(t))^n \text{ where } S = \sum X_i$$

$$M_w(t) = E[e^{tw}] = E\left[e^{t\left(\frac{S-n\mu}{\sqrt{n}\sigma}\right)}\right]$$

$$= e^{-\left(\frac{\sqrt{n}\mu}{\sigma}t\right)} E\left[e^{\frac{t}{\sqrt{n}\sigma}S}\right]$$

$$= e^{-\left(\frac{\sqrt{n}\mu}{\sigma}t\right)} \left(E\left(e^{\frac{t}{\sqrt{n}\sigma}X}\right)\right)^n$$

$$= e^{-\frac{\sqrt{n}\mu}{\sigma}t} \left(M\left(\frac{t}{\sqrt{n}\sigma}\right)\right)^n$$

$$Log_e\, M_w(t) = -\frac{\sqrt{n}\mu}{\sigma}t + n\log_e M\left(\frac{t}{\sqrt{n}\sigma}\right)$$

Note, $\log_e X = 1 + X + \frac{X^2}{2!} + \frac{X^3}{3!} + - - - -$

$$\log_e M_w(t) = 1 + M'(0)t + \frac{M''(0)t^2}{2} + - - - -$$

$$= 1 + \mu t + \frac{(\mu^2 + \sigma^2)}{2}$$

$$In\ M_w(t) = -\frac{\sqrt{n}\mu t}{\sigma} + nIn\left(1 + \frac{\mu t}{\sqrt{n}\sigma} + \frac{\mu^2 + \sigma^2}{2}\frac{t^2}{n\sigma^2} - - - -\right)$$

$$\log_e(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3}$$

$$\Rightarrow InM_w(t) = -\frac{\sqrt{n}\mu t}{\sigma} + n\left[\frac{\mu t}{\sqrt{n}\sigma} + \frac{(\mu^2+\sigma^2)t^2}{2n\sigma^2} - \frac{1}{2}\left(\frac{\mu t}{\sqrt{n}\sigma} + \frac{(\mu^2+\sigma^2)t^2}{2n\sigma^2}\right)^2\right]$$

$$= \frac{\sqrt{n}\mu t}{\sigma} + \frac{\sqrt{n}\mu t}{\sigma} + \frac{(\mu^2+\sigma^2)t^2}{2\sigma^2} - \frac{\mu^2 t^2}{2\sigma^2} - \frac{1}{2}\frac{\mu t^3(\mu^2+\sigma^2)}{\sqrt{n}\sigma^3}$$

$$= t^2\!\!\left/2\right. - \frac{\frac{1}{2}\mu t^3(\mu^2+\sigma^2)}{\sqrt{n}\sigma^3}$$

$$= t^2\!\!\left/2\right., \quad \text{since } N(0,1)$$

Example 10.1

Let $\bar{X}$ denote the mean of a random sample of size $n=15$ from the distribution whose *p.d.f* is $f(x) = \left(3/2\right)x^2, \quad -1 < x < 1$. Approximate $P(0.03 \le \bar{X} \le 0.15)$.

Solution:

$$f(x) = \left(3/2\right)x^2$$

$$E(X) = 3/2\int_{-1}^{1} x^3 dx = 3/2\left[\frac{x^4}{4}\right]_{-1}^{1} = 0$$

$$E(X^2) = 3/2\int_{-1}^{1} x^4 dx = 3/2\left[\frac{x^5}{5}\right]_{-1}^{1} = 3/10[1+1] = 6/10 = 3/5$$

$$V(X) = E(X^2) - \mu^2 = 3/5$$

$$P(0.03 \le \bar{X} \le 0.15) = P\left[\frac{0.03-0}{\sqrt{3/5}\big/\sqrt{15}} \le \frac{\bar{X}-0}{\sqrt{3/5}\big/\sqrt{15}} \le \frac{0.15-0}{\sqrt{3/5}\big/\sqrt{15}}\right]$$

$$= \quad P[0.15 \le W \le 0.75]$$

$$= \quad \phi(0.75) - \phi(0.15)$$

$$= \quad 0.7734 - 0.5596$$

$$= \quad 0.2138$$

OR $n\bar{X} = \sum X_i = Y \sim N\left(0, \tfrac{3}{5}n\right)$

$$= \quad Y - \; N(0,9)$$

$$P(0.03 \le \bar{X} \le 0.15) = P(0.03n \le n\bar{X} \le 0.15n)$$

$$= \quad P(0.45 \le Y \le 2.25)$$

$$= \quad P\left[\frac{0.45}{\sqrt{9}} \le W \le \frac{2.25}{\sqrt{9}}\right]$$

$$= \quad P(0.15 \le W \le 0.75)$$

$$= \quad \Phi(0.75) - \Phi(0.15)$$

$$= \quad 0.2138$$

Example 10.2: Uniform

Let $\bar{X}$ be the mean of a random sample of size 12 from the uniform distribution on the interval (0, 1). Approximate $P\left(\frac{1}{2} \le \bar{X} \le \tfrac{2}{3}\right)$

Solution: $E(X_i) = \tfrac{1}{2}$

$$V(X_i) = \tfrac{1}{12}$$

$$Y = \sum_{i=1}^{n} X_i; \quad \text{then} \quad Y \sim N\left(\frac{n}{2}, \frac{n}{12}\right)$$

$$P(6 \le Y \le 8) = P\left(\frac{6-6}{1} \le \frac{Y-6}{1} \le \frac{8-6}{1}\right)$$

$$= \quad P(0 \le W \le 2)$$

$$= \quad \Phi(2) - \Phi(0)$$

$$= \quad 0.9772 - 0.5$$

$$= \quad 0.4772$$

Example 10.3: Exponential

Let $\bar{X}$ be the mean of a random sample of size 36, from an exponential distribution with mean 3. Approximate $P(2.5 \le \bar{X} \le 4)$.

$$E(X_i) = \theta, \quad V(X_i) = \theta^2$$

$$Y = \sum X_i, \quad Y \sim N(108, 324)$$

$$P(2.5 \le \bar{X} \le 4) \equiv P(90 \le Y \le 144)$$

$$= \quad P\left[\frac{90-108}{18} \le \frac{Y-108}{18} \le \frac{144-108}{18}\right]$$

$$= \quad P[-1 \le W \le 2]$$

$$= \quad \Phi(2) - \Phi(-0)$$

$$= \quad 0.9772 - 0.158 \quad = \quad 0.8185$$

Normal Approximation to Binomial

Definition: Let $X_1, X_2, \ldots, X_n$ be a random sample from a Bernoulli distribution with mean $\mu = P$ and a variance $\sigma^2 = P(1-P)$, $0 < P < 1$. Then $Y = \sum_{1}^{n} X_i$ is $b(n, P)$

The CLT states that the distribution of

$$W = \frac{Y - nP}{\sqrt{nP(1-P)}} = \frac{\bar{X} - nP}{\sqrt{P(1-P)/n}} \quad \sim N(0,1) \text{ as } n \to \infty$$

is $\left(K - \tfrac{1}{2} < y < k + \tfrac{1}{2}\right)$ if $nP \geq 5, n(1-P) \geq 5$

Example 10.4a: Let X_i denote whether or not a randomly selected individual approves of the job the Chairman of their local government is doing. More specifically:

- Let $X_i = 1$, if the person approves of the job the Chairman is doing, with probability p
- Let $X_i = 0$, if the person does not approve of the job the Chairman is doing with probability $1 - p$

Then, recall that X_i is a Bernoulli random variable with mean:
$\mu = E(X) = (0)(1-p) + (1)(p) = p$ and variance:
$\sigma^2 = Var(X) = E\left[(X-p)^2\right] = p(1-p)$

Now, take a random sample of n people, and let:
$Y = X_1 + X_2 + \ldots + X_n$

Then Y is a binomial (n, p) random variable, $y = 0, 1, 2, \ldots, n$, with mean:
$\mu = np$ and variance $\sigma^2 = np(1-p)$

Now, let $n = 10$ and $p = ½$, so that Y is binomial (10, ½). What is the probability that exactly five people approve of the job the Chairman is doing?

Solution:
We can calculate the exact probability using the binomial table with $n = 10$ and $p = ½$.

$$P(Y=5)=P(Y\le 5)-P(Y\le 4)$$
$$=0.6230-0.3770$$
$$=0.2460$$

That is, there is a 24.6% chance that exactly five of the ten people selected approve of the job the Chairman is doing.

Note, however, that Y in the above example is defined as a sum of independent, identically distributed random variables. Therefore, as long as n is sufficiently large, we can use the Central Limit Theorem to calculate probabilities for Y. Specifically, the Central Limit Theorem is given by:

$$Z=\frac{Y-np}{\sqrt{np(1-p)}}\xrightarrow{d}N(0,1).$$

The Central Limit Theorem is a tool that allows using the normal distribution to approximate binomial probabilities.

Example 10.4b: Using the normal distribution to approximate some probabilities for Y. Again, what is the probability that exactly five people approve of the job the Chairman is doing?

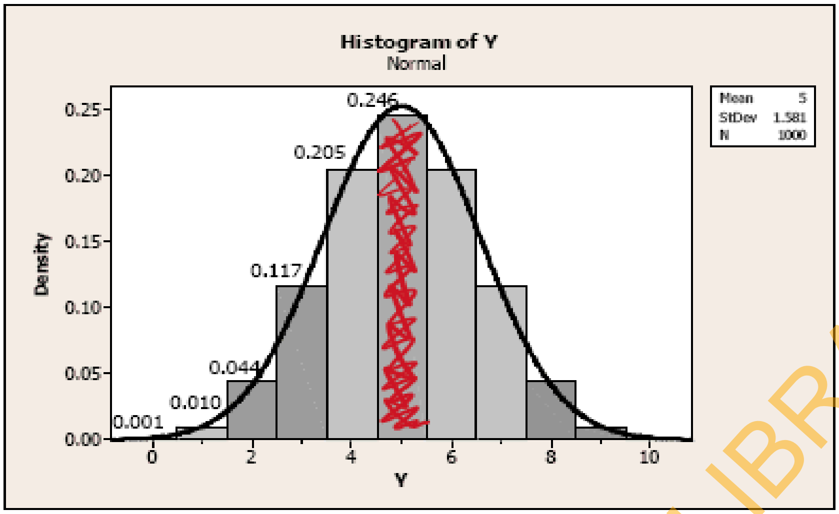
Solution:
First, recognize that the mean is:

$$\mu=np=10\left(\tfrac{1}{2}\right)=5$$

and the variance is:

$$\sigma^2=np(1-p)=10\left(\tfrac{1}{2}\right)\left(\tfrac{1}{2}\right)=2.5$$

Now, the diagram below is the graph of the binomial distribution with the rectangle corresponding to Y = 5 shaded:



It should be observed that we would benefit from making some kind of correction for the fact that we are using a continuous distribution to approximate a discrete distribution. Specifically, it seems that the rectangle Y = 5 really includes any Y greater than 4.5 but less than 5.5. That is:

$$P(Y=5)=P(4.5<Y<5.5)$$

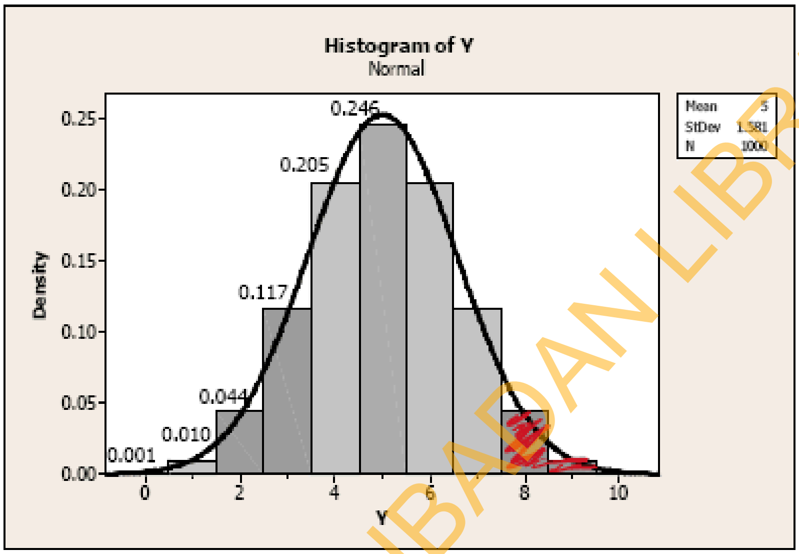
Such an adjustment is called a "***continuity correction***." Once the continuity correction has been made, the calculation reduces to a normal probability calculation

$$P(Y=5)=P(4.5<Y<5.5)=P\left(\frac{4.5-5}{\sqrt{2.5}}<Z<\frac{5.5-5}{\sqrt{2.5}}\right)$$
$$=P(-0.32<Z<0.32)$$
$$=0.6255-0.3745=0.251$$

Now, recall that we previously used the binomial distribution to determine the probability that Y = 5 is exactly 0.246. Here, we used the normal distribution to determine the probability that Y = 5 is approximately 0.251. That's not too shabby of an approximation, in light of the fact that the sample size of n = 10 is relatively small.

Example 10.4c: What is the probability that more than 7, but at most 9, of the ten people sampled approve of the job the Chairman is doing?

Solution:
Looking at a graph of the binomial distribution with the area corresponding to $7 < Y \leq 9$ shaded in red:



Note that the following continuity correction should be made:

$$P(7 < Y \leq 9) = P(7.5 < Y < 9.5)$$

Now again, once we've made the continuity correction, the calculation reduces to a normal probability calculation:

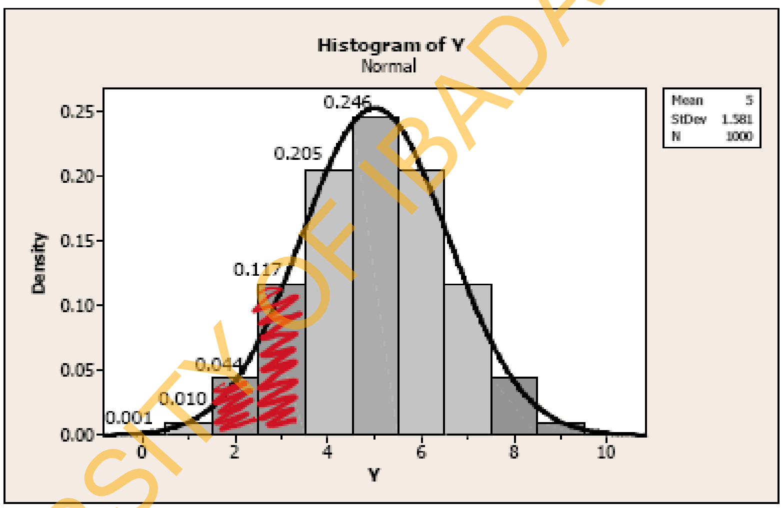
$$\begin{aligned} P(7 < Y \leq 9) &= P(7.5 < Y < 9.5) \\ &= P\left(\frac{7.5-5}{\sqrt{2.5}} < Z < \frac{9.5-5}{\sqrt{2.5}}\right) \\ &= P(1.58 < Z < 2.85) \\ &= 0.9778 - 0.9429 = 0.0549 \end{aligned}$$

It interesting to note that the approximate normal probability is quite close to the exact binomial probability as shown below;
The approximate probability is 0.0549, whereas the following calculation shows that the exact probability (using the binomial table with n = 10 and p = ½) is 0.0537:

$$\begin{aligned} P(7 < Y \leq 9) &= P(Y \leq 9) - (Y \leq 7) \\ &= 0.9990 - 0.9453 \\ &= 0.0537 \end{aligned}$$

Example 10.4d: What is the probability that at least 2, but less than 4, of the ten people sampled approve of the job the Chairman is doing?

Solution:
If we look at a graph of the binomial distribution with the area corresponding to $2 \leq Y < 4$ shaded in red:



we should see that we'll want to make the following continuity correction:

$$P(2 \leq Y < 4) = P(1.5 < Y < 3.5)$$

Again, once we've made the continuity correction, the calculation reduces to a normal probability calculation:

$$\begin{aligned} P(2 \le Y < 4) &= P(1.5 < Y < 3.5) \\ &= P\left(\frac{1.5-5}{\sqrt{2.5}} < Z < \frac{3.5-5}{\sqrt{2.5}}\right) \\ &= P(-2.21 < Z < -0.95) \\ &= 0.1711 - 0.0136 = 0.1575 \end{aligned}$$

By the way, the exact binomial probability is 0.1612, as the following calculation illustrates:

$$\begin{aligned} P(2 \le Y < 4) &= P(Y \le 3) - (Y \le 1) \\ &= 0.1719 - 0.0107 \\ &= 0.1612 \end{aligned}$$

The following comments are worth noting before we close this discussion of the normal approximation to the binomial.

(i) First, to discuss what "sufficiently large" means in terms of when it is appropriate to use the normal approximation to the binomial.

The general rule of thumb is that the sample size n is "sufficiently large" if:

$$np \ge 5 \quad \textbf{and} \quad n(1-p) \ge 5$$

For example, in the above example, in which p = 0.5, the two conditions are met if:

$$np = n(0.5) \ge 5 \quad \textbf{and} \quad n(1-p) = n(0.5) \ge 5$$

Now, both conditions are true if:

$$n \ge 5(10/5) = 10$$

Because the sample size of the example above was at least 10 (well, barely!), this is why our approximations were quite close to the exact probabilities. In general, the farther p is away from 0.5, the larger the sample size n is needed. For example, suppose p = 0.1. Then, the two conditions are met if:

$$np = n(0.1) \geq 5 \quad \textbf{and} \quad n(1-p) = n(0.9) \geq 5$$

Now, the first condition is met if:

$$n \geq 5(10) = 10$$

And, the second condition is met if:

$$n \geq 5(10/9) = 5.5$$

That is, the only way both conditions are met is if $n \geq 50$. So, in summary, when $p = 0.5$, a sample size of $n = 10$ is sufficient. But, if $p = 0.1$, then we need a much larger sample size, namely $n = 50$.

(i) In reality, the Central Limit Theorem is often applied to the sum of independent Bernoulli random variables to help draw conclusions about a true population proportion p. If we take the Z random variable that we've been dealing with above, and divide the numerator by n and the denominator by n (and thereby not changing the overall quantity), we get the following result:

$$Z = \frac{\sum X_i - np}{\sqrt{np(1-p)}} = \frac{\widehat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \xrightarrow{d} N(0,1)$$

The quantity:

$$\widehat{p} = \frac{\sum_{i=1}^{n} X_i}{n}$$

that appears in the numerator is the "**sample proportion**," that is, the proportion in the sample meeting the condition of interest (approving of the Chairman's job, for example).

Normal Approximation to Poisson

Just as the Central Limit Theorem can be applied to the sum of independent Bernoulli random variables, it can be applied to the sum of independent Poisson random variables. Suppose Y denotes the number of events occurring in an interval with mean λ and vaiance λ. Now, if $X_1, X_2, \ldots, X_\lambda$ are independent Poisson random variables with mean 1, then:

$$Y = \sum_{i=1}^{\lambda} X_i$$

is a Poisson random variable with mean λ. So, now that we have written Y as a sum of independent, identically distributed random variables, we can apply the Central Limit Theorem. Specifically, when λ is sufficiently large:

$$Z = \frac{Y-\lambda}{\sqrt{\lambda}} \xrightarrow{d} N(0,1)$$

This result will be used to approximate Poisson probabilities using the normal distribution.

Example 10.5

Let Y be $b(36, \frac{1}{2})$. Then $np = 36 \times \frac{1}{2} = 18 > 5$ and $n(1-P) > 5$

$$P(12 < Y \le 18) = P(12.5 \le Y \le 18.5)$$

$$= P\left(\frac{12.5-18}{\sqrt{9}} \le \frac{Y-18}{\sqrt{9}} \le \frac{18.5-18}{\sqrt{9}}\right)$$

$$= \Phi(0.167) - \Phi(-1.833)$$

$$= 0.5329$$

Note that 12 was increased to 12.5 because $P(Y = 12)$ is not included in the desired probability.

Example 10.6

Let Y have the binomial distribution of $b\left(10, \tfrac{1}{2}\right)$.

Approximate $P\left(3 \le Y < 6\right)$.

$P\left(3 \le Y < 6\right) = P\left(2.5 \le Y \le 5.5\right)$ because $P\left(Y = 6\right)$ is not included in the probability.

$$P\left(\frac{2.5-5}{\sqrt{10/4}} \le \frac{Y-5}{\sqrt{10/4}} \le \frac{5.5-5}{\sqrt{10/4}}\right)$$

$$= \Phi(0.316) - \Phi(-1.581)$$

$$= 0.6240 - 0.0570$$

$$= 0.5670$$

Summary

In this chapter, we have discussed the following:

(1) The theory of central limit theory.

(2) To approximate exact probability distributions for sums of independent random variables.

(3) The central limit theorem also gives a good approximation when the number of random variables summed together is large

Post-Test

(1) A nursery man plants 115 cuttings of ivy in every flat he prepares. Assume the probability that an individual cutting will develop roots is 0.9 and approximate the probability that the average number of rooted cuttings (per flat) in 50 flats is less than 100.

(2) A big city car dealer opens 365 days per year; the number of sales he makes per day is a Poisson random variable with parameter $\mu = 2$, independently from one day to another. Let Y be the number of sales he makes in a year, approximate

(i) $P(Y \geq 700)$

(ii) $P(\leq 800)$

(iii) $P(700 \leq Y \leq 800)$

(3) Let $Y = X_1 + X_2 + ... + X_{15}$ be the sum of a random sample of size 15 from the distribution whose p.d.f. is $f(x) = \frac{3}{2}x^2$, $-1 < x < 1$. Approximate $P(-0.3 \leq \overline{X} \leq 1.5)$

(4) Let $\overline{X}$ be the mean of a random sample of size 36 from an exponential distribution with mean 3. Approximate $P(2.5 \leq \overline{X} \leq 4.0)$.

11
Joint Probability Density Functions

Introduction

Up till now, we have only examined probabilities (outcomes) as a function of one variable. We have also spent time studying the concept of a random variable and have studied some simple models that led to several other frequently used probability distributions. The random variables considered so far are one dimensional, because the observed value for a random variable can be thought of as a single point on a real line. In almost all applications, random variables do not occur singly. We need to develop tools necessary to describe the behavior of two, three or more random variables simultaneously. For example, the hardness and tensile strength of a manufactured piece of steel may be of interest and so a *p.d.f.* may be necessary for experimental outcome. In order to deal with situations such as these, we will extend some definitions as well as give new ones.

We have discussed in chapter five, how expected values can be used to summarize or describe various aspects of one-dimensional probability distributions. These concepts can be extended to the case of two-dimensional variables. Just as was done in chapter four, in this chapter, we will discuss measures that describe the "middle" or the "spread" of probability distribution involving two random variables.

Objectives

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

(1) Define rules that would associate two numbers, or three numbers or even n numbers with experimental outcomes. These rules are examples of what we call two-dimensional, three-dimensional or in general n-dimensional random variables;

(2) Handle problems involving two-dimensional random variables;
(3) Discuss the concepts of joint and marginal probability density functions; and
(4) Discuss when two random variables are either independent or dependent.
(5) Compute the conditional probability density functions of a random variable given another random variable.

Pre-Test

(1) Define probability density function of a random variable X.
(2) Enumerate the properties of a probability density function.
(3) Describe what you understand by a 'two-dimensional' random variable.
(4) Define the conditional probability of an event.
(5) Define independent event.

Contents

Joint Probability Density Function

Definition 11.1

Let X *and* Y be two random variables defined on a discrete probability space. Let $\mathbb{R}$ denote the corresponding two-dimensional space of X *and* Y, the two random variables of the discrete type. The probability that $X = x$, and $Y = y$ is denoted by $f(x, y) = P[X = x,\ Y = y]$. The function $f(x, y)$ is called the joint *p.d.f.* of X *and* Y and has the following properties:

(a) $0 \le f(x, y) \le 1$

(b) $\sum\sum_{(x,y)\leftarrow R} f(x, y) = 1$

(c) $P[(X, Y) \in A] = \sum_{(x,y)\in A} f(x, y)$, where A is a subset of the space $\mathbb{R}$.

Example 11.1
Roll a pair of unbiased die. For each of the 36 sample points with probability $\frac{1}{36}$, Let X denotes the smaller and Y the larger outcome on the dice.

Definition 11.2
The joint probability density function of two continuous-type random variables is an integrable function $f(x, y)$ with the following properties:

(a) $f(x, y) \geq 0$.

(b) $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)dxdy = 1$.

(c) $P[(x, y) \in A] = \iint_{A} f(x, y)dxdy$,

where, $\{(X, Y) \in A\}$ is an event defined in the plane. Property (c) implies that $P[(X, Y) \in A]$ is the volume of the solid over the region A in the $xy - plane$ and bounded by the surface $z = f(x, y)$.

The Marginal Probability Density functions
Definition 11.3
Let X *and* Y have the joint *p.d.f* $f(x, y)$ with space $\mathbb{R}$. The p.d.f of X alone, called the marginal p.d.f of X is defined by

$$f_1(x) = \sum_{y} f(x, y) \qquad x \in \mathbb{R}_1$$

the summation is taken over all possible y values for each given x in the space $\mathbb{R}_1$.

The marginal *p.d.f* of Y is given by

$$f_2(y) = \sum_{x} f(x, y) \qquad y \in \mathbb{R}_2$$

where the summation is taken over all possible x values for each given y in the space $\mathbb{R}_2$.
The random variables X *and* Y are independent if and only if

$$f(x,y) \equiv f_1(x) f_2(y) \qquad x \in \mathbb{R}_1, y \in \mathbb{R}_2$$

otherwise X *and* Y are said to be dependent.

Example 11.2
Let the joint *p.d.f* of X *and* Y be defined by

$$f(x,y) = \frac{x+y}{21}, \; x = 1, 2, 3, \; y = 1, 2$$

Obtain the marginal *p.d.f.*'s of X *and* Y.

Solution:

$$f_1(x) = \sum_y f(x,y) = \sum_{y=1}^{2} \frac{x+y}{21}$$

$$= \frac{x+1}{21} + \frac{x+2}{21} = \frac{2x+3}{21}, \; x = 1, 2, 3;$$

$$f_2(y) = \sum_x f(x,y) = \sum_{x=1}^{3} \frac{x+y}{21} = \frac{6+3y}{21}, \; y = 1, 2.$$

Definition 11.4
Let $f(x,y)$ be the *p.d.f.* of X *and* Y then the respective marginal *p.d.f*'s of continuous-type random variables X *and* Y are given by

$$f(x) = \int_{-\infty}^{\infty} f(x,y)dy, \; x \in \mathbb{R}_1$$

$$f(y) = \int_{-\infty}^{\infty} f(x,y)dx, \; y \in \mathbb{R}_2$$

where $\mathbb{R}_1$ and $\mathbb{R}_2$ are the spaces of X *and* Y.

Example 11.3

Let $f(x,y) = 2\ell^{-x-y}$, $0 \le y \le \infty$ be the joint *p.d.f* of X *and* Y. Find $f(x)$ and $f(y)$, the marginal *p.d.f*'s of X *and* Y respectively.

Solution:

$$f(x) = \int_x^\infty f(x,y)dy = \int_x^\infty 2e^{-x-y}dy = \frac{2e^{-x-y}}{-1}\Bigg|_x^\infty$$

$$= 2e^{-x-y} \qquad\qquad x < y < \infty$$

$$= 2e^{-2x}$$

$$f(y) = \int_0^y f(x,y)dx = \int_0^y 2e^{-x-y}dx = \frac{2e^{-x-y}}{-1}\Bigg|_0^y - 2e^{-2y} + 2e^{-y}$$

$$= \quad 2e^{-y} - 2e^{-2y}$$

$$= \quad 2e^{-y}(1-e^2) \qquad\qquad 0 < y < \infty\, 0$$

Example 11.4

Let $f(x,y) = \frac{1}{4}$, $0 \le x \le 2$, $0 \le y \le 2$ be the joint *p.d.f* of X *and* Y. Find;

(i) $f(x)$;

(ii) $f(y)$; the marginal probability density functions.

(iii) Are the two random variables independent?

Solution:

(i) $f(x) = \int_0^2 f(x,y)dy = \int_0^2 \frac{1}{4}dy = \frac{1}{4}y\Big|_0^2 = \frac{2}{4} = \frac{1}{2} \qquad 0 \le x \le 2$

(ii) $f(y) = \int_0^2 f(x,y)dx = \int_0^2 \frac{1}{4}dx = \frac{1}{4}x\Big|_0^2 = \frac{2}{4} = \frac{1}{2} \qquad 0 \le y \le 2$

(iii) $f(x)f(y) = \frac{1}{2} \text{ x } \frac{1}{2} = \frac{1}{4}$, Yes, the two random variables are independent.

Conditional Distributions and Mathematical Expectations

Let X *and* Y have a joint discrete distribution with *p.d.f* $f(x, y)$ on space $\mathbb{R}$ Also let $f_1(x)$ and $f_2(y)$ be the marginal probability density functions with spaces $\mathbb{R}_1$ and $\mathbb{R}_2$ respectively. Let event $A = \{X = x\}$ and event $B = \{Y = y\}$, $(x, y \in R)$.

Thus $A \cap B = \{X = x, y = y\}$. Since

$$P(A \cap B) = P(X = x, Y = y) = f(x, y)$$

and

$$P(B) = P(Y = y) = f_2(y) > 0 \text{ (since } y \in R_2\text{)},$$

The conditional probability of event A given B is

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{f(x, y)}{f_2(y)}.$$

The following definition is then apparent.

Definition 11.5
The conditional probability density function of X, given that $Y = y$, is defined by

$$g(x|y) = \frac{f(x, y)}{f_2(y)}, \text{provided that } f_2(y) > 0$$

Similarly, the conditional probability density function of Y, given $X = x$, is given by

$$h(y|x) = \frac{f(x, y)}{f_1(x)}, \text{provided that } f_1(x) > 0$$

Moreover, since $h(y|x) \geq 0$, if we sum $h(y|x)$ over y for fixed x, we obtain

$$\sum_y h(y|x) = \sum_y \frac{f(x, y)}{f_1(x)} = \frac{f_1(x)}{f_1(x)} = 1$$

Thus $h(y|x)$ satisfies the conditions of a probability density function, so the conditional probability can be computed as

$$P(a < Y < b | X = x) = \sum_{\{y: a<y<b\}} h(y|x)$$

and the corresponding conditional expectations are given as

$$E[u(Y) | X = x] = \sum_{y} u(y) h(y|x)$$

Written compactly, the conditional mean and conditional variance of Y given $X = x$ are given respectively as

$$\mu_{Y|x} = E(Y|x) = \sum_{y} y h(y|x),$$

and

$$\sigma^2_{Y|x} = E\{[Y - E(Y|x)]^2 | x\} = \sum_{y} [Y - E(Y|x)]^2 h(y|x),$$

which is alternatively written as

$$\sigma^2_{Y|x} = E(Y^2|x) - [E(Y|x)]^2.$$

Similar expressions can be used for conditional mean and conditional variance for X given $Y = y$.

The above definitions also hold for continuous random variables. For continuous random variables, X *and* Y, with joint *p.d.f.* $f(x, y)$ and marginal *p.d.f's* $f_1(x)$ and $f_2(y)$, respectively. The conditional *p.d.f.*, mean and variance of Y, given $X = x$, are, respectively,

$$h(y|x) = \frac{f(x, y)}{f_1(x)}, \text{ provided } f_1(x) > 0$$

$$E(Y|x) = \int_{-\infty}^{\infty} yh(y|x)dy\,,$$

and

$$Var(Y|x) = E\{[Y - E(Y|x)]^2 |x\}$$

$$= \int_{-\infty}^{\infty} [y - E(Y|x)]^2 h(y|x)dy$$

$$= E(Y^2|x) - [E(Y|x)]^2.$$

Expressions for conditional distribution of X *given* $Y = y$ are similar.

Example 11.5
Let X *and* Y have the joint *p.d.f.*

$$f(x, y) = \frac{x+y}{32}, \qquad x = 1, 2, \quad y = 1, 2, 3, 4$$

Find

(i) $g(x|y)$
(ii) $h(y|x)$
(iii) $P(1 \le Y \le 3|X = 1)$,
(iv) $P(Y \le 2|X = 2)$
(v) $P(X = 2|Y = 3)$
(vi) $E(Y|X = 1)$
(vii) $Var(Y|X = 1)$.

Solution:

(i) $g(x|y) = \dfrac{f(x,y)}{f(y)} = \dfrac{(x+y)/32}{(2y+3)/32} = \dfrac{x+y}{2y+3}, x = 1,2$

(ii) $h(y|x) = \dfrac{f(x,y)}{f(x)} = \dfrac{(x+y)/32}{(4x+10)/32} = \dfrac{x+y}{4x+10}, y = 1,2,3,4$

(iii) $P(1 \le Y \le 3|X = 1) = \sum_{y=1}^{3}\left(\dfrac{1+y}{14}\right) = \dfrac{1}{14}(2+3+4) = \dfrac{9}{14}$

(iv) $P(Y \le 2|X = 2) = \sum_{1}^{2}\dfrac{2+y}{18} = \dfrac{3}{18} + \dfrac{4}{18} = \dfrac{7}{18}$

(v) $P(X = 2|Y = 3) = \dfrac{5}{9}$

(vi) $E(Y|X = 1) = \sum_{1}^{4} yh(y|x=1) = \sum_{1}^{4} y\left(\dfrac{1+y}{14}\right)$

$$= \frac{1}{14}[1(2)+2(3)+3(4)+4(5)] = \frac{40}{14} = \frac{20}{7}$$

(vii) $Var(Y|X = 1) = E(Y^2|X = 1) - [E(Y|X = 1)]^2 = \sum_{1}^{4} y^2\left(\dfrac{1+y}{14}\right) - \left(\dfrac{20}{7}\right)^2$

$$= \frac{130}{14} + \frac{400}{49} = \frac{55}{49}$$

Definition 11.6

Let $X_1, X_2, ..., X_n$, be random variables of the discrete type having a joint distribution $f(x,y)$. If $U(X_1, X_2, ..., X_n)$ is a function of n random variables of the discrete type having a joint pdf. $f(x_1, x_2, ..., x_n)$ and space $\mathbb{R}$, then

$$E[U(X_1, X_2, ..., X_n)] = \sum\cdots\sum_{(x_1,...,x_n)} U(x_1, x_2, ..., x_n) f(x_1, x_2, ..., x_n)$$

if it exists and is called the mathematical expectation of $U(X_1, X_2, ..., X_n)$

(a) If $U_1(X_1, X_2, ..., X_n) = X_i$, then

$$E[U_1(X_1, X_2, ..., X_n)] = E(X_i) = \mu_i$$

is called the mean of $X_i, i = 1, 2, ..., n$

(b) If $U_2(X_1, X_2, ..., X_n) = (X_i - \mu_i)^2$, then

$$E[U_2(X_1, X_2, ..., X_n)] = E[(X_i - \mu_i)^2] = \sigma_i^2$$

is called the variance of $X_i, i = 1, 2, ..., n$

Example 11.6

Let the joint *p.d.f* of X_1 *and* X_2 be defined as

$$f(x_1, x_2) = \frac{3 - x_1 - x_2}{8}, \quad x_1 = 0,\ 1 \text{ and } x_2 = 0,\ 1$$

Find the $E(X_1 + X_2)$

Solution:

$$E(X_1 + X_2) = \sum_{x_2=0}^{1} \sum_{x_1=0}^{1} (x_1 + x_2) \frac{3 - x_1 - x_2}{8}$$

$$= \quad 0\left(\frac{3}{8}\right) + 1\left(\frac{2}{8}\right) + 1\left(\frac{2}{8}\right) + 2\left(\frac{1}{8}\right) = \frac{6}{8} = \frac{3}{4}$$

Example 11.7

Let X *and* Y have the joint *p.d.f*

$$f(x, y) = 2 \qquad 0 \le x \le y \le 1$$

Obtain

(i) Marginal *p.d.f's* of X *and* Y.

(ii) $E(X)$, $E(Y)$ and $E(Y^2)$.

Solution:

(i) $f(x) = \int_x^1 f(x,y)dy = \int_x^1 2dy = 2y\Big|_x^1 = 2(1-x)$

$f(y) = \int_0^y f(x,y)dy = \int_0^y 2dx = 2y\Big|_0^y = 2y$

(ii) $E(X) = \int_0^1 \int_x^1 2xdydx = \int_0^1 2x(1-x)dx = \frac{2x^2}{2} - \frac{2x^3}{3}\Big|_0^1 = 1 - \frac{2}{3} = \frac{1}{3}$

$$E(Y) = \int_0^1 \int_0^y 2ydxdy = \int_0^1 2y^2dy = \frac{2y^3}{3}\Big|_0^1 = \frac{2}{3}$$

$$E(Y^2) = \int_0^1 \int_0^y 2y^2dxdy = \int_0^1 2y^3dy = \frac{2y^4}{4}\Big|_0^1 = \frac{1}{2}$$

Summary

In this chapter, we learnt the following:

(1) We discussed that in many situations, we may be interested in observing two characteristics simultaneously by solving problems involving two-dimensional random variables.
(2) The concepts of joint and marginal probability density functions.
(3) The concepts of conditional probability density function of a random variable, given another random variable.
(4) The methods used in obtaining the conditional mean and conditional variance.
(5) The points given above are considered for both discrete and continuous random variables.

Post-Test

(1) Let the joint *p.d.f* of X *and* Y be defined by

$$f(x,y) = \frac{x+y}{32}, \qquad x = 1, 2, \qquad y = 1, 2, 3, 4$$

Find

(a) $f(x)$, the marginal p.d.f of X

(b) $f(y)$, the marginal p.d.f of Y

(c) $P(X > Y)$

(d) $P(Y = 2X)$

(e) $P(X + Y = 3)$

(f) $P(X \leq 3 - Y)$

(g) Are X *and* Y independent or dependent?

(2) Let $f(x,y) = \ell^{-x-y}$, $0 < x < \infty, 0 < y < \infty$, be the joint *p.d.f.* of X *and* Y. Argue that X *and* Y are independent and compute

(i) $P(X < Y)$

(ii) $P(X = Y)$

(iii) $P(X < 2)$

(iv) $P(0 < X < \infty, X/3 < Y < 3X)$

(v) $P(0 < X < \infty, 3X < Y < \infty)$

(3) Let $f_1(x) = \frac{1}{10}$, $x = 0,1,2,\ldots,9$, and

$h(y|x) = \frac{1}{(10-x)}$, $y = x, x+1, \ldots, 9$. Find

(i) $f(x,y)$

(ii) $f_2(y)$

(iii) $E(Y|x)$

(4) Let $f(x, y) = 1/8,\ 0 \leq y \leq 4,\quad 0 \leq x \leq y + 2$, be the joint *p.d.f.* of X and Y.

(iv) Find $f_1(x)$, the marginal *p.d.f.* of X.

(v) Determine $h(y|x)$, the conditional *p.d.f.* of Y, given $X = x$.

(vi) Compute $E(Y|x)$, the conditional mean of Y, given $X = x$.

(vii) Evaluate $Var(Y|x)$.

12
Transformations of Random Variables

Introduction
In this chapter, another important method of constructing models shall be discussed. To understand the depth of theory and application of statistics, transformations of random variables must be taught. An example of such transformations is that of the normal distribution that can be transformed into a standard normal distribution. We will first consider transformations of variables in one dimension in this current chapter. The joint transformations shall be discussed in the next chapter.

Let $u(x)$ be a real-valued function of a real variable x. If the equation $y = u(x)$ can be uniquely solved, say $x = w(y)$, then we have a one-to-one transformation. Discrete and continuous cases shall be considered separately.

Objectives
At the end of this chapter, you should be able to:
(1) Transform random variables of both the discrete and continuous types; and
(2) Determine the distribution of a random variable, *Y*, by the distribution of another random variable, *X*.

Pre-Test
(1) What do you understand by the terms 'monotonic increasing' and 'monotonic decreasing' functions.
(2) What is a 'one-to-one' transformation?
(3) Explain the term 'Jacobian'.

Contents
Transformations of Variables of the Discrete Type
An alternative method of finding the distribution of a function of one or more random variables is called the change-of-variable technique.

Let X have a Poisson *p.d.f.*

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

Let A denote the space $A = \{x: x = 0, 1, 2, \ldots\}$ so that $f(x) > 0$. Define a new random variable Y by $Y = 4X$.

Suppose we wish to find the *p.d.f* of Y by the change-of-variable technique.

Let $y = 4x$, we call $y = 4x$ a transformation from x to y, and we say that the transformation maps the space A on to the space $B = \{y: y = 0, 4, 8, 12, \ldots\}$. The space B is obtained by transforming each point in A in accordance with $y = 4x$ (i.e. every point in A corresponds to one and only one point in B and vice-versa). Therefore, any function $y = U(x)$ that maps a space A onto a space B such that there is a one-to-one correspondence between the points of A and those of B is called one-to-one transformation.

$$y = 4x \Rightarrow x = \frac{1}{4}y$$

The problem is to find $g(y)$ of the discrete type of random variable $Y = 4X$.

$$\text{Now } g(y) = P(Y = y) = \Pr\left(X = \frac{y}{4}\right) = \frac{\lambda^{y/4} e^{-\lambda}}{\left(y/4\right)!}, \quad y = 0, 4, 8, 12, ----$$

$$= 0, \text{ otherwise}$$

Example 12.1
Let X have the binomial *p.d.f*

$$f(x)=\begin{cases}\dfrac{3!}{x!(3-x)!}\left(\dfrac{2}{3}\right)^{x}\left(\dfrac{1}{3}\right)^{3-x}, & x=0,\ 1,\ 2,\ 3\\ 0 = otherwise\end{cases}$$

Find the *p.d.f* of $Y=X^2$.

Solution:

$y=U(x)=x^2$ maps $A=\{x:x=0,\ 1,\ 2,\ 3\}$ onto

$B=\{y:y=0,\ 1,\ 4,\ 9\}$

$$g(y)\frac{3!}{\sqrt{y}!(3-\sqrt{y})!}\left(\frac{2}{3}\right)^{\sqrt{y}}\left(\frac{1}{3}\right)^{3-\sqrt{y}} \qquad y = 0, 1, 4, 9$$

Example 12.2
Let $X \sim GEO(P)$, so that

$$f(x)=pq^{x-1} \qquad x = 1, 2, 3 ----$$

Find *p.d.f.* of $Y=X-1$,

Solution:

$$\begin{aligned} y &= U(x) = x-1, \text{ and } x = y+1\\ g(y) &= P(Y=y) = P(x=y+1)\\ &= Pq^{y+1-1}\\ &= Pq^{y} \quad ; y = 0, 1, 2 --- \end{aligned}$$

Example 12.3

Let X have a p.d.f. $f(x)\begin{cases}1/3, & x=1,\ 2,\ 3\\ 0, & otherwise\end{cases}$

Find the p.d.f. of $Y=2X+1$

Solution: $x = {(y-1)}/{2}$

$$g(y) = P\left(X = {(y-1)}/{2}\right) = \frac{1}{3}, \qquad y = 3, 5, 7$$

Example 12.4

Let X have the *p.d.f.* $f(x) = \begin{cases} \left(\frac{1}{2}\right)^2, & x = 1,\ 2,\ 3,\ \ldots \\ 0, & otherwise \end{cases}$

Find the *p.d.f* of $Y = X^3$.

Solution:

$Y = U(x) = x^3$. Since $y = x^3 \Rightarrow x = 3\sqrt{y}$

$$g(y) = \left(\frac{1}{2}\right)^{3\sqrt{y}}, \quad y = 1, 8, 27 ----$$

Transformations of Variables of the Continuous Type

Let X be a random variable of the continuous type having p.d.f $f(x)$. Let A be the one-dimensional space where $f(x) > 0$. Consider the random variable $Y = U(X)$, where $y = U(x)$ defines a one-to-one transformation that maps the set *A* onto the set *B*.

Let the inverse of $y = U(x)$ be denoted by $x = w(y)$, and let the derivative $\frac{dx}{dy} = w'(y)$ be continuous and not equal to zero for all points *y* in *B*.

Then the *p.d.f* of the random variable $Y = U(X)$ is given by

$$g(y) = \begin{cases} f[w(y)]\left|w'(y)\right|, & \mathrm{y} \in \mathrm{B} \\ 0 = otherwise \end{cases}$$

Where $\frac{dx}{dy} = w'(y)$ is the Jacobian (denoted by *J*) of the transformation.

Example 12.5
Let X be a random variable of the continuous type having *p.d.f*

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

Define the random variable Y by $Y = 8X^3$

Solution: $g(y) = f(w(y))|J|$

$y = 8x^3$ is a transformation from *x* to *y*.

$$\Rightarrow x = \sqrt[3]{\frac{y}{8}} = \frac{1}{2}\sqrt[3]{y}$$

$$= \frac{1}{2}y^{1/3}$$

$$|J| = \frac{dx}{dy} = \frac{1}{6}y^{-2/3}$$

$$g(y) = y^{1/3} \text{ x } \frac{1}{6}y^{-2/3}$$

$$= \frac{1}{6}y^{-1/3}$$

$$= \frac{1}{6y^{1/3}}, \quad 0 < y < 8$$

since, $0 < \frac{1}{2}y^{1/3} < 1 \Rightarrow 0 < y^{1/3} < 2 \Rightarrow 0 < y < 8$

Example 12.6
Let X have the *p.d.f*

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

Find the *p.d.f* of $Y=-2\ln X$.

Solution: The transformation $y=U(x)=-2\ln x$, so that $x=w(y)=e^{-y/2}$

$$J=\frac{dx}{dy}=w(y)=-\frac{1}{2}e^{-y/2}$$

The *p.d.f.* of Y is

$$g(y)=|J|\,f(w(y))=\frac{1}{2}e^{-y/2}.1=\frac{1}{2}e^{-y/2},0<y<\infty$$

Example 12.7
Let the *p.d.f* of X be defined by $f(x)=x^3/4$, $0 < x < 2$. Find the *p.d.f* of $Y=X^2$.

Solution: The transformation $y = U(x) = x^2$, so that $x=\sqrt{y}$ or $y^{1/2}=w(y)$

$$g(y)=f(w(y))|J|$$

$$J=\frac{dx}{dy}=1/2\,y^{-1/2}; \qquad f(w(y))=f\left(y^{1/2}\right)=\frac{y^{3/2}}{4}$$

$$g(y)=\frac{y^{3/2}}{4}\text{x}\ \frac{y^{-1/2}}{2}$$

$$= \quad y/8,\quad 0<y<4 \text{ since, } 0<x<2$$

$$0< \ y^{1/2} \ <2 \ \Rightarrow 0<y<4$$

Example 12.8

Let the *p.d.f* of X be defined by $f(x) = \left(\frac{3}{2}\right)x^2, \quad -1 < x < 1.$

Find the *p.d.f.* of $Y = \frac{(X^3+1)}{2}$

Solution: The transformation $y = U(x) = \frac{(x^3+1)}{2}$, so that $x^3 = 2y - 1$

$$x = \sqrt[3]{2y-1} \quad or \quad (2y-1)^{\frac{1}{3}}$$

$$J = \frac{dx}{dy} = \frac{1}{3}(2y-1)^{-\frac{2}{3}} \text{ x } 2$$

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

$$f(w(y)) = \frac{3}{2}(2y-1)^{\frac{2}{3}}$$

$$g(y) = f(w(y))|J| = \frac{2}{3} \text{ x } \frac{3}{2}(2y-1)^{-\frac{2}{3}+\frac{2}{3}} = 1$$

That is $-1 < x < 1 \Rightarrow -1 < (2y-1)^{\frac{1}{3}} < 1$

$-1 < 2y - 1 < 1$

$0 < 2y < 2$

$0 < y < 1$

Summary

In this chapter, we have discussed the following:

(1) Finding the *p.d.f.* of a random variable from a *p.d.f.* of another random variable using the change-of-variable technique.

(2) How to transform variables concerning either discrete or continuous random variable.

Post-Test

(1) Let the *p.d.f.* of X be defined by $f(x) = \left(\frac{1}{2}\right)^x, x = 1,2,3,\ldots,$ zero elsewhere. Find the *p.d.f.* of $Y = X^3$.

(2) If the random variable X is distributed as $N(\mu, \sigma^2)$, show, by means of a transformation, that the random variable $Y = \left[(X-\mu)/\sigma^2\right]^2$ is distributed as χ_1^2.

(3) If Y has a uniform distribution on the interval $(0,1)$, find the *p.d.f.* of $X = (2Y-1)^{\frac{1}{3}}$.

(4) Let X have *p.d.f.* $f(x) = x^2/24, -2 < x < 4$ and zero elsewhere. Find the *p.d.f.* of $Y = X^2$.

13
Joint Transformations

Introduction

A one-dimensional case that was considered in the last chapter can be extended to a multi-dimensional case with appropriate modifications. Joint transformations of continuous random variables can be accomplished in a similar manner to those considered in our last chapter, but the notion of Jacobian will be generalized.

We shall also consider two probability distributions that are used considerably in certain problems of statistical inference; the *t* distribution and the *F* distribution. Both of these probability distributions occur for certain functions of normal random variables, as will be seen shortly. The *t* distribution is first discovered by W.S. Gosset when he was working for an Irish brewery. The distribution is often known as Student's *t* distribution. However, the *F* distribution was first proposed by George Snedecor to honor R.A. Fisher, who used a modification of this ratio in several statistical applications.

Objectives

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

(1) Appreciate the concept of transformations of jointly distributed random variables.
(2) Know the importance of the *t* and the *F* distributions;
(3) Discuss how the *t* distribution is obtained from the normal distribution;
(4) Discuss how the *F* distribution can be generated from two independent Chi square random variables;
(5) Discuss how the Cauchy distribution can be derived from the *t* distribution; and
(6) Discuss how the two distributions (*t* and *F*) are also useful in tests of hypotheses.

Pre-Test

(1) Define the joint p.d.f. of random variables $X \text{ and } Y$.
(2) What is a marginal probability density function of a random variable?
(3) When are two events said to be independent?
(4) Give the p.d.f. of a standard normal distribution.
(5) List the properties of the normal distribution.
(6) Describe the properties of the Chi square distribution.

Contents

Joint Transformations

This method of finding the p.d.f of a function of one random variable of the continuous type will be extended to functions of two random variables of this type. Let $y_1 - U_1(x_1, x_2)$ and $y_2 - U_2(x_1, x_2)$ define a one-to-one transformation that maps a (two-dimensional) set A in the x_1, x_2 – plane unto a (two-dimensional) set B in the $y_1 y_2 -$ plane. If we express each of x_1 and x_2 in terms of y_1 and y_2 we can write $x_1 = w_1(y_1 y_2)$. The determinant is of order 2

$$\begin{vmatrix} \frac{dx_1}{dy_1} & \frac{dx_1}{dy_2} \\ \frac{dx_2}{dy_1} & \frac{dx_2}{dy_2} \end{vmatrix}$$ is called the Jacobian of the transformation, denoted by *J*.

Example 13.1

Let X_1, X_2 denote a random sample from a distribution $\chi^2_{(2)}$ (note that p.d.f. of X is $f(x) = \frac{1}{\Gamma\left(\frac{r}{2}\right)2^{r/2}} X^{r/2 - 1} e^{-x/2} \qquad 0 \le x < \infty$, then X has a Chi-square distribution with r degrees of freedom abbreviated by X is χ^2_r. .

Find

(i) The joint p.d.f of $Y_1 = X_1$ and $Y_2 = X_2 + X_1$, for $0 < y_1 < y_2 < \infty$.

(ii) The marginal p.d.f of each of Y_1 *and* Y_2.

(iii) Are Y_1 *and* Y_2 independent?

Solution:

(i) $f(x_1) = \frac{1}{2} e^{-x_1/2}$

$f(x_2) = \frac{1}{2} e^{-x_2/2}$

The joint p.d.f of X_1 *and* X_2 is

$$f(x_1)f(x_2) = \frac{1}{4} e^{-\frac{x_1+x_2}{2}}$$

The transformation $y_1 = U_1(x_1, x_2) = x_1$ and

$$y_2 = U_2(x_1, x_2) = x_2 + x_1$$

So that $x_1 = w_1(y_1, y_2) = y_1$ and

$x_2 = w_2(y_1, y_2) = y_2 - y_1$

$$g(y_1, y_2) = f[w_1(y_1, y_2)] f[w_2(y_1, y_2)] |J|$$

$$|J| = \begin{vmatrix} 1 & 0 \\ -1 & 1 \end{vmatrix} = 1$$

The joint p.d.f of Y_1 *and* Y_2 is

$$g(y_1, y_2) = \frac{1}{4} e^{-\frac{(y_1+y_2-y_1)}{2}} = \frac{1}{4} e^{-y_2/2}$$

(ii) The marginal p.d.fs are

$$g_1(y_1) = \int g(y_1, y_2)dy_2 \qquad 0 < y_1 < y_2 < \infty$$

$$= \int_{y_1}^{\infty} \frac{1}{4} e^{-y_2/2} dy_2 \qquad y_1 < y_2 < \infty$$

$$= \frac{1}{4}\left[\frac{e^{-y_2/2}}{-\frac{1}{2}}\right]_{y_1}^{\infty}$$

$$= -\frac{1}{2}\left[-e^{-y_1/2}\right]$$

$$= \frac{1}{2} e^{-y_1/2}$$

$$g_2(y_2) = \int_0^{y_2} \frac{1}{4} e^{-y_2/2} dy_1; \quad 0 < y_1 < y_2$$

$$= \frac{1}{4}\left[y_1 e^{-y_2/2}\right]_0^{y_2}$$

$$= \frac{1}{4} y_2 e^{-y_2/2}$$

(iii) No, since $\frac{1}{2} e^{-y_1/2}$ x $\frac{1}{4} y_2 e^{-y_2/2} = \frac{1}{8} y_2 e^{-(y_1+y_2)/2}$

The t and F Distributions

The t-distribution

The t distribution with n degrees of freedom can be defined as that of a random variable symmetrically distributed about zero whose square has the F distribution with 1 and n degrees of freedom in the numerator and denominator, respectively. Let W denote a random variable that is $N(0, 1)$; let V denote a random variable that is χ^2(r); and let W and V be independent.
Then

$$T = \frac{W}{\sqrt{V/r}}$$

has a t-distribution with r degrees of freedom. Its p.d.f is

$$g(t)=\frac{\Gamma[(r+1)/2]}{\sqrt{\pi r}\,\Gamma\left(r/2\right)\left(1+t^2/r\right)^{(r+1)/2}}, \quad -\infty<t<\infty$$

Proof Since W and V are independent, the joint p.d.f of W and V, say $h(w,\ v)$ is the product of the p.d.fs of W and V.

$$h(w,v)=\begin{cases}\dfrac{1}{\sqrt{2\pi}}e^{-w^2/2}\,\dfrac{1}{\Gamma\left(r/2\right)2^{r/2}}v^{r/2-1}e^{-v/2}, & -\infty<w<\infty \\ & 0<v<\infty \\ 0, \quad otherwise & \end{cases}$$

The change-of-variable technique is used to obtain the p.d.f $g(t)$ of T.

$$t=\frac{w}{\sqrt{v/r}} \quad \text{and} \quad u=v$$

Define a one-to-one transformation that maps
$A=\{(w,\ v):-\infty<w<\infty,\ 0<v<\infty\}$ onto
$B=\{(t,\ \mathrm{u}):-\infty<t<\infty,\ 0<u<\infty\}$.

Since $t=\dfrac{w}{\sqrt{v/r}}\Rightarrow w=t\sqrt{u}/\sqrt{r}, \quad v=u$

$$|J|=\left|\frac{dw}{dt}\right|=\sqrt{u}/\sqrt{r}$$

The joint p.d.f of T and U is given by

$$g(t,u)=h\left(t\sqrt{u}/\sqrt{r},u\right)|J|$$

$$= \frac{1}{\sqrt{2\pi}\,\Gamma\left(r/2\right)2^{r/2}} u^{r/2-1} \exp\left[-\frac{u}{2}\left(1+\frac{t^2}{r}\right)\right]\frac{\sqrt{u}}{\sqrt{r}}$$

$$= \quad 0 \quad \text{elsewhere} \qquad -\infty < t < \infty,\ 0 < u < \infty$$

The marginal p.d.f. of T is

$$g(t) = \int_{-\infty}^{\infty} g(t,u)\,du$$

$$= \int_0^{\infty} \frac{1}{\sqrt{2\pi r}\,\Gamma\left(r/2\right)2^{r/2}} u^{(r+1)/2-1} \exp\left[-\frac{u}{2}\left(1+\frac{t^2}{r}\right)\right]du$$

Let $z = u\left[1+\left(t^2/r\right)\right]/2$ then

$$g(t) = \int^{\infty} \frac{1}{\sqrt{2\pi r}\,\Gamma\left(r/2\right)2^{r/2}} \left(\frac{2z}{1+t^2/r}\right)^{\frac{(r+1)}{2}-1} e^{-Z\left(\frac{2}{1+t^2/r}\right)dZ}$$

$$= \frac{\Gamma\left[(r+1)/2\right]}{\sqrt{\pi r}\,\Gamma\left(r/2\right)} \frac{1}{\left(1+t^2/r\right)^{(r+1)/2}} \qquad -\infty < t < \infty$$

Thus if W is $N(0,1)$, if V is $\chi^2(r)$, and W and V are independent, then

$$T = \frac{W}{\sqrt{V/r}}$$

Properties

(1) The t distribution is symmetric about zero ($E(T)=0$ when $r \geq 2$) and its general shape is similar to that of the standard normal distribution.

(2) The t distribution approaches the standard normal distribution as $r \to \infty$, for smaller v the t distribution is flatter with thicker tails and, in fact, $T \sim CAU(1,0)$ when $r = 1$ (the t distribution is the Cauchy distribution and the mean and thus the variance do not exist for the Cauchy distribution).

(3) The t distribution has more variability than the standard normal distribution since more variation is noticed when Z is divided by another random variable.

The F-distribution

In the problem of comparing normal populations with respect to their variances, as well as in a variety of other problems, it will be necessary to know the distribution of the ratio of two Chi-square random variables. Consider two independent Chi-square random variables U and V having r_1 and r_2 degrees of freedom respectively; then

$$F = \frac{U/r_1}{V/r_2}$$

has an F-distribution with r_1 and r_2 degrees of freedom. Its p.d.f is

$$h(u,v) = \frac{\Gamma[(r_1+r_2)/2](r_1/r_2)^{r_1/2} W^{r_1/2-1}}{\Gamma(r_1/2)\Gamma(r_2/2)(1+r_1 w/r_2)^{(r_1+r_2)/2}}, \qquad 0 < w < \infty$$

Proof: The joint p.d.f of U and V is

$$h(u,v)=\frac{1}{\Gamma\left(\frac{r_1}{2}\right)\Gamma\left(\frac{r_2}{2}\right)2^{(r_1+r_2)}}U^{r_1/2-1}V^{r_2/2-1}e^{-(u+v)/2}$$

$$= 0, \qquad \text{elsewhere} \qquad 0<u<\infty;\ 0<v<\infty$$

To find the p.d.f g(w) of W, the equations

$$w=\frac{u/r_1}{v/r_2} \text{ and}$$

$$z = v$$

define a one-to-one transformation that maps the set $A=\{(u,\ v):0<u<\infty,\ 0<v<\infty\}$ onto the set $B=\{(w,\ z):0<w<\infty,\ 0<z<\infty\}$, Since $u=\left(r_1/r_2\right)zw,\ v = z$ $|J|=\left(r_1/r_2\right)z$. The joint p.d.f $g(w,\ z)$ of the random variables W and $Z=V$ is

$$g(w,z)=\frac{1}{\Gamma\left(\frac{r_1}{2}\right)\Gamma\left(\frac{r_2}{2}\right)2^{(r_1+r_2)/2}}\left(\frac{r_1zw}{r_2}\right)^{\frac{r_1}{2}-1}z^{\frac{r_2}{2}-1}\times\exp\left[-\frac{z}{2}\left(\frac{r_1w}{r_2}+1\right)\right]\frac{r_1z}{r_2},$$

provided that $(w,z)\in\beta$ and zero elsewhere,

The marginal p.d.f $g(w)$ of W is

$$g(w)=\int_{-\infty}^{\infty}g(w,z)dz$$

$$= \int_0^{\infty} \frac{\left(\frac{r_1}{r_2}\right)^{r_1/2} (w)^{\frac{r_1}{2}-1}}{\Gamma\left(\frac{r_1}{2}\right)\Gamma\left(\frac{r_2}{2}\right)2^{(r_1+r_2)/2}} z^{\frac{(r_1+r_2)}{2}-1} \times \exp\left[-\frac{z}{2}\left(\frac{r_1 w}{r_2}+1\right)\right]dz$$

If we change the variable of integration by writing

$$y = \frac{z}{2}\left(\frac{r_1 w}{r_2}+1\right)$$

it can be seen that

$$g_1(w) = \int_0^{\infty} \frac{\left(r_1/r_2\right)^{r_1/2} (w)^{\frac{r_1}{2}-1}}{\Gamma\left(r_1/2\right)\Gamma\left(r_2/2\right)2^{(r_1+r_2)/2}} \left(\frac{2_y}{r_1 w/r_2 + 1}\right)^{\frac{(r_1+r_2)}{2}-1} e^{-y} \text{ x} \left(\frac{2}{r_1 w/r_2 + 1}\right)dy$$

$$= \frac{\Gamma\left[(r_1+r_2)/2\right]\left(r_1/r_2\right)^{\frac{r_1}{2}}}{\Gamma\left(r_1/2\right)\Gamma\left(r_2/2\right)} \frac{(w)^{\frac{r_1}{2}-1}}{\left(1+r_1 w/r_2\right)^{(r_1+r_2)/2}} \qquad 0 < \text{w} < \infty$$

$\therefore$ If U and V are independent Chi-square with r_1 and r_2 degrees of freedom respectively, then

$$W = \frac{U/r_1}{V/r_2}$$

has p.d.f $g(w)$. The distribution of this random variable is usually called an F distribution.

If $X \sim F(r_1, r_2)$, then

$$E(X^r) = \frac{\left(\frac{r_2}{r_1}\right)^r \Gamma\left(\frac{r_1}{2}+r\right)\Gamma\left(\frac{r_2}{2}-r\right)}{\Gamma\left(\frac{r_1}{2}\right)\Gamma\left(\frac{r_2}{2}\right)}, r_2 > 2r$$

For $r = 1$, the mean, $E(X) = \frac{r_2}{r_2 - 2}, 2 < r_2$

The variance is given by

$$Var(X) = \frac{2r_2^2(r_1 + r_2 - 2)}{r_1(r_2 - 2)^2(r_2 - 4)}, \ 4 < r_2.$$

Properties

(1) If F is distributed as F_{r_1, r_2}, then, $1/F$ is distributed as F_{r_2, r_1}.

(2) If X is $N(0,1)$, Y is χ_r^2 and X, Y are independent, so that $T = X/\sqrt{Y/r}$ is distributed as t_r, then T^2 is distributed as $F_{1,r}$, since X^2 is χ_1^2.

Examples 13.2

(1) Let X have a t distribution with r degrees of freedom. Find

(i) $P(X \geq 2.228)$ when $r = 10$

(ii) $P(X \leq 2.228)$ when $r = 10$

(iii) $P(1.330 \leq X \leq 2.552)$ when $r = 18$

(2) Let X have a t distribution with $r = 19$. find c such that

(a) $P(X \geq c) = 0.025$

(b) $P(X \leq c) = 0.95$

Solution:

(1) (i) $P(X \geq 2.228) = 1 - 0.975 = 0.025$

(ii) $P(X \leq 2.228) = 0.975$

(iii) $P(1.330 \leq X \leq 2.552) = 0.99 - 0.90 = 0.09$

(2) (i) $P(X \geq c) = 0.025$, $c = 2.093$

(ii) $P(X \leq c) = 0.95$, $c = 1.729$

Summary

This chapter discussed the following:

(1) Transformations of variables involving jointly distributed random variables.

(2) The last chapter dealt with the notion of a one-to-one transformation and the mapping of one set to another set under that transformation; however, we built on these ideas in this chapter to help us find the distribution of a function of two variables of the discrete type.

(3) We also looked at the same problem raised in 2 above when the random variables are of the continuous type.

(4) Examples of these transformations were given for clearer understanding of the concept.

(5) Computation of marginal p.d.f.s and independent variables re visited.

(6) The t distribution is completely determined by the number r, the degrees of freedom.

(7) Because of the symmetry of the t distribution about $t = 0$, the mean if it exists, must be equal to zero.

(8) The F distribution depends on two parameters, r_1 and r_2 in that order. The first parameter is the number of degrees of freedom in the numerator, and the second is the number of degrees of freedom in the denominator.

(9) The t distribution is very useful in testing hypothesis about the mean.

(10) The F distribution on the other hand, is used to test hypothesis about ratio of two variances or sum of

squares.

Post-Test

(1) If X_1 *and* X_2 denote a random sample of size two from a Poisson distribution, $X_i \sim POI(\lambda)$, find the p.d.f. of $Y = X_1 + X_2$.

(2) Suppose that X_1 *and* X_2 denote a random sample of size two from a gamma distribution, $X_i \sim GAM(2, 1/2)$.

   (i) Find the p.d.f. of $Y = \sqrt{X_1 + X_2}$

   (ii) Find the p.d.f. of $W = X_1 / X_2$

(3) Let X_1 *and* X_2 denote a random sample of size two from a distribution that is $N(\mu, \sigma^2)$. Let $Y_1 = X_1 + X_2$ and $Y_2 = X_1 - X_2$. Find the joint p.d.f. of Y_1 *and* Y_2 and show that these random variables are independent.

(4) Let X_1 *and* X_2 have the joint p.d.f. $h(x_1, x_2) = 8x_1x_2$, $0 < x_1 < x_2 < 1$, zero elsewhere. Find the joint p.d.f. of $Y_1 = X_1 / X_2$ and $Y_2 = X_2$ and argue that Y_1 *and* Y_2 are independent. (Hint: Use the inequalities $0 < y_1y_2 < y_2 < 1$ in considering the mapping from A to B).

(5) Let X have an F distribution with r_1 and r_2 degrees of freedom. Find

   (i) $P(X \geq 3.02)$ when $r_1 = 9, r_2 = 10$

   (ii) $P(X \leq 4.14)$ when $r_1 = 7, r_2 = 15$

   (iii) $P(X \leq 0.1508)$ when $r_1 = 8, r_2 = 5$.

   Hint: $0.1508 = 1/6.63$

   (iv) $P(0.1323 \leq X \leq 2.79)$ when $r_1 = 6, r_2 = 15$

(6) Let X have an F distribution with r_1 and r_2 degrees of freedom. Find

(i) $P(a \le X \le b) = 0.90$ when $r_1 = 8, r_2 = 6$

(ii) $P(a \le X \le b) = 0.98$ when $r_1 = 8, r_2 = 6$

(iii) Let X have a t distribution with r degrees of freedom. Find

(iv) $P(|X| \ge 2.228)$ when $r = 10$

(v) $P(-1.753 \le X \le 2.602)$ when $r = 15$

Bibliography

Bain, L.J. and Engelhardt Max (1989) "Introduction to Probability and Mathematical Statistics." Boston: PWS-KENT Publishing Company.

Bertsekas, P., Dimitri, P. and Tsitsiklis N. John (2000) "Introduction to Probability." Massachusetts Institute of Technology, Cambridge, Massachusetts, UK.

Boslaugh Sarah and Watters Paul Andrew (2008) "Statistics in a Nutshell" O'Reilly Media, CA.

DeGroot Morris H. (1989) "Probability and Statistics" Second Edition. Addison-Wesley Publishing Company, Califonia.

Gray F. James (1962) "Sets, Relations and Functions." Holt, Rinehart and Winston, Inc. New York.

Grinstead M. Charles and Snell J. Laurie (2000) "Introduction to Probability" *http://www.dartmouth.edu/~chance*.

Hamburg Morris (1970) "Statistical Analysis for Decision Making." New York: Harcourt, Brace & World, Inc.

Hogg, R.V. and Craig, A.T. (1970) "Introduction to Mathematical Statistics." 3rd Edition. New York: Macmillan Publishing Co., Inc. London: Collier Macmillan Publishers.

Hogg, R.V. and Craig, A.T. (1995) "Introduction to Mathematical Statistics." 5th Edition. London: Prentice-Hall, Inc.

Hogg, R.V. and Tanis, E.A. (1993) "Probability and Statistical Inference." New York: Macmillan Publishing Company.

Larson Harold J. (1982) "Introduction to Probability Theory and Statistical Inference." Third Edition. New York: John Wiley & Sons.

Lind Douglas A., Marchal William G. and Wathen Samuel A. (2006) "Basic Statistics for Business and Economics" 5th Ed. McGraw-Hill, Boston

Lindgren Bernard W. (1976) "Statistical Theory" Third Edition. New York: Macmillan Publishing Co., Inc.

Montgomery C. Douglas and Runger C. George (2003) "Applied Statistics and Probability for Engineers." 3rd Ed. John Wiley & Sons, Inc. New York.

Mukhopadhyay Nitis (2000) "Probability and Statistical inference" Marcey Dekker, Inc., New York.

Rosen H. Kenneth (1999) "Discrete Mathematics and Its Applications." Fourth Edition. McGraw-Hill, Boston, USA.

Ross, S.M. (1987) "Introduction to Probability and Statistics for Engineers and Scientists." John Wiley & Sons, New York.

Ross, S. (1988) "A First Course in Probability." 3rd Ed., London: Nobern Avocation Publishing Co., Macmillan Publishing Co.

Ross, S. (1994) "A First Course in Probability." 4th Ed., New Jersey: Prentice-Hall, Inc. A Simon & Schuster Company, Englewood Cliffs.

Ross M. Sheldon (2004) "Introduction to Probability and Statistics for Engineers and Scientists." Third Edition. Elsevier Academic Press, San Diego, USA.

Roussas George G. (1972) "A First Course in Mathematical Statistics." Reading Massachusetts: Addisson-Wesley publishing Company.

Schay G´eza (2007) "Introduction to Probability with Statistical Applications." Birkh¨auser, Boston, USA.

Shangodoyin D. K, Olubusoye O. E, Shittu O. I and Adepoju A. A (2002). "Statistical Theory and Methods" Nigeria: Joytal Printing Press, Ibadan. ISBN: 978-2906-23-9.

Shao Jun (2003) "Mathematical Statistics" 2nd Ed. Springer Science+Business Media, LLC., New York, N.Y., 10013, USA

Soong, T.T. (2004) “Fundamental of Probability and Statistics for Engineers” John Wiley & sons, Ltd. England.

Spanos Aris (2003) “Probability Theory and Statistical Inference: Econometric Modeling with Observational Data.” Cambridge University Press, New York.

Spiegel Murray R. (1972) “Theory and problems of Statistics.” SI (metric) Edition. New York: McGraw-Hill Book Company (UK) Limited.

Spiegel Murray R., Schiller John and Srinivasan R. Alu (2001) “Probability and Statistics, Schaum’s Outline Series.” Mcgraw-Hill New York.

Wayne W. Daniel (1976) “Biostatistics: A Foundation for Analysis in the Health Sciences.” John Wiley & Sons, Inc., New York.

Wilcox Rand R. (2009) “Basic Statistics *Understanding Conventional Methods And Modern Insights”* Oxford University Press, Inc., New York

Woolfson, Michael M. (2008) “Everyday Probability And Statistics; Health, Elections, Gambling and War” Imperial College Press, London.

Zwillinger Daniel and Kokoska Stephen (2000) “Standard probability and Statistics tables and formulae” Chapman & Hall/CRC, New York.

Index