Probability and Statistics I

By

Prof. Magdy E. El-Adll & Assoc. Prof. Amany E. Aly

Department of Mathematics

Faculty of Science

Helwan University

2022

Contents

1 Introduction **3**

1.1 What Is Statistics? 3

1.2 Why Study Statistics? 4

1.3 Types of Statistics 5

1.4 Types of variables 6

2 Probability **7**

2.1 Introduction 7

2.2 Basic concepts and definition of probability 8

2.2.1 Definition of Probability 12

2.2.2 Exercises 21

2.3 Combinational techniques 23

2.3.1 Exercises 32

2.4 Conditional Probability 34

2.5 Independent Events 35

2.6 The Multiplication Law of Probability 41

2.6.1 Exercises 51

3 Discrete Distributions **55**

3.1 Random Variables of the Discrete Type 55

3.2 Mathematical Expectation 60

3.3 The Mean, Variance, and Standard Deviation 64

3.4 Bernoulli Trials and Binomial Distribution 73

3.5 Geometric and Negative Binomial Distributions 84

3.5.1 Exercises 91

3.6 Poisson Distribution 92

3.7 Moment-Generating Function 102

4 Continuous Distributions **109**

4.1 Random Variables of the Continuous Type 109

4.2 The Uniform and Exponential Distributions 116

4.3 The Gamma and chi-Square Distributions 126

4.4 The Normal Distribution 132

5 Joint Distribution **145**

5.1 Introduction 145

5.2 Joint Probability Distribution 146

5.3 Statistically Independent 156

5.4 Chebyshev′s Inequality and Convergence in Probability 158

5.5 Exercises 164

6 Sampling Distribution Theory **169**

6.1 Basic Concepts 170

6.2 Sampling Theory 172

6.3 Some Location Statistics 173

References **189**

Chapter 1

Introduction

We begin our study by the following question:

1.1 What Is Statistics?

Statistics is the science whereby inferences are made about specific random phenomena on the basis of relatively limited sample material. The field of statistics can be subdivided into two main areas: Mathematical statistics and applied statistics.

1. Mathematical statistics concerns the development of new methods of statistical inference and requires detailed knowledge of abstract mathematics for its implementation.

2. Applied statistics concerns the application of the methods of mathematical statistics to specific subject area, such as economics, psychology, public health, and biostatistics.

The previous definition is the scientific (academic) definition of statistics, however, the applied statisticians define statistics in a very simple way as:

Definition 1. *Statistics is the science of collecting, organizing, presenting, analyzing, and interpreting data to assist in making decisions.*

1.2 Why Study Statistics?

No matter what line of work you select, you will find yourself faced with decisions where an understanding of data analysis is helpful. In order to make an informed decision, you will need to able to:

1. Determine whether the existing information is sufficient or additional information is required.

2. Gather additional information, if it is needed, in such away that it is does not provide misleading results.

3. Summarize the information in a useful informative manner.

4. Analyze the available information.

5. Draw conclusions and make inferences while assessing the risk of an incorrect conclusion.

In summary, there are at least three reasons for studying statistics:

1. Data is everywhere,

2. Statistical techniques are used to make many decisions that effect our lives,

3. No matter what your future line of work, you will make decisions that involve data.

1.3 Types of Statistics

The study of statistics is usually divided into two categories: descriptive statistics and inferential statistics.

Definition 2. *Descriptive statistics are methods of organizing, summarizing and presenting data in an informative way.*

Definition 3. *Inferential statistics are methods used to determine something about a population, based on a sample.*

Definition 4. *A population is a collection of all possible, objects, or measurements of interest.*

Definition 5. *A sample is a portion, or a part, of the population of interest.*

1.4 Types of variables

There are two types of variables

A. A qualitative variable is nonnumeric.

1. Usually we are interested in the number or percent of the observations in each category.
2. qualitative data are usually summarized in graphs and par charts.

B. There are two types of quantitative variables and they are usually reported numerically.

1. Discrete variables can assume only certain values, and there are usually gaps between values.
2. A continuous variables can assume any value within a specific range or interval.

Chapter 2

Probability

2.1 Introduction

Probability as a general concept can be defined as the chance of an event occurring. Most people are familiar with probability from observing or playing games of chance, such as card games, slot machine, or lotteries. In addition to being used in games of chance, probability theory is used in fields of insurance, investments, and weather forecasting, and in various other areas. For example, predictions are based on probability, and hypotheses are tested by using probability.

The basic concepts of probability are explained in this chapter. These concepts include probability experiments, sample spaces, the addition and multiplication rules, *and the* probabilities of complementary events.

The theory of probability grew out of the study of various games of chance using coins, dice, and cards. Since these devices lend themselves well to the application of concepts of probability, they will be used in this chapter as examples.

2.2 Basic concepts and definition of probability

Processes such as flipping a coin, rolling a dice, or drawing a card from a deck are called *probability experiments.*

Definition 6. *A **random experiment** is a process that leads to well-defined results called outcomes.*

Definition 7. *An **outcome** is the result of a single trial of a random experiment.*

When a coin is tossed, there are two possible outcomes: head or tail. In the roll of a single dice, there are six possible outcomes: 1, 2, 3, 4, 5, or 6. In any experiment, the set of all possible outcomes is called the sample space.

Definition 8. *The sample space is the set of all possible outcomes of a random experiment and denoted by* S.

In referring to probabilities of events, an ***event*** is any set of outcomes of interest.

Some sample spaces for various probability experiments are shown here:

Experiment	**Sample space**
Toss one coin	Head(H), tail(T)
Roll a dice	1, 2, 3, 4, 5, 6
Answer a true-false question	True, false
Toss two coins	HH, HT, TH,TT

The collection of all possible outcomes (the universal set) of a random experiment is denoted by S and is called the outcome space. Given an outcome space S, let A be a part of the collection of outcomes in S, that is, $A \subset S$. Then A is called an event. When the random experiment is performed and the outcome of the experiment is in A, we say that event A has occurred. Since, in studying probability, the words set and event are interchangeable, the reader might want to review algebra of sets. For convenience, however, we remind the reader of a little of that terminology:

- ϕ denotes the null or empty set.

- $A \subset B$ means that A is a subset of B .

- $A \cup B$ is the union of A and B.

- $A \cap B$ is the intersection of A and B.

- A^c is the complement of A (i.e., all elements in S that are not in A).

Special terminology associated with events that is often used by statisticians includes the following:

1 $A_1, A_2, \ldots, A_k$ are *mutually exclusive events* means that $A_i \cap A_j = \phi,\ i \neq j,$ that is, $A_1, A_2, \ldots, A_k$ are disjoint sets.

2 $A_1, A_2, \ldots, A_k$ are *exhaustive events* means that

$$A_1 \cup A_2 \cup \cdots \cup A_k = S.$$

So if $A_1, A_2, \ldots, A_k$ are mutually exclusive and exhaustive events, we know that $A_i \cap A_j = \Phi,\ i \neq j,$ and $A_1 \cup A_2 \cup \cdots \cup A_k = S.$

We are interested in defining what is meant by the probability of A, denoted by $P(A)$, and often called the chance of A occurring. To help us understand what is meant by the probability of A, $P(A)$, consider repeating the experiment a number of times, say n times. Count the number of times that event A actually occurred throughout these n performances; this number is called the frequency of event A and is denoted by $\mathbb{N}(A)$. The ratio $\mathbb{N}(A)/n$. is called the relative frequency of event A in these n repetitions of the experiment. A relative frequency is usually very unstable for small values of n, but it tends to stabilize as n increases. This suggests that we associate with event A a number, say p, that is equal to or approximately equal to the number about which the relative frequency tends to stabilize. This number p can then be taken as the number that the relative frequency of event A will be near in future performances of the experiment. Thus, although we cannot predict the outcome of a random experiment with certainty, we can, for a large value of n, predict fairly accurately the relative frequency associated with event A. The number p assigned to event A is called the *probability* of event A, and is denoted by $P(A)$. That is, $P(A)$ represents the proportion of outcomes of a random experiment that terminate into an event.

To help decide what properties the probability set function should satisfy, consider properties possessed by the relative frequency $\mathbb{N}(A)/n$. For example, $N(A)/n$ is always nonnegative. If $A = S$, the sample space, then the outcome of the experiment will always belong to S, and thus $\mathbb{N}(S)/n = 1$. Also, if A and B are two mutually exclusive events, then

$$\mathbb{N}(A \cup B)/n = \mathbb{N}(A)/n + \mathbb{N}(B)/n.$$

Hopefully, these remarks will help to motivate the following definition.

2.2.1 Definition of Probability

Definition 9. ***Probability*** *of an event is the relative frequency of this set of outcomes over an indefinitely large (or infinite) number of trials. The impossible event will be denoted by* $\emptyset$.

Now we can give the general definition of probability.

Definition 10. ***Probability*** *is a set function* P *that assigns to each event* A *in the sample space* S *a number* $P(A)$, *called the probability of the event* A, *such that the following properties are satisfied:*

(i) $P(A) \geq 0,$

(ii) $P(S) = 1,$

(iii) *If* $A_1, A_2, A_3, \cdots, A_k$ *are events and* $A_i \cap A_j = \emptyset$, $i \neq j$, *then*

$$P(A_1 \cup A_2 \cup \cdots \cup A_k) = P(A_1) + P(A_2) + \cdots + P(A_k),$$

for each positive integer k, *and*

$$P(A_1 \cup A_2 \cup A_3 \cup \cdots) = P(A_1) + P(A_2) + P(A_3) + \cdots$$

Some Useful Probabilistic Notations

Definition 11. *Two events A and B are* ***mutually exclusive*** *if they cannot both happen at the same time.*

Definition 12. *$A \cup B$ is the event that either A or B occurs or they both occur.*

Definition 13. *$A \cap B$ is the event that both A and B occur simultaneously.*

Definition 14. *A^c is the event that A does not occur. It is some times referred to as the* **complement** *of A.*

The following theorems give some other important properties of probability set function P. When one conceder these theorems, it is important to understand the theoretical concepts and proofs. However, if the reader keeps the relative frequency concept in mind, the theorems should have also some intuitive appeal.

Theorem 1. *For each event A,*

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

Proof. We have

$$S = A \cup A^c \quad \text{and} \quad A \cap A^c = \Phi.$$

Thus, from properties (ii) and (iii), it follows that

$$1 = P(A) + P(A^c).$$

Hence

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

□

Theorem 2. $P(\emptyset) = 0$.

Proof. In Theorem 1, take $A = \Phi$ so that $A^c = S$. Thus

$$P(\Phi) = 1 - P(S) = 1 - 1 = 0.$$

□

Theorem 3. *If A and B are two events such that $A \subset B$, then $P(A) < P(B)$.*

Proof. Now

$$B = A \cup (B \cap A^c) \quad \text{and} \quad A \cap (B \cap A^c) = \Phi.$$

Hence, from property (iii),

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

because from property (i),

$$P(B \cap A^c) \geq 0.$$

$\square$

Theorem 4. *For each event* A, $P(A) \leq 1$.

Proof. Since $A \subset S$, we have by Theorem 3, and property (ii) that

$$P(A) \leq P(S) = 1,$$

which gives the required result $\square$

Property (i) along with Theorem 4 shows that, for each event A, $0 \leq P(A) \leq 1$.

Theorem 5 (Addition Law of Probability)**.** *If* A *and* B *are any two events, then*

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

Proof. The event ($A \cup B$) can be represented as a union of two mutually exclusive events, namely,

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

Hence, by property (iii),

$$P(A \cup B) = P(A) + P(A^c \cap B). \tag{2.1}$$

However,

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

which is a union of two mutually exclusive events. Thus,

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

that is,

$$P(A^c \cap B) = P(B) - P(A \cap B). \tag{2.2}$$

If we substitute from Equation 2.2 into Equation 2.1 we obtain

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

which is the desired result. □

Example 2.2.1. *A faculty leader was meeting two students in Paris, one arriving by train from Amsterdam and the other arriving by train from Brussels at approximately the same time. Let A and B be the events that the trains are on time, respectively. Suppose from past experience we know that*

$$P(A) = 0.93, \quad P(B) = 0.89, \quad \textit{and} \quad P(A \cap B) = 0.87.$$

Then

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.93 + 0.89 - 0.87 = 0.95$$

Theorem 6. *If A, B and C any three events, then*

$$\begin{aligned} P(A \cup B \cup C) &= P(A) + P(B) + P(C) \\ &- P(A \cap B) - P(A \cap C) - P(B \cap C) \\ &+ P(A \cap B \cap C) \end{aligned}$$

Proof. Write

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

and apply Theorem 5. The details are left as an exercise. □

Let a probability set function be defined on a sample space S. Let

$$S = \{e_1, e_2, \ldots, e_m\},$$

where each e_i is a possible outcome of the experiment. The integer m is called the total number of ways in which the random experiment can terminate. If each of these outcomes has the same probability of occurring, we say that the m outcomes are equally likely. That is,

$$P(e_j) = \frac{1}{m}, \quad i = 1, 2, ..., m.$$

If the number of outcomes in an event A is h, the integer h is called the number of ways that are favorable to the event A. In this case $P(A)$ is equal to the number of ways favorable to the event A divided by the total number of ways in which the experiment can terminate. That is, under this assumption of equally likely outcomes, we have that

$$P(A) = \frac{h}{m} = \frac{\mathbb{N}(A)}{\mathbb{N}(S)},$$

where $h = \mathbb{N}(A)$ is the number of ways that A can occur and $m = \mathbb{N}(S)$ is the number of ways that S can occur.

It should be emphasized that in order to assign the probability h/m to the event A, we must assume that each of the outcomes $e_1, e_2, \ldots, e_m$ has the same probability $1/m$. This assumption is then an important part of our probability model; if it is not realistic in an application, the probability of the event A cannot be computed this way.

Example 2.2.2 (Hypertension)**.** *Let A be the event that a person has normotensive diastolic blood-pressure (DBP) readings (i.e., DBP < 90), B be the event that a person has borderline DBP readings (i.e., $90 \leq DBP \leq 95$) and let C be such that $75 \leq DBP \leq 100$. Suppose that*

$$P(A) = 0.7, \quad P(B) = 0.1, \quad P(C) = 0.6.$$

Let D be the event that a person has DBP<95. then

$$P(D) = P(A \cup B) = P(A) + P(B) = 0.8,$$

since $A \cap B = \emptyset$. It is clear that

$$P(DBP \geq 90) = P(A^c) = 1 - P(A) = 0.3.$$

Example 2.2.3. *If a family has three children , find the probability that all the children are girls.*

<u>Solution</u>

The sample space for the gender of children for a family has three children is

$$S = \{BBB, BBG, BGB, GBB, GGG, GGB, GBG, BGG\}$$

Since there is one way in eight possibilities for all three children to be girls (GGG), we have

$$P(GGG) = \frac{1}{8}.$$

Example 2.2.4. *In a sample of 50 people, 21 had type O blood, 22 had type A blood, 5 had type B blood and 2 had type AB blood. Set up a frequency distribution and find the following probabilities:*

(a) *A person has type O blood.*

(b) *A person has type A or type B blood.*

(c) *A person has neither type A nor type O blood.*

(d) *A person does not have type AB blood.*

<u>Solution</u>

Type	**Frequency**
A	22
B	5
AB	2
O	21
Total	50

(a) $P(O) = \dfrac{21}{50}$.

(b) $P(A \text{ or } B) = \dfrac{22}{50} + \dfrac{5}{50} = \dfrac{27}{50}$

(Add the frequencies of the two classes.)

(c) $P(\text{neither } A \text{ nor } O) = \dfrac{5}{50} + \dfrac{2}{50} = \dfrac{7}{50}$

(Neither A nor O means that a person has either type B or type AB blood.)

(d) $P(AB)^c = 1 - P(AB) = 1 - \dfrac{2}{50} = \dfrac{48}{50}$.

2.2.2 Exercises

1. A coin is tossed four times, and the sequence of heads and tails is observed.

   (a) List each of the 16 sequences in the sample space S.

   (b) Let events A, B, C, and D be given by

$$\begin{aligned} A &= \{\text{at least 3 heads }\}, \\ B &= \{\text{ at most 2 heads}\}, \\ C &= \{\text{heads on the third toss }\}, \\ D &= \{\text{1 head and 3 tails }\}. \end{aligned}$$

   If the probability set function assigns 1/16 to each outcome in the sample space, find

$$(i)\ P(A), \qquad (ii)\ P(A \cap B), \qquad (iii)\ P(B)$$

$$(v)\ P(D), \qquad (iv)\ P(A \cup C), \qquad (vi)\ P(B \cap D)$$

2. A fair eight-sided die is rolled once. Let

$$A = \{2,\ 4,\ 6,\ 8\}, \quad B = \{3,\ 6\},$$

$$C = \{2,\ 5,\ 7\}, \quad \text{and} \quad D = \{1,\ 3,\ 5,\ 7\}.$$

Assume that each face has the same probability.

(a) Find $P(A)$, $P(B)$, $P(C)$, and $P(D)$.

(b) Find $P(A \cap B)$, $P(B \cap C)$, and $P(C \cap D)$.

(c) Find $P(A \cup B)$, $P(B \cup C)$, and $P(C \cup D)$ using Theorem 5.

3. If $P(A) = 0.4$, $P(B) = 0.5$, and $P(A \cap B) = 0.3$, find (a) $P(A \cup B)$, (b) $P(A \cap B^c)$, and (c) $P(A^c \cup B^c)$.

4. If $S = A \cup B$, $P(A) = 0.7$, and $P(B) = 0.9$, find $P(A \cap B)$.

5. If $P(A) = 0.4$, $P(B) = 0.5$, and $P(A \cup B) = 0.7$, find (a) $P(A \cap B)$ and (b) $P(A^c \cup B^C)$. .

6. The five numbers 1, 2, 3, 4, and 5 are written, respectively, on five disks of the same size and placed in a hat. Two disks are drawn without replacement from the hat, and the numbers written on them are observed.

   (a) List the 10 possible outcomes as pairs of numbers (order not important).

(b) If each of the 10 outcomes has probability 1/10, assign a value to the probability that the sum of the two numbers drawn is (i) 3 and (ii) between 6 and 8 inclusive.

7. Let $S = A_1 \cup A_2 \cup \cdots \cup A_m$ where the events $A_1, A_2, \ldots, A_m$, are mutually exclusive and exhaustive.

(a) If P $(A_1) = P(A_2) = \cdots = P(A_m)$, show that $P(A_i) = l/m$, $i = 1, 2, \ldots, m$.

(b) If $A = A_1 \cup A_2 \cup \cdots \cup A_h$, where $h < m$, and part (a) holds, prove that $P(A) = h/m$.

2.3 Combinational techniques

One of the problems that the statistician must consider and attempt to e-valuate is the element of chance associated with the occurrence of certain events when an experiment is performed. These problems belong in the field of probability. In many cases, we shall be able to solve a probability problem by counting the number of points in the sample space without actually listing each element. The fundamental principle of counting, often referred to as the multiplication rule, is stated in next rule.

Rule 1

If an operation can be performed in n_1 ways, and if for each of these ways a second operation can be performed in n_2 ways, then the two operations can be performed together in $n_1 n_2$ ways.

Example 2.3.1. *How many sample points are there in the sample space when a pair of dice is thrown once?*

Solution :

The first die can land face-up in any one of $n_1 = 6$ ways. For each of these 6 ways, the second die can also land face-up in $n_2 = 6$ ways. Therefore, the pair of dice can land in $n_1 n_2 = (6)(6) = 36$ possible ways.

Example 2.3.2. *If a 22-member club needs to elect a chair and a treasurer, how many different ways can these two to be elected?*

Solution :

For the chair position, there are 22 total possibilities. For each of those 22 possibilities, there are 21 possibilities to elect the treasurer. Using the multiplication rule, we obtain $n_1 n_2 = (22)(21) = 462$ different ways.

The answers to the two preceding examples can be verified by constructing tree diagrams and counting the various paths along the branches. The multiplication rule, may be extended to cover any number of operations. Suppose, for instance, that a customer wishes to buy a new cell phone and can choose from $n_1 = 5$ brands, $n_2 = 5$ sets of capability, and $n_3 = 4$ colors. These three classifications result in $n_1 n_2 n_3 = (5)(5)(4) = 100$ different ways for a customer to order one of these phones. **The generalized multiplication rule** covering k operations is stated in the following.

Rule 2

If an operation can be performed in n_1 ways, and if for each of these a second operation can be performed in n_2 ways, and for each of the first two a third operation can be performed in n_3 ways, and so forth, then the sequence of k operations can be performed in $n_1 n_2 \cdots n_k$ ways.

Example 2.3.3. *Sam is going to assemble a computer by himself. He has the choice of chips from two brands, a hard drive from four, memory from three, and an accessory bundle from five local stores. How many different ways can Sam order the parts?*

Solution :

Since $n_1 = 2,\ n_2 = 4,\ n_3 = 3,$ *and* $n_4 = 5,$ *there are* $n_1 \times n_2 \times n_3 \times n_4 = 2 \times 4 \times 3 \times 5 = 120$ *different ways to order the parts.*

Frequently, we are interested in a sample space that contains as elements all possible orders or arrangements of a group of objects. For example, we may want to know how many different arrangements are possible for sitting 6 people around a table, or we may ask how many different orders are possible for drawing 2 lottery tickets from a total of 20. The different arrangements are called **permutations**.

Definition 15. *A **permutation** is an arrangement of all or part of a set of objects.*

Consider the three letters a, b, and c. The possible permutations are abc, acb, bac, bca, cab, and cba. Thus, we see that there are 6 distinct arrangements. Using Rule 2, we could arrive at the answer 6 without actually listing the different orders by the following arguments: There are $n_1 = 3$ choices for the first position. No matter which letter is chosen, there are always $n_2 = 2$ choices for the second position. No matter which two letters are chosen for the first two positions, there is only $n_3 = 1$ choice for the last position, giving a total of $n_1n_2n_3 = (3)(2)(1) = 6$ permutations by Rule 2. In general, n distinct objects can be arranged in $n(n-1)(n-2)\cdots(3)(2)(1)$ ways. There is a notation for such a number.

Definition 16. *For any non-negative integer n, $n!$, called "n factorial," is defined as $n! = n(n-1)\cdots(2)(1)$, with special case $0! = 1$.*

Using the argument above, we arrive at the following theorem.

Theorem 7. *The number of permutations of* n *objects is* $n!$.

The number of permutations of the four letters a, b, c, and d will be 4 $= 24$. Now consider the number of permutations that are possible by taking two letters at a time from four. These would be ab, ac, ad, ba, bc, bd, ca, cb, cd, da, db, and dc. Using Rule 1 again, we have two positions to fill, with $n_1 = 4$ choices for the first and then $n_2 = 3$ choices for the second, for a total of $n_1\ n_2 = (4)(3) = 12$ permutations. In general, n distinct objects taken r at a time can be arranged in $n(n-1)(n-2)\cdots(n-r+1)$ ways. We represent this product by the symbol $nPr = \dfrac{n!}{(n-r)!}$.

Theorem 8. *The number of permutations of* n *distinct objects taken* r *at a time is*

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

Example 2.3.4. *In one year, three awards (research, teaching, and service) will be given to a class of 25 graduate students in a statistics department. If each student can receive at most one award, how many possible selections are there?*

Solution :

Since the awards are distinguishable, it is a permutation problem. The total *number of sample points is*

$$25P3 = \frac{25!}{(25-3)!} = \frac{25!}{22!} = (25)(24)(23) = 13,800.$$

Example 2.3.5. *A president and a treasurer are to be chosen from a student club consisting of 50 people. How many different choices of officers are possible if there are no restrictions*

Solution :

The total number of choices of officers, without any restrictions, is $50P2 = \frac{50!}{48!} = (50)(49) = 2450.$

Permutations that occur by arranging objects in a circle are called circular permutations. Two circular permutations are not considered different unless corresponding objects in the two arrangements are preceded or followed by a different object as we proceed in a clockwise direction. For example, if 4 people are playing bridge, we do not have a new permutation if they all move one position in a clockwise direction. By considering one person in a fixed position and arranging the other three in 3! ways, we find that there are 6 distinct arrangements for the bridge game.

Theorem 9. *The number of permutations of* n *objects arranged in a circle is* $(n-1)!$.

Theorem 10. *The number of distinct permutations of n things of which n_1 are of one kind, n_2 of a second kind, $\cdots, n_k$ of a k^{th} kind is $\dfrac{n!}{n_1!n_2!\cdots n_k!}$.*

Example 2.3.6. *In a college football training session, the defensive coordinator needs to have 10 players standing in a row. Among these 10 players, there are 1 freshman, 2 sophomores, 4 juniors, and 3 seniors. How many different ways can they be arranged in a row if only their class level will be distinguished?*

Solution :

Directly using Theorem 2.10, we find that the total number of arrangements is $\dfrac{10!}{1!2!4!3!} = 12,600.$

Often we are concerned with the number of ways of partitioning a set of n objects into r subsets called cells. A partition has been achieved if the intersection of every possible pair of the r subsets is the empty set and if the union of all subsets gives the original set. The order of the elements within a cell is of no importance. Consider the set $\{a, e, i, o, u\}$. The possible partitions into two cells in which the first cell contains 4 elements and the second cell 1 element are

$\{(a, e, i, o), (u)\}, \{(a, i, o, u), (e)\}, \{(e, i, o, u), (a)\}, \{(a, e, o, u), (i)\},$

$\{(a, e, i, u), (o)\}$. We see that there are 5 ways to partition a set of 4 elements into two subsets, or cells, containing 4 elements in the first cell and 1 element in the second. The number of partitions for this illustration is denoted by the symbol

$$\binom{5}{4,1} = \frac{5!}{4!1!} = 5,$$

where the top number represents the total number of elements and the bottom numbers represent the number of elements going into each cell.

Theorem 11. *The number of ways of partitioning a set of n objects into r cells with n1 elements in the first cell, n2 elements in the second, and so forth, is*

$$\binom{n}{n_1, n_2, \cdots, n_r} = \frac{n!}{n_1!, n_2!, \cdots, n_r!},$$

where $n_1 + n_2 + \cdots + n_r = n$

Example 2.3.7. *In how many ways can 7 graduate students be assigned to 1 triple and 2 double hotel rooms during a conference?*

Solution :

The total number of possible partitions would be

$$\binom{7}{3,2,2} = \frac{7!}{3!2!2!} = 210,$$

In many problems, we are interested in the number of ways of selecting r objects from n without regard to order. These selections are called combinations. A combination is actually a partition with two cells, the one cell containing the r objects selected and the other cell containing the (n-r) objects that are left. The number of such combinations, denoted by $\binom{n}{r,n-r}$, is usually shortened to $\binom{n}{r}$, since the number of elements in the second cell must be $n-r$.

Theorem 12. *The number of combinations of n distinct objects taken r at a time is*

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

Example 2.3.8. *A young boy asks his mother to get 5 Game-Boy cartridges from his collection of 10 arcade and 5 sports games. How many ways are there that his mother can get 3 arcade and 2 sports games?*

Solution :

The number of ways of selecting 3 cartridges from 10 is

$$\binom{10}{3} = \frac{10!}{3!(10-3)!} = 120.$$

The number of ways of selecting 2 cartridges from 5 is

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

Using the multiplication rule with $n_1 = 120$ *and* $n_2 = 10$*, we have (120)(10) = 1200 ways.*

Example 2.3.9. *How many different letter arrangements can be made from the letters in the word STATISTICS?*

Solution :

Here we have 10 total letters, with 2 letters (S, T) appearing 3 times each, letter I appearing twice, and letters A and C appearing once each. On the other hand, this result can be directly obtained by using Theorem 2.11

$$\binom{10}{3,3,2,1,1} = \frac{10!}{3!3!2!1!1!} = 50,400.$$

2.3.1 Exercises

1. Registrants at a large convention are offered 6 sightseeing tours on each of 3 days. In how many ways can a person arrange to go on a sightseeing tour planned by this convention?

2. If an experiment consists of throwing a die and then drawing a letter at random from the English alphabet, how many points are there in the sample space?

3. A certain brand of shoes comes in 5 different styles, with each style available in 4 distinct colors. If the store wishes to display pairs of these shoes showing all of its various styles and colors, how many different

pairs will the store have on display?

4. A developer of a new subdivision offers a prospective home buyer a choice of 4 designs, 3 different heating systems, a garage or carport, and a patio or screened porch. How many different plans are available to this buyer?

5. In a fuel economy study, each of 3 race cars is tested using 5 different brands of gasoline at 7 test sites located in different regions of the country. If 2 drivers are used in the study, and test runs are made once under each distinct set of conditions, how many test runs are needed?

6. A witness to a hit-and-run accident told the police that the license number contained the letters RLH followed by 3 digits, the first of which was a 5. If the witness cannot recall the last 2 digits, but is certain that all 3 digits are different, find the maximum number of automobile registrations that the police may have to check.

7. How many distinct permutations can be made from the letters of the word INFINITY ?

8. A contractor wishes to build 9 houses, each different in design. In how many ways can he place these houses on a street if 6 lots are on one

side of the street and 3 lots are on the opposite side?

2.4 Conditional Probability

We start the conditional probability by the following definition.

Definition 17. *The* ***conditional probability*** *of an event A given that event B has occurred is defined by*

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

provided that $P(B) > 0$.

Example 2.4.1. *If $P(A) = 0.4$, $P(B) = 0.5$, and $P(A \cap B) = 0.3$, then*

$$P(A|B) = 0.3/0.5 = 0.6$$

and

$$P(B|A) = P(A \cap B)/P(A) = 0.3/0.4 = 0.75.$$

We can think of the "given B" as specifying the new sample space for which we now want to calculate the probability of that part of A that is contained in B to determine $P(A|B)$. The following example illustrates this idea.

Example 2.4.2. *Suppose that*

$$P(A) = 0.7, \quad P(B) = 0.3, \quad \text{and} \quad P(A \cap B) = 0.2.$$

formally by definition, we obtain

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0.2}{0.3} = \frac{2}{3}.$$

2.5 Independent Events

For certain pairs of events, the occurrence of one of them may or may not change the probability of the occurrence of the other. In the latter case they are said to be independent events. However, before giving the formal definition of independence, let us consider an example.

Example 2.5.1. *Flip a coin twice and observe the sequence of heads and tails. The sample space is then*

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

It is reasonable to assign a probability of $1/4$ *to each of these four outcomes.*

Let

$$\begin{aligned} A &= \{\textit{heads on the first flip}\} = \{HH,\ HT\} \\ B &= \{\textit{tails on the second flip}\} = \{HT,\ TT\} \\ C &= \{\textit{tails on both flips}\} = \{TT\}. \end{aligned}$$

Now $P(B) = 2/4 = 1/2$. *However, if we are given that* C *has occurred, then* $P(B|C) = 1$ *because* $C \subset B$. *That is, the knowledge of the occurrence of* C *has changed the probability of* B. *On the other hand, if we are given that* A *has occurred,*

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

So the occurrence of A *has not changed the probability of* B. *Hence the probability of* B *does not depend upon knowledge about event* A, *so we say that* A *and* B *are independent events. That is, events* A *and* B *are independent if the occurrence of one of them does not affect the probability of the occurrence of the other. A more mathematical way of saying this is*

$$P(B|A) = P(B) \quad or \quad P(A|B) = P(A),$$

provided that $P(A) > 0$ *or, in the latter case,* $P(B) > 0$. *With the first of these equalities and the multiplication rule, we have*

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

The second of these equalities, namely $P(A|B) = P(A)$, *gives the same result:*

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

This example motivates the following definition of independent events.

Definition 18. *Events* A *and* B *are independent if and only if*

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

Otherwise, A *and* B *are called dependent events.*

Events that are independent are sometimes called statistically independent, stochastically independent, or independent in a probability sense, but in most instances we use independent without a modifier if there is no possibility of misunderstanding. It is interesting to note that the definition always holds if $P(A) = 0$ or $P(B) = 0$ because then $P(A \cap B) = 0$, since $(A \cap B) \subset A$ and $(A \cap B) \subset B$. Thus the left- and right-hand members of $P(A \cap B) = P(A)P(B)$ are both equal to zero and thus are equal to each other.

Example 2.5.2. *red die and a white dice are rolled. Let event* $A = \{4$ *on the red dice* $\}$ *and event* $B = \{$*sum of dice is odd* $\}$. *Of the* 36 *equally likely outcomes, 6 are favorable to A, 18 are favorable to B, and 3 are favorable to $A \cap B$. Thus*

$$P(A).P(B) = \frac{6}{36}.\frac{18}{36} = \frac{3}{36} = P(A \cap B).$$

Hence A and B are independent by Definition.

Example 2.5.3. *A red die and a white dice are rolled. Let event $C = \{5 \text{ on red dice}\}$ and event $D = \{ \text{ sum of dice is } 11\}$. Of the 36 equally likely outcomes, 6 are favorable to C, 2 are favorable to D, and 1 is favorable to $C \cap D$. Thus*

$$P(C)P(D) = \frac{6}{36}.\frac{2}{36} = \frac{1}{108} \neq \frac{1}{36} = P(C \cap D).$$

Hence C and D are dependent events.

Theorem 13. *If A and B are independent events, then the following pairs of events are also independent:*

(a) *A and B^c.* ***(b)*** *A^c and B.* ***(c)*** *A^c and B^c.*

Proof. We know that conditional probability satisfies the axioms for a probability function. Hence, if $P(A) > 0$, then

$$P(B^C|A) = 1 - P(B|A).$$

Thus

$$\begin{aligned} P(A \cap B^c) &= P(A)P(B^c|A) = P(A)[1 - P(B|A)] \\ &= P(A)[1 - P(B)] \\ &= P(A)P(B^c), \end{aligned}$$

since $P(B|A) = P(B)$ by hypothesis. Thus A and B^c are independent events. The proofs for parts (b) and (c) are left as exercises. □

Before extending the definition of independent events to more than two events, we present the following example.

Example 2.5.4. *An urn contains four bails numbered 1, 2, 3, and 4. One ball is to be drawn at random from the urn. Let the events A, B, and C be defined by*

$$A = \{1,\ 2\}, \quad B = \{1,\ 3\}, \quad C = \{1,\ 4\}.$$

Then

$$P(A) = P(B) = P(C) = 1/2.$$

Furthermore,

$$\begin{aligned} P(A \cap B) &= \frac{1}{4} = P(A)P(B), \\ P(A \cap C) &= \frac{1}{4} = P(A)P(C), \\ P(B \cap C) &= \frac{1}{4} = P(B)P(C), \end{aligned}$$

which implies that A, B, and C are independent in pairs (called pairwise independence). However, since $A \cap B \cap C = \{1\}$, we have

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

That is, something seems to be lacking for the complete independence of A, B, and C.

This example illustrates the reason for the second condition in the following definition.

Definition 19. *Events A, B, and C are mutually independent if and only if the following two conditions hold:*

(a) *They are pairwise independent; that is,*

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

(ii) $P(A \cap C) = P(A)P(C)$,

(iii) $P(B \cap C) = P(B)P(C)$.

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

Definition 19 can be extended to mutual independence of four or more events. In this extension, each pair, triple, quartet, and so on, must satisfy this type of multiplication rule. If there is no possibility of misunderstanding, independent is often used without the modifier mutually when considering several events.

Suppose we want to compute the probability of several events occurring simultaneously. If the events are independent, then the multiplication law of probability can be used to accomplish this. If some of the events are dependent, then some quantitative measure of dependence is needed in order to extend the multiplication law of probability to the case of dependence events.

2.6 The Multiplication Law of Probability

In this section, certain specific types of events are discussed.

Example 2.6.1 (Hypertension, Genetics)**.** *Suppose we are conducting a* *hypertension-screening program in the home. Consider all possible pairs of DBP measurements of the mother and father within a given family, assuming that the mother and father are not genetically related. This space consists of all pairs of numbers of the form* $(X < Y)$, *where* $X > 0$, $Y > 0$. *certain specific events might be of interest in the context. In particular, we might be interested in whether the mother or father is hypertensive, which is described, respectively, by the events*

$$A = \{\textit{mother's } DBP \geq 95\},$$

and

$$B = \{\textit{father's } DBP \geq 95\}.$$

Suppose we know that $P(A) = 0.1$, $P(B) = 0.2$. *What we can say about*

$$P(A \cap B) = P(\textit{both mother and father are hypertensive})?$$

We can say nothing unless we are willing to make certain assumptions.

Example 2.6.2 (Hypertension, Genetics)**.** *Compute the probability that both the mother and father are hypertensive if the events in Example 3. are independent.*

Solution. *If* A *and* B *are independent, then 3*

$$P(A \cap B) = P(A) \times P(B) = 0.1 \times 0.2 = 0.02.$$

Example 2.6.3. *Consider all possible diastolic blood-pressure measurements from a mother and her first-born child. let*

$$A = \{\textit{mother's}\, DBP \geq 95\}$$

and

$$B = \{\textit{first-born child's}\, DBP \geq 80\}.$$

Suppose

$$P(A \cap B) = 0.05, \quad P(A) = 0.1, \quad \textit{and} \quad P(B) = 0.2.$$

Then

$$P(A \cap B) = 0.05 > P(A) \times P(B) = 0.02$$

and the events A *and* B *are dependent.*

This outcome would be expected, since the mother and first-born child both share the same environment and are genetically related. In other words, the first-born child is likely to have elevated blood-pressure in households where the mother is hypertensive than in households where the mother is no hypertensive.

Definition 3.3.1 can be generalized to the case of $(k > 2)$ independent events, in the following theorem.

Theorem 14 (The Multiplication Law of Probability)**.**

If $A_1, A_2, \cdots, A_k$ are mutually independent events, then

$$P(A_1 \cap A_2 \cap \cdots \cap A_k) = P(A_1) \times P(A_2) \times \cdots \times P(A_k).$$

It is easy to show that if A and B are independent events, then

$$P(B|A^c) = P(B|A) = P(B).$$

This relationship leads to the following alternative interpretation of independence in terms of conditional probabilities:

Theorem 15.

(1) *If A and B are independent events, then*

$$P(B|A) = P(B) = P(B|A^c).$$

(2) *If two events A and B are dependent, then*

$$P(B|A) \neq P(B) \neq P(B|A^c)$$

and

$$P(A \cap B) \neq P(A) \times P(B).$$

The conditional $(P(B|A), P(B|A^c))$ and unconditional $(P(B))$ probabilities mentioned previously can be related in the following way:

Theorem 16. *For any events A and B,*

$$P(B) = P(B|A)P(A) + P(B|A^c)P(A^c).$$

Definition 20. *A set of events $A_1, A_2, \ldots, A_k$ is* ***exhaustive*** *if at least one of the events must occur.*

Assume that the events $A_1, A_2, \ldots, A_k$ are mutually exclusive and exhaustive, that is, at least one of the events $A_1, A_2, \ldots, A_k$ must occur and not two events can occur simultaneously. Thus, exactly one of the events $A_1, A_2, \ldots, A_k$ must occur simultaneously. Hence, exactly one of the events $A_1, A_2, \ldots, A_k$ must occurs.

We now formulate the following two important theorems in the general form.

Theorem 17 (Total Law of Probability)**.**

Let $A_1, A_2, \cdots, A_k$ be mutually exclusive and exhaustive events. The unconditional probability of $B, (P(B))$ can then be written as a weighted average of the conditional probabilities of B given $A_i, (P(B|A_i))$ as follows:

$$P(B) = \sum_{i=1}^{k} P(B|A_i) \times P(A_i).$$

Theorem 18 (Bayes Rule)**.** *Let $A_1, A_2, \ldots, A_n$ constitute a partition of the sample space S. That is, $S = A_1 \cup A_2 \cup \ldots \cup A_n$, and $A_i \cap A_j = \Phi \quad \forall\, i \neq j$.*

Then for any event B in the same sample space with $P(B) > 0$, we have

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^{n} P(B|A_j)P(A_j)}.$$

Proof. The event B can be written in the form:

$$\begin{aligned} B = B \cap S &= B \cap (A_1 \cup A_2 \cup \cdots \cup A_n) \\ &= (B \cap A_1) \cup (B \cap A_2) \cup \cdots (B \cap A_n). \end{aligned}$$

Thus

$$P(B) = \sum_{i=1}^{n} P(B \cap A_i) = \sum_{i=1}^{n} P(B|A_i)P(A_i). \tag{2.3}$$

If $P(B) > 0$, we have that

$$P(A_j|B) = \frac{P(A_j \cap B)}{P(B)}, \quad j = 1, 2, \ldots, n. \tag{2.4}$$

Using Equation 2.3 and replacing $P(B)$ in Equation 3.4, we have

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^{n} P(B|A_j)P(A_j)},$$

which was to be proved. □

Example 2.6.4. *In a certain factory, machines A_1, A_2, and A_3 are all producing springs of the same length. Of the production, machines A_1, A_2, and A_3 produce 2%, 1%, and 3% defective springs,respectively. Of the total* *production of springs of the factory, machine* A_1*, produce* 35%*, machine* A_2*, produce* 25%*, machine* A_3*, produce* 40%*. If one spring is selected at random from the total springs produced in a day, what is the probability that it is defective? If the selected spring is defective, find the probability that it was produced by machine* A_3*.*

Let the event D *denote the selected spring is defective. Therefore we have*

$$\begin{aligned} P(D) &= \sum_{i=1}^{3} P(A_i)P(D|A_i) \\ &= (0.35)(0.02) + (0.25)(0.01) + (0.4)(0.03) \\ &= 0.0215. \end{aligned}$$

If the selected spring is defective, the conditional probability that it was produced by machine A_3 *is, by Bayes' formula,*

$$P(A_3|D) = \frac{P(A_3)P(D|A_3)}{P(D)} = \frac{(0.4)(0.03)}{0.0215} = \frac{120}{215}.$$

Theorem 19. *For any events* $A_1, A_2, \ldots, A_n,$ *we have*

$$P(A_1 \cap A_2 \cap \cdots \cap A_n)$$

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

Example 2.6.5. *Au urn contains three red balls, two blue balls, and five white balls. A ball is selected at random and its color noted. Then it is* *replaced. A second ball is selected at random and its color noted. Find the probability of each of the following.*

(a) *Selecting two blue balls.*

(b) *Selecting a blue ball and then a white ball.*

(c) *Selecting a red ball and then a blue ball.*

Solution

Answers: **(a)** $\frac{1}{25}$ **(b)** $\frac{1}{10}$ **(c)** $\frac{3}{50}$

Example 2.6.6. *Solve Example 2.6.5 when the balls are selected without replacement.*

Solution Let

$$B_i = \{\text{the } i^{th} \text{ selected ball is blue, } i = 1, 2\},$$
$$R_i = \{\text{the } i^{th} \text{ selected ball is red, } i = 1, 2\} \text{ and}$$
$$W = \{\text{the } i^{th} \text{ selected ball is white, } i = 1, 2\}.$$

Then

(a) $P(B_1 \, and \, B_2) = P(B_1) \cdot P(B_2|B_1) = \frac{2}{10} \cdot \frac{1}{9} = \frac{1}{45}$.

(b) $P(B_1 \, and \, W_2) = P(B_1) \cdot P(W_2|B_1) = \frac{2}{10} \cdot \frac{5}{9} = \frac{1}{9}$.

(c) $P(R_1\, and\, B_2) = P(R_1) \cdot P(B_2|R_1) = \frac{3}{10} \cdot \frac{2}{9} = \frac{1}{15}$.

Example 2.6.7. *It has been found that 40 % of all people over the age of 85 suffer from Alzheimer's disease. If three people over 85 are selected at random, find the probability that at least one person does not suffer from Alzheimer's disease.*

Solution

The complement of " at least one person" is " no person", and the probability that a person over 85 does not suffer from Alzheimer's disease is 1-0.4=0.6. Hence the probability that at least one person does not suffer from Alzheimer's disease is

$$\begin{aligned} 1 - P(no\, persons\, have) &= 1 - (0.4)^3 \\ &= 1 - 0.064 = 0.936. \end{aligned}$$

Example 2.6.8. *Bowel B_1 contains two red and four white balls; bowel B_2 contains one red and two white balls; and bowel B_3 contains five red and four white balls. Assume that the probabilities of selecting the bowels are not the same but are given by*

$$P(B_1) = \frac{1}{3}, \quad P(B_2) = \frac{1}{6}, \quad and \quad P(B_3) = \frac{1}{2},$$

where B_1, B_2 , *and* B_3 *are the events that bowels* B_1, B_2 , *and* B_3 *are chosen, respectively. The experiment consists of selecting a bowel with these probabilities and then drawing a ball at random from that bowel . Let us compute the probability of event* R, *drawing a red ball, say* $P(R)$. *Note that* $P(R)$ *is dependent first of all on which bowel is selected and then of the probability of drawing a red ball from the selected bowel. That is, the event* R *is the union of the mutually exclusive events* $B_1 \cap R$, $B_2 \cap R$, *and* $B_3 \cap R$. *Thus*

$$\begin{aligned} P(R) &= P(B_1 \cap R) + P(B_2 \cap R) + P(B_3 \cap R) \\ &= P(B_1)P(R|B_1) + P(B_2)P(R|B_2) + P(B_3)P(R|B_3) \\ &= \frac{1}{3}.\frac{2}{6} + \frac{1}{6}.\frac{1}{3} + \frac{1}{2}.\frac{5}{9} = \frac{4}{9} \end{aligned}$$

Suppose now that the outcome of the experiment is red ball, but we do not know from which bowel it was drawn. Accordingly, we compute the conditional probability that the ball was from bowel B_1, *namely* $P(B_1|R)$. *From the definition of the conditional probability and the result above, we* *have that*

$$\begin{aligned}P(B_1|R) &= \frac{P(B_1 \cap R)}{P(R)} \\ &= \frac{P(B_1)P(R|B_1)}{P(B_1)P(R|B_1)+P(B_2)P(R|B_2)+P(B_3)P(R|B_3)} \\ &= \frac{(1/3)(2/6)}{(1/3)(2/6)+(1/6)(1/3)+(1/2)(5/9)} = \frac{5}{8}.\end{aligned}$$

Note that the conditional probabilities $P(R|B_1)$, $P(R|B_2)$, *and* $P(R|B_3)$ *have changes from the original probabilities* $P(B_1)$, $P(B_2)$, *and* $P(B_3)$ *in a way that agrees with your intuition. Namely, once the red ball has been observed, the probability concerning* B_3 *seems more favorable than originally because* B_3 *has a larger percentage of red balls than do* B_1 *and* B_2. *The conditional probabilities* B_1, *and* B_2 *decrease from their original ones once the red ball is observed. Frequently, the original probabilities are called prior probabilities, and the conditional probabilities are called posterior probabilities.*

2.6.1 Exercises

1. Suppose that

$$P(A) = 0.7, \; P(B) = 0.5, \; \text{and} \; P([A \cup B]^c) = 0.1$$

(a) Find $P(A \cap B)$.

(b) Give the value of $P(A|B)$.

(c) Compute $P(B|A)$.

2. Let A and B are independent events with

$$P(A) = 0.7, \quad \text{and} \quad P(B) = 0.2.$$

Compute (a) $P(A \cap B)$, (b) $P(A \cup B)$, and (C) $P(A^c \cup B^c)$.

3. Let $P(A) = 0.3$, and $P(B) = 0.6$.

(a) Find $P(A \cup B)$ when A and B are independent.

(b) Find $P(A|B)$ when A and B are mutually exclusive.

(c) $P(B|A)$ when $A \subset B$.

4. If $P(A) = 0.8$, $P(B) = 0.5$, and $P(A \cup B) = 0.9$, are A and B independent? Why?

5. Each of three persons fires one shot at a target. Let A_i denote the event that the target is hit by person i, $i = 1, 2, 3$. If we assume that A_1, A_2, A_3 are mutually independent and if

$$P(A_1) = 0.7, \quad P(A_2) = 0.9, \quad P(A_3) = 0.8,$$

compute the probability that exactly two persons hit the target.

6. Suppose that A, B, and C are mutually independent events and that

$$P(A) = 0.5, \quad P(B) = 0.8, \quad \text{and} \quad P(C) = 0.9.$$

Find the probabilities that

(a) All three events occur.

(b) Exactly two of the three events occur.

(c) None of the events occur.

7. Let A and B be two events

(a) If the events A and B are mutually exclusive, are A and B always independent? If the answer is no, can they ever be independent? Explain.

(b) If $A \subset B$, can A and B ever be independent events? Explain.

8. Bowel A contains three red balls and two white balls, and bowel B contains four red balls and three white balls. A ball is drawn at random from bowel A and transferred to bowel B. Compute the probability of then drawing a red ball from bowel B.

9. If A, B, and C are mutually independent, show that the following pairs are independent

   A & $(B \cap C)$, A & $(B \cup C)$, and A^c & $(B \cap C^c)$. Also show that A^c, B^c and C^c are mutually independent.

10. Bowel B_1 contains two white balls; bowel B_2 contains two red; bowel B_3 contains two white and two red balls; and bowel B_4 contains three white and one red balls. The probabilities of selecting the bowels B_1, B_2, B_3, and B_4 are $1/2$, $1/4$, $1/8$, and $1/8$, respectively. A bowel is selected using these probabilities, and a ball is drawn at random. Find

    (a) $P(W)$, the probability of drawing white ball.

    (b) $P(B_1|W)$, the conditional probability that bowel B_1 had been selected, given that a white ball was drawn.

Chapter 3

Discrete Distributions

3.1 Random Variables of the Discrete Type

An outcome space S may be difficult to describe if the elements of S are not numbers. We shall now discuss how we can use a rule by which each outcome of a random experiment, an element s of S, may be associated with a real number x. We begin the discussion with an example.

Example 3.1.1. *A rat is selected at random from a cage and its sex is determined. The set of possible outcome is female and male. Thus the outcome space is*

$$S = \{\textit{female,male}\} = \{F, M\}.$$

Let X *be a function that has the outcome space* S *such that*

$$X(F) = 0 \qquad \textit{and} \qquad X(M) = 1.$$

Thus X is a real-valued function that has the outcome space S as its domain and the set of real numbers $\{x : x = 0, 1\}$ as its range. We call X a random variable and, in this example, the space associated with X is $\{x : x = 0, 1\}$.

We now formulate the definition of a random variable.

Definition 21. *Given a random experiment with an outcome space* S, *a function* X *that assigns to each element* s *in* S *one and only one real number* $X(s) = x$ *is called a random variable. The space of* X *is the set of real numbers* $\{x : X(s) = x,\, s \in S\}$, *where* $s \in S$ *means that the element* s *belongs to the set* S.

Remark As we give examples of random variables and their probability distributions, the student soon recognizes that when observing a random experiment the experimenter must take some type of measurement(or measurements). This measurement can be thought of as outcome of a random variable. We would simply like to know the probability of measurement ending in A a subset of the space of X. If this is known for all subsets A, then we know the probability distribution of the random variable. Obviously, in practice, we do not very often know this distribution exactly. Hence statisticians make conjectures about these distributions; that is, we construct probabilistic models for random variables The ability of a statistician to model a real situation appropriately is a valuable trait; some of use are better than others. In this chapter we introduce some probability models in which the space of the random variables consist of sets of integers.

It may be that the set S has elements that are themselves real numbers. In such case we could write $X(s) = s$ so that X is the identity function and the space of X is also X. This is illustrated in the following example.

Example 3.1.2. *Let the random experiment be the cast of a die, observing the number of spots on the side facing up.*

The outcome space associated with this experiment is $S = \{1, 2, 3, 4, 5, 6\}$.

For each $s \in S$, *let* $X(s) = s$. *The space of the random variable* X *is then* $\{1, 2, 3, 4, 5, 6\}$. *If we associated a probability of* $\frac{1}{6}$ *with each outcome, then*

$$P(X = 5) = \frac{1}{6},\ P(2 \le X \le 5) = \frac{4}{6},\ \text{and } P(X \le 2) = \frac{2}{6},$$

for example, seem to be reasonable assignments, where

$$\{2 \le X \le 5\} \quad \text{means} \quad \{X = 2, 3, 4, \text{ or } 5\}$$

and

$$\{X \leq 2\} \quad \textit{means} \quad \{X = 1 \textit{ or } 2\},$$

in this example.

For a random variable X of the discrete type, the probability $P(X = x)$ is frequently denoted by $f(x)$, and this function $f(x)$ is called probability density function. Note that some authors refer to $f(x)$ as the probability function, the frequency function, or the probability mass function. We prefer "probability density function," and it is hereafter abbreviated pdf

Let $f(x)$ be the pdf of the random variable X of the discrete type, and let R be the space of X. Since $f(x) = P(X = x)$, $x \in R$, $f(x)$ must be positive for all $x \in R$ and we want all these probabilities to add to 1 because each $P(X = x)$ represents the fraction of times x can be expected to occur. Moreover, to determine the probability associated with the event $A \in R$, we would sum the probabilities of x values in A. This leads to the following definition.

Definition 22. *Let the pdf $f(x)$ of a discrete random variable X is a function that satisfies the following properties*

(a) $f(x) > 0, \; x \in R.$

(b) $\sum_{x \in R} f(x) = 1.$

(c) $P(X \in A) = \sum_{x \in A} f(x), \quad \textit{where} \quad A \subset R.$

We shall let $f(x) = 0$ when $x \notin R$ and thus the domain of $f(x)$ is the set of real numbers. When we define the pdf $f(x)$ and do not say zero elsewhere, then we tacitly mean that $f(x)$ has been defined at all x's in the space R, and it is assumed that $f(x) = 0$ elsewhere, namely, $f(x) = 0,\ x \notin R$. Since the probability $P(X = x) = f(x) > 0$ when $x \in R$ and since R contains all the probability associated with X, R is sometimes referred to as the support of X as well as the space of X.

Example 3.1.3. *Roll a four-sided die twice and let X equal the larger of the two outcomes if they are different and the common value if they are the same. Find the pdf $f(x)$ and show that it can be written in the form*

$$f(x) = P(X = x) = \frac{2x - 1}{16}, \qquad x = 1, 2, 3, 4. \tag{3.1}$$

solution. *The outcome space for this experiment is*

$$S = \{(d_1, d_2) : d_1 = 1, 2, 3, 4;\ d_2 = 1, 2, 3, 4\},$$

where we assume that each of these 16 points has probability $\frac{1}{16}$. Then

$$P(X = 1) = P[(1, 1)] = \frac{1}{16},$$

$$P(X=2) = P[(1,2),(2,1),(2,2)] = \frac{3}{16},$$

and similarly,

$$P(X=3) = \frac{5}{16} \quad \text{and} \quad P(X=4) = \frac{7}{16}.$$

It is easy to show that the pdf is given by (3.1) by setting $x = 1, 2, 3,$ *and* 4 *in (3.1)*

EXERCISES

(1) For each of the following, determine the constant c so that $f(x)$ satisfies the conditions of being pdf for a random variable X.

$$(a) \quad f(x) = \frac{x}{c}, \qquad x = 1, 2, 3, 4.$$

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

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

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

$$(e) \quad f(x) = \frac{x}{c} \qquad x = 1, 2, 3, \ldots, n.$$

3.2 Mathematical Expectation

Definition 23. *If* $f(x)$ *is the pdf of the random variable* X *of the discrete type with space* R *and the summation*

$$\sum_{x \in R} u(x) f(x), \text{ which is sometimes written } \sum_{R} u(x) f(x),$$

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,*

$$E[u(X)] = \sum u(x)f(x).$$

We can think of the expected value $E[u(X)]$ as a weighted mean of $u(x),\ x \in R$, where the weights are the probabilities

$$f(x) = P(X = x), \qquad x \in R.$$

Remark The usual definition of mathematical expectation of $u(X)$ requires that the sum converges absolutely; that is,

$$\sum_{x \in R} \mid u(x) \mid f(x)$$

converges and is finite.

We now list some useful facts about mathematical expectation in the following theorem.

Theorem 20. *When it exists, mathematical expectation* E *satisfies the following properties:*

(a) *If* c *is a constant, then* $E[c] = c$.

(b) *if* c *is constant and* u *is a function, then*

$$E[cu(X)] = cE[u(X)].$$

(c) *if* c_1 *and* c_2 *are constants and* u_1 *and* u_2 *are functions, then*

$$E[c_1u_1(X) + c_2u_2(X)] = c_1E[u_1(X)] + c_2E[u_2(X)].$$

Proof. First, we have for the proof of property (a) that

$$E(c) = \sum_{x\in R} cf(x) = c\sum_{x\in R} f(x) = c$$

because

$$\sum_{x\in R} f(x) = 1.$$

Next, to prove property (b), we see that

$$\begin{aligned} E[cu(X)] &= \sum_{x\in R} cu(x)f(x) \\ &= c\sum_{x\in R} u(x)f(x) = cE[u(X)]. \end{aligned}$$

Finally, the proof of property (c) is given by

$$\begin{aligned} E[c_1u_1(X) + c_2u_2(X)] &= \sum_{x\in R}[c_1u_1(x) + c_2u_2(x)]f(x) \\ &= \sum_{x\in R} c_1u_1(x)f(x) + \sum_{x\in R} c_2u_2(x)f(x). \end{aligned}$$

By applying property (b), we obtain

$$E[c_1u_1(X) + c_2u_2(X)] = c_1E[u_1(X)] + c_2E[u_2(X)].$$

□

Property (c) can be extended to more than two terms by mathematical induction; that is, we have

$$(c)' \qquad E\left[\sum_{i=1}^{k} c_i u_i(X)\right] = \sum_{i=1}^{k} c_i E[u_i(X)].$$

Because of property (c)', mathematical expectation E is often called a <u>linear or distributive</u> operator.

Example 3.2.1. *Let X have the pdf*

$$f(x) = \frac{x}{10}, \qquad x = 1, 2, 3, 4.$$

Find (i) $E[X]$, (ii) $E[X^2]$ and (iii) $E[X(5-X)]$

<u>*solution*</u> *(i)*

$$\begin{aligned} E[X] &= \sum_{x=1}^{4} x\left(\frac{x}{10}\right) \\ &= (1)\left(\frac{1}{10}\right) + (2)\left(\frac{2}{10}\right) + (3)\left(\frac{3}{10}\right) + (4)\left(\frac{4}{10}\right) \\ &= 3 \end{aligned}$$

(ii)

$$\begin{aligned} E[X^2] &= \sum_{x=1}^{4} x^2\left(\frac{x}{10}\right) \\ &= (1)^2\left(\frac{1}{10}\right) + (2)^2\left(\frac{2}{10}\right) + (3)^2\left(\frac{3}{10}\right) + (4)\left(\frac{4}{10}\right) \\ &= 10 \end{aligned}$$

(iii)

$$E[X(5-X)] = 5E[X] - E[X^2] = 5(3) - 10 = 5.$$

3.3 The Mean, Variance, and Standard Deviation

Certain mathematical expectations are so important that they have special names. In this section we consider two of them: the mean and the variance.

Definition 24. *If X is random variable of the discrete type with pdf $f(x)$ and space $R = x_1, x_2, x_3, \ldots,$ then*

$$\begin{aligned} E[X] &= \sum_{x \in R} x f(x) \\ &= x_1 f(x_1) + x_2 f(x_2) + x_3 f(x_3) + \cdots \end{aligned}$$

it is the weighted average of the numbers belonging to R, where the weights are given by the pdf $f(x)$. We call $E[x]$ the mean of X (or the mean of the distribution) and denote it by μ. That is, $\mu = E[X]$.

Remark In machines, the weighted average of the points $x_1, x_2, x_3, \ldots$ in one dimensional space is called the centroid of the system.

Example 3.3.1. *Let X have the pdf*

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

Fine the mean of X.

<u>solution</u>

The mean of X is

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

The next example shows that if the outcomes of X are equally likely (i.e, each of the outcomes has the same probability), then the mean of X is the arithmetic average of these outcomes.

Example 3.3.2. *Roll a fair die and let X denote the outcome. Thus X has the pdf*

$$f(x) = \frac{1}{6}, \quad x = 1, 2, 3, 4, 5, 6.$$

The mean of X is

$$\mu \;=\; E[X] = \sum_{x=1}^{6} x\left(\frac{1}{6}\right) = \frac{1+2+3+4+5+6}{6} = \frac{7}{2},$$

which is the arithmetic average of the first six positive integers.

We have noted that the mean $\mu = E[X]$ is the centroid of a system of weights or a measure of central location of the probability distribution of X.

Definition 25. *A measure of the dispersion or spread of a distribution is defined as follows. If*

$$u(x) = (x - \mu)^2 \quad \text{and} \quad E[(X - \mu)^2]$$

is finite, the variance*, frequently denoted by* σ^2 *or* $Var(X)$, *is defined by*

$$\sigma^2 = E[(X - \mu)^2] = \sum_{x \in R} (x - \mu)^2 f(x).$$

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

$$\sigma = \sqrt{Var(X)} = \sqrt{E[(X - \mu)^2]}.$$

Example 3.3.3. *Let* X *equal the outcome when rolling a fair die. From the above example we know that* $\mu = \dfrac{7}{2} = 3.5$. *Thus*

$$\begin{aligned}
\sigma^2 &= E[(X - 3.5)^2] = \sum_{x=1}^{6} (x - 3.5)^2 \left(\frac{1}{6}\right) \\
&= \left[(1 - 3.5)^2 + (2 - 3.5)^2 + \cdots + (6 - 3.5)^2\right] \left(\frac{1}{6}\right) \\
&= \frac{35}{12}.
\end{aligned}$$

The standard deviation of X *is* $\sigma = \sqrt{35/12} \approx 1.708$.

It is worthwhile to note that the variance can be computed in another manner. We have

$$\sigma^2 = E[(X - \mu)^2] = E[X^2 - 2\mu X + \mu^2],$$

which is, by distributive property of E, is

$$\begin{aligned}\sigma^2 &= E[X^2] - 2\mu E[X] + \mu^2 \\ &= E[X^2] - 2\mu^2 + \mu^2 = E[X^2] - \mu^2.\end{aligned}$$

Sometimes $\sigma^2 = E[X^2] - \mu^2$ provides an easier way of computing $Var(X)$ than does $\sigma^2 = E[(X-\mu)^2]$. Thus, in Example 3.3.3 we could have first computed

$$\begin{aligned}E[X^2] &= \sum_{x=1}^{6} x^2 \left(\frac{1}{6}\right) \\ &= \frac{1^2+2^2+3^2+4^2+5^2+6^2}{6} = \frac{91}{6}\end{aligned}$$

and then

$$\sigma^2 = Var(X) = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}.$$

Although most students understand that $\mu = E[X]$ is, in some sense, a measure of the middle of the distribution of X, it is difficult to get much of a feeling for the variance and the slandered deviation. The following example illustrates that the standard deviation is a measure of dispersion or spread of points belonging to the space of R.

Example 3.3.4. *Let X have the pdf*

$$f(x) = \frac{1}{3}, \qquad x = -1, 0, 1.$$

Here the mean is

$$\begin{aligned}\mu &= \sum_{x=-1}^{1} xf(x)\\ &= (-1)\left(\frac{1}{3}\right) + (0)\left(\frac{1}{3}\right) + (1)\left(\frac{1}{3}\right) = 0.\end{aligned}$$

Accordingly, the variance, denoted by σ_X^2, *is*

$$\begin{aligned}\sigma_X^2 &= E[(X-0)^2] = \sum_{x=-1}^{1} x^2 f(x)\\ &= (-1)^2\left(\frac{1}{3}\right) + (0)^2\left(\frac{1}{3}\right) + (1)^2\left(\frac{1}{3}\right) = \frac{2}{3},\end{aligned}$$

so the standard deviation is $\sigma_X = \sqrt{2/3}$.

Let another random variable Y have pdf

$$g(y) = \frac{1}{3}, \qquad y = -2, 0, 2.$$

Its mean is also zero, and it is easy to show that $Var(Y) = \frac{8}{3}$, so the standard deviation of Y is $\sigma_Y = 2\sqrt{2/3}$. Here the standard deviation of Y is twice that of the standard deviation of X, reflecting the fact that the probability of Y is spread out twice as much as that of X.

Now let X be a random variable with mean μ_X and variance σ_X^2. Of course, $Y = aX + b$, where a and b are constants, is a random variable, too. The mean of Y is

$$\mu_Y = E[Y] = E[aX + b] = aE[X] + b = a\mu_X + b.$$

Moreover, the variance of Y is

$$\begin{aligned}\sigma_Y^2 &= E[(Y - \mu_Y)^2] = E[(aX + b - a\mu_X - b)^2] \\ &= E[a^2(X - \mu_X)^2] = a^2\sigma_X^2.\end{aligned}$$

Thus $\sigma_Y =\mid a \mid \sigma_X$. For illustration, in Example 3.3.4 we note that the relationship between the two distributions could be explained by

$$Y = 2X \quad \text{so that} \quad \sigma_Y^2 = 4\sigma_X^2 \quad \text{and thus} \quad \sigma_Y = 2\sigma_X,$$

which we had observed there. In addition, we see that adding or subtracting a constant from X does not change the variance. For illustration,

$$Var(X - 1) = Var(X), \quad \text{because} \;\; a = 1 \;\; \text{and} \;\; b = -1.$$

Definition 26. *Let r be a positive integer. If*

$$E[X^r] = \sum_{x \in R} x^r f(x)$$

is finite, it is called the rth moment of the distribution about the origin. The expression moment has its origin in the study of mechanics. In addition, the *expectation*

$$E[(X-b)^r] = \sum_{x \in R} (x-b)^r f(x)$$

is called the rth moment of the distribution about b.

For a given positive integer r,

$$E[(X)_r] = E[X(X-1)(X-2)\cdots(X-r+1)]$$

is called the rth factorial moment. We note that the second factorial moment is equal to the difference of the second and the first moments:

$$E[X(X-1)] = E[X^2] - E[X].$$

Remark We say that $x_1, x_2, \ldots, x_n$ a sample if these were n observations resulting from n independent replications of a random experiment. We create the *empirical distribution* if we replace the weight (probability) of $\frac{1}{n}$ on each of the n observations $x_1, x_2, \ldots, x_n$. Then the mean of this empirical distribution is

$$\sum_{i=1}^{n} x_i \left(\frac{1}{n}\right) = \frac{\sum_{i=1}^{n} x_i}{n} = \overline{x}.$$

The sample mean $\overline{x}$ is the mean of the empirical distribution. We shall see that $\overline{x}$ is usually close in value to $\mu = E[X]$; thus, when μ is unknown, $\overline{x}$ can be used to estimate μ.

Similarly, the variance of the empirical distribution. can be computed.

Let v denote this variance so that it is equal to

$$\begin{aligned} v &= \sum_{i=1}^{n}(x_i - \overline{x})^2 \left(\frac{1}{n}\right) \\ &= \sum_{i=1}^{n} x_i^2 \left(\frac{1}{n}\right) - \overline{x}^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \overline{x}^2. \end{aligned}$$

This last statement is true because, in general.

$$\sigma^2 = E[X^2] - \mu^2.$$

Note that there is a relationship between the sample variance,

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2,$$

and the variance v of the empirical distribution, namely

$$s^2 = nv/(n-1).$$

EXERCISES

(1) Let the pdf of the random variable X is given by

$$f(0) = \frac{3}{10},\ f(1) = \frac{3}{10},\ f(2) = \frac{1}{10},\ \text{and } f(3) = \frac{3}{10}.$$

Compute the mean, variance, and standard deviation of X.

(2) Find the mean, variance, and standard deviation for the following discrete distributions:

$$
\begin{array}{llll}
(a) & f(x) = \dfrac{1}{5}, & x = 5, 10, 15, 20, 25. \\
(b) & f(x) = 1, & x = 5 \\
(c) & f(x) = \dfrac{4-x}{6}, & x = 1, 2, 3.
\end{array}
$$

(3) For the following distribution, find $\mu = E[X]$ and then find $\sigma^2 = E[(X-\mu)^2]$:

(a) $f(x) = \dfrac{1 + \mid x - 3 \mid}{10}, \qquad x = 0, 1, 2, 3.$

(b) $f(0) = \dfrac{8}{27},\ f(1) = \dfrac{12}{27},\ f(2) = \dfrac{6}{27},\ f(3) = \dfrac{1}{27}.$

(4) Given $E[X+4] = 10$ and $E[(X+4)^2] = 116$, determine **(a)** $Var(X+4)$, **(b)** μ and **(c)** σ^2.

(5) Let X equal the large outcome when a pair of four-sided dice is rolled. The pdf of the random variable X is

$$f(x) = \frac{2x-1}{16}, \qquad x = 1, 2, 3, 4.$$

(a) Find the mean, variance, and standard deviation of X.

(b) Calculate the sample mean, sample variance, and sample standard deviation of the following 100 simulation observation of X and compare these answers to those in part (a):

4 4 4 4 2 2 2 3 1 4 3 3 2 3 2 4 4 2 3 4

4 3 4 3 4 3 3 2 4 4 3 2 3 3 3 2 4 4 3 4

1 4 3 4 4 4 3 2 4 4 4 3 1 3 2 4 4 4 4 1

3 4 3 2 4 4 3 3 1 3 3 3 3 2 2 2 3 4 3 3

2 4 2 3 3 2 4 4 3 4 4 4 4 3 4 4 4 4 4 4

3.4 Bernoulli Trials and Binomial Distribution

The probability models for random experiments described in this section occur frequently in applications.

A Bernoulli experiment is a random experiment, the outcome of which can be classified in but one of two mutually exclusive and exhaustive ways, say success or failure (e.g., female or male, life or death, non defective or defective). A sequence of Bernoulli trials occurs when a Bernoulli experiment is performed several independent times so that the probability of success, say p remains the same from trial to trial. That is, in such a sequence we let p denote the probability of success on each trial. In addition, we shall frequently let $q = 1 - p$ denote the probability of failure; that is, we shall use q and $1 - p$ interchangeably.

Example 3.4.1. *Suppose that the probability of germination of a beet seed is* 0.8 *and the germination of a seed is called success. If we plant* 10 *seeds and can assume that the germination of one seed is independent of the germination of another seed, this would correspond to* 10 *Bernoulli trials with* $p = 0.8$.

Example 3.4.2. *In the Michigan daily lottery the probability of wining when placing a six-way boxed bet is* 0.006. *A bet placed on each of* 12 *successive days would correspond to* 12 *Bernoulli trials with* $p = 0.006$.

Let X be a random variable associated with Bernoulli trial by defining it as follows:

$$X(\text{success}) = 1 \qquad \text{and} \qquad X(\text{failure}) = 0.$$

That is, the two outcomes, success and failure, are denoted by one and zero, respectively. The pdf of X can be written as

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

and we say that X has Bernoulli distribution. The expected value of X is

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

and the variance of X is

$$\begin{aligned}\sigma^2 &= Var(X) = \sum_{i=0}^{1}(x-p)^2 p^x (1-p)^{1-x} \\ &= (0-p)^2(1-p) + (1-p^2)(p) = p(1-p) = pq.\end{aligned}$$

It follows that the standard deviation of X is

$$\sigma = \sqrt{p(1-p)} = \sqrt{pq}.$$

In a sequence of n Bernoulli trials, we shall let X_i denote the Bernoulli random variable associated with the ith trial. An observed sequence of n Bernoulli trials will then be an $n-$tuple of zeros and ones, we often call this collection a random sample of size n from a Bernoulli distribution.

Example 3.4.3. *Out of millions of instant lottery tickets, suppose that* 20% *are winners. If five such tickets are purchased,* $(0, 0, 0, 1, 0)$ *is a possible observed sequence in which the fourth ticket is a winner and the other four are losers. Assuming independence among winning and losing tickets, the probability of this outcome is*

$$(0.8)(0.8)(.08)(0.2)(0.8) = (0.2)(0.8)^4.$$

Example 3.4.4. *If five beet seeds are planted in a row, a possible observed sequence would be* $(1, 0, 1, 0, 1)$, *in which the first, third, and fifth seeds* *germinated and the other two did not. If the probability of germination is* $p = 0.8,$ *the probability of this outcome is, assuming independence,* $(0.8)(0.2)(.08)(0.2)(0.8) = (0.8)^3(0.2)^2.$

In a sequence of n Bernoulli trials, we are often interested in the total number of successes and not in the order of their occurrence. If we let the random variable X equal the number of observed successes In n Bernoulli trials, the possible values of X are $0, 1, 2, \ldots, n.$ If x successes occur, where $x = 0, 1, 2, \ldots, n,$ then $n - x$ failures occur. The number of ways of selecting x positions for the x successes in the n trials is

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

Since the trials are independent and since the probabilities of success and failure on each trial are, respectively p and $q = 1 - p,$ the probability of each of these ways is $p^x(1-p)^{n-x}.$ Thus the pdf of $X,$ say $f(x),$ is the sum of the probabilities of these $\binom{n}{x}$ mutually exclusive events; that is,

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

These probabilities are called binomial probabilities, and the random variable X is said to have a binomial distribution.

Summarizing, a binomial experiment satisfies the following properties:

1. A Bernoulli (success-failure) experiment is performed n times.
2. The trials are independent.
3. The probability of success on each trial is a constant p; the probability of failure is $q = 1 - p$.
4. The random variable X equals the number of successes in the n trials.

A binomial distribution will be denoted by the symbol $b(n, p)$ and we say that the distribution of X is $b(n, p)$. The constants n and p are called the parameters of the binomial distribution; they correspond to the number n of independent trials and the probability p of success on each trial. Thus, if we say that the distribution of X is $b(12, 1/4)$, we mean that X is the number of successes in a sample of $n = 12$ from a Bernoulli distribution with $p = 1/4$.

Example 3.4.5. *In the instant lottery with* 20% *winning tickets, if* X *is equal to the number of winning tickets among* $n = 8$ *that are purchased, the probability of purchasing* 2 *winning tickets is*

$$f(2) = P(X = 2) = \binom{8}{2}(0.2)^2(0.8)^6 = 0.2936$$

The distribution of the random variable X is $b(8, 0.2)$.

Example 3.4.6. *In Example 3.4.1, the number of seeds that germinate in $n = 10$ independent trials is $b(10, 0.8)$; that is,*

$$f(x) = \binom{10}{x}(0.8)^x(0.2)^{10-x}, \quad x = 0, 1, 2, \ldots, 10.$$

In particular,

$$\begin{aligned} P(X \le 8) &= 1 - P(X > 8) \\ &= 1 - [P(X = 9) + P(X = 10)] \\ &= 1 - [10(0.8)^9(0.2) + (0.8)^{10}] = 0.6242. \end{aligned}$$

Also, we could find, with a little more work,

$$P(X \le 6) = \sum_{x=0}^{6} \binom{10}{x}(0.8)^x(0.2)^{10-x}.$$

cumulative probabilities like those in the preceding example are often of interest.

Definition 27. *we call the function defined by*

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

the cumulative distribution function, or more simply, the distribution function of the random variable X.

Recall that if n is a positive integer, then

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

Thus the sum of the binomial probabilities, if we use the binomial expansion above with $b = p$ and $a = 1-p$, is

$$\sum_{x=0}^{n} \binom{n}{x} p^x (1-p)^{n-x} = [(1-p)+p]^n = 1$$

a result that had to follow from the fact that $f(x)$ is a pdf

Theorem 21. *Let X be a binomial random variable with parameters n and p that is, X is $b(n,p)$. Then*

(a) *The expectation of X is $\mu = E[X] = np$.*

(b) *The variance of X is given by*

$$\sigma_X^2 = Var(x) = np(1-p) = npq.$$

(c) *The standard deviation of X is $\sigma_X = \sqrt{npq}$.*

Proof. We use the binomial expansion to find the mean and the variance of the binomial random variable X.

(a) The expectation of X is given by

$$\mu = E[X] = \sum_{x=0}^{n} x \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}.$$

Since the first term of this sum is equal to zero, this can be written as

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

Because $x/x! = 1/(x-1)!$ when $x > 0$.

If we let $k = x - 1$ in the latter sum, we obtain

$$\begin{aligned}\mu &= \sum_{k=0}^{n-1} \frac{n!}{k!(n-k-1)!} p^{k+1} (1-p)^{n-k-1} \\ &= np \sum_{k=0}^{n-1} \frac{(n-1)!}{k!(n-k-1)!} p^k (1-p)^{n-1-k} \\ &= np(p + (1-p))^{n-1} = np.\end{aligned}$$

(b) To find the variance, we first find the second factorial moment $E[X(X-1)]$:

$$E[X(X-1)] = \sum_{x=0}^{n} x(x-1) \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}.$$

The first two terms in this summation equal zero, thus we find that

$$E[X(X-1)] = \sum_{x=2}^{n} \frac{n!}{(x-2)!(n-x)!} p^x (1-p)^{n-x}.$$

Let $k = x - 2$, we obtain

$$\begin{aligned}E[X(X-1)] &= \sum_{k=0}^{n-2} \frac{n!}{k!(n-k-2)!} p^{k+2} (1-p)^{n-k-2} \\ &= n(n-1)p^2 \sum_{k=0}^{n-2} \frac{(n-2)!}{k!(n--2-k)!} p^k (1-p)^{n-2-k} \\ &= n(n-1)p^2 (p + (1-p))^{n-2} \\ &= n(n-1)p^2.\end{aligned}$$

Thus

$$\begin{aligned}
\sigma_X^2 &= Var(X) = E[X^2] - (E[X])^2 \\
&= E[X(X-1)] + E[X] - (E[X])^2 \\
&= n(n-1)p^2 + np - (np)^2 \\
&= -np^2 + np \\
&= np(1-p).
\end{aligned}$$

(c) The standard deviation is given by

$$\sigma_X = \sqrt{Var(X)} = \sqrt{np(1-p)},$$

which complete the proof.

□

Example 3.4.7. *Suppose that observation over a large period of time has disclosed that, on the average, one out of* 10 *items produced by a process is defective. Select five items independently from the production line and test them. Let* X *denote the number of defective items among the* $n = 5$ *items. Find* $E[X]$, $Var(X)$, *and the probability of observing at most one defective item.*

<u>solution</u> *It is clear that* X *is* $b(5, 0.1)$. *Thus*

$$\mu_X = E[X] = np = 5(0.1) = 0.5,$$

and

$$\sigma_X^2 = npq = 5(0.1)(0.9) = 0.45.$$

The required probability is

$$\begin{aligned} P(X \le 1) &= \binom{5}{0}(0.1)^0(0.9)^5 + \binom{5}{1}(0.1)^1(0.9)^4 \\ &= 0.9185. \end{aligned}$$

EXERCISES

(1) An urn contains 7 red and 11 white balls. Draw one ball at random from the urn. Let $X = 1$ if a red ball is drawn, and let $X = 0$ if a white ball is drawn. Find the pdf, the mean, variance, and standard deviation of X.

(2) Suppose in exercise 1, $X = 1$ if a red ball is drawn, and let $X = -1$ if a white ball is drawn. Find the mean, variance, and standard deviation of X.

(3) On a six-question multiple-choice test there are five possible answers, of which one is correct (C) and four are incorrect (I). If a student guesses randomly and independently, find the probability of

(a) Being correct only two questions 1 and 4 (i.e., scoring $\mathrm{C,I,I,C,I,I}$).

(b) Being correct on two questions.

(4) Let p equal the proportion of all college and university who would answer "Yes" to the question, "Would you drink from the same glass as your friend if you suspected that this friend were an AIDS virus carrier?" Assume that $p = 0.1$. Let X equal the number of students out of a random sample of size $n = 9$ who would answer "Yes" to this question.

(a) How is X distributed?

(b) Find the values of the mean, variance, and standard deviation of X.

(c) Find (i) $P(X = 2)$ and (ii) $P(X \geq 2)$.

(5) A random variable X has a binomial distribution with mean 6 and variance 3.6. Find $P(X = 4)$.

(6) Suppose That 40% of American homes have a microwave oven. Let X equal the number of American homes in a random sample of $n = 25$ homes that have a microwave oven. Find the probability that

(a) X is at most 11.

(b) X is at least 7.

(c) X is equal to 8.

(d) Give the mean, variance, and standard deviation of X.

3.5 Geometric and Negative Binomial Distributions

To obtain a binomial random variable, we observed a sequence of n Bernoulli trials and counted the number of successes. Suppose now that we do not fix the number of Bernoulli trials in advance but instead continue to observe the sequence of Bernoulli trials until a certain number, say r, of successes occurs. The random variable of interest is the number of trials needed to observe the rth success.

We First discuss this problem when $r = 1$. That is, consider a sequence of Bernoulli trials with probability p of success. This sequence is observer until the first success occurs. Let X denote the trial number on which this first success occurs. For example, if F and S represent failure and success, respectively, and the sequence starts with $F,\ F,\ F, S\ , \ldots,$ Then $X = 4$. More0ver, because the trials are independent, the probability of such sequence is

$$P(X=4)=(q)(q)(q)(p)=q^3p=(1-p)^3p.$$

In general, the pdf, $f(x)=P(X=x)$, of X is given by

$$f(x)=(1-p)^{(x-1)}p, \qquad x=1,2,3,\ldots$$

because there must be $x-1$ failures before the first success that occurs in trial x. We say that X has a geometric distribution.

For a geometric series, the sum is given by

$$\sum_{k=0}^{\infty} ar^k=\sum_{k=1}^{\infty} ar^{k-1}=\frac{a}{1-r}$$

when $|x|<1$. Thus For a geometric series, the sum is given by

$$\sum_{x=1}^{\infty} f(x)=\sum_{x=1}^{\infty}(1-p)^{x-1}p=\frac{p}{1-(1-p)}=1,$$

so that $f(x)$ does satisfy the properties of a pdf

For the sum of a geometric series we also note that when k is an integer,

$$\begin{aligned} P(X>k) &= \sum_{x=k+1}^{\infty}(1-p)^{x-1}p \\ &= \frac{(1-p)^k p}{1-(1-p)} \\ &= (1-p)^k=q^k, \end{aligned}$$

and thus the value of the distribution function at a positive integer k is

$$\begin{aligned} P(X \le k) &= \sum_{x=1}^{k} (1-p)^{x-1} p \\ &= 1 - P(X > k) = 1 - (1-p)^k \\ &= 1 - q^k. \end{aligned}$$

Example 3.5.1. *Some biology students were checking the eye color for a large number of fruit flies. For the individual fly, suppose that the probability of white eyes is 1/4 and the probability of red eyes is 3/4, and that we may treat these observations as having independent Bernoulli trials.*

(a) The probability that at least four flies have to be checked for eye color to observe a white-eyed fly is given by

$$\begin{aligned} P(X \ge 4) &= P(X > 3) = q^3 \\ &= \left(\frac{3}{4}\right)^3 = \frac{27}{64} = 0.4219. \end{aligned}$$

(b) The probability that at most four flies have to be checked for eye color to observe a white-eyed fly is given by

$$\begin{aligned} P(X \le 4) &= 1 - q^4 \\ &= 1 - \left(\frac{3}{4}\right)^4 = \frac{175}{256} = 0.1055. \end{aligned}$$

It is also true that

$$\begin{aligned} P(X=4) &= P(X \le 4) - P(X \le 3) \\ &= \left[1-\left(\frac{3}{4}\right)^4\right] - \left[1-\left(\frac{3}{4}\right)^3\right] \\ &= \left(\frac{3}{4}\right)^3 \left(\frac{1}{4}\right) \end{aligned}$$

(c) In general,

$$f(x) = P(X=x) = \left(\frac{3}{4}\right)^{x-1} \left(\frac{1}{4}\right), \quad x = 1, 2, \ldots .$$

Theorem 22. *Let* X *have a geometric distribution with parameter* $0 < p < 1$. *Then*

(a) *The mean of* X *is* $\mu = E[X] = \dfrac{1}{p}$.

(b) *The variance of* X *is given by*

$$\sigma_X^2 = Var(x) = \frac{1-p}{p^2}.$$

(c) *The standard deviation of* X *is* $\sigma_X = \sqrt{(1-p)/p^2}$.

Proof. To prove the theorem, we will use the following results about the sum and the first and second derivatives of a geometric series. For $-1 < w < 1$, let

$$g(w) = \sum_{k=0}^{\infty} a w^k = \frac{a}{1-w}. \tag{3.2}$$

Then

$$g'(w) = \sum_{k=1}^{\infty} akw^{k-1} = \frac{a}{(1-w)^2} \tag{3.3}$$

and

$$g''(w) = \sum_{k=2}^{\infty} ak(k-1)w^{k-2} = \frac{2a}{(1-w)^3}. \tag{3.4}$$

Then the mean of X is given by

$$E[X] = \sum_{x=1}^{\infty} xq^{x-1}p = \frac{p}{(1-q)^2} = \frac{1}{p}.$$

To find the variance of X, we first find the second factorial moment $E[X(X-1)]$. We have

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

using Equation 3.4 with $a = pq$ and $w = q$. Thus the variance of X is

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

The standard deviation of X is

$$\sigma_X = \sqrt{\frac{1-p}{p^2}},$$

this complete the proof of Theorem 22. □

Example 3.5.2. *Continuing to Example 3.5.1, with* $p = 1/4$, *we obtain*

$$\begin{aligned} \mu &= \frac{1}{1/4} = 4 \\ \sigma^2 &= \frac{3/4}{(1/4)^2} = 12 \\ \sigma &= \sqrt{12} = 3.4641 \end{aligned}$$

We turn now to the more general problem of observing a sequence of Bernoulli trials until exactly r successes occur, where r is a fixed positive integer. Let the random variable X denote the number of trials needed to observe the rth success. That is, X is the trial number on which the rth success is observed. By the multiplication rule of probabilities, the pdf of X, say $g(x)$, equals the product of the probability

$$\binom{x-1}{r-1} p^{r-1}(1-p)^{x-r} = \binom{x-1}{r-1} p^{r-1} q^{x-r}$$

of obtaining exactly $r-1$ successes in the first $x-1$ trials and the probability p of a success on the rth trial. The pdf of X is

$$\begin{aligned} g(x) &= \binom{x-1}{r-1} p^{r}(1-p)^{x-r} \\ &= \binom{x-1}{r-1} p^{r} q^{x-r}, \quad x = r, r+1, \ldots. \end{aligned}$$

We say that X has a negative binomial distribution

Example 3.5.3. *Suppose that the biology student in Example 3.5.1 check the eye color of fruit flies until the third white-eyed fruit fly is observed.*

(a) *With* $p = 1/4$, *the probability of observing the third white-eyed fruit fly on the thirteenth trial is*

$$\binom{13-1}{3-1}\left(\frac{1}{4}\right)^3\left(\frac{3}{4}\right)^{13-3} = 0.0581$$

(b) *The probability that at most* 12 *flies have to be checked to find the third white-eyed fly is*

$$\sum_{x=3}^{12}\binom{x-1}{3-1}\left(\frac{1}{4}\right)^3\left(\frac{3}{4}\right)^{x-3} = 0.6093.$$

The reason of calling this the negative binomial distribution is the following. Consider $h(w) = (1-w)^{-r}$, the binomial $1-w)$ with the negative exponent $-r$. Using Maclaurin's series expansion, we have

$$(1-w)^{-r} = \sum_{k=0}^{\infty}\frac{h^{(k)}(0)}{k!}w^k = \sum_{k=0}^{\infty}\binom{r+k-1}{r-1}w^k. \tag{3.5}$$

If we let $x = k + r$ in the summation, then $k = x - r$ and

$$\begin{aligned}(1-w)^{-r} &= \sum_{x=r}^{\infty}\binom{r+x-r-1}{r-1}w^{x-r}\\ &= \sum_{x=r}^{\infty}\binom{x-1}{r-1}w^{x-r},\end{aligned}$$

Inn particular, we see that the sum of probabilities of for the negative binomial distribution is 1 because

$$\begin{aligned}\sum_{x=r}^{\infty} g(x) &= \sum_{x=r}^{\infty} \binom{x-1}{r-1} p^r q^{x-r} \\ &= p^r (1-q)^{-r} = 1\end{aligned}$$

problem Prove that the mean and the variance of X are given by

$$\mu = \frac{r}{p} \quad \text{and} \quad \sigma_X^2 = \frac{r(1-p)}{p^2}.$$

3.5.1 Exercises

1. If a student selects true or false at random on an examination, assuming independence among answers, determine the probability that

   (a) The first correct answer is that of question 3.

   (b) At most three questions must be answered to obtain the first correct answer.

2. For each question on a multiple-choice test, there are five possible answers of which exactly one is correct. If a student selects answers at random. Find the probability that the first question answered correctly is question 4.

3. Apples are packaged automatically in " 3-pound" bags. Suppose that 4% of the time the bag of apples weights less than 3 pounds. If you select bags randomly and weigh them in order to find one underweight bag of apples, find the probability that the number of bags that must be selected is

   (a) At least 20. **(b)**At most 20 **(c)** Exactly 20

   (d) Find the mean and the variance of the number of bags.

3.6 Poisson Distribution

Some experiments result in counting the number of times particular events occur in given times or on given physical objects. For example, we could count the number of phone calls arriving at switchboard between 9 and 10 A.M., the number of flaws in 100 feet of wire, the number of customers that arrive at a ticket window between 12 noon and 2 P.M., or the number of defects in a 100-foot roll of aluminum screen that 2 feet wide. Each count can be looked upon as a random variable associated with an approximate Poisson process provided the conditions in the following definition are satisfied.

Definition 28. *Let the number of changes that occur in a given continuous interval be counted. We have an approximate Poisson process with parameter*

$\lambda > 0$ *if the following three conditions are satisfied.*

(a) *The number of changes occurring in non-overlapping intervals are independent.*

(b) *The probability of exactly one change in sufficiently short interval of length h is approximately λh.*

(c) *The probability of two or more changes in a sufficiently short interval is essentially zero.*

Remark In the definition, we have modified the usual requirements of a Poisson process by using the words *approximately and essentially* in conditions (b) and (c). We do this to avoid some advanced mathematics. Thus we refer to this as the *approximate* Poisson process.

Suppose that an experiment satisfies the three points of an approximate Poisson process. Let X denote the number of changes in an interval of "length 1" (where "length 1" represents one unit of the quantity under consideration). We would like to find an approximation for $P(X = x)$, where x is a nonnegative integer. To achieve this, we partition the unit interval into n subintervals of length $1/n$. If n is sufficiently large (i.e. much larger than x) we shall approximate the probability that x changes occur in this unit interval by finding the probability that one change occurs in each of exactly x of these n subintervals. The probability of one change occurring in any one subinterval of length $1/n$ ia approximately $\lambda(1/n)$ by condition (b). The Probability of approximately $\lambda(1/n)$. Consider the occurrence or nonoccurrence of a change in each subinterval as Bernoulli trial. By condition (a) we have a sequence of n Bernoulli trials with probability p approximately equal to $\lambda(1/n)$. Thus an approximation of $P(X = x)$ is given by the binomial probability

$$\frac{n!}{x!(n-x)!}\left(\frac{\lambda}{n}\right)^x\left(1-\frac{\lambda}{n}\right)^{n-x}.$$

If n increases without bound, we have that

$$\begin{aligned}
&\lim_{n\to\infty}\frac{n!}{x!(n-x)!}\left(\frac{\lambda}{n}\right)^x\left(1-\frac{\lambda}{n}\right)^{n-x}\\
&= \lim_{n\to\infty}\frac{n(n-1)\cdots(n-x+1)}{n^x}\times\frac{\lambda^x}{x!}\left(1-\frac{\lambda}{n}\right)^n\left(1-\frac{\lambda}{n}\right)^{-x}.
\end{aligned}$$

Now, for fixed x, we have

$$\begin{aligned}
&\lim_{n\to\infty}\frac{n(n-1)\cdots(n-x+1)}{n^x}\\
&= \lim_{n\to\infty}\left[(1)\left(1-\frac{1}{n}\cdots\left(1-\frac{x-1}{n}\right)\right)\right] = 1,\\
&\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n = e^{-\lambda},\\
&\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{-x} = 1.
\end{aligned}$$

Thus

$$\begin{aligned} P(X=x) &= \lim_{n\to\infty} \binom{n}{x} \left(\frac{\lambda}{n}\right)^x \left(1-\frac{\lambda}{n}\right)^{n-x} \\ &= \frac{\lambda^x e^{-\lambda}}{x!}. \end{aligned}$$

The distribution of probability associated with this process has a special name. We say that the random variable X has a Poisson distribution if its pdf is of the form

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

where $\lambda > 0$.

It is easy to see that $f(x)$ enjoys the properties of a pdf because clearly $f(x) \geq 0$ and from the Maclaurin's expansion of e^{λ}, we have

$$\sum_{x=0}^{\infty} \frac{\lambda^x e^{-\lambda}}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = e^{-\lambda} e^{\lambda} = 1.$$

Theorem 23. *If the random variable X has a Poisson distribution with parameter λ, then*

$$\mu = E[X] = Var(X) = \lambda \tag{3.6}$$

Proof. The mean of the Poisson distribution is given by

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

because $(0)f(0) = 0$ and $x/x! = 1/(x-1)!$ when $x > 0$. If we let $k = x - 1$, then

$$E[X] = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^{k+1}}{k!} = \lambda e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = \lambda e^{-\lambda} e^{\lambda} = \lambda.$$

That is, the parameter λ is the mean of the Poisson distribution. We first determine the second factorial moment $E[(X(X-1)]$ in order to find the variance. We have

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

because

$$(0)(0-1)f(0) = 0, \quad (1)(1-1)f(1) = 0.$$

and

$$\frac{x(x-1)}{x!} = \frac{1}{(x-2)!}$$

when $x > 1$ If we let $k = x - 2$, then

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

Thus

$$\begin{aligned} Var(X) &= E[X^2] - (E[X])^2 \\ &= E[X(X-1)] + E[X] - (E[X])^2 \\ &= \lambda^2 + \lambda - \lambda^2 = \lambda. \end{aligned}$$

That is, for Poisson distribution, $\mu = \sigma^2 = \lambda$. □

Table III gives values of the distribution function of a Poisson random variable for selected values of λ. This table is illustrated in the next example.

Example 3.6.1. *Let X have a Poisson distribution with a mean of $\lambda = 5$. Find $P(X \le 6)$, $P(X > 5)$, and $P(X = 6)$.*

Solution

$$\begin{aligned} P(X \le 6) &= \sum_{x=0}^{6} \frac{5^x e^{-5}}{x!} = 0.762 \\ P(X > 5) &= 1 - P(X \le 5) = 1 - 0.616 = 0.384, \end{aligned}$$

and

$$\begin{aligned} P(X = 6) &= P(X \le 6) - P(X \le 5) \\ &= 0.762 - 0.616 = 0.146. \end{aligned}$$

Example 3.6.2. *Flaws (bad records) on a used computer tap occur on the average of one flaw per1200 feet. If one assume that a Poisson distribution,* *what is pdf of* X, *the number of flaws in a 4800-foot roll? Find* $(P(X = 0)$ *and* $P(X \leq 4)$

Solution The expected number of flaws in 4800-foot roll is $E[X] = \lambda = 4800/1200 = 4$ feet. Thus the pdf of X is

$$f(x) = \frac{4^x e^{-4}}{x!}, \qquad x = 0, 1, 2, \ldots,$$

and

$$\begin{aligned} P(X = 0) &= \frac{4^0 e^{-4}}{0!} = e^{-4} = 0.018, \\ P(X \leq 4) &= 0.629, \qquad \text{by Table III.} \end{aligned}$$

Example 3.6.3. *Telephone calls enter a college switchboard on the average of two every 3 minutes If one assumes an approximate Poisson process, what is the probability of five or more calls arriving in a 9-minute period?*

Solution Let X denote the number of calls in a 9-minute period. It is easy to note that $E[X] = 6$; that is, on the average, 6 calls will arrive during a 9-minute period. Thus,

$$\begin{aligned} P(X \geq 5) &= 1 - P(X \leq 4) = 1 - \sum_{x=0}^{4} \frac{6^x e^{-6}}{x!} \\ &= 1 - 0.285 = 0.715 \quad \text{using Table III.} \end{aligned}$$

Not only is the Poisson distribution important in its own right, but it can also be used to approximate probabilities for a binomial distribution. If X has a Poisson distribution with parameter λ, we saw that with n large,

$$P(X = x) \approx \binom{n}{x} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x},$$

where $p = \lambda/n$ so that $\lambda = np$ in the binomial probability above. That is, if X has the binomial distribution $b(n.p)$ with large n and small p, then

$$\frac{(np)^x e^{-np}}{x!} \approx \binom{n}{x} p^x (1-p)^{n-x}.$$

This approximation is reasonably good if n is large. But since λ was fixed constant in that earlier argument, p should be small since $np = \lambda$. In particular, the approximation is quite accurate if $n \geq 20$ and $p \leq 0.05$ or if $n \geq 100$ and $p \leq 0.1$.

Example 3.6.4. *A manufacturer of Christmas tree light bulbs knows that* 2% *of its bulbs are defective. Assume independence, approximate the probability that a box of 100 of these bulbs contains at most three defective bulbs.*

Solution Let X denote the number of defective bulbs in the box. So that X has a binomial distribution with parameters $p = 0.02$, and $n = 100$. Since n is large and p is small, we can approximate the required probability using Poisson distribution with parameter $\lambda = np = 100(0.02) = 2$, which gives

$$P(X \leq 3) \approx \sum_{x=0}^{3} \frac{2^x e^{-2}}{3!} = 0.857$$

from Table III. Using the binomial distribution, we obtain, after some tedious calculations,

$$\sum_{x=0}^{3} \binom{100}{x} (0.02)^x (0.98)^{100-x} = 0.859.$$

Hence, in this case, the Poisson approximation is extremely close to the true value, but much easier to find.

Exercises

1. Let X have a Poisson distribution with a mean of 4. Find

   (a) $P(2 \leq X \leq 5)$ **(b)** $P(X \geq 3)$

   (c) $P(X \leq 3)$.

2. Let X have a Poisson distribution with a variance of 3 find $P(X = 2)$.

3. Customers arrive at a travel agency at a mean rate of 11 per hour. Assuming that the number of arrivals per hour has a Poisson distribution, give the probability that more than 10 customers arrive in a given hour.

4. If X have a Poisson distribution with

$$3P(X = 1) = P(X = 2),$$

find $P(X = 4)$.

5. With probability 0.001, a prize of \$499 is won in the Michigan daily lottery when a \$1 straight bet is placed. Let Y equal the number of \$499 prizes won by a gambler after placing n straight bets. What is the distribution of Y? After placing $n = 2000$ \$bets, a gambler is behind if $\{Y \le 4\}$. Use the Poisson distribution to approximate $P(Y \le 4)$ when $n = 2000$.

6. Let X equal the number of telephone calls that are received at the college switchboard from on-campus during a 25-minute period. The following numbers of calls were received during each of 26 time period:

4	8	5	3	1	3	2	5	6	7	4	4	5
2	3	6	4	1	2	5	6	7	5	7	5	1

(a) Calculate the sample mean and the sample variance for these data. Are they approximately equal to each other?

(b) Assume that $\lambda = 4.2$. Compare $P(X \le 3)$ with the proportion of observations that are less than or equal 3.

(c) Compare $P(X > 5)$ with the proportion of observations that are greater than 5.

(d) Does it look like the Poisson distribution with $\lambda = 4.2$ could be the correct probability model based on these limited data?

3.7 Moment-Generating Function

Definition 29. *Let X be a random variable of the discrete type with pdf $f(x)$ and space R. If there is a positive number h such that*

$$E[e^{tX}] = \sum_{x \in \mathbb{R}} e^{tx} f(x)$$

exits and finite for $-h < t < h$, then the function of t defined by

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

is called the moment-generating function of X.

Example 3.7.1. *Consider the random variable X that has the geometric pdf*

$$f(x) = (1-p)^{x-1} p, \quad x = 1, 2, \ldots.$$

Here

$$M(t) = \sum_{x=1}^{\infty} e^{tx} (1-p)^{x-1} p = pe^t \sum_{x=1}^{\infty} [(1-p)e^t]^{x-1}.$$

The summation is the sum of a geometric series which exists provided that $(1-p)e^t < 1$ *or, equivalently,* $t < -\ln(1-p)$. *That is,*

$$M(t) = \frac{pe^t}{1-(1-p)e^t}, \quad t < -\ln(1-p).$$

For the theory of mathematical analysis, it can be shown that the existence of $M(t)$, for $-h < t < h$, implies that derivatives of $M(t)$ of all orders exist at $t = 0$; moreover, it is permissible to interchange differentiation and summation. Thus

$$\begin{aligned} M'(t) &= \sum_{x\in\mathbb{R}} xe^{tx} f(x) \\ M''(t) &= \sum_{x\in\mathbb{R}} x^2 e^{tx} f(x), \end{aligned}$$

and for each positive integer r,

$$M^{(r)} = \sum_{x\in\mathbb{R}} x^r e^{tx} f(x).$$

Setting $t = 0$, we see that

$$\begin{aligned} M'(0) &= \sum_{x\in\mathbb{R}} xf(x) = E[X], \\ M''(0) &= \sum_{x\in\mathbb{R}} x^2 f(x) = E[X^2], \end{aligned}$$

and, in general,

$$M^{(r)}(0) = \sum_{x\in\mathbb{R}} x^r f(x) = E[X^r].$$

In particular, if the moment-generating function exists,

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

Example 3.7.2. *Continuing with Example 3.7.1, suppose that we take two derivatives of the moment generating function for the geometric distribution. We have*

$$\begin{aligned} M'(t) &= \frac{[1-(1-p)e^t]pe^t - [-(1-p)pe^t]}{[1-(1-p)e^t]^2} \\ &= \frac{pe^t}{[1-(1-p)e^t]^2} \end{aligned}$$

and

$$M''(t) = \frac{pe^t + p(1-p)e^{2t}}{[1-(1-p)e^t]^3}.$$

Thus

$$\mu = M'(0) = \frac{1}{p}.$$

and

$$\begin{aligned} \sigma^2 &= Var(X) = M''(0) - [M'(0)]^2 \\ &= \frac{p+p(1-p)}{p^3} - \left(\frac{1}{p}\right)^2 = \frac{1-p}{p^2}. \end{aligned}$$

Example 3.7.3. *Let X have a binomial $b(n,p)$ with pdf*

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, 2, \ldots, n.$$

The moment-generating function of X *is*

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

Using the formula for the binomial expansion with

$$a = 1 - p \quad \text{and} \quad b = pe^t,$$

we see that

$$M(t) = [(1-p) + pe^t]^n, \quad -\infty < t < \infty$$

Example 3.7.4. *Let* X *have a Poison distribution with pdf*

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

The moment-generating function of X *is*

$$
\begin{aligned}
M(t) = E[e^{tX}] &= \sum_{x=0}^{\infty} e^{tx} \frac{\lambda^x e^{-\lambda}}{x!} \\
&= e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda e^t)^x}{x!}.
\end{aligned}
$$

From the series representation of the exponential function, we have that

$$M(t) = e^{\lambda(e^t - 1)}$$

for all real values of t. *Now the values of the mean and the variance of* X *are*

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

and

$$\sigma^2 = M''(0) - [M'(0)]^2 = \lambda.$$

EXERCISES

1. Find the moment-generating function when the pdf of X.is defined by

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

$$(b)\quad f(x)=1,\qquad x=5.$$

$$(c)\quad f(x)=\frac{5-x}{10},\qquad x=1,2,3,4.$$

$$(d)\quad f(x)=(0.3)^x(0.7)^{1-x},\qquad x=0,1$$

2. Define the pdf and give the values of μ, σ^2 when the moment-generating function of X is defined by

$$(\mathbf{a})\qquad M(t)=\frac{1}{3}+\left(\frac{2}{3}\right)e^t.$$

$$(\mathbf{b})\qquad M(t)=(0.25+0.75e^t)^{12}.$$

$$(\mathbf{c})\qquad M(t)=\exp[4.6(e^t-1)].$$

3. Let the moments of the random variable X is defined by

$$E[X^r]=p,\quad 0<p<1,\quad r=1,2,3,\ldots.$$

Find the distribution of the random variable X.

4. If the moment-generating function of X is

$$M(t)=\frac{2}{5}e^t+\frac{1}{5}e^{2t}+\frac{2}{5}e^{3t},$$

find the mean, and the variance of X.

5. **(i)** Give the name of the distribution of X (if it has a name), **(ii)** find the values of μ, and σ^2 **(iii)** calculate $P(1 \leq X \leq 2)$ when the moment generating function is given by

$$(\mathbf{a}) \qquad M(t) = \frac{0.3e^t}{1-0.7e^t}, \quad t < -\ln(0.7).$$

$$(\mathbf{b}) \qquad M(t) = (0.3 + 0.7e^t)^5.$$

$$(\mathbf{c}) \qquad M(t) = \exp[4(e^t - 1)].$$

$$(\mathbf{d}) \qquad M(t) = 0.1 + 0.3e^t + 0.4e^{2t} + 0.2e^{3t}.$$

Chapter 4

Continuous Distributions

4.1 Random Variables of the Continuous Type

Definition 30. *We say that the integrable function* $f(x)$ *is a probability density function (pdf) of a random variable* X *of the continuous type, with space* R *that is an interval or union of intervals, if the following three conditions are satisfied:*

(a) $f(x) > 0, \qquad \forall\, x \in R.$

(b) $\int_R f(x)dx = 1.$

(c) *The probability of the event* $X \in A$ *is*

$$P(X \in A) = \int_A f(x)dx.$$

Example 4.1.1. *Let the random variable X be the distance in feet between bad records on a used computer tap. Suppose that a reasonable probability model for X is given by the pdf*

$$f(x) = \frac{1}{40}e^{-x/40}, \qquad 0 \le x \le \infty.$$

Show that $f(x)$ is a pdf and find $P(X > 40)$

solution Not that

$$R = \{x : 0 \le x \le \infty\} \quad \text{and} \quad f(x) > 0, \text{ for } x \in R.$$

Also,

$$\begin{aligned}\int_R f(x)dx &= \int_0^\infty \frac{1}{40}e^{-x/40}dx \\ &= \lim_{b\to\infty}\left[-e^{-x/40}\right]_0^b \\ &= 1 - \lim_{b\to\infty} e^{-b/40} = 1.\end{aligned}$$

The probability that the distance between bad records is greater than 40 feet is

$$P(X > 40) = \int_0^\infty \frac{1}{40}e^{-x/40}dx = e^{-1} = 0.368.$$

We extend the definition of the pdf $f(x)$ to the entire set of real numbers by letting it equal zero when $x \notin R$. For example,

$$f(x) = \begin{cases} \frac{1}{40}e^{-x/40}, & 0 \leq x < \infty, \\ 0, & \text{elsewhere}, \end{cases}$$

has the properties of a pdf of a continuous-type random variable X having support $\{x : 0 \leq x < \infty\}$. It always be understood that $f(x) = 0$ when $x \notin R$, even when this is not explicitly written out.

The cumulative distribution function (cdf) of a random variable X of the continuous type, defined in terms of the pdf of X, is given by

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

Example 4.1.2. *Let Y be a continuous random variable with pdf $g(y) = 2y,\ 0 < y < 1$. Find the cdf of Y and*

$$P\left(\frac{1}{2} < Y \leq \frac{3}{4}\right) \quad \text{and} \quad P\left(\frac{1}{4} \leq Y \leq 2\right).$$

solution The cdf of X is given by

$$G(y) = \begin{cases} 0, & y < 0; \\ \int_0^y 2tdt = y^2, & 0 \leq y < 1; \\ 1, & y \geq 1. \end{cases}$$

we can evaluate the required probabilities using the cdf $G(y)$ as follows

$$\begin{aligned} P\left(\frac{1}{2} < Y \leq \frac{3}{4}\right) &= G\left(\frac{3}{4}\right) - G\left(\frac{1}{2}\right) \\ &= \left(\frac{3}{4}\right)^2 - \left(\frac{1}{2}\right)^2 = \frac{5}{16} \end{aligned}$$

and

$$\begin{aligned} P\left(\frac{1}{4} \leq Y \leq 2\right) &= G(2) - G\left(\frac{1}{4}\right) \\ &= 1 - \left(\frac{1}{4}\right)^2 = \frac{15}{16}. \end{aligned}$$

The expected value of X or mean of X is

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

The variance of X is

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

The standard deviation of is

$$\sigma = \sqrt{Var(X)}.$$

The moment-generating function of the continuous random variable X is given by

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

Moreover, results such as

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

are still valid. It is important to note that the moment-generating function, if it is finite for $-h < t < h$ for some $h > 0$, completely determines the distribution.

Example 4.1.3. *Find the mean, variance of the random variable Y of Example 2*

<u>solution</u> The mean is

$$\mu = E[Y] = \int_0^1 y(2y)dy = \frac{2}{3},$$

and the variance is given by

$$\begin{aligned} \sigma_Y^2 &= Var(Y) = E[Y^2] - \mu^2 \\ &= \int_0^1 y^2(2y)dy - \left(\frac{2}{3}\right)^2 = \frac{1}{18}. \end{aligned}$$

Example 4.1.4. *Let X have the pdf*

$$f(x) = \begin{cases} xe^{-x}, & 0 \leq x < \infty; \\ 0, & elsewhere. \end{cases}$$

Find the moment-generating function, the mean and the variance.

solution

$$\begin{aligned} M(t) &= \int_0^\infty e^{tx} x e^{-x} dx = \lim_{b\to\infty} \int_0^b x e^{-(1-t)x} dx \\ &= \frac{1}{(1-t)^2}, \end{aligned}$$

provided that $t < 1$. Note that $M(0) = 1$, which is true for every moment-generating function. Now

$$M'(t) = \frac{2}{(1-t)^3} \quad \text{and} \quad M''(t) = \frac{6}{(1-t)^4}.$$

Thus

$$\mu = E[X] = M'(0) = 2$$

and

$$\sigma^2 = M''(0) - [M'(0)]^2 = 6 - 2^2 = 2.$$

EXERCISES

1. Let the random variable X have the pdf

$$f(x) = \begin{cases} 2(1-x), & 0 \le x \le 1; \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Determine the distribution function of X.

(b) Find

(i) $P\left(0 \le X \le \frac{1}{2}\right)$,

(ii) $P\left(\frac{1}{4} \leq X \leq \frac{3}{4}\right)$,

(iii) $P\left(X = \frac{3}{4}\right)$, and $P\left(X \geq \frac{3}{4}\right)$.

2. For each of the following functions,

(i) find the constant c so that $f(x)$ is a pdf of a random variable X,

(ii) Find the distribution function,

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

$$(a)\quad f(x) = 4x^c, \qquad 0 \leq x \leq 1.$$

$$(b)\quad f(x) = c\sqrt{x}, \qquad 0 \leq x \leq 4.$$

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

3. For each of the distributions in Exercise 3, find μ, σ^2, and σ.

4. The pdf of the random variable Y is

$$g(y) = \begin{cases} \dfrac{d}{y^3}, & 1 < y < \infty; \\ 0, & \text{elsewhere.} \end{cases}$$

(a) calculate the value of d do that $g(y)$ is a pdf

(b) Find the expected value of Y.

(c) Show that the variance of Y is not finite.

5. Let $R(t) = \ln M(t)$, where $M(t)$ is the moment-generating function of a random variable of the continuous type. Show that

$$\textbf{(a)} \quad \mu = R'(0), \qquad \textbf{(b)} \quad \sigma^2 = R''(0).$$

6. If $M(t) = (1-t)^{-2}$, $t < 1$, use $R = \ln M(t)$ and the result in Exercise 5 to find μ, and σ^2.

7. The logistic distribution is associated to the distribution function

$$F(x) = (1 + e^{-x})^{-1}, \quad -\infty < x < \infty.$$

Find the pdf of the logistic distribution and show that its graph is symmetric about $x = 0$.

4.2 The Uniform and Exponential Distributions

Let the random variable X denote the outcome when a point is selected at random from an interval

$$[a, b], \quad -\infty < a < b < \infty.$$

If the experiment is performed in fair manner, it is reasonable to assume that the probability that the point is selected from the interval

$$[a, b], \quad a \le x < b \quad \text{is} \quad \frac{x-a}{b-a}.$$

That is , the probability is proportional to the length of the interval so that the distribution function of X is

$$F(x) = \begin{cases} 0, & x < a; \\ \dfrac{x-a}{b-a}, & a \le x < b; \\ 1, & b \le x. \end{cases}$$

Because X is a continuous-type random variable, $F'(x)$ is equal to the pdf of X whenever $F'(x)$ exists; thus when $a < x < b$, we have

$$f(x) = F'(x) = \frac{1}{b-a}.$$

The random variable X has a uniform distribution if its pdf is equal to a constant on its support. In particular, if the support is the interval $[a, b]$, then

$$f(x) = \frac{1}{b-a}, \qquad a \le x \le b$$

Theorem 24. *The mean, variance, and moment- generating function of the uniform distribution are given by*

$$\mu = \frac{a+b}{2}, \qquad \sigma^2 = \frac{(b-a)^2}{12},$$

and

$$M(t) = \begin{cases} \dfrac{e^{tb} - e^{ta}}{t(b-a)}, & t \neq 0; \\ 1, & t = 0. \end{cases}$$

We turn now to a continuous distribution that is related to the Poisson distribution. when previously observing a process of the (approximate) Poisson type, we counted the number of changes occurring in a given interval. This number was a discrete random variable with a Poisson distribution. But not only is the number of changes a random variable; the waiting times between successive changes are also random variables. However, the latter are of the continuous type, since each of them can assume any positive value. In particular, let W denote the waiting time until the first change occurs when observing a Poisson process in which the mean number of changes in the unit interval is λ. Then W is a continuous-type random variable, and we proceed to find its distribution function.

Because this waiting time is nonnegative, the distribution function $F(w) = 0,\ w < 0$. For $w \geq 0$,

$$\begin{aligned} F(w) &= P(W \leq w) = 1 - P(W > w) \\ &= 1 - P(\text{no changes in}[0, w]) \\ &= 1 - e^{-\lambda w}, \end{aligned}$$

since it is known that $e^{-\lambda w}$ equals the probability of no changes in an interval of length w. That is, if the mean number of changes per unit interval is λ, then the mean number of changes in an interval of length w is proportional to w, namely, λw. Thus, when $w > 0$, the pdf of W is given by

$$F'(w) = f(w) = \lambda e^{-\lambda w},$$

We often let $\lambda = 1/\theta$ and say that the random variable X has an exponential distribution if its pdf is given by

$$f(x) = \frac{1}{\theta} e^{-x/\theta}, \qquad 0 \leq x < \infty,$$

where the parameter $\theta > 0$. Accordingly, the waiting time W until the first change in a Poisson process has an exponential distribution with $\theta = \frac{1}{\lambda}$. To determine the exact meaning of the parameter θ, we first find the moment-generating function. It is

$$\begin{aligned} M(t) &= \int_0^\infty e^{tx} \left(\frac{1}{\theta}\right) e^{-x/\theta} dx \\ &= \lim_{b \to \infty} \int_0^b \left(\frac{1}{\theta}\right) e^{-(1-\theta t)x/\theta} dx \\ &= \frac{1}{1-\theta t}, \quad t < \frac{1}{\theta}. \end{aligned}$$

Thus

$$M'(t) = \frac{\theta}{(1-\theta t)^2} \quad \text{and} \quad M''(t) = \frac{2\theta^2}{(1-\theta t)^3}.$$

Hence, for the exponential distribution, we have

$$\mu = \theta \qquad \text{and} \qquad \sigma^2 = \theta^2.$$

So if λ is the mean number of changes in the unit interval, then $\theta = 1/\lambda$ is the mean waiting time for the first change. In particular, suppose that $\lambda = 5$ is the mean number of changes per minute; then the mean waiting for the first change is $1/5$ of a minute.

Let X have an exponential distribution with mean $\mu = \theta$. Then the distribution function of X is

$$F(x) = \begin{cases} 0, & -\infty < x < 0; \\ 1 - e^{-x/\theta}, & 0 \le x < \infty. \end{cases}$$

The median, m, is found by solving $F(m) = 0.5$. That is,

$$1 - e^{-m/\theta} = 0.5$$

Thus

$$m = -\theta \ln(0.5) = \theta \ln(2).$$

So that with $\theta = 5$, the median is

$$m = -5 \ln(0.5) = 3.466$$

It is useful to note that for the exponential random variable, X, we have that

$$P(X > x) = 1 - F(x) = 1 - (1 - e^{-x/\theta}) = e^{-x/\theta},$$

where $x > 0$.

Example 4.2.1. *Customers arrive in a certain shop according to an approximate Poisson process at a mean rate of* 20 *pre hour. What is the probability that the shopkeeper will have to wait more than* 5 *minutes for the arrival of the first customer?*

solution Let X denote the waiting time in minutes until the first customers arrives and note that $\lambda = 1/3$ is the expected number of arrivals per minute. Thus

$$\theta = \frac{1}{\lambda} = 3$$

and

$$f(x) = \frac{1}{3}e^{-(1/3)x}, \qquad 0 \leq x < \infty.$$

Hence

$$P(X > 5) = \int_5^\infty \frac{1}{3}e^{-(1/3)x}dx = e^{-5/3} = 0.1889.$$

The median time until the first arrival is

$$m = -3\ln(0.5) = 2.0794.$$

Example 4.2.2. *Suppose that the life of a certain type of electronic component has an exponential distribution with mean life of* 500 *hours.*

(a) *Find the probability that this component will be in operation more than c hours ($c > 0$).*

(b) *Suppose that the component has been in operation for* 300 *hours.Find the conditional probability that it will last for another* 600.

<u>solution</u> **(a)** If X denotes the life of this component (or the time to failure of this component), then

$$P(X > c) = \int_c^{\infty} \frac{1}{500} e^{-(1/500)t} dt = e^{-c/500}.$$

(b) The conditional probability that it will last for another 600 hours is

$$\begin{aligned} P(X > 900 | X > 300) &= \frac{P(X > 900)}{P(X > 300)} \\ &= \frac{\exp[-900/500]}{\exp[-300/500]} = e^{-6/5}. \end{aligned}$$

It is important to note that this conditional probability is exactly equal to

$$P(X > 600) = e^{-6/5}.$$

That is, the probability that it will last an additional 600 hours, given that it has operated 300 hours, is the same as the probability that it would last

600 hours when first put into operation. Thus, for such components, an old component is good as a new one, and we say that the failure rate is constant.

EXERCISES

1. Find the mean, variance, and the moment-generating function of the uniform distribution and the exponential distribution.

2. If the moment-generating of X is

$$M(t) = \frac{e^{5t} - e^{4t}}{t}, \quad t \neq 0, \quad \text{and} \quad M(0) = 1,$$

Find $E[X]$, $Var(X)$, and $P(4.2 < X \leq 4.7)$.

3. Let Y have the uniform distribution over $(0, 1)$, that is, X is $U(0, 1)$, and let

$$W = a + (b - a)Y, \quad a < b$$

(a) Find the distribution of W.

(hint: Find $P[a + (b - a)Y \leq w]$)

(b) How is W distributed?

4. Let the pdf of X is given by

$$f(x) = \frac{1}{2} e^{-x/2}, \qquad 0 \leq x < \infty.$$

(a) What are the mean, variance, and the moment-generating function of X?

(b) calculate $P(X > 3)$.

(c) calculate $P(X > 3|X > 2)$.

5. Determine the mean, variance of X if the moment-generating function of X is given by

$$(a) \quad M(t) = \frac{1}{1-3t}, \qquad t < 1/3.$$
$$(b) \quad M(t) = \frac{3}{3-t}, \qquad t < 3.$$

6. Let X equal the number of students who use a college card catalog every 15 minutes. Assume that X has a Poisson distribution of mean 5. Let W equal the time in minutes between two student arrivals.

(a) How is W distributed?

(b) Find $P(W > 6)$.

(c) Find $P(W > 12|W > 6)$.

7. Let X have a logistic distribution with pdf

$$f(x) = \frac{e^{-x}}{(1+e^{-x})^2}, \quad -\infty < x < \infty.$$

Show that $Y = \dfrac{1}{1+e^{-X}}$ has a $U(0,1)$ distribution. Hint: find

$$G(y) = P(Y \le y) = P\left(\frac{1}{1+e^{-X}} \le y\right),$$

when $0 < y < 1$.

4.3 The Gamma and chi-Square Distributions

In the (approximate) Poisson process with mean λ, we have seen that the waiting time until the first change has an exponential distribution. We now let W denote the waiting time until the αth change occurs and find the distribution of W.

The distribution function of W, when $w \geq 0$, is given by

$$\begin{aligned} F(w) &= P(W \leq w) = 1 - P(W > w) \\ &= 1 - P(\text{fewer than } \alpha \text{ changes occur in } [0, w]) \\ &= 1 - \sum_{k=0}^{\alpha-1} \frac{(\lambda w)^k e^{-\lambda w}}{k!} \qquad (4.1) \end{aligned}$$

since the number of changes in the interval $[0, w]$ has a Poisson distribution with mean λw. Because W is a continuous-type random variable, $F'(w)$ is equal to the pdf of W whenever this derivative exists. We have, provided that $w > 0$,

$$\begin{aligned} F'(w) &= \lambda e^{-\lambda w} - e^{-\lambda w} \sum_{k=1}^{\alpha-1} \left[\frac{k(\lambda w)^{k-1}\lambda}{k!} - \frac{(\lambda w)^k \lambda}{k!} \right] \\ &= \lambda e^{-\lambda w} - e^{-\lambda w} \left[\lambda - \frac{\lambda(\lambda w)^{\alpha-1}}{(\alpha-1)!} \right] \\ &= \frac{\lambda(\lambda w)^{\alpha-1}}{(\alpha-1)!} e^{-\lambda w}. \end{aligned}$$

If $w < 0$, then $F(w) = 0$ and $F'(w) = 0$. A pdf of this form is said to be of the gamma type, and the random variable W is said to have a gamma distribution.

Before determining the characteristics of the gamma distribution, let us consider the gamma function for which the distribution is named. The <u>gamma function</u> is defined by

$$\Gamma(t) = \int_0^\infty y^{t-1} e^{-y} dy, \qquad t > 0.$$

This integral is positive for $t > 0$ because the integrand is positive. Values of it are often given in a table of integrals. If $t > 1$, integration of the gamma function of t by parts yields

$$\Gamma(t) = (t-1)\Gamma(t-1).$$

Whenever $t = n$, a positive integer, we have, by repeated application of $\Gamma(t) = (t-1)\Gamma(t-1)$, that

$$\Gamma(n) = (n-1)\Gamma(n-1) = (n-1)(n-2)\cdots(2)(1)\Gamma(1).$$

However,

$$\Gamma(1) = \int_0^\infty e^{-y} dy = 1.$$

Thus, when n is positive integer, we have

$$\Gamma(n) = (n-1)!;$$

and, for this reason, the gamma function is called the generalized factorial. Incidentally, $\Gamma(1)$ corresponds to $0!$, and we have noted that $\Gamma(1) = 1$, which is consistent with the earlier discussions.

Let us now formally define the pdf of the gamma distribution and find its characteristics. The random variable X has a gamma distribution if its pdf is defined by

$$f(x) = \frac{1}{\Gamma(\alpha)\theta^{\alpha}} x^{\alpha-1} e^{-x/\theta}, \qquad 0 \leq x < \infty.$$

Hence, W, the waiting time until the αth change in a Poisson process, has a gamma distribution with parameters α and $\theta 1/\lambda$. To see that $f(x)$ actually has the properties of a pdf, note that $f(x) \geq 0$ and

$$\int_{-\infty}^{\infty} f(x) = \int_{0}^{\infty} \frac{1}{\Gamma(\alpha)\theta^{\alpha}} x^{\alpha-1} e^{-x/\theta} dx = 1$$

by using the change of variables $y = x/\theta$.

Theorem 25. *The moment-generating function of the gamma distribution is*

$$M(t) = (1 - \theta t)^{-\alpha}, \qquad t < \frac{1}{\theta}.$$

The mean and the variance are given by

$$\mu = \alpha\theta \qquad \text{and} \qquad \sigma^2 = \alpha\theta^2.$$

Example 4.3.1. *Suppose that an average of* 30 *customers per hour arrive at a shop in accordance with a Poisson process. That is, if a minute is our unit, then* $\lambda = 1/2$. *What is the probability that the shopkeeper will wait more than* 5 *minutes before both of the first two customers arrive?*

solution If X denotes the waiting time in minutes until the second customer arrives, then X ha a gamma distribution with $\alpha = 2$, $\theta = 1/\lambda = 2$. Hence

$$\begin{aligned} P(X > 5) &= \int_5^\infty \frac{x^{2-1}e^{-x/2}}{\Gamma(2)2^2}\,dx \\ &= \frac{1}{4}\left[(-2)xe^{-x/2} - 4e^{-x/2}\right]_5^\infty \\ &= \frac{7}{2}e^{-5/2} = 0.287. \end{aligned}$$

Example 4.3.2. *Telephone calls arrive at a switchboard at a mean rate of* $\lambda = 2$ *per minute according to a Poisson process. Let* X *denote the waiting time in minutes until the fifth call arrives. The pdf of* X, *with* $\alpha = 5$ *and* $\theta = 1/\lambda$, *is*

$$f(x) = \frac{2^5 x^4}{4!}e^{-2x}, \quad 0 \le x < \infty.$$

The mean and the variance of X *are, respectively,* $\mu = 5/2$ *and* $\sigma^2 = 5/4$.

We now consider a special case of the gamma distribution that plays an important role in statistics. Let X have a gamma distribution with

$\theta = 2$ and $\alpha = r/2$, where r is a positive integer. The pdf of X is

$$f(x) = \frac{1}{\Gamma(r/2)2^{r/2}}\, x^{r/2-1} e^{-x/2}, \quad 0 \leq x < \infty.$$

We say that X has a chi-square distribution with r degrees of freedom, which we abbreviate by saying X is $\chi^2(r)$. The mean and the variance of this chi-square distribution are

$$\mu = \alpha\theta = \left(\frac{r}{2}\right) = r, \quad \text{and} \quad \sigma^2 = \alpha\theta^2 = \left(\frac{r}{2}\right) 2^2 = 2r.$$

That is, the mean equals the number of degrees of freedom, and the variance equals twice the number of degrees of freedom. The moment-generating function is

$$M(t) = (1-2t)^{-r/2}, \qquad t < \frac{1}{2}.$$

Because the chi-square distribution is so important in applications, tables have been prepared giving the values of the distribution function

$$F(x) = \int_0^x \frac{1}{\Gamma(r/2)2^{r/2}}\, w^{r/2-1} e^{-w/2} dw$$

for selected values of r and x. For an example, see Table v.

Example 4.3.3. *Let X have a chi-square distribution with $r = 5$ degrees* *of freedom. Then, using Table IV, we obtain*

$$\begin{aligned} P(1.145 \leq X \leq 12.83) &= F(12.83) - F(1.145) \\ &= 0.975 - 0.050 = 0.925 \end{aligned}$$

and

$$P(X > 15.09) = 1 - F(15.09) = 1 - 0.99 = 0.01.$$

EXERCISES

1. If x has a gamma distribution with $\theta = 4$ and $\alpha = 2$, find $P(X < 5)$.

2. Find the moment-generating function of a gamma distribution with parameters α and θ.

3. Use the moment-generating function of a gamma distribution to show that

$$E[X] = \alpha\theta \quad \text{and} \quad V(X) = \alpha\theta^2.$$

4. If the moment generating function of a random variable W is

$$M(t) = (1 - 7t)^{-20},$$

find the pdf, the mean, and the variance of W.

5. If the moment-generating function of X is

$$M(t) = (1 - 2t)^{-12}, \qquad t < \frac{1}{2},$$

find

(a) $E[X]$.

(b) $Var(X)$.

(c) $P(15.66 < X < 42.98)$.

6. Find the point at which a chi-square pdf obtains its maximum when $r > 2$.

4.4 The Normal Distribution

The normal distribution is perhaps the most important distribution in statistical applications since many measurements have (approximate) normal distributions. One explanation of this fact is the role of the normal distribution in the central limit theorem. One form of this theorem is considered below.

We give the definition of the pdf of the normal distribution. The random variable X has a *normal distribution* if its pdf is defined by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right], \quad -\infty < x < \infty,$$

where μ and σ are parameters satisfying

$$-\infty < \mu < \infty, \quad 0 < \sigma < \infty.$$

The moment-generating function of the normal distribution is

$$M(t) = \exp\left[\mu t + \frac{\sigma^2 t^2}{2}\right].$$

Now

$$M'(t) = (\mu + \sigma^2 t) \exp\left[\mu t + \frac{\sigma^2 t^2}{2}\right]$$

and

$$M''(t) = [(\mu + \sigma^2 t)^2 + \sigma^2] \exp\left[\mu t + \frac{\sigma^2 t^2}{2}\right].$$

Thus

$$\begin{aligned} E[X] &= M'(0) = \mu, \\ Var(X) &= M''(0) - [M'(0)]^2 = \mu^2 + \sigma^2 - \mu^2 = \sigma^2. \end{aligned}$$

That is, the parameters μ and σ^2 in the pdf of X are the mean and the variance of X.

Example 4.4.1. *If the pdf of* X *is*

$$f(x) = \frac{1}{\sqrt{32\pi}} \exp\left[-\frac{(x+7)^2}{32}\right], \quad -\infty < x < \infty,$$

then X *have the normal distribution With mean*

$\mu = -7$ *and variance* $\sigma^2 = 16$ *which can be written as* X *is* $N(\mu, \sigma^2)$,

in this example X *is* $N(-7, 16)$.

If Z is $N(0,1)$, we shall say that Z has a standard normal distribution and the pdf of Z is

$$\phi(z) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{z^2}{2}\right], \quad -\infty < z < \infty.$$

Moreover, the cumulative distribution function of Z is given by

$$\Phi(z) = P(Z \le z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-w^2/2} dw.$$

It is impossible to evaluate this integral exactly. However, numerical approximations for integrals of this type have been tabulated and are given in tables. The function $\phi(z)$ have the following properties:

1. $\phi(z) > 0$ for all $-\infty < z < \infty$.

2. $\phi(z)$ is even function (i.e. $\phi(z)$ is symmetric about the line $z = 0$.)

3. The function $\phi(z)$ accomplished its maximum when $z = 0$.

4. $\lim\limits_{z\to-\infty}\phi(z)=\lim\limits_{z\to\infty}\phi(z)=0.$

5. $\int\limits_{-\infty}^{\infty}\phi(z)dz=1.$

6. $\mu_Z=E[Z]=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}ze^{-z^2/2}dz=0.$

7. $\sigma_Z^2=E[Z^2]-\mu_Z^2=1.$

Theorem 26. *The standard normal random variable Z and its cumulative distribution function $\Phi(z)$ have the following properties:*

1. $P(a\le Z\le b)=P(a<Z<b)=\Phi(b)-\Phi(a),$

2. $P(Z>a)=1-P(Z\le a)=1-\Phi(a),$

3. $\Phi(-a)=1-\Phi(a),$

4. $P(Z>-a)=\Phi(a),$

5. $P(-a\le Z\le -b)=\Phi(a)-\Phi(b),$

6. $P(-a\le Z\le b)=\Phi(b)+\Phi(a)-1.$

Proof. We only prove properties (3), (4), and (5) while the remaining properties are similar. For property (3) :

$$\begin{aligned}\Phi(-a) &= P(Z\le -a)=P(Z\ge a)\\ &= 1-P(Z<a)=1-\Phi(a).\end{aligned}$$

To prove (4):

$$\begin{aligned} P(Z > -a) &= 1 - P(Z \leq -a) = 1 - \Phi(-a) \\ &= 1 - [1 - \Phi(a)] = \Phi(a). \end{aligned}$$

Finally (5) can be proved as follows:

$$\begin{aligned} P(-a \leq Z \leq b) &= \Phi(-b) - \Phi(-a) \\ &= 1 - \Phi(b) - [1 - \Phi(a)] \\ &= \Phi(a) - \Phi(b). \end{aligned}$$

Hence the theorem is proved. $\square$

Example 4.4.2. *If* Z *is* $N(0,1)$, *find the following probabilities:*

(a) $P(0 \leq Z \leq 2)$, (b) $P(1.25 \leq Z \leq 2.75)$,

(c) $P(-1.65 \leq Z \leq 0.60)$, (d) $P(Z > -1.77)$,

(e) $P(\mid Z \mid \leq 1.96)$, (f) $P(Z < -1.65)$.

solution. First note that $\Phi(z_0) = P(Z \leq z_0)$ see table i.

$$\begin{aligned} (a) \qquad P(0 \leq Z \leq 2) &= \Phi(2) - \Phi(0) \\ &= 0.9772 - 0.5000 = 0.4772. \end{aligned}$$

$$(b) \quad P(1.25 \le Z \le 2.75) = \Phi(2.75) - \Phi(1.25)$$
$$= 0.9970 - 0.8944 = 0.1026.$$

$$\begin{aligned}(c) \quad P(-1.65 \le Z \le 0.70) &= \Phi(0.70) - \Phi(-1.65) \\ &= \Phi(0.70) - [1 - \Phi(1.65)] \\ &= 0.7580 - 1 + 0.9505 \\ &= 0.7085.\end{aligned}$$

$$(d) \qquad P(Z \ge -1.77) = \Phi(1.77) = 0.9616.$$

$$\begin{aligned}(e) \qquad P(\mid Z \mid < 1.96) &= P(-1.96 < Z < 1.96) \\ &= \Phi(1.96) - \Phi(-1.96) \\ &= \Phi(1.96) - [1 - \Phi(1.96)] \\ &= 2\Phi(1.96) - 1 = 0.95.\end{aligned}$$

$$\begin{aligned}(f) \qquad P(Z < -1.65) &= 1 - \Phi(1.65) \\ &= 1 - 0.9505 = 0.0495\end{aligned}$$

Example 4.4.3. *If Z is $N(0,1)$, find the constants $a, b,$ and C such that:*

(a) $P(Z \leq a) = 0.9147$, (b) $P(0 \leq Z \leq a) = 0.4147$,

(c) $P(Z > b) = 0.05$, (d) $P(| Z |\leq c) = 0.95$.

solution.

$$(a) \qquad P(Z \leq a) = \Phi(a) = 0.9147$$

find the given probability in the normal table and read off the corresponding value of z. From the table we see that $a = 1.37$.

(b) Since

$$P(0 \leq Z \leq a) = \Phi(a) - \Phi(0) = \Phi(a) - 0.5000 = 0.4147,$$

we have

$$\Phi(a) = 0.9147 \quad \text{which implies} \quad a = 1.37.$$

(c) Clearly,

$$P(Z > b) = 1 - P(Z \leq b) = 1 - \Phi(b),$$

that is

$$1 - \Phi(b) = 0.05 \quad \text{so that} \quad \Phi(b) = 0.950,$$

which implies $b = 1.645$. (verify this result from the standard Normal table)

(d) Since

$$\begin{aligned} P(\mid Z \mid \leq c) &= P(-c \leq Z \leq c) \\ &= \Phi(c) - \Phi(-c) \\ &= 2\Phi(c) - 1 = 0.95, \end{aligned}$$

we have

$$\Phi(c) = 0975, \quad \text{which implies} \quad c = 1.96.$$

Theorem 27. *If* X *is* $N(\mu, \sigma^2)$, *then*

$$Z = \frac{X - \mu}{\sigma} \quad is \quad N(0, 1).$$

Proof. The cumulative distribution function of Z is

$$\begin{aligned} P(Z \leq z) &= P\left(\frac{X - \mu}{\sigma} \leq z\right) \\ &= P(X \leq \sigma z + \mu) \\ &= \int_{-\infty}^{\sigma z + \mu} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x - \mu)^2}{2\sigma}\right] dx. \end{aligned}$$

In the integral representing $P(Z \leq z)$, use the change of variable of integration given by

$$w = \frac{x - \mu}{\sigma} \quad \text{or} \quad x = \sigma w + \mu$$

to obtain

$$P(Z \le z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}}\, e^{-w^2/2}\, dw.$$

But this is the expression for $\Phi(z)$, the standard normal random variable. Hence Z is $N(0,1)$. $\square$

This theorem can be used to find probabilities a bout X, which is $N(\mu, \sigma^2)$, as follows

$$\begin{aligned} P(a \le X \le b) &= \left(\frac{a-\mu}{\sigma} \le \frac{X-\mu}{\sigma} \le \frac{b-\mu}{\sigma}\right) \\ &= \left(\frac{a-\mu}{\sigma} \le Z \le \frac{b-\mu}{\sigma}\right) \\ &= \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right). \end{aligned}$$

since $(X-\mu)/\sigma$ is $N(0,1)$.

Example 4.4.4. *If* X *is* $N(3, 16)$, *find the following probabilities:*

$$(a)\, P(4 \le X \le 8), \quad (b)\, P(0 < X \le 5), \quad (c)\, P(-2 \le X \le 1).$$

<u>*Solution*</u>

$$\begin{aligned} (a) \quad P(4 \le X \le 8) &= \left(\frac{4-3}{4} \le \frac{X-3}{4} \le \frac{8-3}{4}\right) \\ &= P(0.25 \le Z \le 1.25 \\ &= \Phi(1.25) - \Phi(0.25) \\ &= 0.8944 - 0.5987 = 0.2957. \end{aligned}$$

$$(b) \quad P(0 < X \le 5) = \left(\frac{0-3}{4} < \frac{X-3}{4} \le \frac{5-3}{4}\right)$$
$$= P(-0.75 < Z \le 0.5)$$
$$= \Phi(0.5) - \Phi(-0.75)$$
$$= 0.6915 - 0.2266 = 0.4649.$$

$$(c) \quad P(-2 \le X \le 1) = \left(\frac{-2-3}{4} \le \frac{X-3}{4} \le \frac{1-3}{4}\right)$$
$$= P(-1.25 \le Z \le -0.5)$$
$$= \Phi(-0.5) - \Phi(-1.25)$$
$$= 0.3085 - 0.1056 = 0.2029.$$

Example 4.4.5. *If X is $N(25, 36)$, find a constant c such that*

$$P(\mid X - 25 \mid \le c) = 0.9544.$$

Solution since

$$P(\mid X - 25 \mid \le c) = P(-c \le X - 25 \le c),$$

we have

$$P\left(-\frac{c}{6} \le \frac{X-25}{6} \le \frac{c}{6}\right) = P\left(-\frac{c}{6} \le Z \le \frac{c}{6}\right) = 0.9544.$$

Thus

$$\Phi\left(\frac{c}{6}\right) - \left[1 - \Phi\left(\frac{c}{6}\right)\right] = 0.9544$$

and

$$\Phi\left(\frac{c}{6}\right) = 0.9772.$$

Hence $\dfrac{c}{6} = 2$ and $c = 12.$

Clearly, one can note that:

$$\begin{aligned} P(\mu - \sigma < X < \mu + \sigma) &= 0.68 \\ P(\mu - 2\sigma < X < \mu + 2\sigma) &= 0.95 \\ P(\mu - 3\sigma < X < \mu + 3\sigma) &= 0.997 \end{aligned}$$

Exercises

(1) If Z is $N(0,1)$, find

(a) $P(0.53 < Z \leq 2.06)$, (b) $P(-2.13 \leq Z \leq -0.56)$,

(c) $P(Z > -1.77)$, (d) $P(Z > 2.89)$,

(e) $P(\mid Z \mid < 1.96)$, (f) $P(\mid Z \mid < 1)$,

(g) $P(\mid Z \mid < 2)$ (h) $P(\mid Z \mid < 3)$.

(2) If Z is $N(0,1)$, find

(a) $P(0 \leq Z \leq 0.87)$, (b) $P(\mid Z \mid > 1)$,

(c) $(-2.64 \leq Z \leq 0)$, (d) $P(\mid Z \mid > 2)$,

(e) $P(-0.79 \leq Z < 1.52)$, (f) $P(\mid Z \mid > 3)$,

(g) $P(\mid Z \mid > 1.39)$ (h) $P(\mid Z \mid \leq 1.96)$.

(3) If Z is $N(0,1)$, find values of C such that

(a) $P(Z \leq c) = 0.025$.

(b) $(\mid Z \mid \leq c) = 0.95$.

(c) $P(Z > c) = 0.05$.

(4) Let X equal the birth weight (in grams) of babies in Egypt. Assuming that the distribution of X is $N(3315, 575^2)$, find

(a) $P(2584.75 \leq X \leq 4390.25)$.

(b) $P(2619.25 \leq X \leq 3642.75)$.

(c) $P(3119.50 \leq X \leq 3579.50)$.

(5) If the moment-generating function of X is

$$M(t) = \exp[166t + 200t^2],$$

find

(a) The mean of X.

(b) The variance of X.

(c) $P(148 \leq X \leq 172.)$

(d) $P(170 < X < 200)$.

(6) If X is $N(650, 625)$, find

(a) $P(600 \leq X \leq 660)$.

(b) A constant $c > 0$ such that

$$P(|\ X - 650\ | \leq c) = 0.9544.$$

(7) An exclusive college desires to accept only the top 10% of all graduating seniors based on the results of a national placement test. This test has a mean of 500 and a standard deviation of 100. Find the cutoff score for the exam. Assume the scores are normally distributed.

Chapter 5

Joint Distribution

5.1 Introduction

Our study of random variables and their probability distributions is restricted to one-dimensional sample spaces, in that we recorded outcomes of an experiment as values assumed by a single random variable. There will be situations, however, where we may find it desirable to record the simultaneous outcomes of several random variables. For example, we might measure the amount of precipitate P and volume V of gas released from a controlled chemical experiment, giving rise to a two-dimensional sample space consisting of the outcomes (p, v).

5.2 Joint Probability Distribution

If X and Y are two discrete random variables, the probability distribution for their simultaneous occurrence can be represented by a function with values $f(x, y)$ for any pair of values (x, y) within the range of the random variables X and Y . It is customary to refer to this function as the **joint probability distribution** of X and Y. Hence, in the discrete case,

$$f(x, y) = P(X = x, Y = y);$$

that is, the values $f(x, y)$ give the probability that outcomes x and y occur at the same time. For example, if an 18-wheeler is to have its tires serviced and X represents the number of miles these tires have been driven and Y represents the number of tires that need to be replaced, then $f(30000, 5)$ is the probability that the tires are used over 30,000 miles and the truck needs 5 new tires.

Definition 31. *The function $f(x, y)$ is a* **joint probability distribution** *or* **probability mass function** *of the discrete random variables X and Y if*

1. $f(x, y) \geqslant 0$ *for all* (x, y),

2. $\sum_x \sum_y f(x, y) = 1$,

3. $P(X = x, Y = y) = f(x, y)$.

For any region A in the xy plane, $P[(X, Y) \in A] = \sum\sum_{A} f(x, y)$.

Example 5.2.1. *Two ballpoint pens are selected at random from a box that contains 3 blue pens, 2 red pens, and 3 green pens. If X is the number of blue pens selected and Y is the number of red pens selected, find*

1. *the joint probability function $f(x, y)$,*

2. $P[(X, Y) \in A]$, *where A is the region* $\{(x, y)|x + y \leqslant 1\}$.

Solution :

The possible pairs of values (x, y) are (0, 0), (0, 1), (1, 0), (1, 1), (0, 2), and (2, 0).

1. Now, $f(0, 1)$, for example, represents the probability that a red and a green pens are selected. The total number of equally likely ways of selecting any 2 pens from the 8 is $\binom{8}{2} = 28$. The number of ways of selecting 1 red from 2 red pens and 1 green from 3 green pens is $\binom{2}{1}\binom{3}{1} = 6$. Hence, f(0, 1) $= 6/28 = 3/14$. Similar calculations yield the probabilities for the other cases, which are presented in Table 5.1. Note that the probabilities sum to 1. The joint probability distribution

of Table 5.1 can be represented by the formula

$$f(x,y)=\frac{\binom{3}{x}\binom{2}{y}\binom{3}{2-x-y}}{\binom{8}{2}},$$

for $x=0,1,2$; $y=0,1,2$; and $0\leqslant x+y\leqslant 2$.

2. The probability that (X,Y) fall in the region A is

$$\begin{aligned} P[(X,Y)\in A] &= P(X+Y\leqslant 1) \\ &= f(0,0)+f(0,1)+f(1,0) \\ &= \frac{3}{28}+\frac{3}{14}+\frac{9}{28}=\frac{9}{14} \end{aligned}$$

	$f(x,y)$	x = 0	x = 1	x = 2	Row Totals
y	0	$\frac{3}{28}$	$\frac{9}{28}$	$\frac{3}{28}$	$\frac{15}{28}$
	1	$\frac{3}{14}$	$\frac{3}{14}$	0	$\frac{3}{7}$
	2	$\frac{1}{28}$	0	0	$\frac{1}{28}$
Column	Totals	$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$	1

Table 5.1

When X and Y are continuous random variables, the **joint density function** $f(x,y)$ is a surface lying above the xy plane, and $P[(X,Y)\in A]$, where A is any region in the xy plane, is equal to the volume of the right cylinder bounded by the base A and the surface.

Definition 32. *The function* $f(x, y)$ *is a* **joint density function** *of the continuous random variables* X *and* Y *if*

1. $f(x, y) \geqslant 0$, *for all* (x, y),
2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)dxdy = 1$,
3. $P[(X, Y) \in A] = \int \int_A f(x, y)dxdy$, *for any region* A *in the* xy *plane.*

Example 5.2.2. *A privately owned business operates both a drive-in facility and a walk-in facility. On a randomly selected day, let* X *and* Y *, respectively, be the proportions of the time that the drive-in and the walk-in facilities are in use, and suppose that the joint density function of these random variables is*

$$f(x, y) = \begin{cases} \frac{2}{5}(2x + 3y), & 0 \leqslant x \leqslant 1, 0 \leqslant y \leqslant 1; \\ 0, & \text{elsewhere.} \end{cases}$$

1. *Verify condition 2 of Definition 2.*
2. *Find* $P[(X, Y) \in A]$, *where* $A = \{(x, y)|0 < x < \frac{1}{2}, \frac{1}{4} < y < \frac{1}{2}\}$.

Solution :

1. The integration of $f(x,y)$ over the whole region is

$$\begin{aligned}
\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy &= \int_0^1\int_0^1 \frac{2}{5}(2x+3y)dxdy \\
&= \int_0^1 \left[\frac{2x^2}{5}+\frac{6xy}{5}\right]_{x=0}^{x=1} dy \\
&= \int_0^1 \frac{2}{5}+\frac{6y}{5}dy = \left[\frac{2y}{5}+\frac{3y^2}{5}\right]_0^1 = \frac{2}{5}+\frac{3}{5} = 1
\end{aligned}$$

2. To calculate the probability, we use

$$\begin{aligned}
P[(X,Y)\in A] &= P\left(0<X<\frac{1}{2}, \frac{1}{4}<Y<\frac{1}{2}\right) \\
&= \int_{1/4}^{1/2}\int_0^{1/2} \frac{2}{5}(2x+3y)dxdy \\
&= \int_{1/4}^{1/2} \left[\frac{2x^2}{5}+\frac{6xy}{5}\right]_{x=0}^{x=1/2} dy \\
&= \int_{1/4}^{1/2} \left[\frac{1}{10}+\frac{3y}{5}\right] dy = \left[\frac{y}{10}+\frac{3y^2}{10}\right]_{1/4}^{1/2} \\
&= \frac{1}{10}\left[\left(\frac{1}{2}+\frac{3}{4}\right)-\left(\frac{1}{4}+\frac{3}{16}\right)\right] = \frac{13}{160}.
\end{aligned}$$

Given the joint probability distribution $f(x,y)$ of the discrete random variables X and Y , the probability distribution $g(x)$ of X alone is obtained by summing $f(x,y)$ over the values of Y. Similarly, the probability distribution $h(y)$ of Y alone is obtained by summing $f(x,y)$ over the values of X. We define $g(x)$ and $h(y)$ to be the marginal distributions of X and Y, respectively. When X and Y are continuous random variables, summations are replaced by integrals.

Definition 33. *The* **marginal distributions** *of X alone and of Y alone are*

$$g(x) = \sum_y f(x,y) \qquad \textit{and} \qquad h(y) = \sum_x f(x,y)$$

for the discrete case, and

$$g(x) = \int_{-\infty}^{\infty} f(x,y)dy \qquad \textit{and} \qquad h(y) = \int_{-\infty}^{\infty} f(x,y)dx$$

for the continuous case.

The term marginal is used here because, in the discrete case, the values of $g(x)$ and $h(y)$ are just the marginal totals of the respective columns and rows when the values of $f(x,y)$ are displayed in a rectangular table.

Example 5.2.3. *Show that the column and row totals of Table 5.1 give the marginal distribution of X alone and of Y alone.*

Solution : For the random variable X, we see that

$$g(0) = f(0,0) + f(0,1) + f(0,2) = \frac{3}{28} + \frac{3}{14} + \frac{1}{28} = \frac{5}{14},$$

$$g(1) = f(1,0) + f(1,1) + f(1,2) = \frac{9}{28} + \frac{3}{14} + 0 = \frac{15}{28},$$

and

$$g(2) = f(2,0) + f(2,1) + f(2,2) = \frac{3}{28} + 0 + 0 = \frac{3}{28},$$

which are just the column totals of Table 5.1. In a similar manner we could show that the values of $h(y)$ are given by the row totals. In tabular form, these marginal distributions may be written as follows:

x	0	1	2
g(x)	$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$

y	0	1	2
h(y)	$\frac{15}{28}$	$\frac{3}{7}$	$\frac{1}{28}$

Example 5.2.4. *Find $g(x)$ and $h(y)$ for the joint density function of Example 2.*

Solution : By definition,

$$g(x) = \int_{-\infty}^{\infty} f(x,y)dy = \int_0^1 \frac{2}{5}(2x+3y)dy = \left[\frac{4xy}{5} + \frac{6y^2}{10}\right]_{y=0}^{y=1} = \frac{4x+3}{5},$$

for $0 \leqslant x \leqslant 1$, and $g(x) = 0$ elsewhere. Similarly,

$$h(y) = \int_{-\infty}^{\infty} f(x,y)dx = \int_0^1 \frac{2}{5}(2x+3y)dx = \frac{2(1+3y)}{5},$$

for $0 \leqslant y \leqslant 1$, and $h(y) = 0$ elsewhere.

Definition 34. *Let X and Y be two random variables, discrete or continuous. The* **conditional distribution** *of the random variable Y given that $X = x$ is*

$$f(y|x) = \frac{f(x,y)}{g(x)}, \qquad \textit{provided} \qquad g(x) > 0.$$

Similarly, the conditional distribution of X given that $Y = y$ is

$$f(x|y) = \frac{f(x,y)}{h(y)}, \qquad \textit{provided} \qquad h(y) > 0.$$

If we wish to find the probability that the discrete random variable X falls between a and b when it is known that the discrete variable $Y = y$, we evaluate

$$P(a < X < b|Y = y) = \sum_{a<x<b} f(x|y),$$

where the summation extends over all values of X between a and b. When X and Y are continuous, we evaluate

$$P(a < X < b|Y = y) = \int_a^b f(x|y)dx.$$

Example 5.2.5. *Referring to Example 1, find the conditional distribution of X, given that $Y = 1$, and use it to determine*

$P(X = 0|Y = 1)$.

Solution : We need to find $f(x|y)$, where $y = 1$. First, we find that

$$h(1) = \sum_{x=0}^{2} f(x, 1) = \frac{3}{14} + \frac{3}{14} + 0 = \frac{3}{7}.$$

Now

$$f(x|1) = \frac{f(x, 1)}{h(1)} = \frac{7}{3}\, f(x, 1), \quad x = 0, 1, 2.$$

Therefore,

$$f(0|1) = \frac{7}{3} f(0, 1) = \frac{7}{3}\frac{3}{14} = \frac{1}{2}, \qquad f(1|1) = \frac{7}{3} f(1, 1) = \frac{7}{3}\frac{3}{14} = \frac{1}{2},$$

$$f(2|1) = \frac{7}{3} f(2, 1) = \frac{7}{3} 0 = 0$$

and the conditional distribution of X, given that $Y = 1$, is

x	0	1	2
$f(x\|1)$	$\frac{1}{2}$	$\frac{1}{2}$	0

Finally,

$$P(X = 0|Y = 1) = f(0|1) = \frac{1}{2}.$$

Therefore, if it is known that 1 of the 2 pen refills selected is red, we have a probability equal to 1/2 that the other refill is not blue.

Example 5.2.6. *The joint density for the random variables* (X, Y), *where* X *is the unit temperature change and* Y *is the proportion of spectrum shift that a certain atomic particle produces, is*

$$f(x, y) = \begin{cases} 10xy^2, & 0 < x < y < 1, \\ 0, & \textit{elsewhere.} \end{cases}$$

1. *Find the marginal densities* $g(x)$, $h(y)$, *and the conditional density* $f(y|x)$.

2. *Find the probability that the spectrum shifts more than half of the total observations, given that the temperature is increased by 0.25 unit.*

Solution :

1. By definition,

$$\begin{aligned} g(x) &= \int_{-\infty}^{\infty} f(x,y)dy = \int_{x}^{1} 10xy^2 dy \\ &= \frac{10}{3}\, xy^3\Big]_{y=x}^{y=1} = \frac{10}{3}x(1-x^3), \quad 0<x<1, \end{aligned}$$

$$h(y) = \int_{-\infty}^{\infty} f(x,y)dx = \int_{0}^{y} 10xy^2 dx = 5x^2y^2\Big]_{x=0}^{x=y} = 5y^4, \quad 0<y<1,$$

Now

$$f(y|x) = \frac{f(x,y)}{g(x)} = \frac{10xy^2}{\frac{10}{3}x(1-x^3)} = \frac{3y^2}{1-x^3}, \quad 0<x<y<1.$$

2. Therefore,

$$P(Y > \frac{1}{2}|X=0.25) = \int_{1/2}^{1} f(y|x=0.25)dy = \int_{1/2}^{1} \frac{3y^2}{1-0.25^3}dy = \frac{8}{9}.$$

Example 5.2.7. *Given the joint density function*

$$f(x,y) = \begin{cases} \dfrac{x(1+3y^2)}{4}, & 0<x<2, 0<y<1, \\ 0, & \textit{elsewhere,} \end{cases}$$

find $g(x),\ h(y),\ f(x|y)$. *and evaluate* $P(\frac{1}{4} < X < \frac{1}{2}|Y=\frac{1}{3})$.

Solution : By definition of the marginal density. for $0<x<2$,

$$\begin{aligned} g(x) &= \int_{-\infty}^{\infty} f(x,y)dy = \int_{0}^{1} \frac{x(1+3y^2)}{4}dy \\ &= \left[\frac{xy}{4} + \frac{xy^3}{4}\right]_{y=0}^{y=1} = \frac{x}{2}, \end{aligned}$$

and for $0 < y < 1$,

$$\begin{aligned} h(y) &= \int_{-\infty}^{\infty} f(x,y)dx = \int_0^2 \frac{x(1+3y^2)}{4}dx \\ &= \left[\frac{x^2}{8} + \frac{3x^2y^2}{8}\right]_{x=0}^{x=2} = \frac{1+3y^2}{2}, \end{aligned}$$

Therefore, using the conditional density definition, for $0 < x < 2$,

$$f(x|y) = \frac{f(x,y)}{h(y)} = \frac{x(1+3y^2)/4}{(1+3y^2)/2} = \frac{x}{2},$$

and

$$P(\frac{1}{4} < X < \frac{1}{2}|Y = \frac{1}{3}) = \int_{1/4}^{1/2} \frac{x}{2}dx = \frac{3}{64}.$$

5.3 Statistically Independent

If $f(x|y)$ does not depend on y, then $f(x|y) = g(x)$ and $f(x,y) = g(x)h(y)$.

It should make sense that if $f(x|y)$ does not depend on y, then of course the outcome of the random variable Y has no impact on the outcome of the random variable X. In other words, we say that X and Y are independent random variables.

Definition 35. *Let X and Y be two random variables, discrete or continuous, with joint probability distribution $f(x,y)$ and marginal distributions $g(x)$ and $h(y)$, respectively. The random variables X and Y are said to be*

statistically independent *if and only if*

$$f(x, y) = g(x)h(y)$$

for all (x, y) *within their range.*

The continuous random variables of Example 5.2.7 are statistically independent, since the product of the two marginal distributions gives the joint density function. This is obviously not the case, however, for the continuous variables of Example 6. Checking for statistical independence of discrete random variables requires a more thorough investigation, since it is possible to have the product of the marginal distributions equal to the joint probability distribution for some but not all combinations of (x, y). If you can find any point (x, y) for which $f(x, y)$ is defined such that $f(x, y) \neq g(x)h(y)$, the discrete variables X and Y are not statistically independent.

Example 5.3.1. *Show that the random variables of Example 1 are not statistically independent.*

Solution : Let us consider the point $(0, 1)$. From Table 5.1 we find the three probabilities $f(0,1)$, $g(0)$, and $h(1)$ to be

$$\begin{aligned}
f(0,1) &= \frac{3}{14}, \\
g(0) &= \sum_{y=0}^{2} f(0,y) = \frac{3}{28} + \frac{3}{14} + \frac{1}{28} = \frac{5}{14}, \\
h(1) &= \sum_{x=0}^{2} f(x,1) = \frac{3}{14} + \frac{3}{14} + 0 = \frac{3}{7}.
\end{aligned}$$

Clearly,

$$f(0,1) \neq g(0)h(1),$$

5.4 Chebyshev′s Inequality and Convergence in Probability

In this section, we use Chebyshev′s inequality to show, in another sense, that the sample mean, $\overline{X}$, is a good statistic to use to estimate a population mean μ; the relative frequency of success in n independent Bernoulli trials, Y/n, is a good statistic for estimating p. We examine the effect of the sample size n on these estimates. We begin by showing 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. The proof is given for the discrete case, but it holds for the continuous case, with integrals replacing summations.

Theorem 28. (**Chebyshev′s Inequality**) *If the random variable X has a mean μ and variance σ^2, then, for every $k \geqslant 1$,*

$$P(|X-\mu| \geqslant k\sigma) \leqslant \frac{1}{k^2}.$$

Proof

Let $f(x)$ denote the pmf of X. Then

$$\begin{aligned} \sigma^2 &= E[(X-\mu)^2] = \sum_{x\in S}(x-\mu)^2 f(x) \\ &= \sum_{x\in A}(x-\mu)^2 f(x) + \sum_{x\in A'}(x-\mu)^2 f(x), \end{aligned} \tag{5.1}$$

where

$$A = \{x : |x-\mu| \geqslant k\sigma\}.$$

The second term in the right-hand member of Equation 5-1 is the sum of nonnegative numbers and thus is greater than or equal to zero. Hence,

$$\sigma^2 \geqslant \sum_{x\in A}(x-\mu)^2 f(x).$$

However, in A, $|x-\mu| \geqslant k\sigma$; so

$$\sigma^2 \geqslant \sum_{x\in A}(k\sigma)^2 f(x) = k^2\sigma^2 \sum_{x\in A} f(x).$$

But the latter summation equals $P(X \in A)$; thus,

$$\sigma^2 \geqslant k^2\sigma^2 P(X \in A) = k^2\sigma^2 P(|X-\mu| \geqslant k\sigma).$$

That is,

$$P(|X-\mu| \geqslant k\sigma) \leqslant \frac{1}{k^2}.$$

Corollary 1. *If $\varepsilon = k\sigma$, then*

$$P(|X-\mu| \geqslant \varepsilon) \leqslant \frac{\sigma^2}{\varepsilon^2}.$$

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,

$$P(|X-\mu| < k\sigma) \geqslant 1 - \frac{1}{k^2}.$$

From the corollary, it also follows that

$$P(|X-\mu| < \varepsilon) \geqslant 1 - \frac{\sigma^2}{\varepsilon^2}.$$

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 5.4.1. *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*

$$
\begin{aligned}
P(17 < X < 33) &= P(|X - 25| < 8) \\
&= P(|X - 25| < 2\sigma) \geqslant 1 - \frac{1}{4} = 0.75
\end{aligned}
$$

and an upper bound for $P(|X - 25| \geqslant 12)$ is found to be

$$
P(|X - 25| \geqslant 12) = P(|X - \mu| \geqslant 3\sigma) \leqslant \frac{1}{9}
$$

Note that the results of the last example hold for any distribution with mean 25 and standard deviation 4. But, even stronger, the probability that any random variable X differs from its mean by 3 or more standard deviations is at most 1/9, which may be seen by letting $k = 3$ in the theorem. Also, the probability that any random variable X differs from its mean by less than two standard deviations is at least 3/4, which may be seen by letting $k = 2$.

The following consideration partially indicates the value of Cheb- yshev′s inequality in theoretical discussions: If Y is the number of successes in n independent 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 its mean p. To gain some insight into the closeness of Y/n to p, we shall use Chebyshev′s inequality. With $\varepsilon > 0$, we note from the previous Corollary that, since $Var(Y/n) = pq/n$, it follows that

$$P\left(\left|\frac{Y}{n} - p\right| \geqslant \varepsilon\right) \leqslant \frac{pq/n}{\varepsilon^2}$$

or, equivalently,

$$P\left(\left|\frac{Y}{n} - p\right| < \varepsilon\right) \geqslant 1 - \frac{pq}{n\varepsilon^2} \tag{5.2}$$

On the one hand, when p is completely unknown, we can use the fact that $pq = p(1-p)$ is a maximum when $p = 1/2$ in order to find a lower bound for the probability in Equation 5.2. That is,

$$1 - \frac{pq}{n\varepsilon^2} \geqslant 1 - \frac{(1/2)(1/2)}{n\varepsilon^2}.$$

For example, if $\varepsilon = 0.05$ and $n = 400$, then

$$P\left(\left|\frac{Y}{400} - p\right| < 0.05\right) \geqslant 1 - \frac{(1/2)(1/2)}{400(0.0025)} = 0.75.$$

On the other hand, if it is known that p is equal to $1/10$, we would have

$$P\left(\left|\frac{Y}{400} - p\right| < 0.05\right) \geqslant 1 - \frac{(0.1)(0.9)}{400(0.0025)} = 0.91.$$

Note that Chebyshev's inequality is applicable to all distributions with a finite variance, and thus the bound is not always a tight one; that is, the bound is not necessarily close to the true probability.

In general, however, it should be noted that, with fixed $\varepsilon > 0$ and $0 < p < 1$, we have

$$\lim_{n\to\infty} P\left(\left|\frac{Y}{400} - p\right| < \varepsilon\right) \geqslant \lim_{n\to\infty}\left(1 - \frac{pq}{n\varepsilon^2}\right) = 1.$$

But since the probability of every event is less than or equal to 1, it must be that

$$\lim_{n\to\infty} P\left(\left|\frac{Y}{400} - p\right| < \varepsilon\right) = 1.$$

That is, the probability that the relative frequency Y/n is within ε of p is arbitrarily close to 1 when n is large enough. This is one form of the **law of large numbers**, and we say that Y/n **converges in probability** to p.

A more general form of the law of large numbers is found by considering the mean X of a random sample from a distribution with mean μ and variance σ^2. This form of the law is more general because the relative frequency Y/n can be thought of as X when the sample arises from a Bernoulli distribution. To derive it, we note that

$$E(\overline{X}) = \mu \qquad \text{and} \qquad Var(\overline{X}) = \frac{\sigma^2}{n}.$$

Thus, from Corollary 1, for every $\varepsilon > 0$, we have

$$P[|\overline{X} - \mu| \geqslant \varepsilon] \leqslant \frac{\sigma^2/n}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2}$$

Since probability is nonnegative, it follows that

$$\lim_{n\to\infty} P(|\overline{X}-\mu| \geqslant \varepsilon) \leqslant \lim_{n\to\infty} \frac{\sigma^2}{n\varepsilon^2} = 0,$$

which implies that

$$\lim_{n\to\infty} P(|\overline{X}-\mu| \geqslant \varepsilon) = 0,$$

or, equivalently,

$$\lim_{n\to\infty} P(|\overline{X}-\mu| < \varepsilon) = 1,$$

The preceding discussion shows that the probability associated with the distribution of X becomes concentrated in an arbitrarily small interval centered at μ as n increases. This is a more general form of the law of large numbers, and we say that X converges in probability to μ.

5.5 Exercises

1. From a sack of fruit containing 3 oranges, 2 apples, and 3 bananas, a random sample of 4 pieces of fruit is selected. If X is the number of oranges and Y is the number of apples in the sample, find

   (a) the joint probability distribution of X and Y;

   (b) $P[(X,Y) \in A]$, where A is the region that is given by $\{(x,y)|x+y \leqslant 2\}$.

2. Let X denote the reaction time, in seconds, to a certain stimulus and Y denote the temperature at which a certain reaction starts to take place. Suppose that two random variables X and Y have the joint density

$$f(x,y) = \begin{cases} 4xy, & 0 < x < 1, 0 < y < 1, \\ 0, & \text{elsewhere.} \end{cases}$$

Find

   (a) $P(0 \leqslant X \leqslant \frac{1}{2} \ \textit{and} \ \frac{1}{4} \leqslant Y \leqslant \frac{1}{2})$;

   (b) $P(X < Y)$.

3. A fast-food restaurant operates both a drive-through facility and a walk-in facility. On a randomly selected day, let X and Y , respectively, be the proportions of the time that the drive-through and walk-in facilities are in use, and suppose that the joint density function of these random variables is

$$f(x,y) = \begin{cases} \frac{2}{3}(x+2y), & 0 \leqslant x \leqslant 1, 0 \leqslant y \leqslant 1, \\ 0, & \text{elsewhere.} \end{cases}$$

Find

   (a) Find the marginal density of X.

   (b) Find the marginal density of Y.

(c) Find the probability that the drive-through facility is busy less than one-half of the time.

4. The joint density function of the random variables X and Y is

$$f(x,y) = \begin{cases} 6x, & 0 < x < 1, 0 < y < 1-x,, \\ 0, & \text{elsewhere.} \end{cases}$$

   (a) Show that X and Y are not independent.

   (b) Find $(X > 0.3|Y = 0.5)$.

5. The amount of kerosene, in thousands of liters, in a tank at the beginning of any day is a random amount Y from which a random amount X is sold during that day. Suppose that the tank is not resupplied during the day so that $x \leqslant y$, and assume that the joint density function of these variables is

$$f(x,y) = \begin{cases} 2, & 0 < x \leqslant y < 1, \\ 0, & \text{elsewhere.} \end{cases}$$

   (a) Determine if X and Y are independent.

   (b) Find $P(1/4 < X < 1/2|Y = 3/4)$.

6. 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 $P(|X - 33| \geqslant 14)$.

7. If $E(X) = 17$ and $E(X^2) = 298$, use Chebyshev′s inequality to determine

   (a) A lower bound for $P(10 < X < 24)$.

   (b) An upper bound for $P(|X - 17| \geqslant 16)$.

8. If the distribution of Y is $b(n, 0.5)$, give a lower bound for $P(|Y/n - 0.5| < 0.08)$ when

   (a) $n = 100$.

   (b) $n = 500$.

   (c) $n = 1000$.

9. If the distribution of Y is $b(n, 0.25)$, give a lower bound for $P(|Y/n - 0.25| < 0.05)$ when

   (a) $n = 100$.

   (b) $n = 500$.

   (c) $n = 1000$.

10. Let X be the mean of a random sample of size $n = 15$ from a distribution with mean $\mu = 80$ and variance $\sigma^2 = 60$. Use Chebyshev′s inequality to find a lower bound for $P(75 < X < 85)$.

Chapter 6

Sampling Distribution Theory

Progress in science is ascribed to experimentation. The research worker performs an experiment and obtains some data. On the basis of this data, certain conclusions are drawn. In other words, research worker may generalize from a particular experiments to the class of all similar experiments. This sort of extension from the particular to the general is called *statistical inference.* Such inference is used to find new knowledge in the empirical sciences.

6.1 Basic Concepts

Definition 36. *The totality of elements which are under discussion and about which information is desired will be called the population. Population does not mean living beings alone, but we speak of a population of births, weights, heights, prices of wheat and so on. The population may be finite or infinite.*

For any statistical investigation, it is often important to an analyze the whole population because of the restrictions of times and costs. In order to overcome the difficulties involved in the studding the whole population, we use the techniques of sampling.

Definition 37. *The sample is a finite subset of the population. The number of individuals (objects) in the sample is called the sample size (the size of the sample). The process of obtaining suitable sample from a population is called sampling.*

In drawing a sample, our main aim is to choose a representative sample to the population, so that from that sample we can obtain maximum information about the population with minimum effort and to measure and control the introduced error.

In what follows, when we speak about a "binomial population", a "normal population", or, in general, the "population $F(x)$" we shall mean a population whose observations are possible values of a random variable having a binomial distribution, a normal distribution, or the distribution $F(x)$ respectively. Hence, the mean and the variance of a random variable or distribution are also referred to as the mean and the variance of the corresponding population.

The statisticians are interested in arriving at a conclusions concerning unknown population parameters. For example, in a normal population parameters, μ and σ^2 may be unknown and are to be estimated from the information of a selected sample from that population. This leads us into the theory of sampling. If our inference are to be accurate, we must understand the relationship between the sample and its population. Certainly, the sample should be representative of the population. It should be a *random sample* in the sense that the observations are made independently and at random. The concept of a random sample is formulated in the following definition.

Definition 38. *Let* $X_1, X_2, \ldots, X_n$ *be* n *independent and identically distributed random variables each of which have the same cumulative distri-*

bution function $F(x)$, that is,

$$F(x) = P(X_i \leq x), \quad \forall\ i = 1, 2, \ldots, n.$$

Then we define $X_1, X_2, \ldots, X_n$ to be a random sample of size n from the population $F(x)$.

6.2 Sampling Theory

Sampling theory study the relationship between samples and their population. Statistical inference depends mainly on the sampling theory.

A variable which computed from a sample is called a *statistic*. Since many random samples can be selected from the same population, we would expect that the statistic vary randomly from sample to sample. Thus a statistic is a random variable.

Definition 39. *A statistic is a random variable that depends only on the observed random sample.*

Now, we define some statistics that describe some important measures of a random sample.

6.3 Some Location Statistics

Definition 40. *If* $X_1, X_2, \ldots, X_n$ *represent a random sample of size* n*, then the* *sample mean* *is defined by the statistic*

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

Note that the statistic $\overline{X}$ *is a random variable assumes the value* $\overline{x} = \sum_{i=1}^{n} x_i$ *when* X_i *assumes the value* x_i, $i = 1, 2, \ldots, n$.

Definition 41. *If* $X_1, X_2, \ldots, X_n$ *represent a random sample of size* n*, arranged in nondecreasing order of magnitude, then the* *sample median* *is defined by the statistic*

$$\tilde{m} = \begin{cases} X_{\left(\frac{n+1}{2}\right)}, & \text{if } n \text{ is odd,} \\[2ex] \frac{1}{2}\left[X_{\left(\frac{n}{2}\right)} + X_{\left(\frac{n}{2}+1\right)}\right], & \text{if } n \text{ is even.} \end{cases}$$

Definition 42. *If* $X_1, X_2, \ldots, X_n$ *represent a random sample of size* n*, then the* *mode* M *is the value of the sample that occurs with the greatest frequency. The mode may not exist, and when it exists, it is not necessarily unique.*

Example 6.3.1. *In a random sample of size* $n = 18$, *the following observations were recorded*

$$1,\ 3,\ 4,\ 0,\ 4,\ 2,\ 3,\ 1,\ 2,\ 3,\ 0,\ 4,\ 1,\ 1,\ 1,\ 5,\ 1,\ 0.$$

Find the mean, median, and the mode.

<u>solution</u> The observed value $\overline{x}$ of the statistic $\overline{X}$ is

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{36}{18} = 2.$$

To find the median we must arrange the observations in nondecreasing order of magnitude as follows:

$$0,\ 0,\ 0,\ 1,\ 1,\ 1,\ 1,\ 1,\ 1,\ 2,\ 2,\ 3,\ 3,\ 3,\ 4,\ 4,\ 4,\ 5.$$

Since $n = 18$ is even, the observed value of the statistic $\widetilde{m}$ is

$$\widetilde{m} = \frac{1}{2}\left[x_{(9)} + x_{(10)}\right] = \frac{1}{2}[1+2] = 1.5.$$

The mode of the observations exists, and the observed value is 1.

Some Statistics of Dispersion

Definition 43. *The* <u>range</u> *of the random sample* $X_1, X_2, \ldots, X_n$ *is defined by the statistic*

$$R = X_{(n)} - X_{(1)},$$

where $X_{(n)}$ *is the maximum of the sample and* $X_{(1)}$ *is the minimum of that sample.*

Definition 44. *If* $X_1, X_2, \ldots, X_n$ *is a random sample of size* n, *then the* *sample variance* *is defined by the statistic*

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

A more useful formula is

$$S^2 = \frac{n\sum\limits_{i=1}^{n} X_i^2 - \left(\sum\limits_{i=1}^{n} X_i\right)^2}{n(n-1)}.$$

The *sample standard deviation*, *denoted by* S *is defined to be the positive square root of the sample variance.*

Example 6.3.2. *Find the range, the variance, and the standard deviation of the sample whose observations are:*

$$3, \quad 4, \quad 7, \quad 5, \quad 6, \quad 3.$$

Solution The sample observation arranged in nondecreasing order of magnitude are

$$3, \quad 4, \quad 5, \quad 6, \quad 6, \quad 7.$$

The observed value r of the statistic R is

$$r = 7 - 3 = 4.$$

Table 6.1:

i	x_i	x_i^2
1	3	9
2	4	16
3	5	25
4	6	36
5	6	36
6	7	49
–	31	171

To find the value of the sample variance, we form the following table:

$$s^2 = \frac{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}{n(n-1)} = \frac{6(171) - (31)^2}{6(6-1)} = 2.167.$$

The sample standard deviation is $s = \sqrt{2.167} = 1.472$.

Definition 45. *The distribution of a statistic is called a sampling distribution.*

Definition 46. *The standard deviation of the sampling distribution of a statistic is called a standard error of the statistic.*

The Sampling Distribution of the Mean

Suppose that a random sample of size n is drawn from a normal population with mean μ and variance σ^2. Hence, by the reproductive property of the normal distribution, we conclude that

$$\boxed{\overline{X} \quad \text{has distribution} \quad N\left(\mu, \sigma^2/n\right)}$$

If we are sampling from a population with unknown distribution, the sampling distribution of $\overline{X}$ will be approximately normal with mean μ and variance σ^2/n provided that the sample size is large, that is $n \geq 30$. This result is an immediate consequence of the <u>*central limit theorem*</u>:

Theorem 29 (Central Limit Theorem)**.** *If $\overline{X}$ is the mean of a random sample $X_1, X_2, \ldots, X_n$ of size n from a distribution with finite mean μ and finite positive variance σ^2, then the distribution of*

$$W = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} = \frac{\sum\limits_{i=1}^{n} X_i - n\mu}{\sqrt{n}\,\sigma}$$

is $N(0,1)$ in limit as $n \to \infty$.

Example 6.3.3. *Let $\overline{X}$ denote the mean of a random sample of size $n = 15$ from the distribution whose pdf is*

$$f(x) = \begin{cases} \dfrac{3}{2}x^2, & -1 < x < 1; \\ 0, & \textit{elsewhere.} \end{cases}$$

Approximate $P(0.03 \leq \overline{X} \leq 0.15)$.

Solution It is easy to show that $\mu = 0$, and $\sigma^2 = 3/5$. Thus

$$
\begin{aligned}
& P(0.03 \leq \overline{X} \leq 0.15) \\
= & \ P\left(\frac{0.03-0}{\sqrt{3/5}/\sqrt{15}} \leq \frac{\overline{X}-0}{\sqrt{3/5}/\sqrt{15}} \leq \frac{0.15-0}{\sqrt{3/5}/\sqrt{15}}\right) \\
= & \ P(0.15 \leq W \leq 0.75) \\
\approx & \ \Phi(0.75) - \Phi(0.15) \\
= & \ 0.7734 - 0.5596 = 0.2138.
\end{aligned}
$$

Example 6.3.4. *Let* $X_1, X_2, \ldots, X_{20}$ *be a random sample of size 20 from the uniform distribution* $U(0,1)$. *Find* $E[X_i]$, *and* $Var(X_i)$ $, i = 1, 2, \ldots, 20$, *and approximate* $P(X_1 + X_2 + \cdots + X_n \leq 9.1)$.

Solution It is easy to show that

$$E[X_i] = \frac{1}{2}, \quad \text{and} \quad Var(X_i) = \frac{1}{12}, \quad \forall\ , i = 1, 2, \ldots, 20.$$

Then, if we set $Y = X_1 + X_2 + \cdots + X_n$, we have, by the Central Limit Theorem,

$$
\begin{aligned}
P(Y \leq 9.1) &= P\left(\frac{Y - (20)(1/2)}{\sqrt{20/12}} \leq \frac{9.1 - (20)(1/2)}{\sqrt{20/12}}\right) \\
&= P(W \leq -0.697) \approx \Phi(-0.697) = 0.2423.
\end{aligned}
$$

Exercises

1. Let $X_1, X_2, \ldots, X_{16}$ be a random sample from a normal distribution $N(77, 25)$. Compute

   (a) $P(77 < \overline{X} < 79.5)$,

   (b) $P(74.5 < \overline{X} < 78.5)$.

2. Let X equal the force required to pull a stud out of a window that is to be inserted into an automobile. Assume that the distribution of X is $N(147.8, 12.3^2)$.

   (a) Find $P(X < 163.3)$

   (b) If $\overline{X}$ is the mean of a random sample of size 25 from this distribution, determine

   $P(\overline{X} \leq 150.9)$.

3. Let $\overline{X}$ be the mean of a random sample of size 12 from the uniform distribution on the interval $(0, 1)$. Approximate $P\left(\frac{1}{2} \leq \overline{X} \leq \frac{2}{3}\right)$.

4. Let $Y = X_1 + X_2 + \cdots + X_{15}$ be the sum of a random sample of size 15 from the distribution whose pdf is $f(x) = \frac{3}{2}x^2, \quad -1 < x < 1$ and 0 elsewhere. Approximate $P(-0.3 \leq Y \leq 1.5)$.

5. Let $\overline{X}$ be the mean of a random sample of size 36 from the an exponential distribution with mean 3. Approximate $P(2.5 \leq \overline{X} \leq 4)$.

6. A random sample of size $n = 18$ is taken from the distribution with pdf $f(x) = 1 - x/2,\ 0 \leq x \leq 2.$ and 0 otherwise. Find

   (a) μ and σ^2.

   (b) Find, approximately, $P(2/3 \leq \overline{X} \leq 5/6)$.

7. Let X and Y equal the number of miles per gallon for compact cars and minimized cars, respectively, as reported in fuel economy ratings. Assume that $\mu_X = 24.5,\ \sigma_X = 3.8,\ \mu_Y = 21.3.$ and $\sigma_Y = 2.7.$ Let $\overline{X}$ and $\overline{Y}$ be the sample means of independent random samples of eight observations of X and Y, respectively.

   (a) What are the values of the means and variances of $\overline{X}$ and $\overline{Y}$?

   (b) Assuming that $\overline{X}$ and $\overline{Y}$ are each (approximately) normally distributed, how is $\overline{X} - \overline{Y}$ distributed?

   (c) Find the (approximate) probability, $P(\overline{X} > \overline{Y})$.

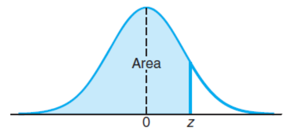

Table A.3 Areas under the Normal Curve

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
−3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
−3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
−3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
−3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
−2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
−2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
−2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
−2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
−2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
−2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
−2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
−2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
−2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
−2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
−1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
−1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
−1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
−1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
−1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
−1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
−1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
−1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
−1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
−1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
−0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
−0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
−0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
−0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
−0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
−0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
−0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
−0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
−0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
−0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641

Table A.3 (continued) Areas under the Normal Curve

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

Table A.4 Student t-Distribution Probability Table

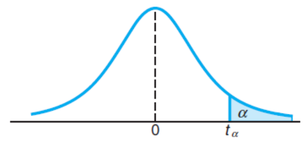

Table A.4 Critical Values of the t-Distribution

	α						
v	**0.40**	**0.30**	**0.20**	**0.15**	**0.10**	**0.05**	**0.025**
1	0.325	0.727	1.376	1.963	3.078	6.314	12.706
2	0.289	0.617	1.061	1.386	1.886	2.920	4.303
3	0.277	0.584	0.978	1.250	1.638	2.353	3.182
4	0.271	0.569	0.941	1.190	1.533	2.132	2.776
5	0.267	0.559	0.920	1.156	1.476	2.015	2.571
6	0.265	0.553	0.906	1.134	1.440	1.943	2.447
7	0.263	0.549	0.896	1.119	1.415	1.895	2.365
8	0.262	0.546	0.889	1.108	1.397	1.860	2.306
9	0.261	0.543	0.883	1.100	1.383	1.833	2.262
10	0.260	0.542	0.879	1.093	1.372	1.812	2.228
11	0.260	0.540	0.876	1.088	1.363	1.796	2.201
12	0.259	0.539	0.873	1.083	1.356	1.782	2.179
13	0.259	0.538	0.870	1.079	1.350	1.771	2.160
14	0.258	0.537	0.868	1.076	1.345	1.761	2.145
15	0.258	0.536	0.866	1.074	1.341	1.753	2.131
16	0.258	0.535	0.865	1.071	1.337	1.746	2.120
17	0.257	0.534	0.863	1.069	1.333	1.740	2.110
18	0.257	0.534	0.862	1.067	1.330	1.734	2.101
19	0.257	0.533	0.861	1.066	1.328	1.729	2.093
20	0.257	0.533	0.860	1.064	1.325	1.725	2.086
21	0.257	0.532	0.859	1.063	1.323	1.721	2.080
22	0.256	0.532	0.858	1.061	1.321	1.717	2.074
23	0.256	0.532	0.858	1.060	1.319	1.714	2.069
24	0.256	0.531	0.857	1.059	1.318	1.711	2.064
25	0.256	0.531	0.856	1.058	1.316	1.708	2.060
26	0.256	0.531	0.856	1.058	1.315	1.706	2.056
27	0.256	0.531	0.855	1.057	1.314	1.703	2.052
28	0.256	0.530	0.855	1.056	1.313	1.701	2.048
29	0.256	0.530	0.854	1.055	1.311	1.699	2.045
30	0.256	0.530	0.854	1.055	1.310	1.697	2.042
40	0.255	0.529	0.851	1.050	1.303	1.684	2.021
60	0.254	0.527	0.848	1.045	1.296	1.671	2.000
120	0.254	0.526	0.845	1.041	1.289	1.658	1.980
∞	0.253	0.524	0.842	1.036	1.282	1.645	1.960

Table A.4 (continued) Critical Values of the t-Distribution

	α						
v	**0.02**	**0.015**	**0.01**	**0.0075**	**0.005**	**0.0025**	**0.0005**
1	15.894	21.205	31.821	42.433	63.656	127.321	636.578
2	4.849	5.643	6.965	8.073	9.925	14.089	31.600
3	3.482	3.896	4.541	5.047	5.841	7.453	12.924
4	2.999	3.298	3.747	4.088	4.604	5.598	8.610
5	2.757	3.003	3.365	3.634	4.032	4.773	6.869
6	2.612	2.829	3.143	3.372	3.707	4.317	5.959
7	2.517	2.715	2.998	3.203	3.499	4.029	5.408
8	2.449	2.634	2.896	3.085	3.355	3.833	5.041
9	2.398	2.574	2.821	2.998	3.250	3.690	4.781
10	2.359	2.527	2.764	2.932	3.169	3.581	4.587
11	2.328	2.491	2.718	2.879	3.106	3.497	4.437
12	2.303	2.461	2.681	2.836	3.055	3.428	4.318
13	2.282	2.436	2.650	2.801	3.012	3.372	4.221
14	2.264	2.415	2.624	2.771	2.977	3.326	4.140
15	2.249	2.397	2.602	2.746	2.947	3.286	4.073
16	2.235	2.382	2.583	2.724	2.921	3.252	4.015
17	2.224	2.368	2.567	2.706	2.898	3.222	3.965
18	2.214	2.356	2.552	2.689	2.878	3.197	3.922
19	2.205	2.346	2.539	2.674	2.861	3.174	3.883
20	2.197	2.336	2.528	2.661	2.845	3.153	3.850
21	2.189	2.328	2.518	2.649	2.831	3.135	3.819
22	2.183	2.320	2.508	2.639	2.819	3.119	3.792
23	2.177	2.313	2.500	2.629	2.807	3.104	3.768
24	2.172	2.307	2.492	2.620	2.797	3.091	3.745
25	2.167	2.301	2.485	2.612	2.787	3.078	3.725
26	2.162	2.296	2.479	2.605	2.779	3.067	3.707
27	2.158	2.291	2.473	2.598	2.771	3.057	3.689
28	2.154	2.286	2.467	2.592	2.763	3.047	3.674
29	2.150	2.282	2.462	2.586	2.756	3.038	3.660
30	2.147	2.278	2.457	2.581	2.750	3.030	3.646
40	2.123	2.250	2.423	2.542	2.704	2.971	3.551
60	2.099	2.223	2.390	2.504	2.660	2.915	3.460
120	2.076	2.196	2.358	2.468	2.617	2.860	3.373
∞	2.054	2.170	2.326	2.432	2.576	2.807	3.290

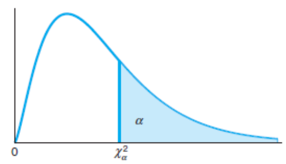

Table A.5 Critical Values of the Chi-Squared Distribution

	α									
v	**0.995**	**0.99**	**0.98**	**0.975**	**0.95**	**0.90**	**0.80**	**0.75**	**0.70**	**0.50**
1	0.0^4393	0.0^3157	0.0^3628	0.0^3982	0.00393	0.0158	0.0642	0.102	0.148	0.455
2	0.0100	0.0201	0.0404	0.0506	0.103	0.211	0.446	0.575	0.713	1.386
3	0.0717	0.115	0.185	0.216	0.352	0.584	1.005	1.213	1.424	2.366
4	0.207	0.297	0.429	0.484	0.711	1.064	1.649	1.923	2.195	3.357
5	0.412	0.554	0.752	0.831	1.145	1.610	2.343	2.675	3.000	4.351
6	0.676	0.872	1.134	1.237	1.635	2.204	3.070	3.455	3.828	5.348
7	0.989	1.239	1.564	1.690	2.167	2.833	3.822	4.255	4.671	6.346
8	1.344	1.647	2.032	2.180	2.733	3.490	4.594	5.071	5.527	7.344
9	1.735	2.088	2.532	2.700	3.325	4.168	5.380	5.899	6.393	8.343
10	2.156	2.558	3.059	3.247	3.940	4.865	6.179	6.737	7.267	9.342
11	2.603	3.053	3.609	3.816	4.575	5.578	6.989	7.584	8.148	10.341
12	3.074	3.571	4.178	4.404	5.226	6.304	7.807	8.438	9.034	11.340
13	3.565	4.107	4.765	5.009	5.892	7.041	8.634	9.299	9.926	12.340
14	4.075	4.660	5.368	5.629	6.571	7.790	9.467	10.165	10.821	13.339
15	4.601	5.229	5.985	6.262	7.261	8.547	10.307	11.037	11.721	14.339
16	5.142	5.812	6.614	6.908	7.962	9.312	11.152	11.912	12.624	15.338
17	5.697	6.408	7.255	7.564	8.672	10.085	12.002	12.792	13.531	16.338
18	6.265	7.015	7.906	8.231	9.390	10.865	12.857	13.675	14.440	17.338
19	6.844	7.633	8.567	8.907	10.117	11.651	13.716	14.562	15.352	18.338
20	7.434	8.260	9.237	9.591	10.851	12.443	14.578	15.452	16.266	19.337
21	8.034	8.897	9.915	10.283	11.591	13.240	15.445	16.344	17.182	20.337
22	8.643	9.542	10.600	10.982	12.338	14.041	16.314	17.240	18.101	21.337
23	9.260	10.196	11.293	11.689	13.091	14.848	17.187	18.137	19.021	22.337
24	9.886	10.856	11.992	12.401	13.848	15.659	18.062	19.037	19.943	23.337
25	10.520	11.524	12.697	13.120	14.611	16.473	18.940	19.939	20.867	24.337
26	11.160	12.198	13.409	13.844	15.379	17.292	19.820	20.843	21.792	25.336
27	11.808	12.878	14.125	14.573	16.151	18.114	20.703	21.749	22.719	26.336
28	12.461	13.565	14.847	15.308	16.928	18.939	21.588	22.657	23.647	27.336
29	13.121	14.256	15.574	16.047	17.708	19.768	22.475	23.567	24.577	28.336
30	13.787	14.953	16.306	16.791	18.493	20.599	23.364	24.478	25.508	29.336
40	20.707	22.164	23.838	24.433	26.509	29.051	32.345	33.66	34.872	39.335
50	27.991	29.707	31.664	32.357	34.764	37.689	41.449	42.942	44.313	49.335
60	35.534	37.485	39.699	40.482	43.188	46.459	50.641	52.294	53.809	59.335

Table A.5 (continued) Critical Values of the Chi-Squared Distribution

	α									
v	**0.30**	**0.25**	**0.20**	**0.10**	**0.05**	**0.025**	**0.02**	**0.01**	**0.005**	**0.001**
1	1.074	1.323	1.642	2.706	3.841	5.024	5.412	6.635	7.879	10.827
2	2.408	2.773	3.219	4.605	5.991	7.378	7.824	9.210	10.597	13.815
3	3.665	4.108	4.642	6.251	7.815	9.348	9.837	11.345	12.838	16.266
4	4.878	5.385	5.989	7.779	9.488	11.143	11.668	13.277	14.860	18.466
5	6.064	6.626	7.289	9.236	11.070	12.832	13.388	15.086	16.750	20.515
6	7.231	7.841	8.558	10.645	12.592	14.449	15.033	16.812	18.548	22.457
7	8.383	9.037	9.803	12.017	14.067	16.013	16.622	18.475	20.278	24.321
8	9.524	10.219	11.030	13.362	15.507	17.535	18.168	20.090	21.955	26.124
9	10.656	11.389	12.242	14.684	16.919	19.023	19.679	21.666	23.589	27.877
10	11.781	12.549	13.442	15.987	18.307	20.483	21.161	23.209	25.188	29.588
11	12.899	13.701	14.631	17.275	19.675	21.920	22.618	24.725	26.757	31.264
12	14.011	14.845	15.812	18.549	21.026	23.337	24.054	26.217	28.300	32.909
13	15.119	15.984	16.985	19.812	22.362	24.736	25.471	27.688	29.819	34.527
14	16.222	17.117	18.151	21.064	23.685	26.119	26.873	29.141	31.319	36.124
15	17.322	18.245	19.311	22.307	24.996	27.488	28.259	30.578	32.801	37.698
16	18.418	19.369	20.465	23.542	26.296	28.845	29.633	32.000	34.267	39.252
17	19.511	20.489	21.615	24.769	27.587	30.191	30.995	33.409	35.718	40.791
18	20.601	21.605	22.760	25.989	28.869	31.526	32.346	34.805	37.156	42.312
19	21.689	22.718	23.900	27.204	30.144	32.852	33.687	36.191	38.582	43.819
20	22.775	23.828	25.038	28.412	31.410	34.170	35.020	37.566	39.997	45.314
21	23.858	24.935	26.171	29.615	32.671	35.479	36.343	38.932	41.401	46.796
22	24.939	26.039	27.301	30.813	33.924	36.781	37.659	40.289	42.796	48.268
23	26.018	27.141	28.429	32.007	35.172	38.076	38.968	41.638	44.181	49.728
24	27.096	28.241	29.553	33.196	36.415	39.364	40.270	42.980	45.558	51.179
25	28.172	29.339	30.675	34.382	37.652	40.646	41.566	44.314	46.928	52.619
26	29.246	30.435	31.795	35.563	38.885	41.923	42.856	45.642	48.290	54.051
27	30.319	31.528	32.912	36.741	40.113	43.195	44.140	46.963	49.645	55.475
28	31.391	32.620	34.027	37.916	41.337	44.461	45.419	48.278	50.994	56.892
29	32.461	33.711	35.139	39.087	42.557	45.722	46.693	49.588	52.335	58.301
30	33.530	34.800	36.250	40.256	43.773	46.979	47.962	50.892	53.672	59.702
40	44.165	45.616	47.269	51.805	55.758	59.342	60.436	63.691	66.766	73.403
50	54.723	56.334	58.164	63.167	67.505	71.420	72.613	76.154	79.490	86.660
60	65.226	66.981	68.972	74.397	79.082	83.298	84.58	88.379	91.952	99.608

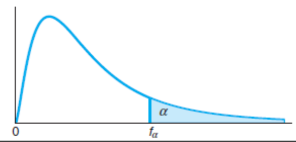

Table A.6 Critical Values of the *F*-Distribution

	$f_{0.05}(v_1, v_2)$								
	v_1								
v_2	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88

Reproduced from Table 18 of *Biometrika Tables for Statisticians*, Vol. I, by permission of E.S. Pearson and the Biometrika Trustees.

Table A.6 (continued) Critical Values of the *F*-Distribution

	$f_{0.05}(v_1, v_2)$									
	v_1									
v_2	**10**	**12**	**15**	**20**	**24**	**30**	**40**	**60**	**120**	∞
1	241.88	243.91	245.95	248.01	249.05	250.10	251.14	252.20	253.25	254.31
2	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
3	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36
6	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
29	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
∞	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

Table A.6 (continued) Critical Values of the *F*-Distribution

	$f_{0.01}(v_1, v_2)$								
	v_1								
v_2	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
1	4052.18	4999.50	5403.35	5624.58	5763.65	5858.99	5928.36	5981.07	6022.47
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41

Table A.6 (continued) Critical Values of the F-Distribution

$f_{0.01}(v_1, v_2)$

v_2 \ v_1	**10**	**12**	**15**	**20**	**24**	**30**	**40**	**60**	**120**	∞
1	6055.85	6106.32	6157.28	6208.73	6234.63	6260.65	6286.78	6313.03	6339.39	6365.86
2	99.40	99.42	99.43	99.45	99.46	99.47	99.47	99.48	99.49	99.50
3	27.23	27.05	26.87	26.69	26.60	26.50	26.41	26.32	26.22	26.13
4	14.55	14.37	14.20	14.02	13.93	13.84	13.75	13.65	13.56	13.46
5	10.05	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02
6	7.87	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	6.88
7	6.62	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65
8	5.81	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.86
9	5.26	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31
10	4.85	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	3.91
11	4.54	4.40	4.25	4.10	4.02	3.94	3.86	3.78	3.69	3.60
12	4.30	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36
13	4.10	3.96	3.82	3.66	3.59	3.51	3.43	3.34	3.25	3.17
14	3.94	3.80	3.66	3.51	3.43	3.35	3.27	3.18	3.09	3.00
15	3.80	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.96	2.87
16	3.69	3.55	3.41	3.26	3.18	3.10	3.02	2.93	2.84	2.75
17	3.59	3.46	3.31	3.16	3.08	3.00	2.92	2.83	2.75	2.65
18	3.51	3.37	3.23	3.08	3.00	2.92	2.84	2.75	2.66	2.57
19	3.43	3.30	3.15	3.00	2.92	2.84	2.76	2.67	2.58	2.49
20	3.37	3.23	3.09	2.94	2.86	2.78	2.69	2.61	2.52	2.42
21	3.31	3.17	3.03	2.88	2.80	2.72	2.64	2.55	2.46	2.36
22	3.26	3.12	2.98	2.83	2.75	2.67	2.58	2.50	2.40	2.31
23	3.21	3.07	2.93	2.78	2.70	2.62	2.54	2.45	2.35	2.26
24	3.17	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	2.21
25	3.13	2.99	2.85	2.70	2.62	2.54	2.45	2.36	2.27	2.17
26	3.09	2.96	2.81	2.66	2.58	2.50	2.42	2.33	2.23	2.13
27	3.06	2.93	2.78	2.63	2.55	2.47	2.38	2.29	2.20	2.10
28	3.03	2.90	2.75	2.60	2.52	2.44	2.35	2.26	2.17	2.06
29	3.00	2.87	2.73	2.57	2.49	2.41	2.33	2.23	2.14	2.03
30	2.98	2.84	2.70	2.55	2.47	2.39	2.30	2.21	2.11	2.01
40	2.80	2.66	2.52	2.37	2.29	2.20	2.11	2.02	1.92	1.80
60	2.63	2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73	1.60
120	2.47	2.34	2.19	2.03	1.95	1.86	1.76	1.66	1.53	1.38
∞	2.32	2.18	2.04	1.88	1.79	1.70	1.59	1.47	1.32	1.00

Bibliography

[1] Prem S. Mann, and Christopher Jay Lacke (2013). Introductory statistics. Eighth Edition, John Wiley & Sons.

[2] Robert V. Hogg, Elliot A. Tanis and Dale L. Zimmerman (2015). Probability and statistical inference. Ninth Edition, Pearson Education.

[3] Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers, and Keying Ye (2012). Probability & Statistics for Engineers & Scientists. Ninth Edition, Pearson Education.

[4] Hodges, J. L., and Lehmann, E. L. (2005). Basic Concepts of Probability and Statistics, 2nd ed. Philadelphia: Society for Industrial and Applied Mathematics.