Probability and Statistics

Chapter 1

Dr. Raed Q. Alathamneh
Department of Industrial Engineering

١

1

Chapter 1: Introduction to Statistics and Data Analysis

٢

2

1.1 Overview: Statistical Inference, Samples, Populations, and Experimental Design

-Use of Scientific Data
-Variability in Scientific Data
-The Role of Probability

٣

3

Example 1.2 Often the nature of the scientific study will dictate the role that probability and deductive reasoning play in statistical inference. Exercise 9.40 on page 297 provides data associated with a study conducted at the Virginia Polytechnic Institute and State University on the development, of a relationship between the roots of trees and the action of a fungus. Minerals are transferred from the fungus to the trees and sugars from the trees to the fungus. Two samples of 10 northern red oak seedlings are planted in a greenhouse, one containing seedlings treated with nitrogen and one containing no nitrogen. All other environmental conditions are held constant. All seedlings contain the fungus *Pisolithus tinctorus.* More details are supplied in Chapter 9. The stem weights in grams were recorded after the end of 140 days. The data are given in Table 1.1.

٤

4

Table 1.1: Data Set for Example 1.2

No Nitrogen	Nitrogen
0.32	0.26
0.53	0.43
0.28	0.47
0.37	0.49
0.47	0.52
0.43	0.75
0.36	0.79
0.42	0.86
0.38	0.62
0.43	0.46

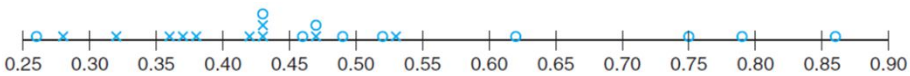


Figure 1.1: A dot plot of stem weight data.

٥

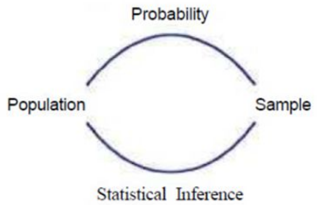


Figure 1.2: Fundamental relationship between probability and inferential statistics.

٦

1.3 Sampling Procedures; Collection of Data

Simple Random Sampling
Experimental Design

Example 1.3:1 A corrosion study was made in order to determine whether corrosion of an aluminum metal coated with a corrosion retardation substance reduced the amount of corrosion. The coating is a protectant that is advertised to minimize fatigue damage in this type of material. Also of interest is the influence of humidity on the amount of corrosion. A corrosion measurement can be expressed in thousands of cycles to failure. Two levels of coating, no coating and chemical corrosion coating, were used. In addition, the two relative humidity levels are 20% relative humidity and 80% relative humidity.

7

Table 1.2: Data for Example **1.3**

Coating	Humidity	Average Corrosion in Thousands of Cycles to Failure
Uncoated	20%	975
	80%	350
Chemical Corrosion	20%	1750
	80%	1550

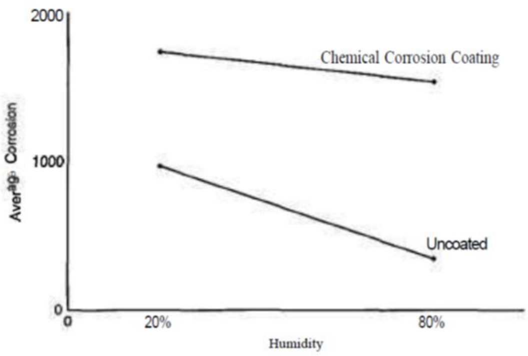
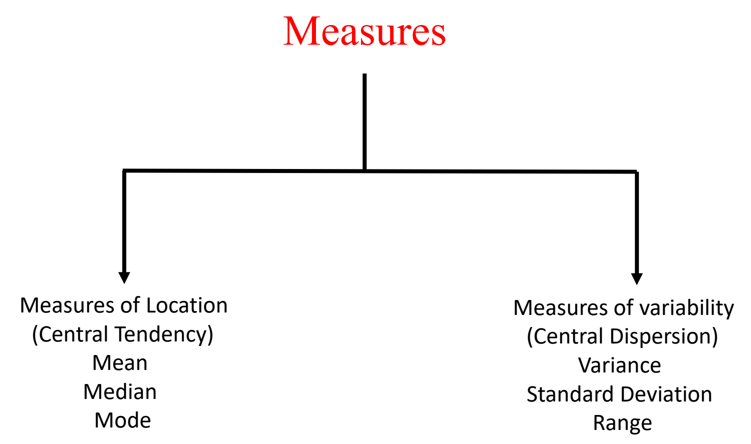


Figure 1.3: Corrosion results for Example 1.3.

8



٩

9

1.4 Measures of Location: The Sample Mean and Median

Definition 1.1: Suppose that the observations in a sample are $x_1, x_2, \ldots, x_n$. The **sample mean**, denoted by $\bar{x}$, is

$$\bar{x} = \sum_{i=1}^{n} \frac{x_i}{n} = \frac{x_1 + x_2 + \cdots + x_n}{n}.$$

Definition 1.2: Given that the observations in a sample are $x_1, x_2, \ldots, x_n$, arranged in **increasing order** of magnitude, the sample median is

$$\tilde{x} = \begin{cases} x_{(n+1)/2}, & \text{if } n \text{ is odd,} \\ \frac{1}{2}(x_{n/2} + x_{n/2+1}), & \text{if } n \text{ is even.} \end{cases}$$

As an example, suppose the data set is the following: 1.7, 2.2, 3.9, 3.11, and 14.7. The sample mean and median are, respectively,

$$\bar{x} = 5.12, \quad \tilde{x} = 3.9.$$

١٠

10

Clearly, the mean is influenced considerably by the presence of the extreme observation, 14.7, whereas the median places emphasis on the true "center" of the data set. In the case of the two-sample data set of Example 1.2, the two measures of central tendency for the individual samples are

$$\begin{aligned}
\bar{x}\text{ (no nitrogen)} &= 0.399 \text{ gram},\\
\tilde{x}\text{ (no nitrogen)} &= \frac{0.38+0.42}{2} = 0.400 \text{ gram},\\
\bar{x}\text{ (nitrogen)} &= 0.565 \text{ gram},\\
\tilde{x}\text{ (nitrogen)} &= \frac{0.49+0.52}{2} = 0.505 \text{ gram}.
\end{aligned}$$

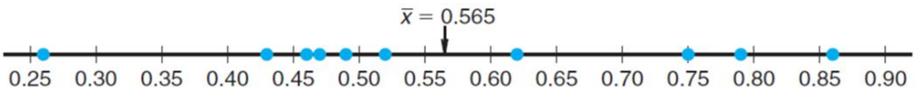


Figure 1.4: Sample mean as a centroid of the with-nitrogen stem weight.

11

Other Measures of Locations

-Trimmed Mean

e.g., in computing 10% trimmed mean, we cancel the highest 10% and the lowers 10% of our data

-Benefit:

1)Having a mean close to median

2) Reduce the effect of very high and very low value

size is 10 for each sample. So for the without-nitrogen group the 10% trimmed mean is given by

$$\bar{x}_{\text{tr}(10)} = \frac{0.32+0.37+0.47+0.43+0.36+0.42+0.38+0.43}{8} = 0.39750,$$

and for the 10% trimmed mean for the with-nitrogen group we have

$$\bar{x}_{\text{tr}(10)} = \frac{0.43+0.47+0.49+0.52+0.75+0.79+0.62+0.46}{8} = 0.56625.$$

12

Measures of Variability
-Variance and Standard deviation
-Range

The **sample variance**, denoted by s^2, is given by

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

The **sample standard deviation**, denoted by s, is the positive square root of s^2, that is,

$$s = \sqrt{s^2}.$$

١٣

13

Example 1.4: In an example discussed extensively in Chapter 10, an engineer is interested in testing the "bias" in a pH meter. Data are collected on the meter by measuring the pH of a neutral substance (pH = 7.0). A sample of size 10 is taken, with results given by

7.07 7.00 7.10 6.97 7.00 7.03 7.01 7.01 6.98 7.08.

The sample mean $\bar{x}$ is given by

$$\bar{x} = \frac{7.07 + 7.00 + 7.10 + \cdots + 7.08}{10} = 7.0250.$$

The sample variance s^2 is given by

$$s^2 = \frac{1}{9}[(7.07 - 7.025)^2 + (7.00 - 7.025)^2 + (7.10 - 7.025)^2 + \cdots + (7.08 - 7.025)^2] = 0.001939.$$

As a result, the sample standard deviation is given by

$$s = \sqrt{0.001939} = 0.044.$$

So the sample standard deviation is 0.0440 with $n - 1 = 9$ degrees of freedom.

١٤

14

Units for Standard Deviation and Variance????
SD= same data unit
S=same data unit squire

Discrete and Continuous Data (countable and uncountable)

Example on Discrete Data
-Family size
-number of class student
-Failure rate (number of failure per unit time)
Example on continues Data
-High, weight, width
-Hand strength
-Human lifespan

١٥

15

Statistical Modeling, Scientific Inspection, and Graphical Diagnostics

1) Scatter Plot

At times the model postulated may take on a somewhat complicated form. Consider, for example, a textile manufacturer who designs an experiment where cloth specimen that contain various percentages of cotton are produced. Consider the data in Table 1.3.

Table 1.3: Tensile Strength

Cotton Percentage	Tensile Strength
15	7, 7, 9, 8, 10
20	19, 20, 21, 20, 22
25	21, 21, 17, 19, 20
30	8, 7, 8, 9, 10

١٦

16

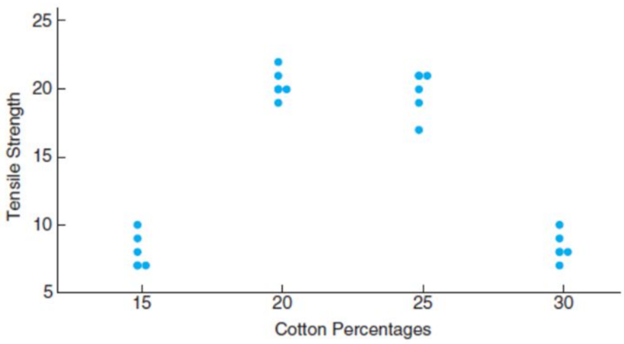


Figure 1.5: Scatter plot of tensile strength and cotton percentages.

١٧

17

2) Stem-and-Leaf Plot

e.g., To illustrate the construction of a stem-and-leaf plot, consider the data of Table 1.4, which specifies the "life" of 40 similar car batteries recorded to the nearest tenth of a year.

Table 1.4: Car Battery Life

2.2	4.1	3.5	4.5	3.2	3.7	3.0	2.6
3.4	1.6	3.1	3.3	3.8	3.1	4.7	3.7
2.5	4.3	3.4	3.6	2.9	3.3	3.9	3.1
3.3	3.1	3.7	4.4	3.2	4.1	1.9	3.4
4.7	3.8	3.2	2.6	3.9	3.0	4.2	3.5

Table 1.5: Stem-and-Leaf Plot of Battery Life

Stem	Leaf	Frequency
1	69	2
2	25669	5
3	0011112223334445567778899	25
4	11234577	8

١٨

18

Table 1.6: Double-Stem-and-Leaf Plot of Battery Life

Stem	Leaf	Frequency
1·	69	2
2⋆	2	1
2·	5669	4
3⋆	001111222333444	15
3·	5567778899	10
4⋆	11234	5
4·	577	3

١٩

19

3) Histogram

Table 1.7: Relative Frequency Distribution of Battery Life

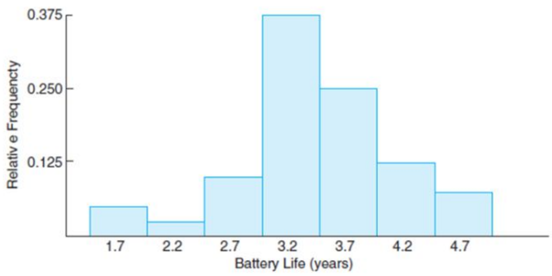
Class Interval	Class Midpoint	Frequency, f	Relative Frequency
1.5–1.9	1.7	2	0.050
2.0–2.4	2.2	1	0.025
2.5–2.9	2.7	4	0.100
3.0–3.4	3.2	15	0.375
3.5–3.9	3.7	10	0.250
4.0–4.4	4.2	5	0.125
4.5–4.9	4.7	3	0.075

Mean of Histogram

Mean= $\frac{\sum Class\ Midpoint \times f}{\sum f}$

$= \frac{1.7 \times 2 + 2.2 \times 1 + \cdots + 4.7 \times 3}{2 + 1 + \cdots + 3}$

=3.4125



٢٠

Figure 1.6: Relative frequency histogram.

20

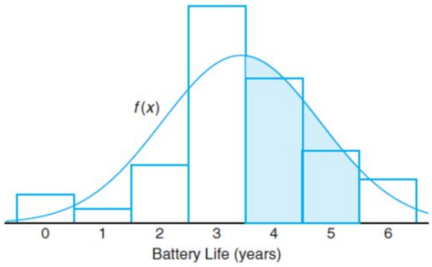


Figure 1.7: Estimating frequency distribution.

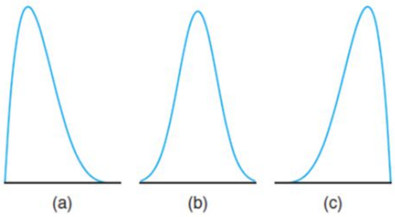


Figure 1.8: Skewness of data.

٢١

21

First Quartile and Third Quartile

Definitions:

- The **lower half** of a data set is the set of all values that are to the left of the median value when the data has been put into increasing order.
- The **upper half** of a data set is the set of all values that are to the right of the median value when the data has been put into increasing order.
- The **first quartile,** denoted by Q_1 , is the median of the *lower half* of the data set. This means that about 25% of the numbers in the data set lie below Q_1 and about 75% lie above Q_1 .
- The **third quartile,** denoted by Q_3 , is the median of the *upper half* of the data set. This means that about 75% of the numbers in the data set lie below Q_3 and about 25% lie above Q_3 .

٢٢

22

First, we write data iExample 1: Find the first and third quartiles of the data set {3, 7, 8, 5, 12, 14, 21, 13, 16, 18}.

in increasing order: 3, 5, 7, 8, 12, 13, 14, 16, 18, 21.

Location of Q1: (10+1)*0.25=2.75
Interpolation
Q1=value of location 2+0.75*(value of location 2- value of location 3)
Q1=5+0.75*(7-5)=6.5
Location of Q2: (10+1)*0.5=5.5
Q2= (12+13)/2=12.5
Location of Q3: (10+1)*0.75=8.25
Q3=value of location 8+0.25*(value of location 9- value of location 8)
Q3=16+0.25*(18-16)=16.5

Inter quartile range (IQR)= Q3-Q1

٢٣ IQR =16.5-6.5=10

4) Box-and-Whisker Plot or Box Plot
-You have to know to estimate the percentile and quartile

e.g., Nicotine content was measured in a random sample of 40 cigarettes. The data are displayed in Table 1.8.

Table 1.8: Nicotine Data for Example 1.5

1.09	1.92	2.31	1.79	2.28	1.74	1.47	1.97
0.85	1.24	1.58	2.03	1.70	2.17	2.55	2.11
1.86	1.90	1.68	1.51	1.64	0.72	1.69	1.85
1.82	1.79	2.46	1.88	2.08	1.67	1.37	1.93
1.40	1.64	2.09	1.75	1.63	2.37	1.75	1.69

In order

0.72	**0.85**	**1.09**	**1.24**	**1.37**	**1.4**	**1.47**	**1.51**	**1.58**	**1.63**
1.64	**1.64**	**1.67**	**1.68**	**1.69**	**1.69**	**1.7**	**1.74**	**1.75**	**1.75**
1.79	**1.79**	**1.82**	**1.85**	**1.86**	**1.88**	**1.9**	**1.92**	**1.93**	**1.97**
2.03	**2.08**	**2.09**	**2.11**	**2.17**	**2.28**	**2.31**	**2.37**	**2.46**	**2.55**

٢٤

Location of Q1: (40+1)*0.25=10.25
Q1: 1.63+0.25*(1.64-1.63)=1.6325

Q2=(1.75+1.79)/2=1.77

Location of Q3: (40+1)*0.75=30.75
Q3: 1.97+0.75*(2.03-1.97)=2.015

IQR=2.015-1.6325=0.3825

1.5IQR=0.57375

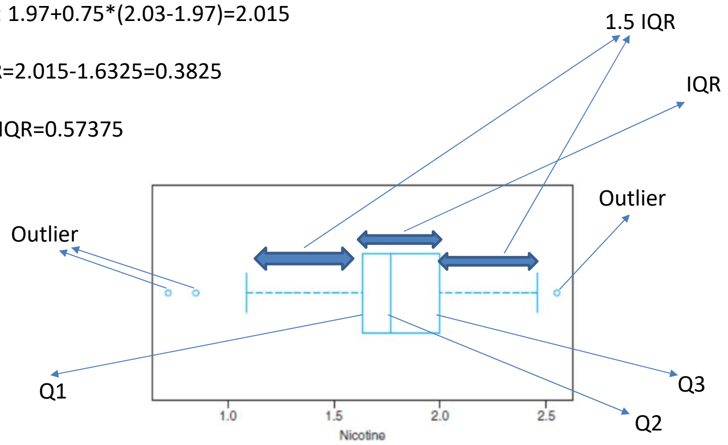


Figure 1.9: Box-and-whisker plot for Example 1.5.

٢٥

25

Percentile

- Find 62% percentile of Example on slides number 23

Location of 62% percentile is (n+1)*0.62=6.82

62% percentile=Value of location 6+0.82*(Value of location 7-Value of location 6)=13+0.82*(14-13)=13.82

- Find 29% percentile of Example on slides number 24

Location of 29% percentile is (40+1)*0.29=11.89

29% percentile=Value of location 11+0.89*(Value of location 12-Value of location 11)=1.64+0.89*(1.64-1.64)=1.64

٢٦

26

Probability and Statistics

Chapter 2

Dr. Raed Al Athamneh

Department of Industrial Engineering

Definition 2.1: The set of all possible outcomes of a statistical experiment is called the **sample space** and is represented by the symbol S.

Each outcome in a sample space is called an **element** or a **member** of the sample space, or simply a **sample point**. If the sample space has a finite number of elements, we may *list* the members separated by commas and enclosed in braces. Thus, the sample space S, of possible outcomes when a coin is flipped, may be written

$$S = \{H, T\},$$

where H and T correspond to heads and tails, respectively.

Example 2.1: Consider the experiment of tossing a die. If we are interested in the number that shows on the top face, the sample space is

$$S_1 = \{1, 2, 3, 4, 5, 6\}.$$

If we are interested only in whether the number is even or odd, the sample space is simply

$$S_2 = \{\text{even}, \text{odd}\}.$$

٢

Example 2.2: An experiment consists of flipping a coin and then flipping it a second time if a head occurs. If a tail occurs on the first flip, then a die is tossed once. To list the elements of the sample space providing the most information, we construct the tree diagram of Figure 2.1. The various paths along the branches of the tree give the distinct sample points. Starting with the top left branch and moving to the right along the first path, we get the sample point *HH*, indicating the possibility that heads occurs on two successive flips of the coin. Likewise, the sample point *T*3 indicates the possibility that the coin will show a tail followed by a 3 on the toss of the die. By proceeding along all paths, we see that the sample space is

$$S = \{HH,\ HT,\ T1,\ T2,\ T3,\ T4,\ T5,\ T6\}.$$

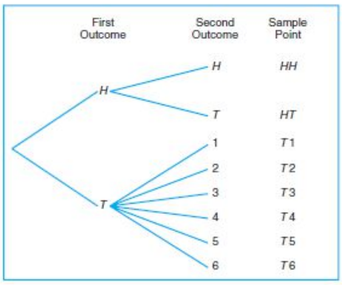


Figure 2.1: Tree diagram for Example 2.2.

٣

3

Example 2.3: Suppose that three items are selected at random from a manufacturing process. Each item is inspected and classified defective, *D*, or nondefective, *N*. To list the elements of the sample space providing the most information, we construct the tree diagram of Figure 2.2. Now, the various paths along the branches of the tree give the distinct sample points. Starting with the first path, we get the sample point *DDD*, indicating the possibility that all three items inspected are defective. As we proceed along the other paths, we see that the sample space is

$$S = \{DDD,\ DDN,\ DND,\ DNN,\ NDD,\ NDN,\ NND,\ NNN\}.$$

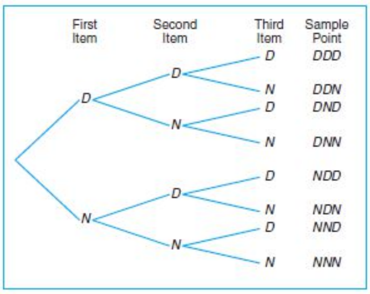


Figure 2.2: Tree diagram for Example 2.3.

٤

4

Definition 2.2: An **event** is a subset of a sample space.

Example 2.4: Given the sample space $S = \{t \mid t \geq 0\}$, where t is the life in years of a certain electronic component, then the event A that the component fails before the end of the fifth year is the subset $A = \{t \mid 0 \leq t < 5\}$.

Definition 2.3: The **complement** of an event A with respect to S is the subset of all elements of S that are not in A. We denote the complement of A by the symbol A'.

Example 2.5: Let R be the event that a red card is selected from an ordinary deck of 52 playing cards, and let S be the entire deck. Then R' is the event that the card selected from the deck is not a red card but a black card.

Example 2.6: Consider the sample space

$$S = \{\text{book, cell phone, mp3, paper, stationery, laptop}\}.$$

Let $A = \{\text{book, stationery, laptop, paper}\}$. Then the complement of A is $A' = \{\text{cell phone, mp3}\}$.

٥

5

Definition 2.4: The **intersection** of two events A and B, denoted by the symbol $A \cap B$, is the event containing all elements that are common to A and B.

Example 2.7: Let E be the event that a person selected at random in a classroom is majoring in engineering, and let F be the event that the person is female. Then $E \cap F$ is the event of all female engineering students in the classroom.

Example 2.8: Let $V = \{a, e, i, o, u\}$ and $C = \{l, r, s, t\}$; then it follows that $V \cap C = \phi$. That is, V and C have no elements in common and, therefore, cannot both simultaneously occur.

For certain statistical experiments it is by no means unusual to define two events, A and B, that cannot both occur simultaneously. The events A and B are then said to be **mutually exclusive**. Stated more formally, we have the following definition:

Definition 2.5: Two events A and B are **mutually exclusive**, or **disjoint**, if $A \cap B = \phi$, that is, if A and B have no elements in common.

Definition 2.6: The **union** of the two events A and B, denoted by the symbol $A \cup B$, is the event containing all the elements that belong to A or B or both.

Example 2.10: Let $A = \{a, b, c\}$ and $B = \{b, c, d, e\}$; then $A \cup B = \{a, b, c, d, e\}$.

٦

6

Example 2.11: Let P be the event that an employee selected at random from an oil drilling company smokes cigarettes. Let Q be the event that the employee selected drinks alcoholic beverages. Then the event $P \cup Q$ is the set of all employees who either drink or smoke or do both.

Example 2.12: If $M = \{x \mid 3 < x < 9\}$ and $N = \{y \mid 5 < y < 12\}$, then

$$M \cup N = \{z \mid 3 < z < 12\}.$$

The relationship between events and the corresponding sample space can be illustrated graphically by means of **Venn diagrams**. In a Venn diagram we let the sample space be a rectangle and represent events by circles drawn inside the rectangle. Thus, in Figure 2.3, we see that

$$A \cap B = \text{ regions 1 and 2,}$$
$$B \cap C = \text{ regions 1 and 3,}$$

Y

7

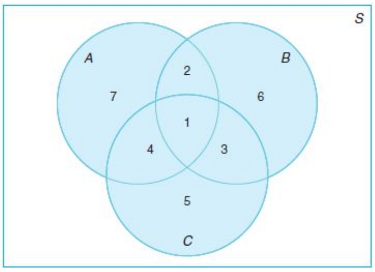


Figure 2.3: Events represented by various regions.

$$A \cup C = \text{ regions 1, 2, 3, 4, 5, and 7,}$$
$$B' \cap A = \text{ regions 4 and 7,}$$
$$A \cap B \cap C = \text{ region 1,}$$
$$(A \cup B) \cap C' = \text{ regions 2, 6, and 7,}$$

A

8

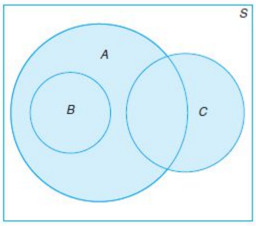


Figure 2.4: Events of the sample space S.

In Figure 2.4, we see that events A, B, and C are all subsets of the sample space S. It is also clear that event B is a subset of event A; event $B \cap C$ has no elements and hence B and C are mutually exclusive; event $A \cap C$ has at least one element; and event $A \cup B = A$. Figure 2.4 might, therefore, depict a situation where we select a card at random from an ordinary deck of 52 playing cards and observe whether the following events occur:

A: the card is red,

B: the card is the jack, queen, or king of diamonds,

C: the card is an ace.

Clearly, the event $A \cap C$ consists of only the two red aces.

Several results that follow from the foregoing definitions, which may easily be verified by means of Venn diagrams, are as follows:

1. $A \cap \phi = \phi$.
2. $A \cup \phi = A$.
3. $A \cap A' = \phi$.
4. $A \cup A' = S$.
5. $S' = \phi$.
6. $\phi' = S$.
7. $(A')' = A$.
8. $(A \cap B)' = A' \cup B'$.
9. $(A \cup B)' = A' \cap B'$.

٩

9

Rule 2.1: 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.

xample 2.13: 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 2.14: A developer of a new subdivision offers prospective home buyers a choice of Tudor, rustic, colonial, and traditional exterior styling in ranch, two-story, and split-level floor plans. In how many different ways can a buyer order one of these homes?

Solution:

Since $n_1 = 4$ and $n_2 = 3$, a buyer must choose from

$$n_1 n_2 = (4)(3) = 12 \text{ possible homes.}$$

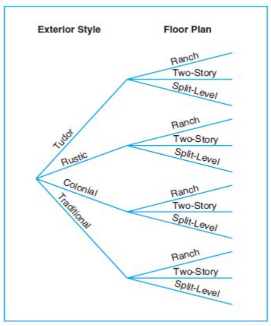


Figure 2.6: Tree diagram for Example 2.14.

١٠

10

Example 2.15: 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 $n_1 \times n_2 = 22 \times 21 = 462$ different ways.

The multiplication rule, Rule 2.1 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 2.2: 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 2.16: 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.

١١

11

Example 2.17: How many even four-digit numbers can be formed from the digits 0, 1, 2, 5, 6, and 9 if each digit can be used only once?

***Solution*:** Since the number must be even, we have only $n_1 = 3$ choices for the units position. However, for a four-digit number the thousands position cannot be 0. Hence, we consider the units position in two parts, 0 or not 0. If the units position is 0 (i.e., $n_1 = 1$), we have $n_2 = 5$ choices for the thousands position, $n_3 = 4$ for the hundreds position, and $n_4 = 3$ for the tens position. Therefore, in this case we have a total of

$$n_1 n_2 n_3 n_4 = (1)(5)(4)(3) = 60$$

even four-digit numbers. On the other hand, if the units position is not 0 (i.e., $n_1 = 2$), we have $n_2 = 4$ choices for the thousands position, $n_3 = 4$ for the hundreds position, and $n_4 = 3$ for the tens position. In this situation, there are a total of

$$n_1 n_2 n_3 n_4 = (2)(4)(4)(3) = 96$$

Definition 2.7: 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 2.2, 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

$$n_1 n_2 n_3 = (3)(2)(1) = 6 \text{ permutations}$$

by Rule 2.2. In general, n distinct objects can be arranged in

$$n(n-1)(n-2)\cdots(3)(2)(1) \text{ ways.}$$

١٢

There is a notation for such a number.

12

Definition 2.8: For any non-negative integer n, $n!$, called "n factorial," is defined as

$$n! = n(n-1)\cdots(2)(1),$$

with special case $0! = 1$.

Theorem 2.1: The number of permutations of n objects is $n!$.

Theorem 2.2: The number of permutations of n distinct objects taken r at a time is

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

Example 2.18: 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

$${}_{25}P_3 = \frac{25!}{(25-3)!} = \frac{25!}{22!} = (25)(24)(23) = 13,800.$$

١٣

13

Example 2.19: 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

(a) there are no restrictions;

(b) A will serve only if he is president;

(c) B and C will serve together or not at all;

(d) D and E will not serve together?

Solution: (a) The total number of choices of officers, without any restrictions, is

$${}_{50}P_2 = \frac{50!}{48!} = (50)(49) = 2450.$$

(b) Since A will serve only if he is president, we have two situations here: (i) A is selected as the president, which yields 49 possible outcomes for the treasurer's position, or (ii) officers are selected from the remaining 49 people without A, which has the number of choices ${}_{49}P_2 = (49)(48) = 2352$. Therefore, the total number of choices is $49 + 2352 = 2401$.

(c) The number of selections when B and C serve together is 2. The number of selections when both B and C are not chosen is ${}_{48}P_2 = 2256$. Therefore, the total number of choices in this situation is $2 + 2256 = 2258$.

(d) The number of selections when D serves as an officer but not E is $(2)(48) = 96$, where 2 is the number of positions D can take and 48 is the number of selections of the other officer from the remaining people in the club except E. The number of selections when E serves as an officer but not D is also $(2)(48) = 96$. The number of selections when both D and E are not chosen is ${}_{48}P_2 = 2256$. Therefore, the total number of choices is $(2)(96) + 2256 = 2448$. This problem also has another short solution: Since D and E can only serve together in 2 ways, the answer is $2450 - 2 = 2448$.

١٤

14

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 2.3: The number of permutations of n objects arranged in a circle is $(n-1)!$.

So far we have considered permutations of distinct objects. That is, all the objects were completely different or distinguishable. Obviously, if the letters b and c are both equal to x, then the 6 permutations of the letters a, b, and c become axx, axx, xax, xax, xxa, and xxa, of which only 3 are distinct. Therefore, with 3 letters, 2 being the same, we have $3!/2! = 3$ distinct permutations. With 4 different letters a, b, c, and d, we have 24 distinct permutations. If we let $a = b = x$ and $c = d = y$, we can list only the following distinct permutations: $xxyy$, $xyxy$, $yxxy$, $yyxx$, $xyyx$, and $yxyx$. Thus, we have $4!/(2!\ 2!) = 6$ distinct permutations.

Theorem 2.4: The number of distinct permutations of n things of which n_1 are of one kind, n_2 of a second kind, $\ldots$, n_k of a kth kind is

$$\frac{n!}{n_1!n_2!\cdots n_k!}.$$

١٥

15

Example 2.20: 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 2.4, we find that the total number of arrangements is

$$\frac{10!}{1!\ 2!\ 4!\ 3!} = 12,600.$$

Theorem 2.5: The number of ways of partitioning a set of n objects into r cells with n_1 elements in the first cell, n_2 elements in the second, and so forth, is

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

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

Example 2.21: 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

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

١٦

16

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

$$\binom{n}{r, n-r}, \text{ is usually shortened to } \binom{n}{r},$$

since the number of elements in the second cell must be $n-r$.

Theorem 2.6: The number of combinations of n distinct objects taken r at a time is

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

Example 2.22: 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 (Rule 2.1) with $n_1 = 120$ and $n_2 = 10$, we have $(120)(10) = 1200$ ways.

١٧

17

Example 2.23: How many different letter arrangements can be made from the letters in the word *STATISTICS*?

***Solution*:** Using the same argument as in the discussion for Theorem 2.6, in this example we can actually apply Theorem 2.5 to obtain

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

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 2.4.

١٨

18

Definition 2.9: The **probability** of an event A is the sum of the weights of all sample points in A. Therefore,

$$0 \leq P(A) \leq 1, \quad P(\phi) = 0, \quad \text{and} \quad P(S) = 1.$$

Furthermore, if $A_1, A_2, A_3, \ldots$ is a sequence of mutually exclusive events, then

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

Example 2.24: A coin is tossed twice. What is the probability that at least 1 head occurs?

Solution: The sample space for this experiment is

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

If the coin is balanced, each of these outcomes is equally likely to occur. Therefore, we assign a probability of ω to each sample point. Then $4\omega = 1$, or $\omega = 1/4$. If A represents the event of at least 1 head occurring, then

$$A = \{HH, HT, TH\} \text{ and } P(A) = \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = \frac{3}{4}.$$

١٩

Example 2.25: A die is loaded in such a way that an even number is twice as likely to occur as an odd number. If E is the event that a number less than 4 occurs on a single toss of the die, find $P(E)$.

Solution: The sample space is $S = \{1, 2, 3, 4, 5, 6\}$. We assign a probability of w to each odd number and a probability of $2w$ to each even number. Since the sum of the probabilities must be 1, we have $9w = 1$ or $w = 1/9$. Hence, probabilities of 1/9 and 2/9 are assigned to each odd and even number, respectively. Therefore,

$$E = \{1, 2, 3\} \text{ and } P(E) = \frac{1}{9} + \frac{2}{9} + \frac{1}{9} = \frac{4}{9}.$$

٢٠

Example 2.26: In Example 2.25, let A be the event that an even number turns up and let B be the event that a number divisible by 3 occurs. Find $P(A \cup B)$ and $P(A \cap B)$.

***Solution*:** For the events $A = \{2, 4, 6\}$ and $B = \{3, 6\}$, we have

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

By assigning a probability of 1/9 to each odd number and 2/9 to each even number, we have

$$P(A \cup B) = \frac{2}{9} + \frac{1}{9} + \frac{2}{9} + \frac{2}{9} = \frac{7}{9} \quad \text{and} \quad P(A \cap B) = \frac{2}{9}.$$

If the sample space for an experiment contains N elements, all of which are equally likely to occur, we assign a probability equal to $1/N$ to each of the N points. The probability of any event A containing n of these N sample points is then the ratio of the number of elements in A to the number of elements in S.

Rule 2.3: If an experiment can result in any one of N different equally likely outcomes, and if exactly n of these outcomes correspond to event A, then the probability of event A is

$$P(A) = \frac{n}{N}.$$

Example 2.27: A statistics class for engineers consists of 25 industrial, 10 mechanical, 10 electrical, and 8 civil engineering students. If a person is randomly selected by the instructor to answer a question, find the probability that the student chosen is (a) an industrial engineering major and (b) a civil engineering or an electrical engineering major.

***Solution*:** Denote by I, M, E, and C the students majoring in industrial, mechanical, electrical, and civil engineering, respectively. The total number of students in the class is 53, all of whom are equally likely to be selected.

(a) Since 25 of the 53 students are majoring in industrial engineering, the probability of event I, selecting an industrial engineering major at random, is

$$P(I) = \frac{25}{53}.$$

(b) Since 18 of the 53 students are civil or electrical engineering majors, it follows that

$$P(C \cup E) = \frac{18}{53}.$$

Theorem 2.7: If A and B are two events, then

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

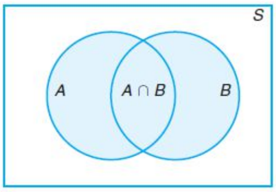


Figure 2.7: Additive rule of probability.

***Proof*:** Consider the Venn diagram in Figure 2.7. The $P(A \cup B)$ is the sum of the probabilities of the sample points in $A \cup B$. Now $P(A) + P(B)$ is the sum of all the probabilities in A plus the sum of all the probabilities in B. Therefore, we have added the probabilities in $(A \cap B)$ twice. Since these probabilities add up to $P(A \cap B)$, we must subtract this probability once to obtain the sum of the probabilities in $A \cup B$.

٢٣

23

Corollary 2.1: If A and B are mutually exclusive, then

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

Corollary 2.1 is an immediate result of Theorem 2.7, since if A and B are mutually exclusive, $A \cap B = 0$ and then $P(A \cap B) = P(\phi) = 0$. In general, we can write Corollary 2.2.

Corollary 2.2: If $A_1, A_2, \ldots, A_n$ are mutually exclusive, then

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = P(A_1) + P(A_2) + \cdots + P(A_n).$$

A collection of events $\{A_1, A_2, \ldots, A_n\}$ of a sample space S is called a **partition** of S if $A_1, A_2, \ldots, A_n$ are mutually exclusive and $A_1 \cup A_2 \cup \cdots \cup A_n = S$. Thus, we have

Corollary 2.3: If $A_1, A_2, \ldots, A_n$ is a partition of sample space S, then

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = P(A_1) + P(A_2) + \cdots + P(A_n) = P(S) = 1.$$

As one might expect, Theorem 2.7 extends in an analogous fashion.

Theorem 2.8: For three events A, B, and C,

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

٢٤

24

Example 2.30: What is the probability of getting a total of 7 or 11 when a pair of fair dice is tossed?

Solution: Let A be the event that 7 occurs and B the event that 11 comes up. Now, a total of 7 occurs for 6 of the 36 sample points, and a total of 11 occurs for only 2 of the sample points. Since all sample points are equally likely, we have $P(A) = 1/6$ and $P(B) = 1/18$. The events A and B are mutually exclusive, since a total of 7 and 11 cannot both occur on the same toss. Therefore,

$$P(A \cup B) = P(A) + P(B) = \frac{1}{6} + \frac{1}{18} = \frac{2}{9}.$$

This result could also have been obtained by counting the total number of points for the event $A \cup B$, namely 8, and writing

$$P(A \cup B) = \frac{n}{N} = \frac{8}{36} = \frac{2}{9}.$$

Example 2.31: If the probabilities are, respectively, 0.09, 0.15, 0.21, and 0.23 that a person purchasing a new automobile will choose the color green, white, red, or blue, what is the probability that a given buyer will purchase a new automobile that comes in one of those colors?

Solution: Let G, W, R, and B be the events that a buyer selects, respectively, a green, white, red, or blue automobile. Since these four events are mutually exclusive, the probability is

$$P(G \cup W \cup R \cup B) = P(G) + P(W) + P(R) + P(B)$$
$$= 0.09 + 0.15 + 0.21 + 0.23 = 0.68.$$

Often it is more difficult to calculate the probability that an event occurs than it is to calculate the probability that the event does not occur. Should this be the case for some event A, we simply find $P(A')$ first and then, using Theorem 2.7, find $P(A)$ by subtraction.

٢٥

25

Theorem 2.9: If A and A' are complementary events, then

$$P(A) + P(A') = 1.$$

Proof: Since $A \cup A' = S$ and the sets A and A' are disjoint,

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

Example 2.32: If the probabilities that an automobile mechanic will service 3, 4, 5, 6, 7, or 8 or more cars on any given workday are, respectively, 0.12, 0.19, 0.28, 0.24, 0.10, and 0.07, what is the probability that he will service at least 5 cars on his next day at work?

Solution: Let E be the event that at least 5 cars are serviced. Now, $P(E) = 1 - P(E')$, where E' is the event that fewer than 5 cars are serviced. Since

$$P(E') = 0.12 + 0.19 = 0.31,$$

it follows from Theorem 2.9 that

$$P(E) = 1 - 0.31 = 0.69.$$

٢٦

26

Example 2.33: Suppose the manufacturer's specifications for the length of a certain type of computer cable are 2000 ± 10 millimeters. In this industry, it is known that small cable is just as likely to be defective (not meeting specifications) as large cable. That is, the probability of randomly producing a cable with length exceeding 2010 millimeters is equal to the probability of producing a cable with length smaller than 1990 millimeters. The probability that the production procedure meets specifications is known to be 0.99.

(a) What is the probability that a cable selected randomly is too large?

(b) What is the probability that a randomly selected cable is larger than 1990 millimeters?

Solution: Let M be the event that a cable meets specifications. Let S and L be the events that the cable is too small and too large, respectively. Then

(a) $P(M) = 0.99$ and $P(S) = P(L) = (1 - 0.99)/2 = 0.005$.

(b) Denoting by X the length of a randomly selected cable, we have

$$P(1990 \le X \le 2010) = P(M) = 0.99.$$

Since $P(X \ge 2010) = P(L) = 0.005$,

$$P(X \ge 1990) = P(M) + P(L) = 0.995.$$

This also can be solved by using Theorem 2.9:

$$P(X \ge 1990) + P(X < 1990) = 1.$$

Thus, $P(X \ge 1990) = 1 - P(S) = 1 - 0.005 = 0.995$.

٢٧

27

Definition 2.10: The conditional probability of B, given A, denoted by $P(B|A)$, is defined by

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

Table 2.1: Categorization of the Adults in a Small Town

	Employed	Unemployed	Total
Male	460	40	500
Female	140	260	400
Total	600	300	900

$$P(M|E) = \frac{n(E \cap M)}{n(E)} = \frac{n(E \cap M)/n(S)}{n(E)/n(S)} = \frac{P(E \cap M)}{P(E)},$$

$$P(E) = \frac{600}{900} = \frac{2}{3} \quad \text{and} \quad P(E \cap M) = \frac{460}{900} = \frac{23}{45}.$$

$$P(M|E) = \frac{460}{600} = \frac{23}{30}.$$

٢٨

28

Example 2.34: The probability that a regularly scheduled flight departs on time is $P(D) = 0.83$; the probability that it arrives on time is $P(A) = 0.82$; and the probability that it departs and arrives on time is $P(D \cap A) = 0.78$. Find the probability that a plane (a) arrives on time, given that it departed on time, and (b) departed on time, given that it has arrived on time.

Solution: Using Definition 2.10, we have the following.

(a) The probability that a plane arrives on time, given that it departed on time, is

$$P(A|D) = \frac{P(D \cap A)}{P(D)} = \frac{0.78}{0.83} = 0.94.$$

(b) The probability that a plane departed on time, given that it has arrived on time, is

$$P(D|A) = \frac{P(D \cap A)}{P(A)} = \frac{0.78}{0.82} = 0.95.$$

٢٩

29

Example 2.35: The concept of conditional probability has countless uses in both industrial and biomedical applications. Consider an industrial process in the textile industry in which strips of a particular type of cloth are being produced. These strips can be defective in two ways, length and nature of texture. For the case of the latter, the process of identification is very complicated. It is known from historical information on the process that 10% of strips fail the length test, 5% fail the texture test, and only 0.8% fail both tests. If a strip is selected randomly from the process and a quick measurement identifies it as failing the length test, what is the probability that it is texture defective?

Solution: Consider the events

L: length defective, T: texture defective.

Given that the strip is length defective, the probability that this strip is texture defective is given by

$$P(T|L) = \frac{P(T \cap L)}{P(L)} = \frac{0.008}{0.1} = 0.08.$$

٣٠

30

Definition 2.11: Two events A and B are **independent** if and only if

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

assuming the existences of the conditional probabilities. Otherwise, A and B are **dependent**.

Theorem 2.10: If in an experiment the events A and B can both occur, then

$$P(A \cap B) = P(A)P(B|A), \text{ provided } P(A) > 0.$$

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

Example 2.36: Suppose that we have a fuse box containing 20 fuses, of which 5 are defective. If 2 fuses are selected at random and removed from the box in succession without replacing the first, what is the probability that both fuses are defective?

Solution: We shall let A be the event that the first fuse is defective and B the event that the second fuse is defective; then we interpret $A \cap B$ as the event that A occurs and then B occurs after A has occurred. The probability of first removing a defective fuse is $1/4$; then the probability of removing a second defective fuse from the remaining 4 is $4/19$. Hence,

$$P(A \cap B) = \left(\frac{1}{4}\right)\left(\frac{4}{19}\right) = \frac{1}{19}.$$

Example 2.37: One bag contains 4 white balls and 3 black balls, and a second bag contains 3 white balls and 5 black balls. One ball is drawn from the first bag and placed unseen in the second bag. What is the probability that a ball now drawn from the second bag is black?

Solution: Let B_1, B_2, and W_1 represent, respectively, the drawing of a black ball from bag 1, a black ball from bag 2, and a white ball from bag 1. We are interested in the union of the mutually exclusive events $B_1 \cap B_2$ and $W_1 \cap B_2$. The various possibilities and their probabilities are illustrated in Figure 2.8. Now

$$\begin{aligned} P[(B_1 \cap B_2) \text{ or } (W_1 \cap B_2)] &= P(B_1 \cap B_2) + P(W_1 \cap B_2) \\ &= P(B_1)P(B_2|B_1) + P(W_1)P(B_2|W_1) \\ &= \left(\frac{3}{7}\right)\left(\frac{6}{9}\right) + \left(\frac{4}{7}\right)\left(\frac{5}{9}\right) = \frac{38}{63}. \end{aligned}$$

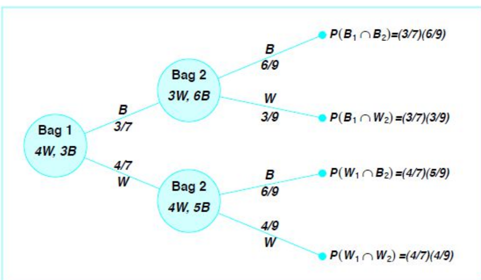


Figure 2.8: Tree diagram for Example 2.37.

Theorem 2.11: Two events A and B are independent if and only if

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

Therefore, to obtain the probability that two independent events will both occur, we simply find the product of their individual probabilities.

Example 2.38: A small town has one fire engine and one ambulance available for emergencies. The probability that the fire engine is available when needed is 0.98, and the probability that the ambulance is available when called is 0.92. In the event of an injury resulting from a burning building, find the probability that both the ambulance and the fire engine will be available, assuming they operate independently.

Solution: Let A and B represent the respective events that the fire engine and the ambulance are available. Then

$$P(A \cap B) = P(A)P(B) = (0.98)(0.92) = 0.9016.$$

٣٣

33

Example 2.39: An electrical system consists of four components as illustrated in Figure 2.9. The system works if components A and B work and either of the components C or D works. The reliability (probability of working) of each component is also shown in Figure 2.9. Find the probability that (a) the entire system works and (b) the component C does not work, given that the entire system works. Assume that the four components work independently.

Solution: In this configuration of the system, A, B, and the subsystem C and D constitute a serial circuit system, whereas the subsystem C and D itself is a parallel circuit system.

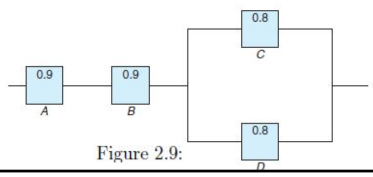
(a) Clearly the probability that the entire system works can be calculated as

$$\begin{aligned} P[A \cap B \cap (C \cup D)] &= P(A)P(B)P(C \cup D) = P(A)P(B)[1 - P(C' \cap D')] \\ &= P(A)P(B)[1 - P(C')P(D')] \\ &= (0.9)(0.9)[1 - (1 - 0.8)(1 - 0.8)] = 0.7776. \end{aligned}$$

The equalities above hold because of the independence among the four components.

(b) To calculate the conditional probability in this case, notice that

$$\begin{aligned} P &= \frac{P(\text{the system works but } C \text{ does not work})}{P(\text{the system works})} \\ &= \frac{P(A \cap B \cap C' \cap D)}{P(\text{the system works})} = \frac{(0.9)(0.9)(1 - 0.8)(0.8)}{0.7776} = 0.1667. \end{aligned}$$



٣٤

Figure 2.9:

34

Theorem 2.12: If, in an experiment, the events $A_1, A_2, \dots, A_k$ can occur, then

$$P(A_1 \cap A_2 \cap \cdots \cap A_k) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P(A_k|A_1 \cap A_2 \cap \cdots \cap A_{k-1}).$$

If the events $A_1, A_2, \dots, A_k$ are independent, then

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

Example 2.40: Three cards are drawn in succession, without replacement, from an ordinary deck of playing cards. Find the probability that the event $A_1 \cap A_2 \cap A_3$ occurs, where A_1 is the event that the first card is a red ace, A_2 is the event that the second card is a 10 or a jack, and A_3 is the event that the third card is greater than 3 but less than 7.

Solution: First we define the events

A_1: the first card is a red ace,

A_2: the second card is a 10 or a jack,

٣٥

A_3: the third card is greater than 3 but less than 7.

Now

$$P(A_1) = \frac{2}{52}, \quad P(A_2|A_1) = \frac{8}{51}, \quad P(A_3|A_1 \cap A_2) = \frac{12}{50},$$

and hence, by Theorem 2.12,

$$P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) = \left(\frac{2}{52}\right)\left(\frac{8}{51}\right)\left(\frac{12}{50}\right) = \frac{8}{5525}.$$

The property of independence stated in Theorem 2.11 can be extended to deal with more than two events. Consider, for example, the case of three events A, B, and C. It is not sufficient to only have that $P(A \cap B \cap C) = P(A)P(B)P(C)$ as a definition of independence among the three. Suppose $A = B$ and $C = \phi$, the null set. Although $A \cap B \cap C = \phi$, which results in $P(A \cap B \cap C) = 0 = P(A)P(B)P(C)$, events A and B are not independent. Hence, we have the following definition.

Definition 2.12: A collection of events $\mathcal{A} = \{A_1, \dots, A_n\}$ are mutually independent if for any subset of $\mathcal{A}$, $A_{i_1}, \dots, A_{i_k}$, for $k \le n$, we have

$$P(A_{i_1} \cap \cdots \cap A_{i_k}) = P(A_{i_1}) \cdots P(A_{i_k}).$$

٣٦

Bayes' Rule

$$P(A) = P[(E \cap A) \cup (E' \cap A)] = P(E \cap A) + P(E' \cap A)$$
$$= P(E)P(A|E) + P(E')P(A|E').$$

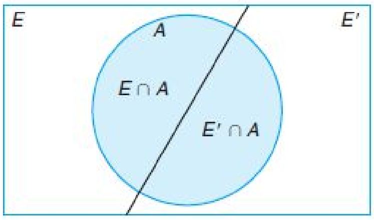


Figure 2.12: Venn diagram for the events A, E, and E'.

Example

$$P(E) = \frac{600}{900} = \frac{2}{3}, \quad P(A|E) = \frac{36}{600} = \frac{3}{50},$$

$$P(E') = \frac{1}{3}, \quad P(A|E') = \frac{12}{300} = \frac{1}{25}.$$

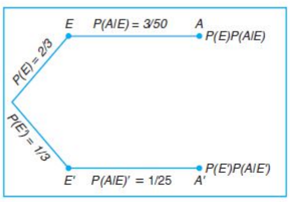


Figure 2.13: Tree diagram for the data on page 63, using additional information on page 72.

the probability $P(E')P(A|E')$, it follows that

$$P(A) = \left(\frac{2}{3}\right)\left(\frac{3}{50}\right) + \left(\frac{1}{3}\right)\left(\frac{1}{25}\right) = \frac{4}{75}.$$

Theorem 2.13: If the events $B_1, B_2, \ldots, B_k$ constitute a partition of the sample space S such that $P(B_i) \neq 0$ for $i = 1, 2, \ldots, k$, then for any event A of S,

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

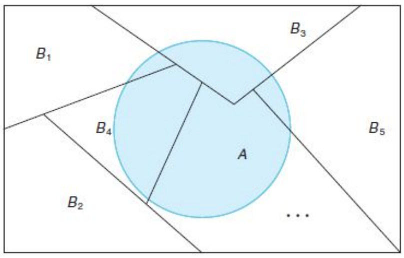


Figure 2.14: Partitioning the sample space S.

٣٩

***Proof*:** Consider the Venn diagram of Figure 2.14. The event A is seen to be the union of the mutually exclusive events

$$B_1 \cap A,\ B_2 \cap A,\ \ldots,\ B_k \cap A;$$

that is,

$$A = (B_1 \cap A) \cup (B_2 \cap A) \cup \cdots \cup (B_k \cap A).$$

Using Corollary 2.2 of Theorem 2.7 and Theorem 2.10, we have

$$\begin{aligned} P(A) &= P[(B_1 \cap A) \cup (B_2 \cap A) \cup \cdots \cup (B_k \cap A)] \\ &= P(B_1 \cap A) + P(B_2 \cap A) + \cdots + P(B_k \cap A) \\ &= \sum_{i=1}^{k} P(B_i \cap A) \\ &= \sum_{i=1}^{k} P(B_i)P(A|B_i). \end{aligned}$$

٤٠

Example 2.41: In a certain assembly plant, three machines, B_1, B_2, and B_3, make 30%, 45%, and 25%, respectively, of the products. It is known from past experience that 2%, 3%, and 2% of the products made by each machine, respectively, are defective. Now, suppose that a finished product is randomly selected. What is the probability that it is defective?

***Solution*:** Consider the following events:

A: the product is defective,

B_1: the product is made by machine B_1,

B_2: the product is made by machine B_2,

B_3: the product is made by machine B_3.

Applying the rule of elimination, we can write

$$P(A) = P(B_1)P(A|B_1) + P(B_2)P(A|B_2) + P(B_3)P(A|B_3).$$

Referring to the tree diagram of Figure 2.15, we find that the three branches give the probabilities

$$P(B_1)P(A|B_1) = (0.3)(0.02) = 0.006,$$
$$P(B_2)P(A|B_2) = (0.45)(0.03) = 0.0135,$$
$$P(B_3)P(A|B_3) = (0.25)(0.02) = 0.005,$$

and hence

$$P(A) = 0.006 + 0.0135 + 0.005 = 0.0245.$$

٤١

41

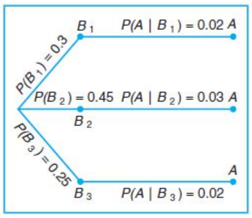


Figure 2.15: Tree diagram for Example 2.41.

٤٢

42

Theorem 2.14: **(Bayes' Rule)** If the events $B_1, B_2, \ldots, B_k$ constitute a partition of the sample space S such that $P(B_i) \neq 0$ for $i = 1, 2, \ldots, k$, then for any event A in S such that $P(A) \neq 0$,

$$P(B_r|A) = \frac{P(B_r \cap A)}{\sum_{i=1}^{k} P(B_i \cap A)} = \frac{P(B_r)P(A|B_r)}{\sum_{i=1}^{k} P(B_i)P(A|B_i)} \quad \text{for } r = 1, 2, \ldots, k.$$

***Proof*:** By the definition of conditional probability,

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

and then using Theorem 2.13 in the denominator, we have

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

which completes the proof.

Example 2.42: With reference to Example 2.41, if a product was chosen randomly and found to be defective, what is the probability that it was made by machine B_3?

***Solution*:** Using Bayes' rule to write

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

and then substituting the probabilities calculated in Example 2.41, we have

$$P(B_3|A) = \frac{0.005}{0.006 + 0.0135 + 0.005} = \frac{0.005}{0.0245} = \frac{10}{49}.$$

In view of the fact that a defective product was selected, this result suggests that it probably was not made by machine B_3.

٤٣

43

Example 2.43: A manufacturing firm employs three analytical plans for the design and development of a particular product. For cost reasons, all three are used at varying times. In fact, plans 1, 2, and 3 are used for 30%, 20%, and 50% of the products, respectively. The defect rate is different for the three procedures as follows:

$$P(D|P_1) = 0.01, \quad P(D|P_2) = 0.03, \quad P(D|P_3) = 0.02,$$

where $P(D|P_j)$ is the probability of a defective product, given plan j. If a random product was observed and found to be defective, which plan was most likely used and thus responsible?

***Solution*:** From the statement of the problem

$$P(P_1) = 0.30, \quad P(P_2) = 0.20, \quad \text{and} \quad P(P_3) = 0.50,$$

we must find $P(P_j|D)$ for $j = 1, 2, 3$. Bayes' rule (Theorem 2.14) shows

$$P(P_1|D) = \frac{P(P_1)P(D|P_1)}{P(P_1)P(D|P_1) + P(P_2)P(D|P_2) + P(P_3)P(D|P_3)}$$
$$= \frac{(0.30)(0.01)}{(0.3)(0.01) + (0.20)(0.03) + (0.50)(0.02)} = \frac{0.003}{0.019} = 0.158.$$

Similarly,

$$P(P_2|D) = \frac{(0.03)(0.20)}{0.019} = 0.316 \text{ and } P(P_3|D) = \frac{(0.02)(0.50)}{0.019} = 0.526.$$

The conditional probability of a defect given plan 3 is the largest of the three; thus a defective for a random product is most likely the result of the use of plan 3.

Using Bayes' rule, a statistical methodology called the Bayesian approach has attracted a lot of attention in applications. An introduction to the Bayesian method will be discussed in Chapter 18.

٤٤

44

Probability and Statistics

Chapter 3

Dr. Raed Al athamneh
Department of Industrial Engineering

١

1

Chapter 3

Random Variables and Probability Distributions

٢

2

Definition 3.1: A **random variable** is a function that associates a real number with each element in the sample space.

Example 3.1: Two balls are drawn in succession without replacement from an urn containing 4 red balls and 3 black balls. The possible outcomes and the values y of the random variable Y, where Y is the number of red balls, are

Sample Space	y
RR	2
RB	1
BR	1
BB	0

Example 3.2: A stockroom clerk returns three safety helmets at random to three steel mill employees who had previously checked them. If Smith, Jones, and Brown, in that order, receive one of the three hats, list the sample points for the possible orders of returning the helmets, and find the value m of the random variable M that represents the number of correct matches.

Solution: If S, J, and B stand for Smith's, Jones's, and Brown's helmets, respectively, then the possible arrangements in which the helmets may be returned and the number of correct matches are

Sample Space	m
SJB	3
SBJ	1
BJS	1
JSB	1
JBS	0
BSJ	0

3

Example 3.3: Consider the simple condition in which components are arriving from the production line and they are stipulated to be defective or not defective. Define the random variable X by

$$X = \begin{cases} 1, & \text{if the component is defective,} \\ 0, & \text{if the component is not defective.} \end{cases}$$

Example 3.4: Statisticians use **sampling plans** to either accept or reject batches or lots of material. Suppose one of these sampling plans involves sampling independently 10 items from a lot of 100 items in which 12 are defective.

Let X be the random variable defined as the number of items found defective in the sample of 10. In this case, the random variable takes on the values $0, 1, 2, \ldots, 9, 10$.

Example 3.5: Suppose a sampling plan involves sampling items from a process until a defective is observed. The evaluation of the process will depend on how many consecutive items are observed. In that regard, let X be a random variable defined by the number of items observed before a defective is found. With N a nondefective and D a defective, sample spaces are $S = \{D\}$ given $X = 1$, $S = \{ND\}$ given $X = 2$, $S = \{NND\}$ given $X = 3$, and so on.

4

Example 3.6: Interest centers around the proportion of people who respond to a certain mail order solicitation. Let X be that proportion. X is a random variable that takes on all values x for which $0 \leq x \leq 1$.

Definition 3.2: If a sample space contains a finite number of possibilities or an unending sequence with as many elements as there are whole numbers, it is called a **discrete sample space**.

Definition 3.3: If a sample space contains an infinite number of possibilities equal to the number of points on a line segment, it is called a **continuous sample space**.

٥

5

Discrete Probability Distributions

m	0	1	3
$P(M = m)$	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{1}{6}$

Definition 3.4: The set of ordered pairs $(x, f(x))$ is a **probability function**, **probability mass function**, or **probability distribution** of the discrete random variable X if, for each possible outcome x,

1. $f(x) \geq 0$,
2. $\sum_x f(x) = 1$,
3. $P(X = x) = f(x)$.

٦

6

Example 3.8: A shipment of 20 similar laptop computers to a retail outlet contains 3 that are defective. If a school makes a random purchase of 2 of these computers, find the probability distribution for the number of defectives.

Solution: Let X be a random variable whose values x are the possible numbers of defective computers purchased by the school. Then x can only take the numbers 0, 1, and

2. Now

$$f(0) = P(X = 0) = \frac{\binom{3}{0}\binom{17}{2}}{\binom{20}{2}} = \frac{68}{95}, \quad f(1) = P(X = 1) = \frac{\binom{3}{1}\binom{17}{1}}{\binom{20}{2}} = \frac{51}{190},$$

$$f(2) = P(X = 2) = \frac{\binom{3}{2}\binom{17}{0}}{\binom{20}{2}} = \frac{3}{190}.$$

Thus, the probability distribution of X is

x	0	1	2
$f(x)$	$\frac{68}{95}$	$\frac{51}{190}$	$\frac{3}{190}$

٧

7

Example 3.9: If a car agency sells 50% of its inventory of a certain foreign car equipped with side airbags, find a formula for the probability distribution of the number of cars with side airbags among the next 4 cars sold by the agency.

Solution: Since the probability of selling an automobile with side airbags is 0.5, the $2^4 = 16$ points in the sample space are equally likely to occur. Therefore, the denominator for all probabilities, and also for our function, is 16. To obtain the number of ways of selling 3 cars with side airbags, we need to consider the number of ways of partitioning 4 outcomes into two cells, with 3 cars with side airbags assigned to one cell and the model without side airbags assigned to the other. This can be done in $\binom{4}{3}$= 4 ways. In general, the event of selling x models with side airbags and $4 - x$ models without side airbags can occur in $\binom{4}{x}$ ways, where x can be 0, 1, 2, 3, or 4. Thus, the probability distribution $f(x) = P(X = x)$ is

$$f(x) = \frac{1}{16}\binom{4}{x}, \quad \text{for } x = 0, 1, 2, 3, 4.$$

There are many problems where we may wish to compute the probability that the observed value of a random variable X will be less than or equal to some real number x. Writing $F(x) = P(X \leq x)$ for every real number x, we define $F(x)$ to be the **cumulative distribution function** of the random variable X.

٨

8

Definition 3.5: The **cumulative distribution function** $F(x)$ of a discrete random variable X with probability distribution $f(x)$ is

$$F(x) = P(X \le x) = \sum_{t \le x} f(t), \quad \text{for } -\infty < x < \infty.$$

For the random variable M, the number of correct matches in Example 3.2, we have

$$F(2) = P(M \le 2) = f(0) + f(1) = \frac{1}{3} + \frac{1}{2} = \frac{5}{6}.$$

The cumulative distribution function of M is

$$F(m) = \begin{cases} 0, & \text{for } m < 0, \\ \frac{1}{3}, & \text{for } 0 \le m < 1, \\ \frac{5}{6}, & \text{for } 1 \le m < 3, \\ 1, & \text{for } m \ge 3. \end{cases}$$

٩

9

Example 3.10: Find the cumulative distribution function of the random variable X in Example 3.9. Using $F(x)$, verify that $f(2) = 3/8$.

Solution: Direct calculations of the probability distribution of Example 3.9 give $f(0) = 1/16$, $f(1) = 1/4$, $f(2) = 3/8$, $f(3) = 1/4$, and $f(4) = 1/16$. Therefore,

$$f(x) = \frac{1}{16}\binom{4}{x}, \quad \text{for } x = 0, 1, 2, 3, 4.$$

$$F(0) = f(0) = \frac{1}{16},$$
$$F(1) = f(0) + f(1) = \frac{5}{16},$$
$$F(2) = f(0) + f(1) + f(2) = \frac{11}{16},$$
$$F(3) = f(0) + f(1) + f(2) + f(3) = \frac{15}{16},$$
$$F(4) = f(0) + f(1) + f(2) + f(3) + f(4) = 1.$$

Hence,

$$F(x) = \begin{cases} 0, & \text{for } x < 0, \\ \frac{1}{16}, & \text{for } 0 \le x < 1, \\ \frac{5}{16}, & \text{for } 1 \le x < 2, \\ \frac{11}{16}, & \text{for } 2 \le x < 3, \\ \frac{15}{16}, & \text{for } 3 \le x < 4, \\ 1 & \text{for } x \ge 4. \end{cases}$$

Now

$$f(2) = F(2) - F(1) = \frac{11}{16} - \frac{5}{16} = \frac{3}{8}.$$

10

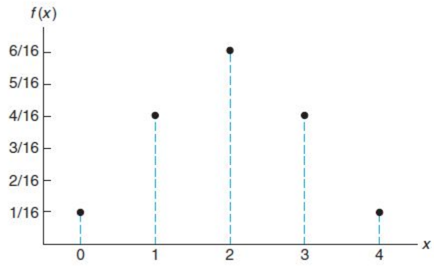


Figure 3.1: Probability mass function plot.

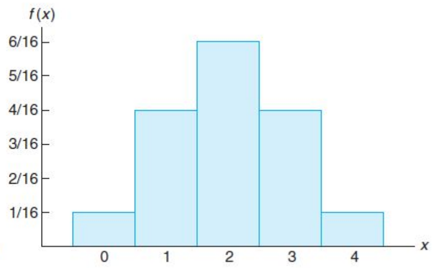


Figure 3.2: Probability histogram.

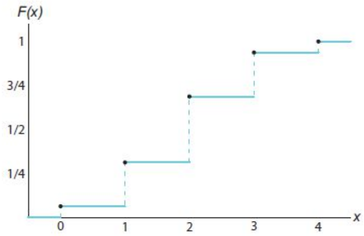


Figure 3.3: Discrete cumulative distribution function.

١١

11

Continuous Probability Distributions

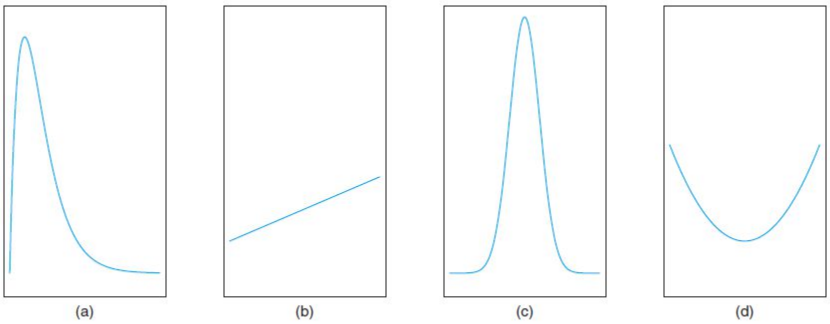
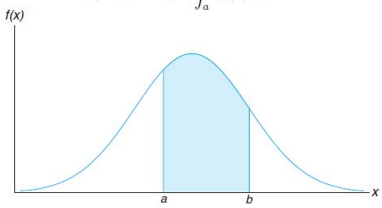


Figure 3.4: Typical density functions.

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



١٢

Figure 3.5: $P(a < X < b)$.

12

Definition 3.6: The function $f(x)$ is a **probability density function** (pdf) for the continuous random variable X, defined over the set of real numbers, if

1. $f(x) \geq 0$, for all $x \in R$.
2. $\int_{-\infty}^{\infty} f(x)\, dx = 1$.
3. $P(a < X < b) = \int_a^b f(x)\, dx$.

Example 3.11: Suppose that the error in the reaction temperature, in °C, for a controlled laboratory experiment is a continuous random variable X having the probability density function

$$f(x) = \begin{cases} \frac{x^2}{3}, & -1 < x < 2, \\ 0, & \text{elsewhere}. \end{cases}$$

.

(a) Verify that $f(x)$ is a density function.

(b) Find $P(0 < X \leq 1)$.

***Solution*:** We use Definition 3.6.

(a) Obviously, $f(x) \geq 0$. To verify condition 2 in Definition 3.6, we have

$$\int_{-\infty}^{\infty} f(x)\, dx = \int_{-1}^{2} \frac{x^2}{3} dx = \frac{x^3}{9}\Big|_{-1}^{2} = \frac{8}{9} + \frac{1}{9} = 1.$$

(b) Using formula 3 in Definition 3.6, we obtain

١٣

$$P(0 < X \leq 1) = \int_0^1 \frac{x^2}{3} dx = \frac{x^3}{9}\Big|_0^1 = \frac{1}{9}.$$

Definition 3.7: The **cumulative distribution function** $F(x)$ of a continuous random variable X with density function $f(x)$ is

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

As an immediate consequence of Definition 3.7, one can write the two results

$$P(a < X < b) = F(b) - F(a) \text{ and } f(x) = \frac{dF(x)}{dx},$$

if the derivative exists.

Example 3.12: For the density function of Example 3.11, find $F(x)$, and use it to evaluate $P(0 < X \leq 1)$.

***Solution*:** For $-1 < x < 2$,

$$F(x) = \int_{-\infty}^{x} f(t)\, dt = \int_{-1}^{x} \frac{t^2}{3} dt = \frac{t^3}{9}\Big|_{-1}^{x} = \frac{x^3 + 1}{9}.$$

Therefore,

$$F(x) = \begin{cases} 0, & x < -1, \\ \frac{x^3+1}{9}, & -1 \leq x < 2, \\ 1, & x \geq 2. \end{cases}$$

The cumulative distribution function $F(x)$ is expressed in Figure 3.6. Now

$$P(0 < X \leq 1) = F(1) - F(0) = \frac{2}{9} - \frac{1}{9} = \frac{1}{9},$$

which agrees with the result obtained by using the density function in Example 3.11.

Example 3.13: The Department of Energy (DOE) puts projects out on bid and generally estimates what a reasonable bid should be. Call the estimate b. The DOE has determined that the density function of the winning (low) bid is

$$f(y) = \begin{cases} \frac{5}{8b}, & \frac{2}{5}b \le y \le 2b, \\ 0, & \text{elsewhere.} \end{cases}$$

Find $F(y)$ and use it to determine the probability that the winning bid is less than the DOE's preliminary estimate b.

Solution: For $2b/5 \le y \le 2b$,

$$F(y) = \int_{2b/5}^{y} \frac{5}{8b} dy = \left. \frac{5t}{8b} \right|_{2b/5}^{y} = \frac{5y}{8b} - \frac{1}{4}.$$

١٥

15

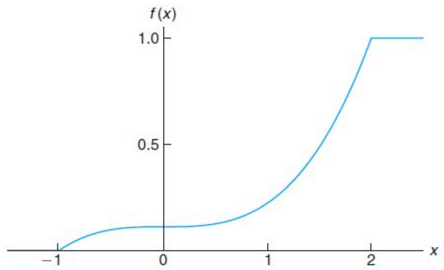


Figure 3.6: Continuous cumulative distribution function.

Thus,

$$F(y) = \begin{cases} 0, & y < \frac{2}{5}b, \\ \frac{5y}{8b} - \frac{1}{4}, & \frac{2}{5}b \le y < 2b, \\ 1, & y \ge 2b. \end{cases}$$

To determine the probability that the winning bid is less than the preliminary bid estimate b, we have

$$P(Y \le b) = F(b) = \frac{5}{8} - \frac{1}{4} = \frac{3}{8}.$$

١٦

16

Definition 3.8: The function $f(x, y)$ is a **joint probability distribution** or **probability mass function** of the discrete random variables X and Y if

1. $f(x, y) \geq 0$ 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 3.14: 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

(a) the joint probability function $f(x, y)$,

(b) $P[(X, Y) \in A]$, where A is the region $\{(x, y)|x + y \leq 1\}$.

Solution: The possible pairs of values (x, y) are $(0, 0)$, $(0, 1)$, $(1, 0)$, $(1, 1)$, $(0, 2)$, and $(2, 0)$.

(a) 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 3.1. Note that the probabilities sum to 1. In Chapter

١٧

5, it will become clear that the joint probability distribution of Table 3.1 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 $0 \leq x + y \leq 2$.

(b) The probability that (X, Y) fall in the region A is

$$P[(X, Y) \in A] = P(X + Y \leq 1) = f(0, 0) + f(0, 1) + f(1, 0)$$
$$= \frac{3}{28} + \frac{3}{14} + \frac{9}{28} = \frac{9}{14}.$$

Table 3.1: Joint Probability Distribution for Example 3.14

$f(x, y)$		$x=0$	$x=1$	$x=2$	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

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.

١٨

Definition 3.9: The function $f(x, y)$ is a **joint density function** of the continuous random variables X and Y if

1. $f(x, y) \geq 0$, for all (x, y),
2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\, dx\, dy = 1$,
3. $P[(X, Y) \in A] = \int\int_A f(x, y)\, dx\, dy$, for any region A in the xy plane.

Example 3.15: 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

$$f(x, y) = \begin{cases} \frac{2}{5}(2x + 3y), & 0 \leq x \leq 1, 0 \leq y \leq 1, \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Verify condition 2 of Definition 3.9.

(b) Find $P[(X, Y) \in A]$, where $A = \{(x, y) \mid 0 < x < \frac{1}{2}, \frac{1}{4} < y < \frac{1}{2}\}$.

١٩

Solution: (a) The integration of $f(x, y)$ over the whole region is

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\, dx\, dy = \int_0^1 \int_0^1 \frac{2}{5}(2x + 3y)\, dx\, dy$$

$$= \int_0^1 \left(\frac{2x^2}{5} + \frac{6xy}{5} \right) \Bigg|_{x=0}^{x=1} dy$$

$$= \int_0^1 \left(\frac{2}{5} + \frac{6y}{5} \right) dy = \left(\frac{2y}{5} + \frac{3y^2}{5} \right) \Bigg|_0^1 = \frac{2}{5} + \frac{3}{5} = 1.$$

(b) To calculate the probability, we use

$$P[(X, Y) \in A] = P\left(0 < X < \frac{1}{2}, \frac{1}{4} < Y < \frac{1}{2}\right)$$

$$= \int_{1/4}^{1/2} \int_0^{1/2} \frac{2}{5}(2x + 3y)\, dx\, dy$$

$$= \int_{1/4}^{1/2} \left(\frac{2x^2}{5} + \frac{6xy}{5} \right) \Bigg|_{x=0}^{x=1/2} dy = \int_{1/4}^{1/2} \left(\frac{1}{10} + \frac{3y}{5} \right) dy$$

$$= \left(\frac{y}{10} + \frac{3y^2}{10} \right) \Bigg|_{1/4}^{1/2}$$

$$= \frac{1}{10} \left[\left(\frac{1}{2} + \frac{3}{4} \right) - \left(\frac{1}{4} + \frac{3}{16} \right) \right] = \frac{13}{160}.$$

٢٠

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. We can now make the following general definition.

Definition 3.10: The **marginal distributions** of X alone and of Y alone are

$$g(x) = \sum_y f(x, y) \quad \text{and} \quad h(y) = \sum_x f(x, y)$$

for the discrete case, and

$$g(x) = \int_{-\infty}^{\infty} f(x, y)\, dy \quad \text{and} \quad h(y) = \int_{-\infty}^{\infty} f(x, y)\, dx$$

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.

٢١

21

Example 3.16: Show that the column and row totals of Table 3.1 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 3.1. 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
$h(y)$	$\frac{15}{28}$	$\frac{3}{7}$	$\frac{1}{28}$

٢٢

22

Example 3.17: Find $g(x)$ and $h(y)$ for the joint density function of Example 3.15.

Solution: By definition,

$$g(x) = \int_{-\infty}^{\infty} f(x,y)\,dy = \int_0^1 \frac{2}{5}(2x+3y)\,dy = \left(\frac{4xy}{5} + \frac{6y^2}{10}\right)\Big|_{y=0}^{y=1} = \frac{4x+3}{5},$$

for $0 \le x \le 1$, 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 $0 \le y \le 1$, and $h(y) = 0$ elsewhere.

The fact that the marginal distributions $g(x)$ and $h(y)$ are indeed the probability distributions of the individual variables X and Y alone can be verified by showing that the conditions of Definition 3.4 or Definition 3.6 are satisfied. For example, in the continuous case

$$\int_{-\infty}^{\infty} g(x)\,dx = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dy\,dx = 1,$$

and

$$P(a < X < b) = P(a < X < b, -\infty < Y < \infty)$$
$$= \int_a^b \int_{-\infty}^{\infty} f(x,y)\,dy\,dx = \int_a^b g(x)\,dx.$$

In Section 3.1, we stated that the value x of the random variable X represents an event that is a subset of the sample space. If we use the definition of conditional probability as stated in Chapter 2,

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

٢٣

23

where A and B are now the events defined by $X = x$ and $Y = y$, respectively, then

$$P(Y = y \mid X = x) = \frac{P(X = x, Y = y)}{P(X = x)} = \frac{f(x,y)}{g(x)}, \text{ provided } g(x) > 0,$$

where X and Y are discrete random variables.

It is not difficult to show that the function $f(x,y)/g(x)$, which is strictly a function of y with x fixed, satisfies all the conditions of a probability distribution. This is also true when $f(x,y)$ and $g(x)$ are the joint density and marginal distribution, respectively, of continuous random variables. As a result, it is extremely important that we make use of the special type of distribution of the form $f(x,y)/g(x)$ in order to be able to effectively compute conditional probabilities. This type of distribution is called a **conditional probability distribution**; the formal definition follows.

Definition 3.11: Let X and Y be two random variables, discrete or continuous. The **conditional distribution** of the random variable Y given that $X = x$ is

$$f(y|x) = \frac{f(x,y)}{g(x)}, \text{ provided } g(x) > 0.$$

Similarly, the conditional distribution of X given that $Y = y$ is

$$f(x|y) = \frac{f(x,y)}{h(y)}, \text{ provided } h(y) > 0.$$

٢٤

24

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

$$P(a < X < b \mid Y = y) = \sum_{a<x<b} f(x|y),$$

where the summation extends over all values of X between a and b. When X and Y are continuous, we evaluate

$$P(a < X < b \mid Y = y) = \int_a^b f(x|y)\, dx.$$

Example 3.18: Referring to Example 3.14, find the conditional distribution of X, given that $Y = 1$, and use it to determine $P(X = 0 \mid Y = 1)$.

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

$$f(x|1) = \frac{f(x,1)}{h(1)} = \left(\frac{7}{3}\right) f(x, 1), \quad x = 0, 1, 2.$$

Therefore,

$$f(0|1) = \left(\frac{7}{3}\right) f(0,1) = \left(\frac{7}{3}\right)\left(\frac{3}{14}\right) = \frac{1}{2}, \quad f(1|1) = \left(\frac{7}{3}\right) f(1,1) = \left(\frac{7}{3}\right)\left(\frac{3}{14}\right) = \frac{1}{2},$$

$$f(2|1) = \left(\frac{7}{3}\right) f(2,1) = \left(\frac{7}{3}\right)(0) = 0,$$

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,

$$P(X = 0 \mid Y = 1) = f(0|1) = \frac{1}{2}.$$

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.

٢٥

25

Example 3.19: 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

$$f(x, y) = \begin{cases} 10xy^2, & 0 < x < y < 1, \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Find the marginal densities $g(x)$, $h(y)$, and the conditional density $f(y|x)$.

(b) 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: (a) By definition,

$$g(x) = \int_{-\infty}^{\infty} f(x, y)\, dy = \int_x^1 10xy^2\, dy$$
$$= \frac{10}{3} xy^3 \Big|_{y=x}^{y=1} = \frac{10}{3} x(1 - x^3), \quad 0 < x < 1,$$
$$h(y) = \int_{-\infty}^{\infty} f(x, y)\, dx = \int_0^y 10xy^2\, dx = 5x^2y^2 \Big|_{x=0}^{x=y} = 5y^4, \quad 0 < y < 1.$$

Now

$$f(y|x) = \frac{f(x, y)}{g(x)} = \frac{10xy^2}{\frac{10}{3}x(1 - x^3)} = \frac{3y^2}{1 - x^3}, \quad 0 < x < y < 1.$$

(b) Therefore,

$$P\left(Y > \frac{1}{2} \,\middle|\, X = 0.25\right) = \int_{1/2}^1 f(y \mid x = 0.25)\, dy = \int_{1/2}^1 \frac{3y^2}{1 - 0.25^3}\, dy = \frac{8}{9}.$$

Example 3.20: Given the joint density function

$$f(x, y) = \begin{cases} \frac{x(1+3y^2)}{4}, & 0 < x < 2,\ 0 < y < 1, \\ 0, & \text{elsewhere,} \end{cases}$$

٢٦

26

Example 3.20: Given the joint density function

$$f(x,y) = \begin{cases} \frac{x(1+3y^2)}{4}, & 0 < x < 2,\ 0 < y < 1, \\ 0, & \text{elsewhere,} \end{cases}$$

find $g(x)$, $h(y)$, $f(x|y)$, and evaluate $P(\frac{1}{4} < X < \frac{1}{2} \mid Y = \frac{1}{3})$.

Solution: By definition of the marginal density. for $0 < x < 2$,

$$g(x) = \int_{-\infty}^{\infty} f(x,y)\,dy = \int_0^1 \frac{x(1+3y^2)}{4} dy$$
$$= \left(\frac{xy}{4} + \frac{xy^3}{4}\right)\Big|_{y=0}^{y=1} = \frac{x}{2},$$

and for $0 < y < 1$,

$$h(y) = \int_{-\infty}^{\infty} f(x,y)\,dx = \int_0^2 \frac{x(1+3y^2)}{4} dx$$
$$= \left(\frac{x^2}{8} + \frac{3x^2y^2}{8}\right)\Big|_{x=0}^{x=2} = \frac{1+3y^2}{2}.$$

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

$$P\left(\frac{1}{4} < X < \frac{1}{2} \,\Big|\, Y = \frac{1}{3}\right) = \int_{1/4}^{1/2} \frac{x}{2} dx = \frac{3}{64}.$$

٢٧

Statistical Independence

If $f(x|y)$ does not depend on y, as is the case for Example 3.20, then $f(x|y) = g(x)$ and $f(x,y) = g(x)h(y)$. The proof follows by substituting

$$f(x,y) = f(x|y)h(y)$$

into the marginal distribution of X. That is,

$$g(x) = \int_{-\infty}^{\infty} f(x,y)\,dy = \int_{-\infty}^{\infty} f(x|y)h(y)\,dy.$$

If $f(x|y)$ does not depend on y, we may write

$$g(x) = f(x|y)\int_{-\infty}^{\infty} h(y)\,dy.$$

Now

$$\int_{-\infty}^{\infty} h(y)\,dy = 1,$$

since $h(y)$ is the probability density function of Y. Therefore,

$$g(x) = f(x|y) \quad \text{and then} \quad f(x,y) = g(x)h(y).$$

٢٨

Definition 3.12: 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.

Example 3.21: Show that the random variables of Example 3.14 are not statistically independent.

Proof: Let us consider the point $(0, 1)$. From Table 3.1 we find the three probabilities $f(0, 1)$, $g(0)$, and $h(1)$ to be

$$f(0, 1) = \frac{3}{14},$$

$$g(0) = \sum_{y=0}^{2} f(0, y) = \frac{3}{28} + \frac{3}{14} + \frac{1}{28} = \frac{5}{14},$$

$$h(1) = \sum_{x=0}^{2} f(x, 1) = \frac{3}{14} + \frac{3}{14} + 0 = \frac{3}{7}.$$

Clearly,

$$f(0, 1) \neq g(0)h(1),$$

and therefore X and Y are not statistically independent.

All the preceding definitions concerning two random variables can be generalized to the case of n random variables. Let $f(x_1, x_2, \ldots, x_n)$ be the joint probability function of the random variables $X_1, X_2, \ldots, X_n$. The marginal distribution of X_1, for example, is

$$g(x_1) = \sum_{x_2} \cdots \sum_{x_n} f(x_1, x_2, \ldots, x_n)$$

٢٩

for the discrete case, and

$$g(x_1) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, x_2, \ldots, x_n)\, dx_2\, dx_3 \cdots dx_n$$

for the continuous case. We can now obtain **joint marginal distributions** such as $g(x_1, x_2)$, where

$$g(x_1, x_2) = \begin{cases} \sum\limits_{x_3} \cdots \sum\limits_{x_n} f(x_1, x_2, \ldots, x_n) & \text{(discrete case)}, \\ \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, x_2, \ldots, x_n)\, dx_3\, dx_4 \cdots dx_n & \text{(continuous case)}. \end{cases}$$

We could consider numerous conditional distributions. For example, the **joint conditional distribution** of X_1, X_2, and X_3, given that $X_4 = x_4, X_5 = x_5, \ldots, X_n = x_n$, is written

$$f(x_1, x_2, x_3 \mid x_4, x_5, \ldots, x_n) = \frac{f(x_1, x_2, \ldots, x_n)}{g(x_4, x_5, \ldots, x_n)},$$

where $g(x_4, x_5, \ldots, x_n)$ is the joint marginal distribution of the random variables $X_4, X_5, \ldots, X_n$.

A generalization of Definition 3.12 leads to the following definition for the mutual statistical independence of the variables $X_1, X_2, \ldots, X_n$.

٣٠

Definition 3.13: Let $X_1, X_2, \ldots, X_n$ be n random variables, discrete or continuous, with joint probability distribution $f(x_1, x_2, \ldots, x_n)$ and marginal distribution $f_1(x_1), f_2(x_2), \ldots, f_n(x_n)$, respectively. The random variables $X_1, X_2, \ldots, X_n$ are said to be mutually **statistically independent** if and only if

$$f(x_1, x_2, \ldots, x_n) = f_1(x_1)f_2(x_2)\cdots f_n(x_n)$$

for all $(x_1, x_2, \ldots, x_n)$ within their range.

Example 3.22: Suppose that the shelf life, in years, of a certain perishable food product packaged in cardboard containers is a random variable whose probability density function is given by

$$f(x) = \begin{cases} e^{-x}, & x > 0, \\ 0, & \text{elsewhere.} \end{cases}$$

Let X_1, X_2, and X_3 represent the shelf lives for three of these containers selected independently and find $P(X_1 < 2, 1 < X_2 < 3, X_3 > 2)$.

Solution: Since the containers were selected independently, we can assume that the random variables X_1, X_2, and X_3 are statistically independent, having the joint probability density

$$f(x_1, x_2, x_3) = f(x_1)f(x_2)f(x_3) = e^{-x_1}e^{-x_2}e^{-x_3} = e^{-x_1-x_2-x_3},$$

for $x_1 > 0$, $x_2 > 0$, $x_3 > 0$, and $f(x_1, x_2, x_3) = 0$ elsewhere. Hence

$$P(X_1 < 2, 1 < X_2 < 3, X_3 > 2) = \int_2^\infty \int_1^3 \int_0^2 e^{-x_1-x_2-x_3}\, dx_1\, dx_2\, dx_3$$
$$= (1 - e^{-2})(e^{-1} - e^{-3})e^{-2} = 0.0372.$$

٣١

31

Mathematical Expectation

Dr. Raed Al athamneh

4.1 Mean of a Random Variable

In Chapter 1, we discussed the sample mean, which is the arithmetic mean of the data. Now consider the following. If two coins are tossed 16 times and X is the number of heads that occur per toss, then the values of X are 0, 1, and 2. Suppose that the experiment yields no heads, one head, and two heads a total of 4, 7, and 5 times, respectively. The average number of heads per toss of the two coins is then

$$\frac{(0)(4) + (1)(7) + (2)(5)}{16} = 1.06.$$

This is an average value of the data and yet it is not a possible outcome of $\{0, 1, 2\}$. Hence, an average is not necessarily a possible outcome for the experiment. For instance, a salesman's average monthly income is not likely to be equal to any of his monthly paychecks.

Let us now restructure our computation for the average number of heads so as to have the following equivalent form:

$$(0)\left(\frac{4}{16}\right) + (1)\left(\frac{7}{16}\right) + (2)\left(\frac{5}{16}\right) = 1.06.$$

Assuming that 1 fair coin was tossed twice, we find that the sample space for our experiment is

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

Since the 4 sample points are all equally likely, it follows that

$$P(X = 0) = P(TT) = \frac{1}{4}, \quad P(X = 1) = P(TH) + P(HT) = \frac{1}{2},$$

and

$$P(X = 2) = P(HH) = \frac{1}{4},$$

where a typical element, say TH, indicates that the first toss resulted in a tail followed by a head on the second toss. Now, these probabilities are just the relative frequencies for the given events in the long run. Therefore,

$$\mu = E(X) = (0)\left(\frac{1}{4}\right) + (1)\left(\frac{1}{2}\right) + (2)\left(\frac{1}{4}\right) = 1.$$

3

Definition 4.1: Let X be a random variable with probability distribution $f(x)$. The **mean**, or **expected value**, of X is

$$\mu = E(X) = \sum_x x f(x)$$

if X is discrete, and

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

if X is continuous.

Example 4.1: A lot containing 7 components is sampled by a quality inspector; the lot contains 4 good components and 3 defective components. A sample of 3 is taken by the inspector. Find the expected value of the number of good components in this sample.

Solution: Let X represent the number of good components in the sample. The probability distribution of X is

$$f(x) = \frac{\binom{4}{x}\binom{3}{3-x}}{\binom{7}{3}}, \qquad x = 0, 1, 2, 3.$$

Simple calculations yield $f(0) = 1/35$, $f(1) = 12/35$, $f(2) = 18/35$, and $f(3) = 4/35$. Therefore,

$$\mu = E(X) = (0)\left(\frac{1}{35}\right) + (1)\left(\frac{12}{35}\right) + (2)\left(\frac{18}{35}\right) + (3)\left(\frac{4}{35}\right) = \frac{12}{7} = 1.7.$$

Thus, if a sample of size 3 is selected at random over and over again from a lot of 4 good components and 3 defective components, it will contain, on average, 1.7 good components.

4

Example 4.2: A salesperson for a medical device company has two appointments on a given day. At the first appointment, he believes that he has a 70% chance to make the deal, from which he can earn \$1000 commission if successful. On the other hand, he thinks he only has a 40% chance to make the deal at the second appointment, from which, if successful, he can make \$1500. What is his expected commission based on his own probability belief? Assume that the appointment results are independent of each other.

***Solution*:** First, we know that the salesperson, for the two appointments, can have 4 possible commission totals: \$0, \$1000, \$1500, and \$2500. We then need to calculate their associated probabilities. By independence, we obtain

$$f(\$0) = (1 - 0.7)(1 - 0.4) = 0.18, \quad f(\$2500) = (0.7)(0.4) = 0.28,$$
$$f(\$1000) = (0.7)(1 - 0.4) = 0.42, \text{ and } f(\$1500) = (1 - 0.7)(0.4) = 0.12.$$

Therefore, the expected commission for the salesperson is

$$\begin{aligned} E(X) &= (\$0)(0.18) + (\$1000)(0.42) + (\$1500)(0.12) + (\$2500)(0.28) \\ &= \$1300. \end{aligned}$$

Example 4.3: Let X be the random variable that denotes the life in hours of a certain electronic device. The probability density function is

$$f(x) = \begin{cases} \frac{20,000}{x^3}, & x > 100, \\ 0, & \text{elsewhere.} \end{cases}$$

Find the expected life of this type of device.

***Solution*:** Using Definition 4.1, we have

$$\mu = E(X) = \int_{100}^{\infty} x \frac{20,000}{x^3}\, dx = \int_{100}^{\infty} \frac{20,000}{x^2}\, dx = 200.$$

Therefore, we can expect this type of device to last, *on average*, 200 hours.

Now let us consider a new random variable $g(X)$, which depends on X; that is, each value of $g(X)$ is determined by the value of X. For instance, $g(X)$ might be X^2 or $3X - 1$, and whenever X assumes the value 2, $g(X)$ assumes the value $g(2)$. In particular, if X is a discrete random variable with probability distribution $f(x)$, for $x = -1, 0, 1, 2$, and $g(X) = X^2$, then

$$\begin{aligned} P[g(X) = 0] &= P(X = 0) = f(0), \\ P[g(X) = 1] &= P(X = -1) + P(X = 1) = f(-1) + f(1), \\ P[g(X) = 4] &= P(X = 2) = f(2), \end{aligned}$$

and so the probability distribution of $g(X)$ may be written

$g(x)$	0	1	4
$P[g(X) = g(x)]$	$f(0)$	$f(-1) + f(1)$	$f(2)$

By the definition of the expected value of a random variable, we obtain

$$\begin{aligned} \mu_{g(X)} &= E[g(x)] = 0f(0) + 1[f(-1) + f(1)] + 4f(2) \\ &= (-1)^2 f(-1) + (0)^2 f(0) + (1)^2 f(1) + (2)^2 f(2) = \sum_x g(x)f(x). \end{aligned}$$

Theorem 4.1: Let X be a random variable with probability distribution $f(x)$. The expected value of the random variable $g(X)$ is

$$\mu_{g(X)} = E[g(X)] = \sum_x g(x)f(x)$$

if X is discrete, and

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

if X is continuous.

Example 4.4: Suppose that the number of cars X that pass through a car wash between 4:00 P.M. and 5:00 P.M. on any sunny Friday has the following probability distribution:

x	4	5	6	7	8	9
$P(X = x)$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{6}$	$\frac{1}{6}$

Let $g(X) = 2X - 1$ represent the amount of money, in dollars, paid to the attendant by the manager. Find the attendant's expected earnings for this particular time period.

Solution: By Theorem 4.1, the attendant can expect to receive

$$\begin{aligned} E[g(X)] = E(2X - 1) &= \sum_{x=4}^{9} (2x - 1)f(x) \\ &= (7)\left(\frac{1}{12}\right) + (9)\left(\frac{1}{12}\right) + (11)\left(\frac{1}{4}\right) + (13)\left(\frac{1}{4}\right) \\ &+ (15)\left(\frac{1}{6}\right) + (17)\left(\frac{1}{6}\right) = \$12.67. \end{aligned}$$

Example 4.5: Let X be a random variable with density function

$$f(x) = \begin{cases} \frac{x^2}{3}, & -1 < x < 2, \\ 0, & \text{elsewhere.} \end{cases}$$

Find the expected value of $g(X) = 4X + 3$.

Solution: By Theorem 4.1, we have

$$E(4X + 3) = \int_{-1}^{2} \frac{(4x + 3)x^2}{3}\ dx = \frac{1}{3}\int_{-1}^{2} (4x^3 + 3x^2)\ dx = 8.$$

We shall now extend our concept of mathematical expectation to the case of two random variables X and Y with joint probability distribution $f(x, y)$.

Definition 4.2: Let X and Y be random variables with joint probability distribution $f(x, y)$. The mean, or expected value, of the random variable $g(X, Y)$ is

$$\mu_{g(X,Y)} = E[g(X, Y)] = \sum_x \sum_y g(x, y)f(x, y)$$

if X and Y are discrete, and

$$\mu_{g(X,Y)} = E[g(X, Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x, y)f(x, y)\ dx\ dy$$

if X and Y are continuous.

Example 4.6: Let X and Y be the random variables with joint probability distribution indicated in Table 3.1 on page 96. Find the expected value of $g(X, Y) = XY$. The table is reprinted here for convenience.

	$f(x, y)$	$x=0$	$x=1$	$x=2$	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

***Solution*:** By Definition 4.2, we write

$$\begin{aligned} E(XY) &= \sum_{x=0}^{2}\sum_{y=0}^{2} xyf(x, y) \\ &= (0)(0)f(0, 0) + (0)(1)f(0, 1) \\ &\quad + (1)(0)f(1, 0) + (1)(1)f(1, 1) + (2)(0)f(2, 0) \\ &= f(1, 1) = \frac{3}{14}. \end{aligned}$$

Example 4.7: Find $E(Y/X)$ for the density function

$$f(x, y) = \begin{cases} \frac{x(1+3y^2)}{4}, & 0 < x < 2,\ 0 < y < 1, \\ 0, & \text{elsewhere.} \end{cases}$$

***Solution*:** We have

$$E\left(\frac{Y}{X}\right) = \int_0^1 \int_0^2 \frac{y(1+3y^2)}{4}\, dx dy = \int_0^1 \frac{y + 3y^3}{2}\, dy = \frac{5}{8}.$$

Note that if $g(X, Y) = X$ in Definition 4.2, we have

$$E(X) = \begin{cases} \sum_x \sum_y xf(x, y) = \sum_x xg(x) & \text{(discrete case)}, \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xf(x, y)\, dy\, dx = \int_{-\infty}^{\infty} xg(x)\, dx & \text{(continuous case)}, \end{cases}$$

where $g(x)$ is the marginal distribution of X. Therefore, in calculating $E(X)$ over a two-dimensional space, one may use either the joint probability distribution of X and Y or the marginal distribution of X. Similarly, we define

$$E(Y) = \begin{cases} \sum_y \sum_x yf(x, y) = \sum_y yh(y) & \text{(discrete case)}, \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} yf(x, y)\, dxdy = \int_{-\infty}^{\infty} yh(y)\, dy & \text{(continuous case)}, \end{cases}$$

where $h(y)$ is the marginal distribution of the random variable Y.

4.2 Variance and Covariance of Random Variables

The mean, or expected value, of a random variable X is of special importance in statistics because it describes where the probability distribution is centered. By itself, however, the mean does not give an adequate description of the shape of the distribution. We also need to characterize the variability in the distribution. In Figure 4.1, we have the histograms of two discrete probability distributions that have the same mean, $\mu = 2$, but differ considerably in variability, or the dispersion of their observations about the mean.

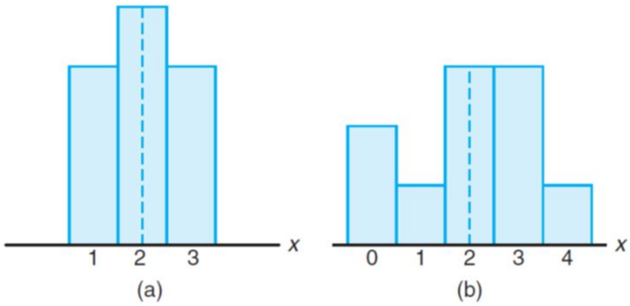


Figure 4.1: Distributions with equal means and unequal dispersions.

Definition 4.3: Let X be a random variable with probability distribution $f(x)$ and mean μ. The variance of X is

$$\sigma^2 = E[(X-\mu)^2] = \sum_x (x-\mu)^2 f(x), \qquad \text{if } X \text{ is discrete, and}$$

$$\sigma^2 = E[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx, \qquad \text{if } X \text{ is continuous.}$$

The positive square root of the variance, σ, is called the **standard deviation** of X.

Example 4.8: Let the random variable X represent the number of automobiles that are used for official business purposes on any given workday. The probability distribution for company A [Figure 4.1(a)] is

x	1	2	3
$f(x)$	0.3	0.4	0.3

and that for company B [Figure 4.1(b)] is

x	0	1	2	3	4
$f(x)$	0.2	0.1	0.3	0.3	0.1

Show that the variance of the probability distribution for company B is greater than that for company A.

Solution: For company A, we find that

$$\mu_A = E(X) = (1)(0.3) + (2)(0.4) + (3)(0.3) = 2.0,$$

and then

$$\sigma_A^2 = \sum_{x=1}^{3} (x-2)^2 = (1-2)^2(0.3) + (2-2)^2(0.4) + (3-2)^2(0.3) = 0.6.$$

For company B, we have

$$\mu_B = E(X) = (0)(0.2) + (1)(0.1) + (2)(0.3) + (3)(0.3) + (4)(0.1) = 2.0,$$

and then

$$\sigma_B^2 = \sum_{x=0}^{4} (x-2)^2 f(x)$$
$$= (0-2)^2(0.2) + (1-2)^2(0.1) + (2-2)^2(0.3)$$
$$+ (3-2)^2(0.3) + (4-2)^2(0.1) = 1.6.$$

calculations, is stated in the following theorem.

Theorem 4.2: The variance of a random variable X is

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

***Proof*:** For the discrete case, we can write

$$\sigma^2 = \sum_x (x-\mu)^2 f(x) = \sum_x (x^2 - 2\mu x + \mu^2) f(x)$$
$$= \sum_x x^2 f(x) - 2\mu \sum_x x f(x) + \mu^2 \sum_x f(x).$$

Since $\mu = \sum_x x f(x)$ by definition, and $\sum_x f(x) = 1$ for any discrete probability distribution, it follows that

$$\sigma^2 = \sum_x x^2 f(x) - \mu^2 = E(X^2) - \mu^2.$$

For the continuous case the proof is step by step the same, with summations replaced by integrations.

Example 4.9: Let the random variable X represent the number of defective parts for a machine when 3 parts are sampled from a production line and tested. The following is the probability distribution of X.

x	0	1	2	3
$f(x)$	0.51	0.38	0.10	0.01

Using Theorem 4.2, calculate σ^2.

***Solution*:** First, we compute

$$\mu = (0)(0.51) + (1)(0.38) + (2)(0.10) + (3)(0.01) = 0.61.$$

Now,

$$E(X^2) = (0)(0.51) + (1)(0.38) + (4)(0.10) + (9)(0.01) = 0.87.$$

Therefore,

$$\sigma^2 = 0.87 - (0.61)^2 = 0.4979.$$

Example 4.10: The weekly demand for a drinking-water product, in thousands of liters, from a local chain of efficiency stores is a continuous random variable X having the probability density

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

Find the mean and variance of X.

$$\mu = E(X) = 2\int_1^2 x(x-1)\,dx = \frac{5}{3}$$

and

$$E(X^2) = 2\int_1^2 x^2(x-1)\,dx = \frac{17}{6}.$$

Therefore,

$$\sigma^2 = \frac{17}{6} - \left(\frac{5}{3}\right)^2 = \frac{1}{18}.$$

Theorem 4.3: Let X be a random variable with probability distribution $f(x)$. The variance of the random variable $g(X)$ is

$$\sigma^2_{g(X)} = E\{[g(X) - \mu_{g(X)}]^2\} = \sum_x [g(x) - \mu_{g(X)}]^2 f(x)$$

if X is discrete, and

$$\sigma^2_{g(X)} = E\{[g(X) - \mu_{g(X)}]^2\} = \int_{-\infty}^{\infty} [g(x) - \mu_{g(X)}]^2 f(x)\, dx$$

if X is continuous.

Proof: Since $g(X)$ is itself a random variable with mean $\mu_{g(X)}$ as defined in Theorem 4.1, it follows from Definition 4.3 that

$$\sigma^2_{g(X)} = E\{[g(X) - \mu_{g(X)}]\}.$$

Now, applying Theorem 4.1 again to the random variable $[g(X) - \mu_{g(X)}]^2$ completes the proof.

Example 4.11: Calculate the variance of $g(X) = 2X + 3$, where X is a random variable with probability distribution

x	0	1	2	3
$f(x)$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{2}$	$\frac{1}{8}$

$$\mu_{2X+3} = E(2X + 3) = \sum_{x=0}^{3} (2x + 3) f(x) = 6.$$

Now, using Theorem 4.3, we have

$$\sigma^2_{2X+3} = E\{[(2X + 3) - \mu_{2x+3}]^2\} = E[(2X + 3 - 6)^2]$$
$$= E(4X^2 - 12X + 9) = \sum_{x=0}^{3} (4x^2 - 12x + 9) f(x) = 4.$$

Example 4.12: Let X be a random variable having the density function given in Example 4.5 on page 115. Find the variance of the random variable $g(X) = 4X + 3$.

***Solution*:** In Example 4.5, we found that $\mu_{4X+3} = 8$. Now, using Theorem 4.3,

$$\sigma^2_{4X+3} = E\{[(4X+3) - 8]^2\} = E[(4X-5)^2]$$
$$= \int_{-1}^{2} (4x-5)^2 \frac{x^2}{3}\, dx = \frac{1}{3}\int_{-1}^{2} (16x^4 - 40x^3 + 25x^2)\, dx = \frac{51}{5}.$$

If $g(X,Y) = (X-\mu_X)(Y-\mu_Y)$, where $\mu_X = E(X)$ and $\mu_Y = E(Y)$, Definition 4.2 yields an expected value called the **covariance** of X and Y, which we denote by σ_{XY} or $\text{Cov}(X,Y)$.

Definition 4.4: Let X and Y be random variables with joint probability distribution $f(x,y)$. The covariance of X and Y is

$$\sigma_{XY} = E[(X-\mu_X)(Y-\mu_Y)] = \sum_x \sum_y (x-\mu_X)(y-\mu_y) f(x,y)$$

if X and Y are discrete, and

$$\sigma_{XY} = E[(X-\mu_X)(Y-\mu_Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\mu_X)(y-\mu_y) f(x,y)\, dx\, dy$$

if X and Y are continuous.

Theorem 4.4: The covariance of two random variables X and Y with means μ_X and μ_Y, respectively, is given by

$$\sigma_{XY} = E(XY) - \mu_X\mu_Y.$$

***Proof*:** For the discrete case, we can write

$$\sigma_{XY} = \sum_x \sum_y (x-\mu_X)(y-\mu_Y) f(x,y)$$
$$= \sum_x \sum_y xy f(x,y) - \mu_X \sum_x \sum_y y f(x,y)$$
$$- \mu_Y \sum_x \sum_y x f(x,y) + \mu_X\mu_Y \sum_x \sum_y f(x,y).$$

Since

$$\mu_X = \sum_x x f(x,y), \quad \mu_Y = \sum_y y f(x,y), \text{ and } \sum_x \sum_y f(x,y) = 1$$

for any joint discrete distribution, it follows that

$$\sigma_{XY} = E(XY) - \mu_X\mu_Y - \mu_Y\mu_X + \mu_X\mu_Y = E(XY) - \mu_X\mu_Y.$$

For the continuous case, the proof is identical with summations replaced by integrals.

Example 4.13: Example 3.14 on page 95 describes a situation involving the number of blue refills X and the number of red refills Y. Two refills for a ballpoint pen are selected at random from a certain box, and the following is the joint probability distribution:

	$f(x,y)$	$x=0$	$x=1$	$x=2$	$h(y)$
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	$\frac{1}{28}$
	$g(x)$	$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$	1

Find the covariance of X and Y.

Solution: From Example 4.6, we see that $E(XY) = 3/14$. Now

$$\mu_X = \sum_{x=0}^{2} xg(x) = (0)\left(\frac{5}{14}\right) + (1)\left(\frac{15}{28}\right) + (2)\left(\frac{3}{28}\right) = \frac{3}{4},$$

and

$$\mu_Y = \sum_{y=0}^{2} yh(y) = (0)\left(\frac{15}{28}\right) + (1)\left(\frac{3}{7}\right) + (2)\left(\frac{1}{28}\right) = \frac{1}{2}.$$

Therefore,

$$\sigma_{XY} = E(XY) - \mu_X\mu_Y = \frac{3}{14} - \left(\frac{3}{4}\right)\left(\frac{1}{2}\right) = -\frac{9}{56}.$$

Example 4.14: The fraction X of male runners and the fraction Y of female runners who compete in marathon races are described by the joint density function

$$f(x,y) = \begin{cases} 8xy, & 0 \le y \le x \le 1, \\ 0, & \text{elsewhere.} \end{cases}$$

Find the covariance of X and Y.

Solution: We first compute the marginal density functions. They are

$$g(x) = \begin{cases} 4x^3, & 0 \le x \le 1, \\ 0, & \text{elsewhere,} \end{cases}$$

and

$$h(y) = \begin{cases} 4y(1-y^2), & 0 \le y \le 1, \\ 0, & \text{elsewhere.} \end{cases}$$

From these marginal density functions, we compute

$$\mu_X = E(X) = \int_0^1 4x^4\, dx = \frac{4}{5} \text{ and } \mu_Y = \int_0^1 4y^2(1-y^2)\, dy = \frac{8}{15}.$$

From the joint density function given above, we have

$$E(XY) = \int_0^1 \int_y^1 8x^2y^2\, dx\, dy = \frac{4}{9}.$$

Then

$$\sigma_{XY} = E(XY) - \mu_X\mu_Y = \frac{4}{9} - \left(\frac{4}{5}\right)\left(\frac{8}{15}\right) = \frac{4}{225}.$$

Definition 4.5: Let X and Y be random variables with covariance σ_{XY} and standard deviations σ_X and σ_Y, respectively. The correlation coefficient of X and Y is

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}.$$

Example 4.15: Find the correlation coefficient between X and Y in Example 4.13.

***Solution*:** Since

$$E(X^2) = (0^2)\left(\frac{5}{14}\right) + (1^2)\left(\frac{15}{28}\right) + (2^2)\left(\frac{3}{28}\right) = \frac{27}{28}$$

and

$$E(Y^2) = (0^2)\left(\frac{15}{28}\right) + (1^2)\left(\frac{3}{7}\right) + (2^2)\left(\frac{1}{28}\right) = \frac{4}{7},$$

we obtain

$$\sigma_X^2 = \frac{27}{28} - \left(\frac{3}{4}\right)^2 = \frac{45}{112} \text{ and } \sigma_Y^2 = \frac{4}{7} - \left(\frac{1}{2}\right)^2 = \frac{9}{28}.$$

Therefore, the correlation coefficient between X and Y is

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{-9/56}{\sqrt{(45/112)(9/28)}} = -\frac{1}{\sqrt{5}}.$$

Example 4.16: Find the correlation coefficient of X and Y in Example 4.14.

***Solution*:** Because

$$E(X^2) = \int_0^1 4x^5\, dx = \frac{2}{3} \text{ and } E(Y^2) = \int_0^1 4y^3(1-y^2)\, dy = 1 - \frac{2}{3} = \frac{1}{3},$$

we conclude that

$$\sigma_X^2 = \frac{2}{3} - \left(\frac{4}{5}\right)^2 = \frac{2}{75} \text{ and } \sigma_Y^2 = \frac{1}{3} - \left(\frac{8}{15}\right)^2 = \frac{11}{225}.$$

Hence,

$$\rho_{XY} = \frac{4/225}{\sqrt{(2/75)(11/225)}} = \frac{4}{\sqrt{66}}.$$

4.3 Means and Variances of Linear Combinations of Random Variables

Theorem 4.5: If a and b are constants, then

$$E(aX + b) = aE(X) + b.$$

***Proof*:** By the definition of expected value,

$$E(aX + b) = \int_{-\infty}^{\infty} (ax + b)f(x)\,dx = a\int_{-\infty}^{\infty} xf(x)\,dx + b\int_{-\infty}^{\infty} f(x)\,dx.$$

The first integral on the right is $E(X)$ and the second integral equals 1. Therefore, we have

$$E(aX + b) = aE(X) + b.$$

Corollary 4.1: Setting $a = 0$, we see that $E(b) = b$.

Corollary 4.2: Setting $b = 0$, we see that $E(aX) = aE(X)$.

Example 4.17: Applying Theorem 4.5 to the discrete random variable $f(X) = 2X - 1$, rework Example 4.4 on page 115.

***Solution*:** According to Theorem 4.5, we can write

$$E(2X - 1) = 2E(X) - 1.$$

Now

$$\mu = E(X) = \sum_{x=4}^{9} xf(x)$$
$$= (4)\left(\frac{1}{12}\right) + (5)\left(\frac{1}{12}\right) + (6)\left(\frac{1}{4}\right) + (7)\left(\frac{1}{4}\right) + (8)\left(\frac{1}{6}\right) + (9)\left(\frac{1}{6}\right) = \frac{41}{6}.$$

Therefore,

$$\mu_{2X-1} = (2)\left(\frac{41}{6}\right) - 1 = \$12.67,$$

as before.

Example 4.18: Applying Theorem 4.5 to the continuous random variable $g(X) = 4X + 3$, rework Example 4.5 on page 115.

***Solution*:** For Example 4.5, we may use Theorem 4.5 to write

$$E(4X + 3) = 4E(X) + 3.$$

Now

$$E(X) = \int_{-1}^{2} x\left(\frac{x^2}{3}\right) dx = \int_{-1}^{2} \frac{x^3}{3}\, dx = \frac{5}{4}.$$

Therefore,

$$E(4X + 3) = (4)\left(\frac{5}{4}\right) + 3 = 8,$$

as before.

Theorem 4.6: The expected value of the sum or difference of two or more functions of a random variable X is the sum or difference of the expected values of the functions. That is,

$$E[g(X) \pm h(X)] = E[g(X)] \pm E[h(X)].$$

***Proof*:** By definition,

$$\begin{aligned} E[g(X) \pm h(X)] &= \int_{-\infty}^{\infty} [g(x) \pm h(x)] f(x)\, dx \\ &= \int_{-\infty}^{\infty} g(x) f(x)\, dx \pm \int_{-\infty}^{\infty} h(x) f(x)\, dx \\ &= E[g(X)] \pm E[h(X)]. \end{aligned}$$

Example 4.19: Let X be a random variable with probability distribution as follows:

x	0	1	2	3
$f(x)$	$\frac{1}{3}$	$\frac{1}{2}$	0	$\frac{1}{6}$

Find the expected value of $Y = (X - 1)^2$.

***Solution*:** Applying Theorem 4.6 to the function $Y = (X - 1)^2$, we can write

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

From Corollary 4.1, $E(1) = 1$, and by direct computation,

$$E(X) = (0)\left(\frac{1}{3}\right) + (1)\left(\frac{1}{2}\right) + (2)(0) + (3)\left(\frac{1}{6}\right) = 1 \text{ and}$$

$$E(X^2) = (0)\left(\frac{1}{3}\right) + (1)\left(\frac{1}{2}\right) + (4)(0) + (9)\left(\frac{1}{6}\right) = 2.$$

Hence,

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

Example 4.20: The weekly demand for a certain drink, in thousands of liters, at a chain of convenience stores is a continuous random variable $g(X) = X^2 + X - 2$, where X has the density function

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

Find the expected value of the weekly demand for the drink.

***Solution*:** By Theorem 4.6, we write

$$E(X^2 + X - 2) = E(X^2) + E(X) - E(2).$$

From Corollary 4.1, $E(2) = 2$, and by direct integration,

$$E(X) = \int_1^2 2x(x-1)\,dx = \frac{5}{3} \text{ and } E(X^2) = \int_1^2 2x^2(x-1)\,dx = \frac{17}{6}.$$

Now

$$E(X^2 + X - 2) = \frac{17}{6} + \frac{5}{3} - 2 = \frac{5}{2},$$

Theorem 4.7: The expected value of the sum or difference of two or more functions of the random variables X and Y is the sum or difference of the expected values of the functions. That is,

$$E[g(X,Y) \pm h(X,Y)] = E[g(X,Y)] \pm E[h(X,Y)].$$

***Proof*:** By Definition 4.2,

$$\begin{aligned} E[g(X,Y) \pm h(X,Y)] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [g(x,y) \pm h(x,y)] f(x,y)\,dx\,dy \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f(x,y)\,dx\,dy \pm \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x,y)f(x,y)\,dx\,dy \\ &= E[g(X,Y)] \pm E[h(X,Y)]. \end{aligned}$$

Corollary 4.3: Setting $g(X,Y) = g(X)$ and $h(X,Y) = h(Y)$, we see that

$$E[g(X) \pm h(Y)] = E[g(X)] \pm E[h(Y)].$$

Corollary 4.4: Setting $g(X,Y) = X$ and $h(X,Y) = Y$, we see that

$$E[X \pm Y] = E[X] \pm E[Y].$$

If X represents the daily production of some item from machine A and Y the daily production of the same kind of item from machine B, then $X + Y$ represents the total number of items produced daily by both machines. Corollary 4.4 states that the average daily production for both machines is equal to the sum of the average daily production of each machine.

Theorem 4.8: Let X and Y be two independent random variables. Then

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

***Proof*:** By Definition 4.2,

$$E(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x,y)\ dx\ dy.$$

Since X and Y are independent, we may write

$$f(x,y) = g(x)h(y),$$

where $g(x)$ and $h(y)$ are the marginal distributions of X and Y, respectively. Hence,

$$E(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyg(x)h(y)\ dx\ dy = \int_{-\infty}^{\infty} xg(x)\ dx \int_{-\infty}^{\infty} yh(y)\ dy$$
$$= E(X)E(Y).$$

Corollary 4.5: Let X and Y be two independent random variables. Then $\sigma_{XY} = 0$.

***Proof*:** The proof can be carried out by using Theorems 4.4 and 4.8.

Example 4.21: It is known that the ratio of gallium to arsenide does not affect the functioning of gallium-arsenide wafers, which are the main components of microchips. Let X denote the ratio of gallium to arsenide and Y denote the functional wafers retrieved during a 1-hour period. X and Y are independent random variables with the joint density function

$$f(x,y) = \begin{cases} \frac{x(1+3y^2)}{4}, & 0 < x < 2,\ 0 < y < 1, \\ 0, & \text{elsewhere.} \end{cases}$$

Show that $E(XY) = E(X)E(Y)$, as Theorem 4.8 suggests.

***Solution*:** By definition,

$$E(XY) = \int_0^1\int_0^2 \frac{x^2y(1+3y^2)}{4}\ dxdy = \frac{5}{6},\ E(X) = \frac{4}{3},\ \text{and } E(Y) = \frac{5}{8}.$$

Hence,

$$E(X)E(Y) = \left(\frac{4}{3}\right)\left(\frac{5}{8}\right) = \frac{5}{6} = E(XY).$$

We conclude this section by proving one theorem and presenting several corollaries that are useful for calculating variances or standard deviations.

Theorem 4.9: If X and Y are random variables with joint probability distribution $f(x, y)$ and a, b, and c are constants, then

$$\sigma^2_{aX+bY+c} = a^2\sigma^2_X + b^2\sigma^2_Y + 2ab\sigma_{XY}.$$

Proof: By definition, $\sigma^2_{aX+bY+c} = E\{[(aX + bY + c) - \mu_{aX+bY+c}]^2\}$. Now

$$\mu_{aX+bY+c} = E(aX + bY + c) = aE(X) + bE(Y) + c = a\mu_X + b\mu_Y + c,$$

by using Corollary 4.4 followed by Corollary 4.2. Therefore,

$$\begin{aligned}\sigma^2_{aX+bY+c} &= E\{[a(X - \mu_X) + b(Y - \mu_Y)]^2\}\\ &= a^2E[(X - \mu_X)^2] + b^2E[(Y - \mu_Y)^2] + 2abE[(X - \mu_X)(Y - \mu_Y)]\\ &= a^2\sigma^2_X + b^2\sigma^2_Y + 2ab\sigma_{XY}.\end{aligned}$$

Using Theorem 4.9, we have the following corollaries.

Corollary 4.6: Setting $b = 0$, we see that

$$\sigma^2_{aX+c} = a^2\sigma^2_X = a^2\sigma^2.$$

Corollary 4.7: Setting $a = 1$ and $b = 0$, we see that

$$\sigma^2_{X+c} = \sigma^2_X = \sigma^2.$$

Corollary 4.8: Setting $b = 0$ and $c = 0$, we see that

$$\sigma^2_{aX} = a^2\sigma^2_X = a^2\sigma^2.$$

Corollaries 4.6 and 4.7 state that the variance is unchanged if a constant is added to or subtracted from a random variable. The addition or subtraction of a constant simply shifts the values of X to the right or to the left but does not change their variability. However, if a random variable is multiplied or divided by a constant, then Corollaries 4.6 and 4.8 state that the variance is multiplied or divided by the square of the constant.

Corollary 4.9: If X and Y are independent random variables, then

$$\sigma^2_{aX+bY} = a^2\sigma^2_X + b^2\sigma^2_Y.$$

The result stated in Corollary 4.9 is obtained from Theorem 4.9 by invoking Corollary 4.5.

Corollary 4.10: If X and Y are independent random variables, then

$$\sigma^2_{aX-bY} = a^2\sigma^2_X + b^2\sigma^2_Y.$$

Corollary 4.10 follows when b in Corollary 4.9 is replaced by $-b$. Generalizing to a linear combination of n independent random variables, we have Corollary 4.11.

Corollary 4.11: If $X_1, X_2, \ldots, X_n$ are independent random variables, then

$$\sigma^2_{a_1X_1+a_2X_2+\cdots+a_nX_n} = a_1^2\sigma^2_{X_1} + a_2^2\sigma^2_{X_2} + \cdots + a_n^2\sigma^2_{X_n}.$$

Example 4.22: If X and Y are random variables with variances $\sigma^2_X = 2$ and $\sigma^2_Y = 4$ and covariance $\sigma_{XY} = -2$, find the variance of the random variable $Z = 3X - 4Y + 8$.

Solution:

$$\begin{aligned}\sigma^2_Z &= \sigma^2_{3X-4Y+8} = \sigma^2_{3X-4Y} && \text{(by Corollary 4.6)}\\ &= 9\sigma^2_X + 16\sigma^2_Y - 24\sigma_{XY} && \text{(by Theorem 4.9)}\\ &= (9)(2) + (16)(4) - (24)(-2) = 130.\end{aligned}$$

Example 4.23: Let X and Y denote the amounts of two different types of impurities in a batch of a certain chemical product. Suppose that X and Y are independent random variables with variances $\sigma^2_X = 2$ and $\sigma^2_Y = 3$. Find the variance of the random variable $Z = 3X - 2Y + 5$.

Solution:

$$\begin{aligned}\sigma^2_Z &= \sigma^2_{3X-2Y+5} = \sigma^2_{3X-2Y} && \text{(by Corollary 4.6)}\\ &= 9\sigma^2_x + 4\sigma^2_y && \text{(by Corollary 4.10)}\\ &= (9)(2) + (4)(3) = 30.\end{aligned}$$

What If the Function Is Nonlinear?

pproximation of $E[g(X)]$

$$E[g(X)] \approx g(\mu_X) + \left.\frac{\partial^2 g(x)}{\partial x^2}\right|_{x=\mu_X} \frac{\sigma^2_X}{2}.$$

Example 4.24: Given the random variable X with mean μ_X and variance σ^2_X, give the second-order approximation to $E(e^X)$.

Solution: Since $\frac{\partial e^x}{\partial x} = e^x$ and $\frac{\partial^2 e^x}{\partial x^2} = e^x$, we obtain $E(e^X) \approx e^{\mu_X}(1 + \sigma^2_X/2)$.

Similarly, we can develop an approximation for $\mathrm{Var}[g(x)]$ by taking the variance of both sides of the first-order Taylor series expansion of $g(x)$.

Approximation of $\mathrm{Var}[g(X)]$

$$\mathrm{Var}[g(X)] \approx \left[\frac{\partial g(x)}{\partial x}\right]^2_{x=\mu_X} \sigma^2_X.$$

Example 4.25: Given the random variable X as in Example 4.24, give an approximate formula for $\mathrm{Var}[g(x)]$.

Solution: Again $\frac{\partial e^x}{\partial x} = e^x$; thus, $\mathrm{Var}(X) \approx e^{2\mu_X}\sigma^2_X$.

These approximations can be extended to nonlinear functions of more than one random variable.

Given a set of independent random variables $X_1, X_2, \ldots, X_k$ with means $\mu_1, \mu_2, \ldots, \mu_k$ and variances $\sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2$, respectively, let

$$Y = h(X_1, X_2, \ldots, X_k)$$

be a nonlinear function; then the following are approximations for $E(Y)$ and $\mathrm{Var}(Y)$:

$$E(Y) \approx h(\mu_1, \mu_2, \ldots, \mu_k) + \sum_{i=1}^{k} \frac{\sigma_i^2}{2} \left.\left[\frac{\partial^2 h(x_1, x_2, \ldots, x_k)}{\partial x_i^2}\right]\right|_{x_i=\mu_i,\ 1\le i\le k},$$

$$\mathrm{Var}(Y) \approx \sum_{i=1}^{k} \left.\left[\frac{\partial h(x_1, x_2, \ldots, x_k)}{\partial x_i}\right]^2\right|_{x_i=\mu_i,\ 1\le i\le k} \sigma_i^2.$$

Example 4.26: Consider two independent random variables X and Z with means μ_X and μ_Z and variances σ_X^2 and σ_Z^2, respectively. Consider a random variable

$$Y = X/Z.$$

Give approximations for $E(Y)$ and $\mathrm{Var}(Y)$.

***Solution*:** For $E(Y)$, we must use $\frac{\partial y}{\partial x} = \frac{1}{z}$ and $\frac{\partial y}{\partial z} = -\frac{x}{z^2}$. Thus,

$$\frac{\partial^2 y}{\partial x^2} = 0 \text{ and } \frac{\partial^2 y}{\partial z^2} = \frac{2x}{z^3}.$$

As a result,

$$E(Y) \approx \frac{\mu_X}{\mu_Z} + \frac{\mu_X}{\mu_Z^3}\sigma_Z^2 = \frac{\mu_X}{\mu_Z}\left(1 + \frac{\sigma_Z^2}{\mu_Z^2}\right),$$

and the approximation for the variance of Y is given by

$$\mathrm{Var}(Y) \approx \frac{1}{\mu_Z^2}\sigma_X^2 + \frac{\mu_X^2}{\mu_Z^4}\sigma_Z^2 = \frac{1}{\mu_Z^2}\left(\sigma_X^2 + \frac{\mu_X^2}{\mu_Z^2}\sigma_Z^2\right).$$

Probability and Statistics

Chapter 5

Department of Industrial Engineering

١

1

Chapter 5

Some Discrete Probability Distributions

٢

2

The Bernoulli Process

Strictly speaking, the Bernoulli process must possess the following properties:

1. The experiment consists of repeated trials.
2. Each trial results in an outcome that may be classified as a success or a failure.
3. The probability of success, denoted by p, remains constant from trial to trial.
4. The repeated trials are independent.

Consider the set of Bernoulli trials where three items are selected at random from a manufacturing process, inspected, and classified as defective or nondefective. A defective item is designated a success. The number of successes is a random variable X assuming integral values from 0 through 3. The eight possible outcomes and the corresponding values of X are

Outcome	NNN	NDN	NND	DNN	NDD	DND	DDN	DDD
x	0	1	1	1	2	2	2	3

Since the items are selected independently and we assume that the process produces 25% defectives, we have

$$P(NDN) = P(N)P(D)P(N) = \left(\frac{3}{4}\right)\left(\frac{1}{4}\right)\left(\frac{3}{4}\right) = \frac{9}{64}.$$

Similar calculations yield the probabilities for the other possible outcomes. The probability distribution of X is therefore

x	0	1	2	3
$f(x)$	$\frac{27}{64}$	$\frac{27}{64}$	$\frac{9}{64}$	$\frac{1}{64}$

٣

3

Binomial Distribution

The number X of successes in n Bernoulli trials is called a **binomial random variable**. The probability distribution of this discrete random variable is called the **binomial distribution**, and its values will be denoted by $b(x; n, p)$ since they depend on the number of trials and the probability of a success on a given trial. Thus, for the probability distribution of X, the number of defectives is

$$P(X = 2) = f(2) = b\left(2; 3, \frac{1}{4}\right) = \frac{9}{64}.$$

Binomial Distribution A Bernoulli trial can result in a success with probability p and a failure with probability $q = 1 - p$. Then the probability distribution of the binomial random variable X, the number of successes in n independent trials, is

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

Note that when $n = 3$ and $p = 1/4$, the probability distribution of X, the number of defectives, may be written as

$$b\left(x; 3, \frac{1}{4}\right) = \binom{3}{x}\left(\frac{1}{4}\right)^x \left(\frac{3}{4}\right)^{3-x}, \quad x = 0, 1, 2, 3,$$

rather than in the tabular form on page 144.

٤

4

Example 5.2: The probability that a patient recovers from a rare blood disease is 0.4. If 15 people are known to have contracted this disease, what is the probability that (a) at least 10 survive, (b) from 3 to 8 survive, and (c) exactly 5 survive?

Solution: Let X be the number of people who survive.

(a) $P(X \geq 10) =$ P(X=10)+P(X=11)+P(X=12)+P(X=13) +P(X=14)+P(X=15)

$= 0.0338$

(b) $P(3 \leq X \leq 8) =$ P(X=3)+P(X=4)+P(X=5)+P(X=6) +P(X=7)+P(X=8)

$=$ 0.8779

(c) $P(X = 5) = b(5; 15, 0.4) =$ 0.1859

٥

5

Example 5.3: A large chain retailer purchases a certain kind of electronic device from a manufacturer. The manufacturer indicates that the defective rate of the device is 3%.

(a) The inspector randomly picks 20 items from a shipment. What is the probability that there will be at least one defective item among these 20?

(b) Suppose that the retailer receives 10 shipments in a month and the inspector randomly tests 20 devices per shipment. What is the probability that there will be exactly 3 shipments each containing at least one defective device among the 20 that are selected and tested from the shipment?

Solution: (a) Denote by X the number of defective devices among the 20. Then X follows a $b(x; 20, 0.03)$ distribution. Hence,

$$P(X \geq 1) = 1 - P(X = 0) = 1 - b(0; 20, 0.03)$$
$$= 1 - (0.03)^0(1 - 0.03)^{20-0} = 0.4562.$$

(b) In this case, each shipment can either contain at least one defective item or not. Hence, testing of each shipment can be viewed as a Bernoulli trial with $p = 0.4562$ from part (a). Assuming independence from shipment to shipment and denoting by Y the number of shipments containing at least one defective item, Y follows another binomial distribution $b(y; 10, 0.4562)$. Therefore,

$$P(Y = 3) = \binom{10}{3} 0.4562^3(1 - 0.4562)^7 = 0.1602.$$

٦

6

Theorem 5.1: The mean and variance of the binomial distribution $b(x; n, p)$ are

$$\mu = np \text{ and } \sigma^2 = npq.$$

***Proof*:** Let the outcome on the jth trial be represented by a Bernoulli random variable I_j, which assumes the values 0 and 1 with probabilities q and p, respectively. Therefore, in a binomial experiment the number of successes can be written as the sum of the n independent indicator variables. Hence,

$$X = I_1 + I_2 + \cdots + I_n.$$

The mean of any I_j is $E(I_j) = (0)(q) + (1)(p) = p$. Therefore, using Corollary 4.4 on page 131, the mean of the binomial distribution is

$$\mu = E(X) = E(I_1) + E(I_2) + \cdots + E(I_n) = \underbrace{p + p + \cdots + p}_{n \text{ terms}} = np.$$

The variance of any I_j is $\sigma_{I_j}^2 = E(I_j^2) - p^2 = (0)^2(q) + (1)^2(p) - p^2 = p(1-p) = pq$. Extending Corollary 4.11 to the case of n independent Bernoulli variables gives the variance of the binomial distribution as

$$\sigma_X^2 = \sigma_{I_1}^2 + \sigma_{I_2}^2 + \cdots + \sigma_{I_n}^2 = \underbrace{pq + pq + \cdots + pq}_{n \text{ terms}} = npq.$$

7

Example 5.4: It is conjectured that an impurity exists in 30% of all drinking wells in a certain rural community. In order to gain some insight into the true extent of the problem, it is determined that some testing is necessary. It is too expensive to test all of the wells in the area, so 10 are randomly selected for testing.

(a) Using the binomial distribution, what is the probability that exactly 3 wells have the impurity, assuming that the conjecture is correct?

(b) What is the probability that more than 3 wells are impure?

***Solution*:** (a) We require

$$b(3; 10, 0.3) = \sum_{x=0}^{3} b(x; 10, 0.3) - \sum_{x=0}^{2} b(x; 10, 0.3) = 0.6496 - 0.3828 = 0.2668.$$

(b) In this case, $P(X > 3) = 1 - 0.6496 = 0.3504$.

Example 5.5: Find the mean and variance of the binomial random variable of Example 5.2, and then use Chebyshev's theorem (on page 137) to interpret the interval $\mu \pm 2\sigma$.

***Solution*:** Since Example 5.2 was a binomial experiment with $n = 15$ and $p = 0.4$, by Theorem 5.1, we have

$$\mu = (15)(0.4) = 6 \text{ and } \sigma^2 = (15)(0.4)(0.6) = 3.6.$$

Taking the square root of 3.6, we find that $\sigma = 1.897$. Hence, the required interval is $6 \pm (2)(1.897)$, or from 2.206 to 9.794. Chebyshev's theorem states that the number of recoveries among 15 patients who contracted the disease has a probability of at least 3/4 of falling between 2.206 and 9.794 or, because the data are discrete, between 2 and 10 inclusive.

There are solutions in which the computation of binomial probabilities may allow us to draw a scientific inference about population after data are collected. An illustration is given in the next example.

8

Example 5.6: Consider the situation of Example 5.4. The notion that 30% of the wells are impure is merely a conjecture put forth by the area water board. Suppose 10 wells are randomly selected and 6 are found to contain the impurity. What does this imply about the conjecture? Use a probability statement.

***Solution*:** We must first ask: "If the conjecture is correct, is it likely that we would find 6 or more impure wells?"

$$P(X \geq 6) = \sum_{x=0}^{10} b(x; 10, 0.3) - \sum_{x=0}^{5} b(x; 10, 0.3) = 1 - 0.9527 = 0.0473.$$

As a result, it is very unlikely (4.7% chance) that 6 or more wells would be found impure if only 30% of all are impure. This casts considerable doubt on the conjecture and suggests that the impurity problem is much more severe.

As the reader should realize by now, in many applications there are more than two possible outcomes. To borrow an example from the field of genetics, the color of guinea pigs produced as offspring may be red, black, or white. Often the "defective" or "not defective" dichotomy is truly an oversimplification in engineering situations. Indeed, there are often more than two categories that characterize items or parts coming off an assembly line.

٩

9

Hypergeometric Distribution

Hypergeometric Distribution: The probability distribution of the hypergeometric random variable X, the number of successes in a random sample of size n selected from N items of which k are labeled **success** and $N - k$ labeled **failure**, is

$$h(x; N, n, k) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}, \quad \max\{0, n - (N - k)\} \leq x \leq \min\{n, k\}.$$

Example 5.9: Lots of 40 components each are deemed unacceptable if they contain 3 or more defectives. The procedure for sampling a lot is to select 5 components at random and to reject the lot if a defective is found. What is the probability that exactly 1 defective is found in the sample if there are 3 defectives in the entire lot?

***Solution*:** Using the hypergeometric distribution with $n = 5$, $N = 40$, $k = 3$, and $x = 1$, we find the probability of obtaining 1 defective to be

$$h(1; 40, 5, 3) = \frac{\binom{3}{1}\binom{37}{4}}{\binom{40}{5}} = 0.3011.$$

Once again, this plan is not desirable since it detects a bad lot (3 defectives) only about 30% of the time.

١٠

10

Theorem 5.2: The mean and variance of the hypergeometric distribution $h(x; N, n, k)$ are

$$\mu = \frac{nk}{N} \text{ and } \sigma^2 = \frac{N-n}{N-1} \cdot n \cdot \frac{k}{N}\left(1 - \frac{k}{N}\right).$$

Example 5.10: Let us now reinvestigate Example 3.4 on page 83. The purpose of this example was to illustrate the notion of a random variable and the corresponding sample space. In the example, we have a lot of 100 items of which 12 are defective. What is the probability that in a sample of 10, 3 are defective?

***Solution*:** Using the hypergeometric probability function, we have

$$h(3; 100, 10, 12) = \frac{\binom{12}{3}\binom{88}{7}}{\binom{100}{10}} = 0.08.$$

١١

Example 5.11: Find the mean and variance of the random variable of Example 5.9 and then use Chebyshev's theorem to interpret the interval $\mu \pm 2\sigma$.

***Solution*:** Since Example 5.9 was a hypergeometric experiment with $N = 40$, $n = 5$, and $k = 3$, by Theorem 5.2, we have

$$\mu = \frac{(5)(3)}{40} = \frac{3}{8} = 0.375,$$

and

$$\sigma^2 = \left(\frac{40-5}{39}\right)(5)\left(\frac{3}{40}\right)\left(1 - \frac{3}{40}\right) = 0.3113.$$

Taking the square root of 0.3113, we find that $\sigma = 0.558$. Hence, the required interval is $0.375 \pm (2)(0.558)$, or from -0.741 to 1.491. Chebyshev's theorem states that the number of defectives obtained when 5 components are selected at random from a lot of 40 components of which 3 are defective has a probability of at least 3/4 of falling between -0.741 and 1.491. That is, at least three-fourths of the time, the 5 components include fewer than 2 defectives.

١٢

Theorem 5.2: The mean and variance of the hypergeometric distribution $h(x; N, n, k)$ are

$$\mu = \frac{nk}{N} \text{ and } \sigma^2 = \frac{N-n}{N-1} \cdot n \cdot \frac{k}{N}\left(1 - \frac{k}{N}\right).$$

The proof for the mean is shown in Appendix A.24.

Example 5.10: Let us now reinvestigate Example 3.4 on page 83. The purpose of this example was to illustrate the notion of a random variable and the corresponding sample space. In the example, we have a lot of 100 items of which 12 are defective. What is the probability that in a sample of 10, 3 are defective?

***Solution*:** Using the hypergeometric probability function, we have

$$h(3; 100, 10, 12) = \frac{\binom{12}{3}\binom{88}{7}}{\binom{100}{10}} = 0.08.$$

١٣

13

Example 5.11: Find the mean and variance of the random variable of Example 5.9 and then use Chebyshev's theorem to interpret the interval $\mu \pm 2\sigma$.

***Solution*:** Since Example 5.9 was a hypergeometric experiment with $N = 40$, $n = 5$, and $k = 3$, by Theorem 5.2, we have

$$\mu = \frac{(5)(3)}{40} = \frac{3}{8} = 0.375,$$

and

$$\sigma^2 = \left(\frac{40-5}{39}\right)(5)\left(\frac{3}{40}\right)\left(1 - \frac{3}{40}\right) = 0.3113.$$

Taking the square root of 0.3113, we find that $\sigma = 0.558$. Hence, the required interval is $0.375 \pm (2)(0.558)$, or from -0.741 to 1.491. Chebyshev's theorem states that the number of defectives obtained when 5 components are selected at random from a lot of 40 components of which 3 are defective has a probability of at least 3/4 of falling between -0.741 and 1.491. That is, at least three-fourths of the time, the 5 components include fewer than 2 defectives.

١٤

14

Negative Binomial Distribution If repeated independent trials can result in a success with probability p and a failure with probability $q = 1 - p$, then the probability distribution of the random variable X, the number of the trial on which the kth success occurs, is

$$b^*(x; k, p) = \binom{x-1}{k-1} p^k q^{x-k}, \quad x = k, k+1, k+2, \ldots.$$

Example 5.14: In an NBA (National Basketball Association) championship series, the team that wins four games out of seven is the winner. Suppose that teams A and B face each other in the championship games and that team A has probability 0.55 of winning a game over team B.

(a) What is the probability that team A will win the series in 6 games?

(b) What is the probability that team A will win the series?

(c) If teams A and B were facing each other in a regional playoff series, which is decided by winning three out of five games, what is the probability that team A would win the series?

Solution: (a) $b^*(6; 4, 0.55) = \binom{5}{3} 0.55^4 (1 - 0.55)^{6-4} = 0.1853$

(b) P(team A wins the championship series) is

$$b^*(4; 4, 0.55) + b^*(5; 4, 0.55) + b^*(6; 4, 0.55) + b^*(7; 4, 0.55)$$
$$= 0.0915 + 0.1647 + 0.1853 + 0.1668 = 0.6083.$$

(c) P(team A wins the playoff) is

$$b^*(3; 3, 0.55) + b^*(4; 3, 0.55) + b^*(5; 3, 0.55)$$
$$= 0.1664 + 0.2246 + 0.2021 = 0.5931.$$

١٥

15

Geometric Distribution If repeated independent trials can result in a success with probability p and a failure with probability $q = 1 - p$, then the probability distribution of the random variable X, the number of the trial on which the first success occurs, is

$$g(x; p) = pq^{x-1}, \quad x = 1, 2, 3, \ldots.$$

Example 5.15: For a certain manufacturing process, it is known that, on the average, 1 in every 100 items is defective. What is the probability that the fifth item inspected is the first defective item found?

Solution: Using the geometric distribution with $x = 5$ and $p = 0.01$, we have

$$g(5; 0.01) = (0.01)(0.99)^4 = 0.0096.$$

Example 5.16: At a "busy time," a telephone exchange is very near capacity, so callers have difficulty placing their calls. It may be of interest to know the number of attempts necessary in order to make a connection. Suppose that we let $p = 0.05$ be the probability of a connection during a busy time. We are interested in knowing the probability that 5 attempts are necessary for a successful call.

Solution: Using the geometric distribution with $x = 5$ and $p = 0.05$ yields

$$P(X = x) = g(5; 0.05) = (0.05)(0.95)^4 = 0.041.$$

Quite often, in applications dealing with the geometric distribution, the mean and variance are important. For example, in Example 5.16, the *expected* number of calls necessary to make a connection is quite important. The following theorem states without proof the mean and variance of the geometric distribution.

١٦

16

Theorem 5.3: The mean and variance of a random variable following the geometric distribution are

$$\mu = \frac{1}{p} \text{ and } \sigma^2 = \frac{1-p}{p^2}.$$

١٧

17

Poisson Distribution The probability distribution of the Poisson random variable X, representing the number of outcomes occurring in a given time interval or specified region denoted by t, is

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

where λ is the average number of outcomes per unit time, distance, area, or volume and $e = 2.71828\ldots$.

Table A.2 contains Poisson probability sums,

$$P(r; \lambda t) = \sum_{x=0}^{r} p(x; \lambda t),$$

for selected values of λt ranging from 0.1 to 18.0. We illustrate the use of this table with the following two examples.

Example 5.17: During a laboratory experiment, the average number of radioactive particles passing through a counter in 1 millisecond is 4. What is the probability that 6 particles enter the counter in a given millisecond?

Solution: Using the Poisson distribution with $x = 6$ and $\lambda t = 4$ and referring to Table A.2, we have

$$p(6; 4) = \frac{e^{-4}4^6}{6!} = \sum_{x=0}^{6} p(x; 4) - \sum_{x=0}^{5} p(x; 4) = 0.8893 - 0.7851 = 0.1042.$$

١٨

18

Example 5.19: In a certain industrial facility, accidents occur infrequently. It is known that the probability of an accident on any given day is 0.005 and accidents are independent of each other.

(a) What is the probability that in any given period of 400 days there will be an accident on one day?

(b) What is the probability that there are at most three days with an accident?

***Solution*:** Let X be a binomial random variable with $n = 400$ and $p = 0.005$. Thus, $np = 2$. Using the Poisson approximation,

(a) $P(X = 1) = e^{-2}2^1 = 0.271$ and

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

Example 5.20: In a manufacturing process where glass products are made, defects or bubbles occur, occasionally rendering the piece undesirable for marketing. It is known that, on average, 1 in every 1000 of these items produced has one or more bubbles. What is the probability that a random sample of 8000 will yield fewer than 7 items possessing bubbles?

***Solution*:** This is essentially a binomial experiment with $n = 8000$ and $p = 0.001$. Since p is very close to 0 and n is quite large, we shall approximate with the Poisson distribution using

$$\mu = (8000)(0.001) = 8.$$

Hence, if X represents the number of bubbles, we have

$$P(X < 7) = \sum_{x=0}^{6} b(x; 8000, 0.001) \approx p(x; 8) = 0.3134.$$

19

Chapter 6

Some Continuous Probability Distributions

١

1

Uniform Distribution The density function of the continuous uniform random variable X on the interval $[A, B]$ is

$$f(x; A, B) = \begin{cases} \frac{1}{B-A}, & A \leq x \leq B, \\ 0, & \text{elsewhere.} \end{cases}$$

Example 6.1: Suppose that a large conference room at a certain company can be reserved for no more than 4 hours. Both long and short conferences occur quite often. In fact, it can be assumed that the length X of a conference has a uniform distribution on the interval [0, 4].

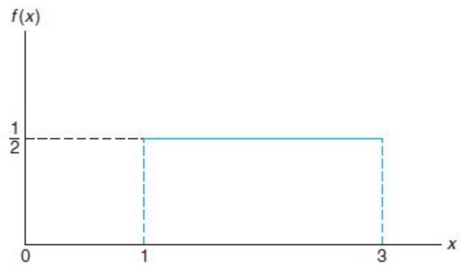


Figure 6.1: The density function for a random variable on the interval [1, 3].

٢

2

(a) What is the probability density function?

(b) What is the probability that any given conference lasts at least 3 hours?

Solution: (a) The appropriate density function for the uniformly distributed random variable X in this situation is

$$f(x) = \begin{cases} \frac{1}{4}, & 0 \le x \le 4, \\ 0, & \text{elsewhere.} \end{cases}$$

(b) $P[X \ge 3] = \int_3^4 \frac{1}{4}\, dx = \frac{1}{4}.$

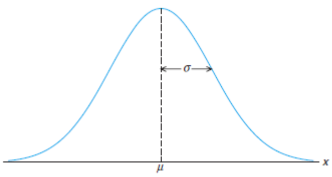
Theorem 6.1: The mean and variance of the uniform distribution are

$$\mu = \frac{A+B}{2} \text{ and } \sigma^2 = \frac{(B-A)^2}{12}.$$

٣

3

Normal Distribution



Normal Distribution The density of the normal random variable X, with mean μ and variance σ^2, is

$$n(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}, \quad -\infty < x < \infty,$$

where $\pi = 3.14159\ldots$ and $e = 2.71828\ldots$.

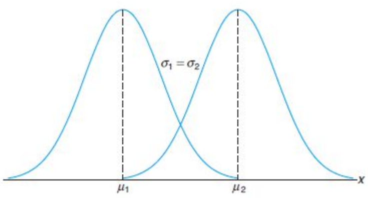


Figure 6.3: Normal curves with $\mu_1 < \mu_2$ and $\sigma_1 = \sigma_2$.

٤

4

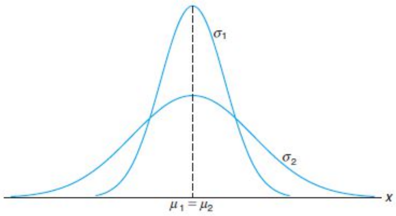


Figure 6.4: Normal curves with $\mu_1 = \mu_2$ and $\sigma_1 < \sigma_2$.

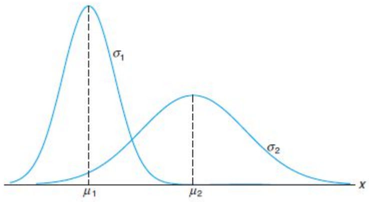


Figure 6.5: Normal curves with $\mu_1 < \mu_2$ and $\sigma_1 < \sigma_2$.

٥

5

Theorem 6.2: The mean and variance of $n(x;\mu,\sigma)$ are μ and σ^2, respectively. Hence, the standard deviation is σ.

Areas under the Normal Curve

The curve of any continuous probability distribution or density function is constructed so that the area under the curve bounded by the two ordinates $x = x_1$ and $x = x_2$ equals the probability that the random variable X assumes a value between $x = x_1$ and $x = x_2$. Thus, for the normal curve in Figure 6.6,

$$P(x_1 < X < x_2) = \int_{x_1}^{x_2} n(x;\mu,\sigma)\,dx = \frac{1}{\sqrt{2\pi}\sigma}\int_{x_1}^{x_2} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}dx$$

is represented by the area of the shaded region.

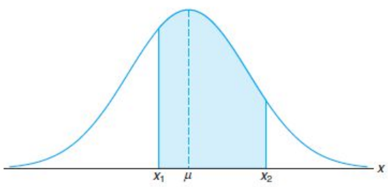


Figure 6.6: $P(x_1 < X < x_2)$ = area of the shaded region.

٦

6

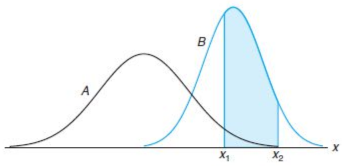
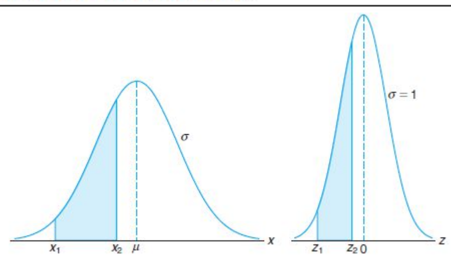


Figure 6.7: $P(x_1 < X < x_2)$ for different normal curves.

of a normal random variable Z with mean 0 and variance 1. This can be done by means of the transformation

$$Z = \frac{X - \mu}{\sigma}.$$

Definition 6.1: The distribution of a normal random variable with mean 0 and variance 1 is called a **standard normal distribution**.



٧

Figure 6.8: The original and transformed normal distributions.

Example 6.2: Given a standard normal distribution, find the area under the curve that lies

(a) to the right of $z = 1.84$ and

(b) between $z = -1.97$ and $z = 0.86$.

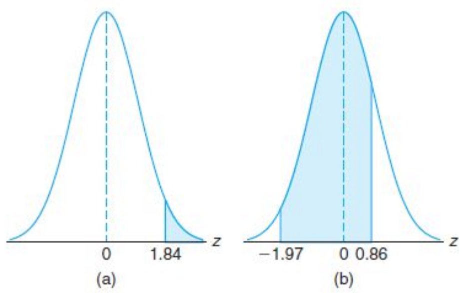


Figure 6.9: Areas for Example 6.2.

Solution: See Figure 6.9 for the specific areas.

(a) The area in Figure 6.9(a) to the right of $z = 1.84$ is equal to 1 minus the area in Table A.3 to the left of $z = 1.84$, namely, $1 - 0.9671 = 0.0329$.

(b) The area in Figure 6.9(b) between $z = -1.97$ and $z = 0.86$ is equal to the area to the left of $z = 0.86$ minus the area to the left of $z = -1.97$. From Table A.3 we find the desired area to be $0.8051 - 0.0244 = 0.7807$.

٨

Example 6.3: Given a standard normal distribution, find the value of k such that

(a) $P(Z > k) = 0.3015$ and
(b) $P(k < Z < -0.18) = 0.4197$.

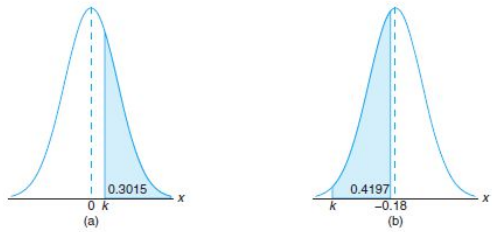


Figure 6.10: Areas for Example 6.3.

Solution: Distributions and the desired areas are shown in Figure 6.10.

(a) In Figure 6.10(a), we see that the k value leaving an area of 0.3015 to the right must then leave an area of 0.6985 to the left. From Table A.3 it follows that $k = 0.52$.

(b) From Table A.3 we note that the total area to the left of -0.18 is equal to 0.4286. In Figure 6.10(b), we see that the area between k and -0.18 is 0.4197, so the area to the left of k must be $0.4286 - 0.4197 = 0.0089$. Hence, from Table A.3, we have $k = -2.37$.

٩

Example 6.4: Given a random variable X having a normal distribution with $\mu = 50$ and $\sigma = 10$, find the probability that X assumes a value between 45 and 62.

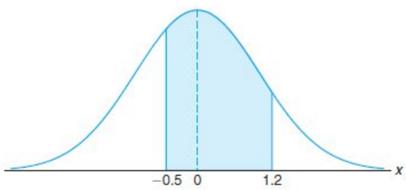


Figure 6.11: Area for Example 6.4.

Solution: The z values corresponding to $x_1 = 45$ and $x_2 = 62$ are

$$z_1 = \frac{45 - 50}{10} = -0.5 \text{ and } z_2 = \frac{62 - 50}{10} = 1.2.$$

Therefore,

$$P(45 < X < 62) = P(-0.5 < Z < 1.2).$$

$P(-0.5 < Z < 1.2)$ is shown by the area of the shaded region in Figure 6.11. This area may be found by subtracting the area to the left of the ordinate $z = -0.5$ from the entire area to the left of $z = 1.2$. Using Table A.3, we have

$$P(45 < X < 62) = P(-0.5 < Z < 1.2) = P(Z < 1.2) - P(Z < -0.5)$$
$$= 0.8849 - 0.3085 = 0.5764.$$

١٠

Example 6.5: Given that X has a normal distribution with $\mu = 300$ and $\sigma = 50$, find the probability that X assumes a value greater than 362.

Solution: The normal probability distribution with the desired area shaded is shown in Figure 6.12. To find $P(X > 362)$, we need to evaluate the area under the normal curve to the right of $x = 362$. This can be done by transforming $x = 362$ to the corresponding z value, obtaining the area to the left of z from Table A.3, and then subtracting this area from 1. We find that

$$z = \frac{362 - 300}{50} = 1.24.$$

Hence,

$$P(X > 362) = P(Z > 1.24) = 1 - P(Z < 1.24) = 1 - 0.8925 = 0.1075.$$

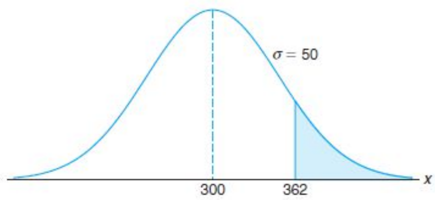


Figure 6.12: Area for Example 6.5.

١١

11

Example 6.6: Given a normal distribution with $\mu = 40$ and $\sigma = 6$, find the value of x that has

(a) 45% of the area to the left and

(b) 14% of the area to the right.

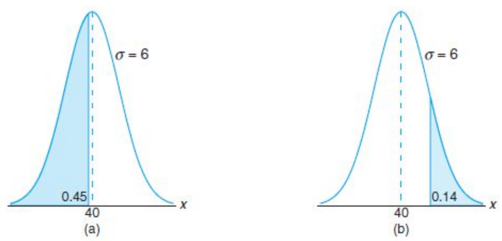


Figure 6.13: Areas for Example 6.6.

Solution: (a) An area of 0.45 to the left of the desired x value is shaded in Figure 6.13(a). We require a z value that leaves an area of 0.45 to the left. From Table A.3 we find $P(Z < -0.13) = 0.45$, so the desired z value is -0.13. Hence,

$$x = (6)(-0.13) + 40 = 39.22.$$

(b) In Figure 6.13(b), we shade an area equal to 0.14 to the right of the desired x value. This time we require a z value that leaves 0.14 of the area to the right and hence an area of 0.86 to the left. Again, from Table A.3, we find $P(Z < 1.08) = 0.86$, so the desired z value is 1.08 and

$$x = (6)(1.08) + 40 = 46.48.$$

١٢

12

Example 6.7: A certain type of storage battery lasts, on average, 3.0 years with a standard deviation of 0.5 year. Assuming that battery life is normally distributed, find the probability that a given battery will last less than 2.3 years.

Solution: First construct a diagram such as Figure 6.14, showing the given distribution of battery lives and the desired area. To find $P(X < 2.3)$, we need to evaluate the area under the normal curve to the left of 2.3. This is accomplished by finding the area to the left of the corresponding z value. Hence, we find that

$$z = \frac{2.3 - 3}{0.5} = -1.4,$$

and then, using Table A.3, we have

$$P(X < 2.3) = P(Z < -1.4) = 0.0808.$$

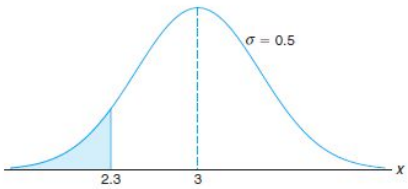


Figure 6.14: Area for Example 6.7.

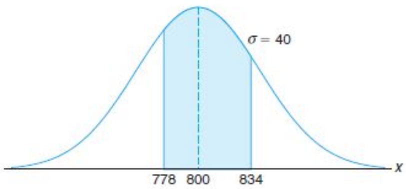


Figure 6.15: Area for Example 6.8.

١٣

Example 6.8: An electrical firm manufactures light bulbs that have a life, before burn-out, that is normally distributed with mean equal to 800 hours and a standard deviation of 40 hours. Find the probability that a bulb burns between 778 and 834 hours.

Solution: The distribution of light bulb life is illustrated in Figure 6.15. The z values corresponding to $x_1 = 778$ and $x_2 = 834$ are

$$z_1 = \frac{778 - 800}{40} = -0.55 \text{ and } z_2 = \frac{834 - 800}{40} = 0.85.$$

Hence,

$$P(778 < X < 834) = P(-0.55 < Z < 0.85) = P(Z < 0.85) - P(Z < -0.55)$$
$$= 0.8023 - 0.2912 = 0.5111.$$

Example 6.10: Gauges are used to reject all components for which a certain dimension is not within the specification $1.50 \pm d$. It is known that this measurement is normally distributed with mean 1.50 and standard deviation 0.2. Determine the value d such that the specifications "cover" 95% of the measurements.

Solution: From Table A.3 we know that

$$P(-1.96 < Z < 1.96) = 0.95.$$

Therefore,

$$1.96 = \frac{(1.50 + d) - 1.50}{0.2},$$

from which we obtain

$$d = (0.2)(1.96) = 0.392.$$

An illustration of the specifications is shown in Figure 6.17.

Example 6.9: In an industrial process, the diameter of a ball bearing is an important measurement. The buyer sets specifications for the diameter to be 3.0 ± 0.01 cm. The implication is that no part falling outside these specifications will be accepted. It is known that in the process the diameter of a ball bearing has a normal distribution with mean $\mu = 3.0$ and standard deviation $\sigma = 0.005$. On average, how many manufactured ball bearings will be scrapped?

Solution: The distribution of diameters is illustrated by Figure 6.16. The values corresponding to the specification limits are $x_1 = 2.99$ and $x_2 = 3.01$. The corresponding z values are

$$z_1 = \frac{2.99 - 3.0}{0.005} = -2.0 \text{ and } z_2 = \frac{3.01 - 3.0}{0.005} = +2.0.$$

Hence,

$$P(2.99 < X < 3.01) = P(-2.0 < Z < 2.0).$$

From Table A.3, $P(Z < -2.0) = 0.0228$. Due to symmetry of the normal distribution, we find that

$$P(Z < -2.0) + P(Z > 2.0) = 2(0.0228) = 0.0456.$$

As a result, it is anticipated that, on average, 4.56% of manufactured ball bearings will be scrapped.

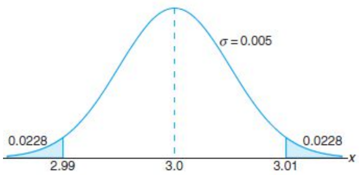


Figure 6.16: Area for Example 6.9.

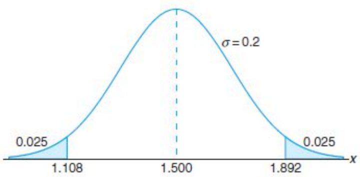


Figure 6.17: Specifications for Example 6.10.

١٥

Example 6.11: A certain machine makes electrical resistors having a mean resistance of 40 ohms and a standard deviation of 2 ohms. Assuming that the resistance follows a normal distribution and can be measured to any degree of accuracy, what percentage of resistors will have a resistance exceeding 43 ohms?

Solution: A percentage is found by multiplying the relative frequency by 100%. Since the relative frequency for an interval is equal to the probability of a value falling in the interval, we must find the area to the right of $x = 43$ in Figure 6.18. This can be done by transforming $x = 43$ to the corresponding z value, obtaining the area to the left of z from Table A.3, and then subtracting this area from 1. We find

$$z = \frac{43 - 40}{2} = 1.5.$$

Therefore,

$$P(X > 43) = P(Z > 1.5) = 1 - P(Z < 1.5) = 1 - 0.9332 = 0.0668.$$

Hence, 6.68% of the resistors will have a resistance exceeding 43 ohms.

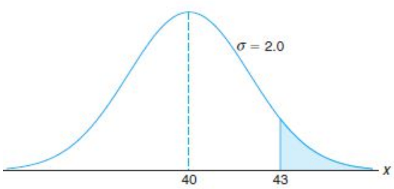


Figure 6.18: Area for Example 6.11.

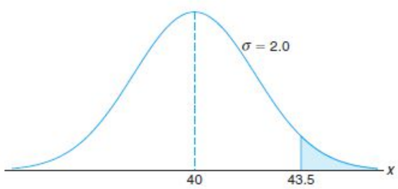


Figure 6.19: Area for Example 6.12.

١٦

Example 6.12: Find the percentage of resistances exceeding 43 ohms for Example 6.11 if resistance is measured to the nearest ohm.

Solution: This problem differs from that in Example 6.11 in that we now assign a measurement of 43 ohms to all resistors whose resistances are greater than 42.5 and less than 43.5. We are actually approximating a discrete distribution by means of a continuous normal distribution. The required area is the region shaded to the right of 43.5 in Figure 6.19. We now find that

$$z = \frac{43.5 - 40}{2} = 1.75.$$

Hence,

$$P(X > 43.5) = P(Z > 1.75) = 1 - P(Z < 1.75) = 1 - 0.9599 = 0.0401.$$

Therefore, 4.01% of the resistances exceed 43 ohms when measured to the nearest ohm. The difference 6.68% − 4.01% = 2.67% between this answer and that of Example 6.11 represents all those resistance values greater than 43 and less than 43.5 that are now being recorded as 43 ohms.

١٧

Example 6.13: The average grade for an exam is 74, and the standard deviation is 7. If 12% of the class is given *A*s, and the grades are curved to follow a normal distribution, what is the lowest possible *A* and the highest possible *B*?

Solution: In this example, we begin with a known area of probability, find the z value, and then determine x from the formula $x = \sigma z + \mu$. An area of 0.12, corresponding to the fraction of students receiving *A*s, is shaded in Figure 6.20. We require a z value that leaves 0.12 of the area to the right and, hence, an area of 0.88 to the left. From Table A.3, $P(Z < 1.18)$ has the closest value to 0.88, so the desired z value is 1.18. Hence,

$$x = (7)(1.18) + 74 = 82.26.$$

Therefore, the lowest *A* is 83 and the highest *B* is 82.

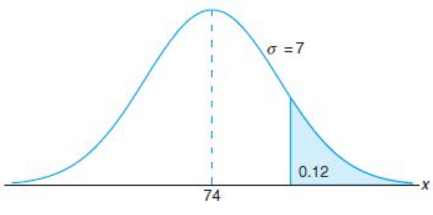


Figure 6.20: Area for Example 6.13.

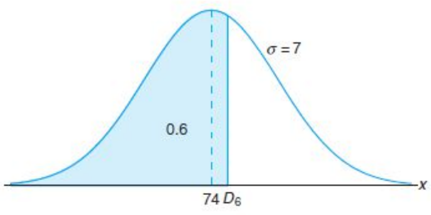


Figure 6.21: Area for Example 6.14.

١٨

Gamma and Exponential Distributions

Definition 6.2: The **gamma function** is defined by

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x}\, dx, \qquad \text{for } \alpha > 0.$$

The following are a few simple properties of the gamma function.

(a) $\Gamma(n) = (n-1)(n-2)\cdots(1)\Gamma(1)$, for a positive integer n.

To see the proof, integrating by parts with $u = x^{\alpha-1}$ and $dv = e^{-x}\, dx$, we obtain

$$\Gamma(\alpha) = -e^{-x}\, x^{\alpha-1}\Big|_0^\infty + \int_0^\infty e^{-x}(\alpha-1)x^{\alpha-2}\, dx = (\alpha-1)\int_0^\infty x^{\alpha-2}e^{-x}\, dx,$$

for $\alpha > 1$, which yields the recursion formula

$$\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1).$$

Gamma Distribution: The continuous random variable X has a **gamma distribution**, with parameters α and β, if its density function is given by

$$f(x;\alpha,\beta) = \begin{cases} \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}, & x > 0, \\ 0, & \text{elsewhere}, \end{cases}$$

where $\alpha > 0$ and $\beta > 0$.

١٩

19

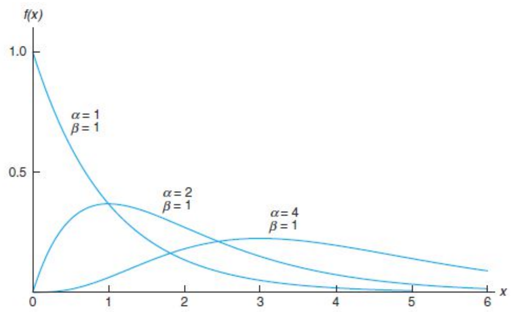


Figure 6.28: Gamma distributions.

Exponential Distribution: The continuous random variable X has an **exponential distribution**, with parameter β, if its density function is given by

$$f(x;\beta) = \begin{cases} \frac{1}{\beta} e^{-x/\beta}, & x > 0, \\ 0, & \text{elsewhere}, \end{cases}$$

where $\beta > 0$.

٢٠

20

Theorem 6.4: The mean and variance of the gamma distribution are

$$\mu = \alpha\beta \text{ and } \sigma^2 = \alpha\beta^2.$$

The proof of this theorem is found in Appendix A.26.

Corollary 6.1: The mean and variance of the exponential distribution are

$$\mu = \beta \text{ and } \sigma^2 = \beta^2.$$

Example 6.17: Suppose that a system contains a certain type of component whose time, in years, to failure is given by T. The random variable T is modeled nicely by the exponential distribution with mean time to failure $\beta = 5$. If 5 of these components are installed in different systems, what is the probability that at least 2 are still functioning at the end of 8 years?

***Solution*:** The probability that a given component is still functioning after 8 years is given by

$$P(T > 8) = \frac{1}{5}\int_8^\infty e^{-t/5}\,dt = e^{-8/5} \approx 0.2.$$

٢١

The Memoryless Property and Its Effect on the Exponential Distribution

$$P(X \geq t) = P(X \geq t_0 + t \mid X \geq t_0)$$

Chi-Squared Distribution The continuous random variable X has a **chi-squared distribution**, with v **degrees of freedom**, if its density function is given by

$$f(x; v) = \begin{cases} \frac{1}{2^{v/2}\Gamma(v/2)} x^{v/2-1} e^{-x/2}, & x > 0, \\ 0, & \text{elsewhere}, \end{cases}$$

Theorem 6.5: The mean and variance of the chi-squared distribution are

$$\mu = v \text{ and } \sigma^2 = 2v.$$

٢٢

Definition 6.3: A **beta function** is defined by

$$B(\alpha,\beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}, \text{ for } \alpha,\beta > 0,$$

where $\Gamma(\alpha)$ is the gamma function.

Beta Distribution The continuous random variable X has a **beta distribution** with parameters $\alpha > 0$ and $\beta > 0$ if its density function is given by

$$f(x) = \begin{cases} \frac{1}{B(\alpha,\beta)}x^{\alpha-1}(1-x)^{\beta-1}, & 0 < x < 1, \\ 0, & \text{elsewhere.} \end{cases}$$

Note that the uniform distribution on $(0,1)$ is a beta distribution with parameters $\alpha = 1$ and $\beta = 1$.

Theorem 6.6: The mean and variance of a beta distribution with parameters α and β are

$$\mu = \frac{\alpha}{\alpha+\beta} \text{ and } \sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)},$$

respectively.

For the uniform distribution on $(0,1)$, the mean and variance are

$$\mu = \frac{1}{1+1} = \frac{1}{2} \text{ and } \sigma^2 = \frac{(1)(1)}{(1+1)^2(1+1+1)} = \frac{1}{12},$$

٢٣

Lognormal Distribution The continuous random variable X has a **lognormal distribution** if the random variable $Y = \ln(X)$ has a normal distribution with mean μ and standard deviation σ. The resulting density function of X is

$$f(x;\mu,\sigma) = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma x}e^{-\frac{1}{2\sigma^2}[\ln(x)-\mu]^2}, & x \ge 0, \\ 0, & x < 0. \end{cases}$$

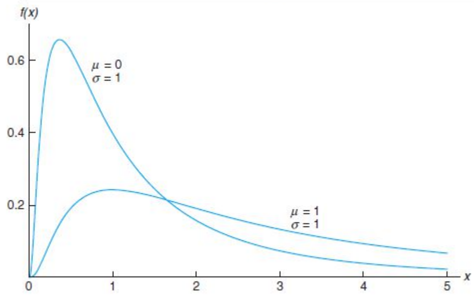


Figure 6.29: Lognormal distributions.

The mean and variance of the lognormal distribution are

$$\mu = e^{\mu+\sigma^2/2} \text{ and } \sigma^2 = e^{2\mu+\sigma^2}(e^{\sigma^2}-1).$$

٢٤

Example 6.22: Concentrations of pollutants produced by chemical plants historically are known to exhibit behavior that resembles a lognormal distribution. This is important when one considers issues regarding compliance with government regulations. Suppose it is assumed that the concentration of a certain pollutant, in parts per million, has a lognormal distribution with parameters $\mu = 3.2$ and $\sigma = 1$. What is the probability that the concentration exceeds 8 parts per million?

Solution: Let the random variable X be pollutant concentration. Then

$$P(X > 8) = 1 - P(X \leq 8).$$

Since $\ln(X)$ has a normal distribution with mean $\mu = 3.2$ and standard deviation $\sigma = 1$,

$$P(X \leq 8) = \Phi\left[\frac{\ln(8) - 3.2}{1}\right] = \Phi(-1.12) = 0.1314.$$

٢٥

25

Weibull Distribution The continuous random variable X has a **Weibull distribution**, with parameters α and β, if its density function is given by

$$f(x;\alpha,\beta) = \begin{cases} \alpha\beta x^{\beta-1}e^{-\alpha x^{\beta}}, & x > 0, \\ 0, & \text{elsewhere,} \end{cases}$$

where $\alpha > 0$ and $\beta > 0$.

Theorem 6.8: The mean and variance of the Weibull distribution are

$$\mu = \alpha^{-1/\beta}\Gamma\left(1 + \frac{1}{\beta}\right) \text{ and } \sigma^2 = \alpha^{-2/\beta}\left\{\Gamma\left(1 + \frac{2}{\beta}\right) - \left[\Gamma\left(1 + \frac{1}{\beta}\right)\right]^2\right\}.$$

٢٦

26

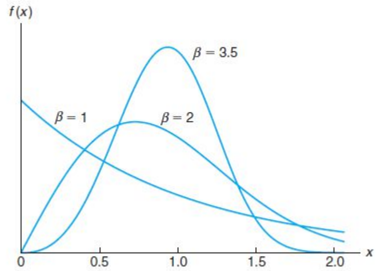


Figure 6.30: Weibull distributions ($\alpha = 1$).

cdf for Weibull Distribution

The **cumulative distribution function for the Weibull distribution** is given by

$$F(x) = 1 - e^{-\alpha x^{\beta}}, \qquad \text{for } x \geq 0,$$

for $\alpha > 0$ and $\beta > 0$.

Example 6.24: The length of life X, in hours, of an item in a machine shop has a Weibull distribution with $\alpha = 0.01$ and $\beta = 2$. What is the probability that it fails before eight hours of usage?

Solution: $P(X < 8) = F(8) = 1 - e^{-(0.01)8^2} = 1 - 0.527 = 0.473.$

٢٧

27

End of Chapter 6

٢٨

28