Probability & Statistics

Chapter 1 Introduction to Statistics

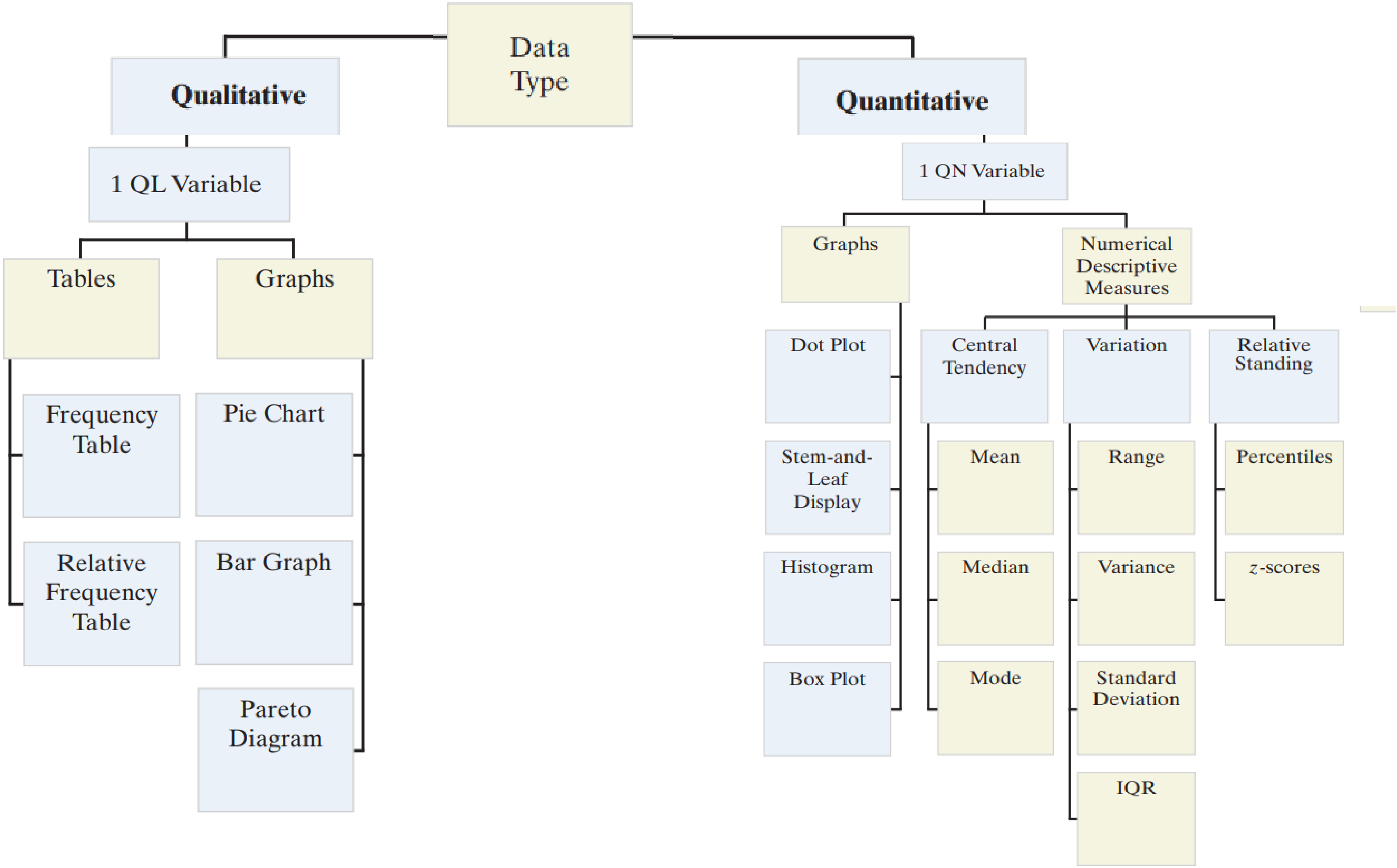
Statistics

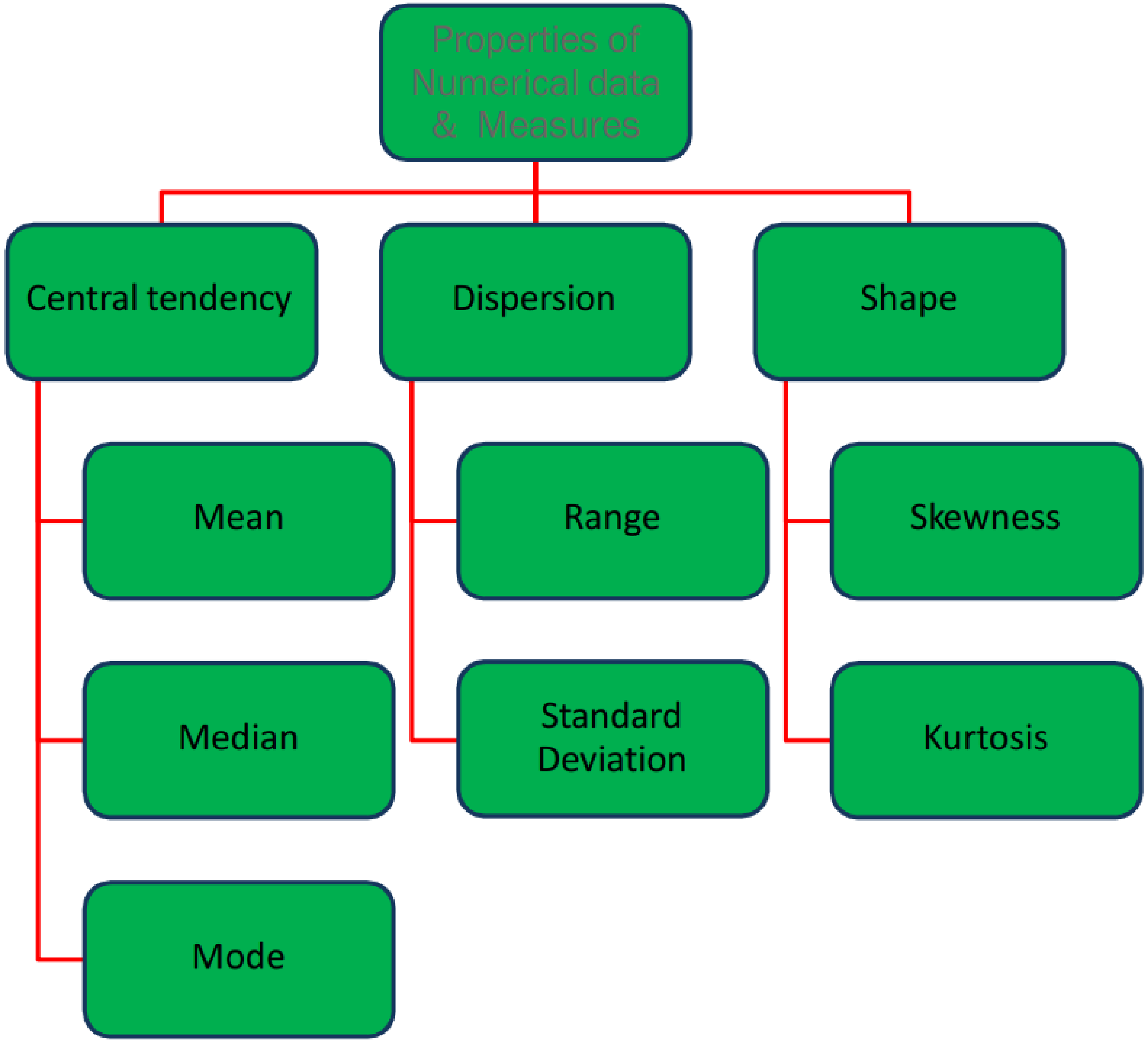
Statistics is the art of learning from data. It is concerned with the collection of data, its subsequent description, and its analysis, which often leads to the drawing of conclusions.

Describing Data Set

The numerical findings of a study should be presented clearly, concisely, and in such a manner that an observer can quickly obtain a feel for the essential characteristics of the data. Over the years it has been found that tables and graphs are particularly useful ways of presenting data, often revealing important features such as the range, the degree of concentration, and the symmetry of the data.

Guide to Selecting the Data Description Method

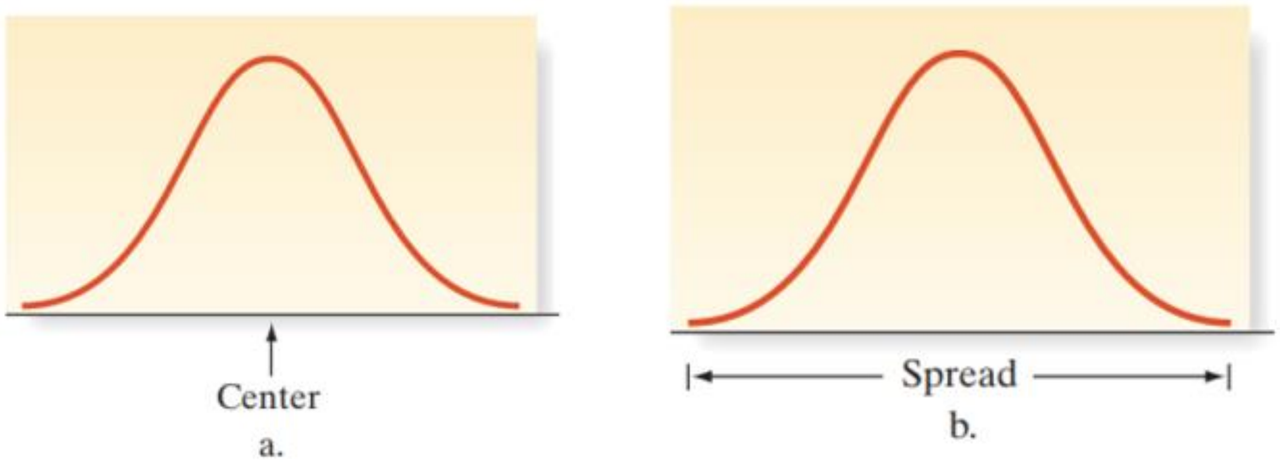



Properties of Numerical data & Measures
Central tendency
Dispersion
Shape
Mean
Range
Skewness
Median
Standard Deviation
Kurtosis
Mode

Descriptive Measures

A large number of numerical methods are available to describe quantitative data sets. Most of these methods measure one of two data characteristics

- The **central tendency** of the set of measurements-that is, the tendency of the data to cluster, or center, about certain numerical values.
- The **variability** of the set of measurements-that is, the spread of the data



Measures of Central Tendency

The most popular and best understood measure of central tendency for a quantitative data set is the arithmetic mean (or simply the mean) of the data set.

Simple Mean

The arithmetic mean-or, more succinctly, the mean-is defined as the sum of the observations divided by sample size

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Sometimes it is preferable to use the sample median as a descriptive measure of the center, or location, of a set of data. This is particularly true if it is desired to minimize the calculations or if it is desired to eliminate the effect of extreme (very large or very small) values

Sample median

The arithmetic mean-or, more succinctly, the mean-is defined as the sum of the observations divided by sample size

Order the n observations from smallest to largest.

$$\begin{aligned} \text{sample median} &= \text{observation in position } \frac{n+1}{2}, && \text{if } n \text{ odd.} \\ &= \text{average of two observations in} \\ &\quad \text{positions } \frac{n}{2} \text{ and } \frac{n+2}{2}, && \text{if } n \text{ even.} \end{aligned}$$

Example

Calculation of the sample mean and median

A sample of five university students responded to the question "How much time, in minutes, did you spend on the social network site yesterday?"

$$100 \quad 45 \quad 60 \quad 130 \quad 30$$

Find the mean and the median.

Solution

The mean is

$$\bar{x} = \frac{100 + 45 + 60 + 130 + 30}{5} = 73 \text{ minutes}$$

and, ordering the data from smallest to largest

$$30 \quad 45 \quad \underbrace{60} \quad 100 \quad 130$$

the median is the third largest value, namely, 60 minutes.

The two very large values cause the mean to be much larger than the median. ■

Example

Calculation of the sample median with even sample size

An engineering group receives e-mail requests for technical information from sales and service. The daily numbers of e-mails for six days are

$$11 \quad 9 \quad 17 \quad 19 \quad 4 \quad 15$$

Find the mean and the median.

Solution

The mean is

$$\bar{x} = \frac{11 + 9 + 17 + 19 + 4 + 15}{6} = 12.5 \text{ requests}$$

and, ordering the data from the smallest to largest

$$4 \quad 9 \quad \underbrace{11 \quad 15} \quad 17 \quad 19$$

the median, the mean of the third and fourth largest values, is 13 requests. ■

Measures of Variability

Measures of central tendency provide only a partial description of a quantitative data set. The description is incomplete without a measure of the variability, or spread, of the data set. Knowledge of the data set's variability, along with knowledge of its center, can help us visualize the shape of the data set as well as its extreme values. Perhaps the simplest measure of the variability of a quantitative data set is its range

Range

The range of a quantitative data set is equal to the largest measurement minus the smallest measurement
Sample Range $= X_{max} - X_{min}$

Sample Variance and Sample standard deviation

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,

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

- The standard deviation is by far the most generally useful measure of variation.
- Its advantage over the variance is that it is expressed in the same units as the observations
- the sample standard deviation measures variability in linear units whereas the sample variance is measured in squared units

Example

Calculation of sample variance

The delay times (handling, setting, and positioning the tools) for cutting 6 parts on an engine lathe are 0.6, 1.2, 0.9, 1.0, 0.6, and 0.8 minutes. Calculate s^2.

Solution

First we calculate the mean:

$$\bar{x} = \frac{0.6 + 1.2 + 0.9 + 1.0 + 0.6 + 0.8}{6} = 0.85$$

To find $\sum (x_i - \bar{x})^2$, we set up the table:

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
0.6	−0.25	0.0625
1.2	0.35	0.1225
0.9	0.05	0.0025
1.0	0.15	0.0225
0.6	−0.25	0.0625
0.8	−0.05	0.0025
5.1	0.00	0.2750

where the total of the third column $0.2750 = \sum(x_i - \bar{x})^2$.

We divide 0.2750 by $6 - 1 = 5$ to obtain

$$s^2 = \frac{0.2750}{5} = 0.055 \text{ (minute)}^2$$

Calculation of sample standard deviation

$s^2 = 0.055$. Take the square root and get

$$s = \sqrt{0.055} = 0.23 \text{ minute}$$

Example Calculate the variance and standard deviation of the following sample: 2, 3, 3, 3, 4.

Solution If you use the formula in the box to compute s^2 and s, you first need to find $\bar{x}$. From Table 2.6, we see that $\Sigma x = 15$. Thus, $\bar{x} = \frac{\Sigma x}{n} = \frac{15}{5} = 3$. Now, for each measurement, find $(x - \bar{x})$ and $(x - \bar{x})^2$, as shown.

Table 2.6 Calculating s^2

x	$(x - \bar{x})$	$(x - \bar{x})^2$
2	−1	1
3	0	0
3	0	0
3	0	0
4	1	1
$\Sigma x = 15$		$\Sigma(x - \bar{x})^2 = 2$

Then we use[†]

$$s^2 = \frac{\Sigma(x - \bar{x})^2}{n - 1} = \frac{2}{5 - 1} = \frac{2}{4} = .5$$

$$s = \sqrt{.5} = .71$$

Quartiles and Percentiles

Quartiles and Percentiles

In addition to the median, which divides a set of data into halves, we can consider other division points. When an ordered data set is divided into quarters, the resulting division points are called sample **quartiles**. The *first quartile*, Q_1, is a value that has one-fourth, or 25%, of the observations below its value. The first quartile is also the sample 25th **percentile** $P_{0.25}$. More generally, we define the sample 100 pth percentile as follows.

Sample percentiles

> The sample 100 pth percentile is a value such that at least $100p\%$ of the observations are at or below this value, and at least $100(1 - p)\%$ are at or above this value.

Quartiles and Percentiles

The following rule simplifies the calculation of sample percentiles.

Calculating the sample 100 pth percentile:

1. Order the n observations from smallest to largest.

2. Determine the product np.

If np is not an integer, round it up to the next integer and find the corresponding ordered value.

If np is an integer, say k, calculate the mean of the kth and $(k+1)$st ordered observations.

Sample quartiles

first quartile	$Q_1 = 25\text{th percentile}$
second quartile	$Q_2 = 50\text{th percentile}$
third quartile	$Q_3 = 75\text{th percentile}$

Example Obtain the quartiles and the 10th percentile

136	143	147	151	158	160
161	163	165	167	173	174
181	181	185	188	190	205

Solution According to our calculation rule, $np = 18\left(\frac{1}{4}\right) = 4.5$, which we round up to 5. The first quartile is the 5th ordered observation

$$Q_1 = 158 \text{ MPa}$$

Since $p = \frac{1}{2}$ for the second quartile, or median,

$$np = 18\left(\frac{1}{2}\right) = 9$$

which is an integer. Therefore, we average the 9th and 10th ordered values

$$Q_2 = \frac{165 + 167}{2} = 166 \text{ MPa}$$

The third quartile is the 14th observation, $Q_3 = 181$ seconds. We could also have started at the largest value and counted down to the 5th position.

To obtain the 10th percentile, we determine that $np = 18 \times 0.10 = 1.8$, which we round up to 2. Counting to the 2nd position, we obtain

$$P_{0.10} = 143 \text{ MPa}$$

$$\textbf{range} = \text{maximum} - \text{minimum}$$

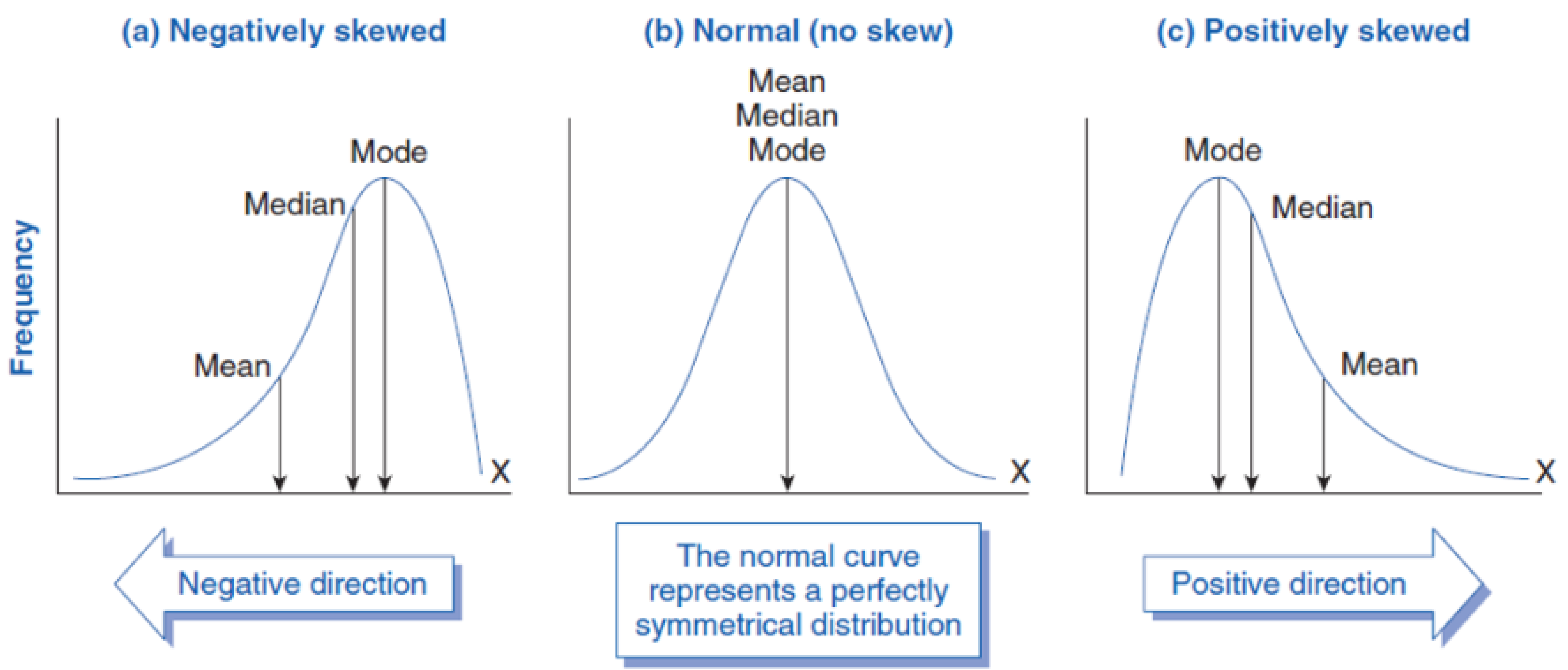
$$\textbf{interquartile range} = \text{third quartile} - \text{first quartile} = Q_3 - Q_1$$

The minimum $= 136$. From the previous example, the maximum $= 205$, $Q_1 = 158$, and $Q_3 = 181$.

$$\text{range} = \text{maximum} - \text{minimum} = 205 - 136 = 69\,\text{MPa}$$

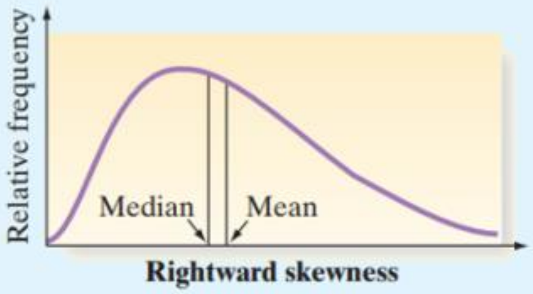
$$\text{interquartile range} = Q_3 - Q_1 = 181 - 158 = 23\,\text{MPa}$$

■

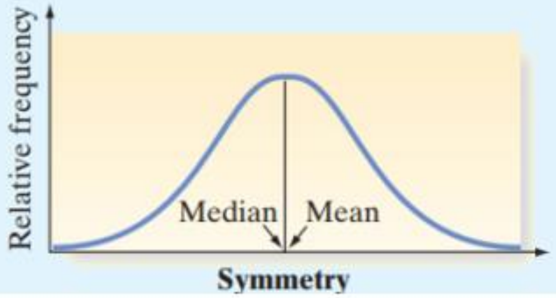


Detecting Skewness by Comparing the Mean and the Median

If the data set is skewed to the right, then typically the median is less than the mean.



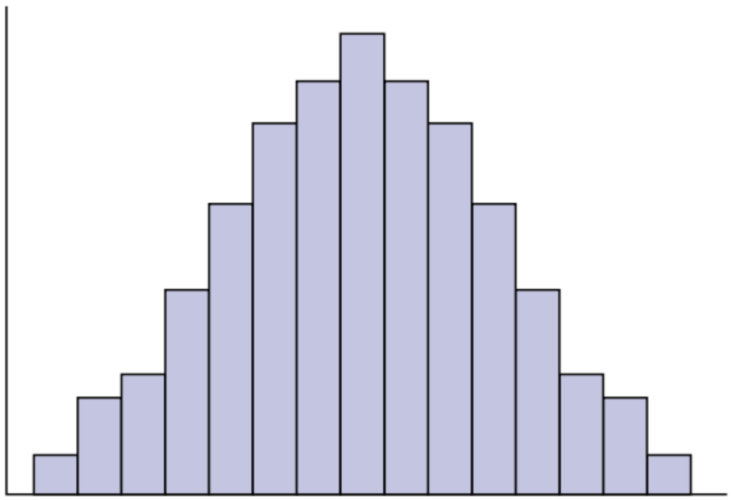
If the data set is symmetric, then the mean equals the median.



NORMAL DATA SETS AND THE EMPIRICAL RULE

Definition *A data set is said to be* normal *if a histogram describing it has the following properties:*

1. *It is highest at the middle interval.*
2. *Moving from the middle interval in either direction, the height decreases in such a way that the entire histogram is bell-shaped.*
3. *The histogram is symmetric about its middle interval.*



Histogram of a normal data set.

Empirical Rule

If a data set is approximately normal with sample mean $\bar{x}$ and sample standard deviation s, then the following are true.

Normal distributions follow the **empirical rule**, also called the **68-95-99.7 rule**

1. Approximately 68 percent of the observations lie within

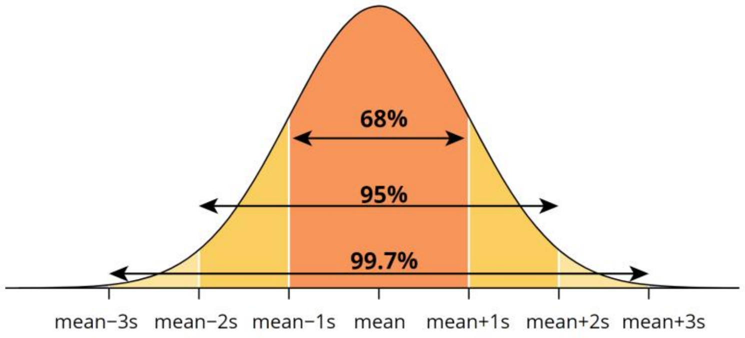
$$\bar{x} \pm s$$

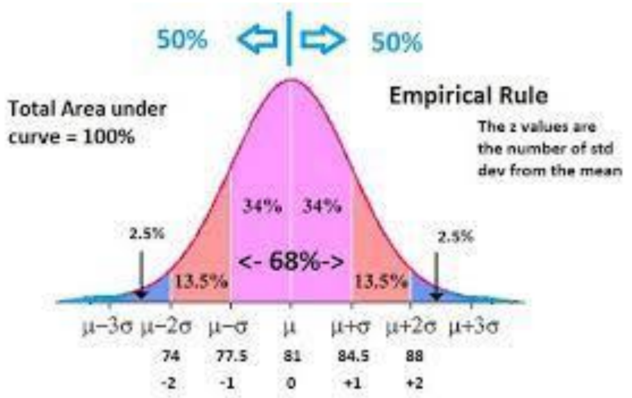
2. Approximately 95 percent of the observations lie within

$$\bar{x} \pm 2s$$

3. Approximately 99.7 percent of the observations lie within

$$\bar{x} \pm 3s$$





Z-scores

A z-**score** tells you the number of standard deviations a point is from the mean. To calculate a z-score for normally distributed data (normal distributions) we use the

$$z = \frac{x - \mu}{\sigma}$$

where x is the data point, μ is the mean, and σ is the standard deviation.

Frequency Tables & Graphs

A data set having a relatively small number of distinct values can be conveniently presented in a frequency table. For instance, Table 1 is a frequency table for a data set consisting of the starting yearly salaries (to the nearest thousand dollars) of 42 recently graduated students with B.S. degrees in computer science. Table 1 tells us, among other things, that the lowest starting salary of \$47,000 was received by four of the graduates, whereas the highest salary of \$60,000 was received by a single student. The most common starting salary was \$52,000, and was received by 10 of the students.

Table 1 *Starting Yearly Salaries*

Starting Salary	Frequency
47	4
48	1
49	3
50	5
51	8
52	10
53	0
54	5
56	2
57	3
60	1

Data from a frequency table can be graphically represented by a line graph that plots the distinct data values on the horizontal axis and indicates their frequencies by the heights of vertical lines. A line graph of the data presented in Table 1 is shown in Figure 1.

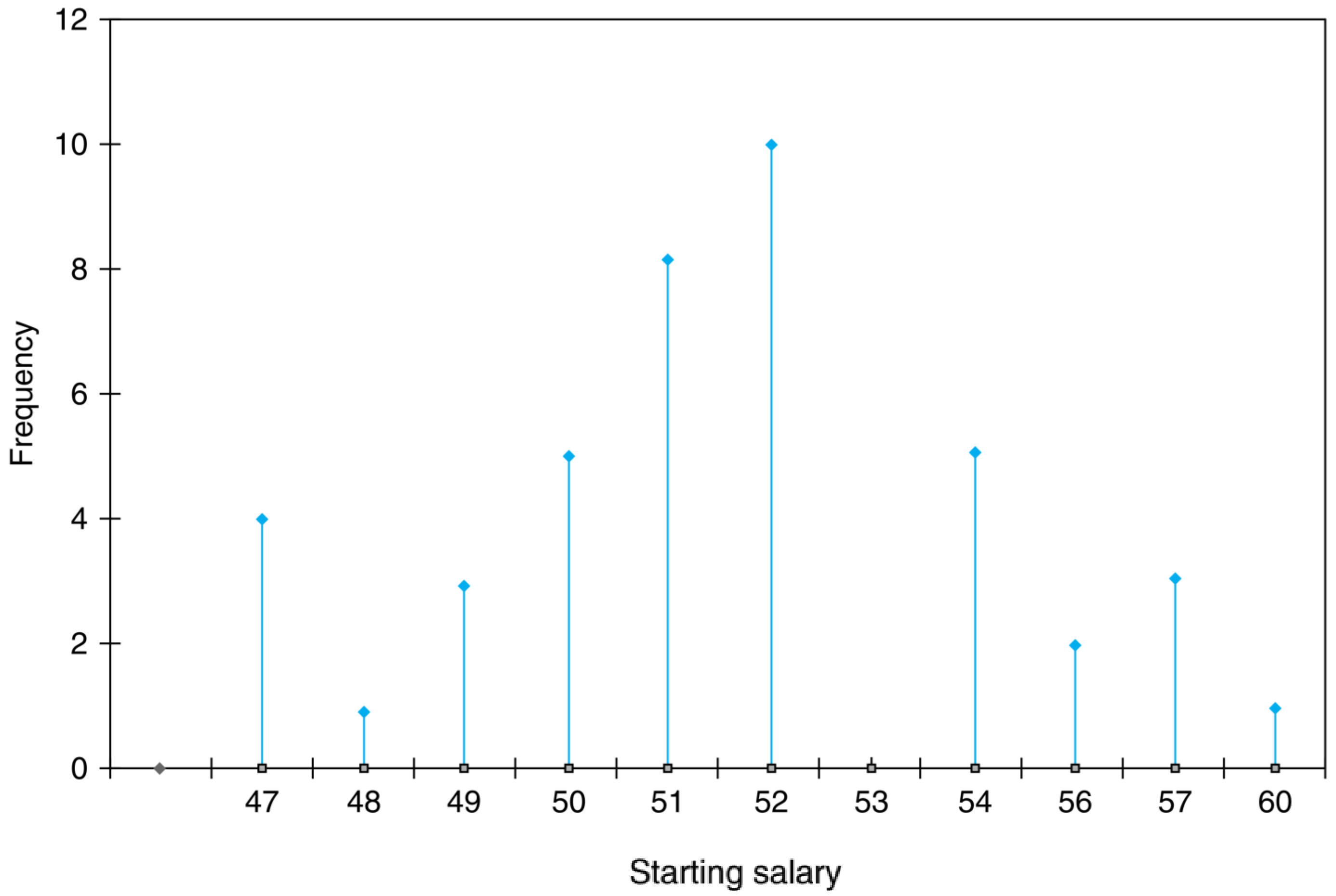


Figure 1

When the lines in a line graph are given added thickness, the graph is called a bar graph. Figure 2 presents a bar graph.

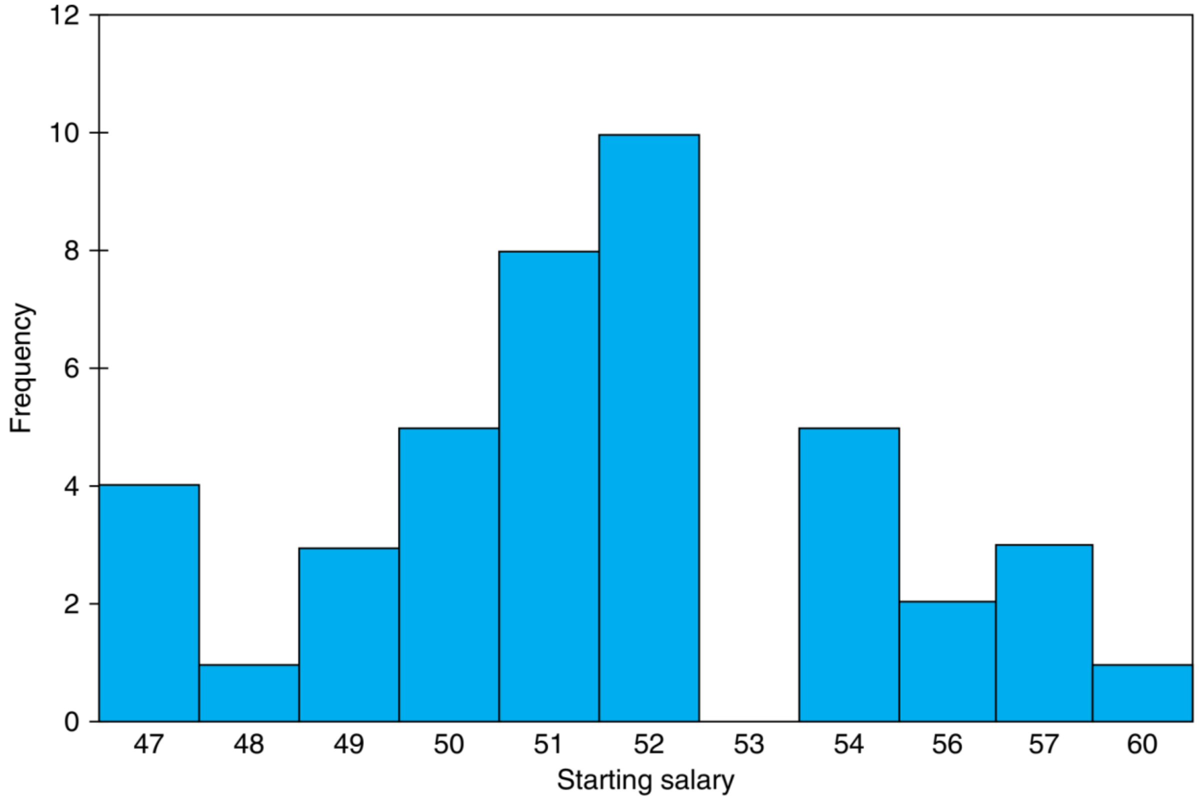


Figure 2 *Bar graph for starting salary data.*

Another type of graph used to represent a frequency table is the frequency polygon, which plots the frequencies of the different data values on the vertical axis, and then connects the plotted points with straight lines. Figure 3 presents a frequency polygon for the data of Table 1.

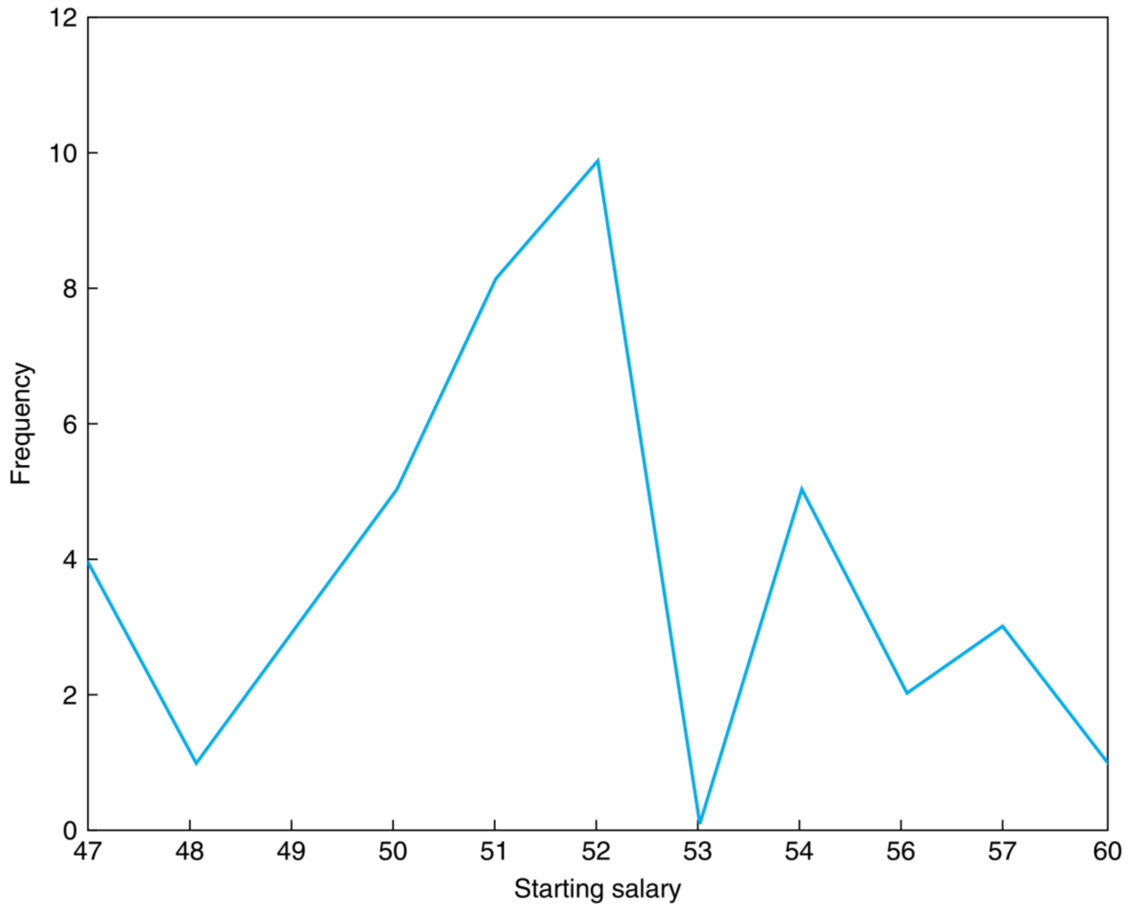


Figure 3

Frequency polygon for starting salary data.

Relative Frequency Tables & Graphs

Consider a data set consisting of n values. If f is the frequency of a particular value, then the ratio f/n is called its relative frequency. That is, the relative frequency of a data value is the proportion of the data that have that value. The relative frequencies can be represented graphically by a relative frequency line or bar graph or by a relative frequency polygon. Indeed, these relative frequency graphs will look like the corresponding graphs of the absolute frequencies except that the labels on the vertical axis are now the old labels (that gave the frequencies) divided by the total number of data points.

Table 4 is a relative frequency table for the data of Table 1. The relative frequencies are obtained by dividing the corresponding frequencies of Table 1 by 42, the size of the data set.

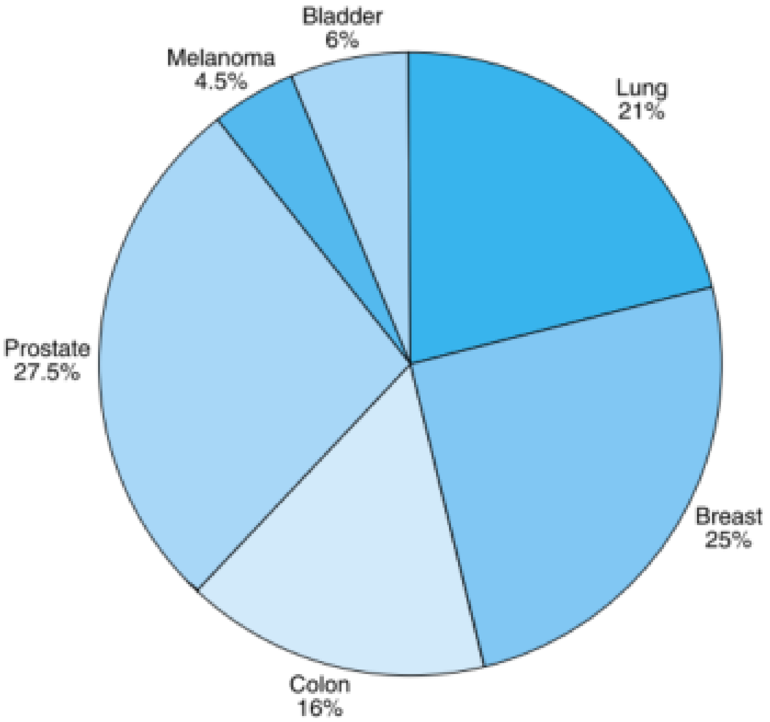
Starting Salary	Frequency
47	$4/42 = .0952$
48	$1/42 = .0238$
49	3/42
50	5/42
51	8/42
52	10/42
53	0
54	5/42
56	2/42
57	3/42
60	1/42

Figure 4

Pie Chart

We can construct pie chart by dividing a circle into various sections or slices. It should be used when we want to compare individual categories with the whole. If you want to compare the values of categories with each other, a bar chart may be more useful.

Type of Cancer	Number of New Cases	Relative Frequency
Lung	42	.21
Breast	50	.25
Colon	32	.16
Prostate	55	.275
Melanoma	9	.045
Bladder	12	.06



Problem

The following table shows the yearly budget of a family

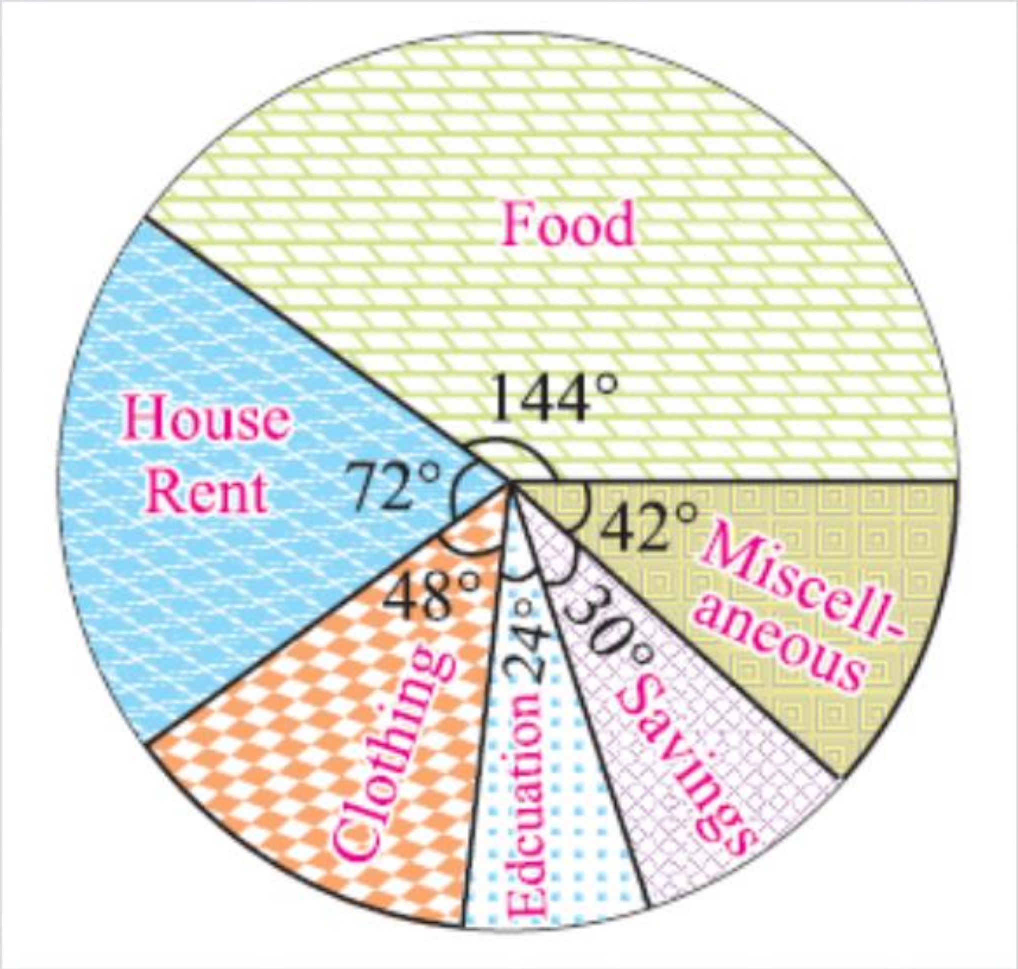
Particulars	Food	House Rent	Clothing	Education	Savings	Miscella-neous
Expenses (in $)	4800	2400	1600	800	1000	1400

Draw a pie chart to represent the above information.

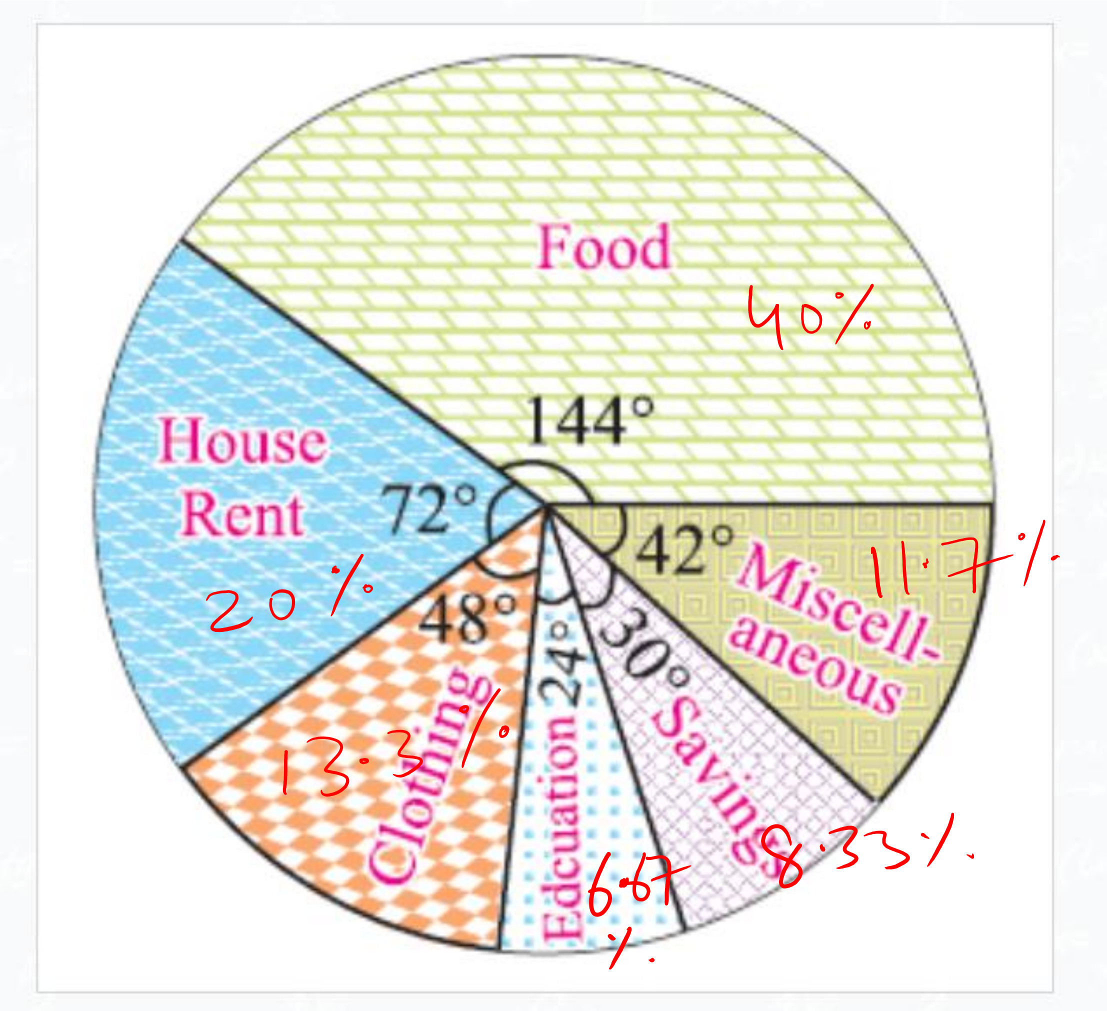
Solution

Particulars	Expenses ($)	Central angle
Food	4800	$\frac{4800}{12000} \times 360^\circ = 144^\circ$
House rent	2400	$\frac{2400}{12000} \times 360^\circ = 72^\circ$
Clothing	1600	$\frac{1600}{12000} \times 360^\circ = 48^\circ$
Education	800	$\frac{800}{12000} \times 360^\circ = 24^\circ$
Savings	1000	$\frac{1000}{12000} \times 360^\circ = 30^\circ$
Miscellaneous	1400	$\frac{1400}{12000} \times 360^\circ = 42^\circ$
Total	**12000**	**360°**

From the table, we obtain the required pie chart as shown below.



Particulars	Expenses ($)	%
Food	4800	$\frac{4800}{12000} \times 100 = 40$
House rent	2400	$\frac{2400}{12000} \times 100 = 20$
Clothing	1600	$\frac{1600}{12000} \times 100 = 13.3$
Education	800	6.67
Savings	1000	8.33
Miscellaneous	1400	11.7
Total	**12000**	100


Food
40%
144°
House
Rent
72°
20%
42°
Miscell-
aneous
11.7%
48°
24°
30°
Clothing
13.3%
Edcuation
6.67
%
Savings
8.33%

Graphical Methods for Describing Quantitative Data

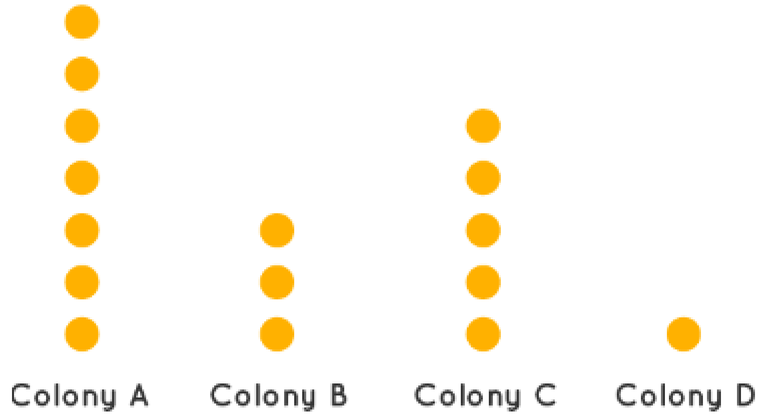
Quantitative data sets consist of data that are recorded on a meaningful numerical scale

To describe, summarize, and detect patterns in such data, we can use three graphical methods: dot plots, stem-and-leaf displays, and histograms

Dot Plot

A dot plot is used to represent any data in the form of dots or small circles.
It is similar to a simplified histogram or a bar graph as the height of the bar formed with dots represents the numerical value of each variable.
Dot plots are used to represent small amounts of data.

Colony	A	B	C	D
Number of babies vaccinated	7	3	5	1



Histogram

Dividing each class frequency by the total number of observations, we obtain the proportion of the set of observations in each of the classes. A table listing relative frequencies is called a **relative frequency distribution**

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

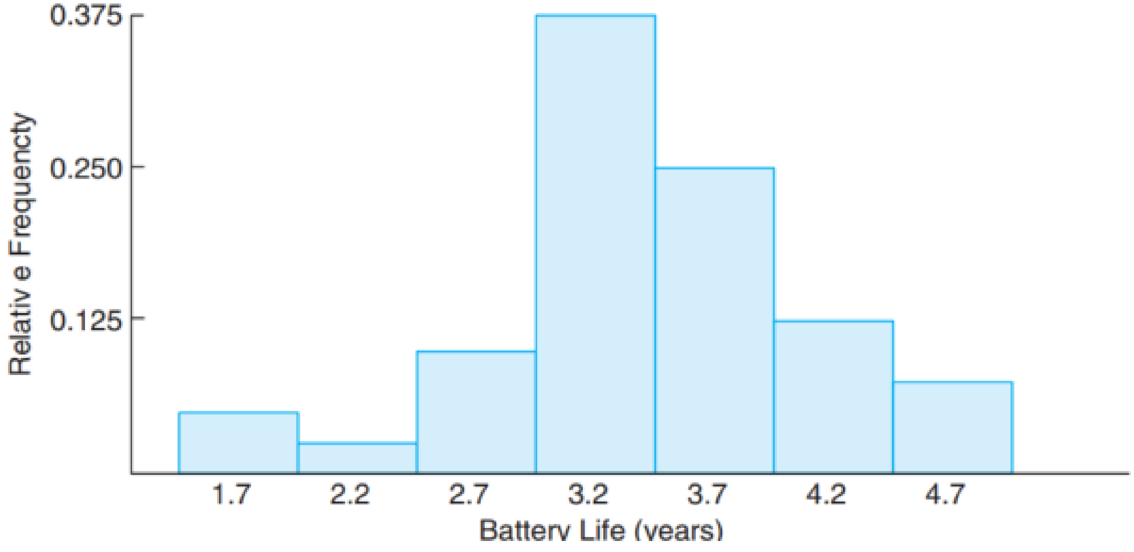


Figure Relative frequency histogram.

Stem and leaf Plot

An efficient way of organizing a small- to moderate-sized data set is to utilize a stem and leaf plot. Such a plot is obtained by first dividing each data value into two parts —its stem and its leaf. For instance, if the data are all two-digit numbers, then we could let the stem part of a data value be its tens digit and let the leaf be its ones digit. Thus, for instance, the value 62 is expressed as

Stem	Leaf
6	2

and the two data values 62 and 67 can be represented as

Stem	Leaf
6	2, 7

Example

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

Example

consider the following humidity readings rounded to the nearest percent:

29	44	12	53	21	34	39	25	48	23
17	24	27	32	34	15	42	21	28	37

Proceeding as in Section 2.2, we might group these data into the following distribution:

Humidity Readings	Frequency
10–19	3
20–29	8
30–39	5
40–49	3
50–59	1

If we wanted to avoid the loss of information inherent in the preceding table, we could keep track of the last digits of the readings within each class, getting

10–19	2 7 5
20–29	9 1 5 3 4 7 1 8
30–39	4 9 2 4 7
40–49	4 8 2
50–59	3

This can also be written as

1	2 7 5		1	2 5 7
2	9 1 5 3 4 7 1 8		2	1 1 3 4 5 7 8 9
3	4 9 2 4 7	or	3	2 4 4 7 9
4	4 8 2		4	2 4 8
5	3		5	3

Probability , Chapter 3

Probability

The term probability refers to the study of randomness and uncertainty. Probability theory is a branch of mathematics that has been developed to deal with uncertainty. The theory of probability has been developed as a scientific tool dealing with chance.

Experiment

An experiment is any process that produces an observation or outcome

Sample Space

A set of all possible outcomes of an experiment is called a sample space and it is denoted by **S**.

Event

Any subset of a sample space S of a random experiment is called an event.
Events will be denoted by the capital letters A, B, and C.

Simple & Compound Events

An event is simple if it consists of exactly one outcome and compound if

it consists of more than one outcome

Occurrence of an Event

When an experiment is performed, a particular event A is said to occur if the resulting experimental outcome is contained in A.

MUTUALLY EXCLUSIVE EVENTS or Disjoint

Two events A and B of a single experiment are said to be mutually exclusive or disjoint if and only if they cannot both occur at the same time i.e. they have no points in common

$$A \cap B = \varnothing$$

Example

When we toss a coin, we get either a head or a tail, but not both at the same time. The two events head and tail are therefore mutually exclusive

EXHAUSTIVE EVENTS

Events are said to be collectively exhaustive when the union of mutually exclusive events is equal to the entire sample space S

Example

- In the coin-tossing experiment, ‘head’ and ‘tail’ are collectively exhaustive events
- In the die-tossing experiment, ‘even number’ and ‘odd number’ are collectively exhaustive events

PARTITION OF THE SAMPLE SPACE

A group of mutually exclusive and exhaustive events belonging to a sample space is called a partition of the sample space. With reference to any sample space S, events A and $\bar{A}$ form a partition as they are mutually exclusive and their union is the entire sample space

RANDOM EXPERIMENT

An experiment that produces different results even though it is repeated a large number of times under essentially similar conditions is called a Random experiment.

Example

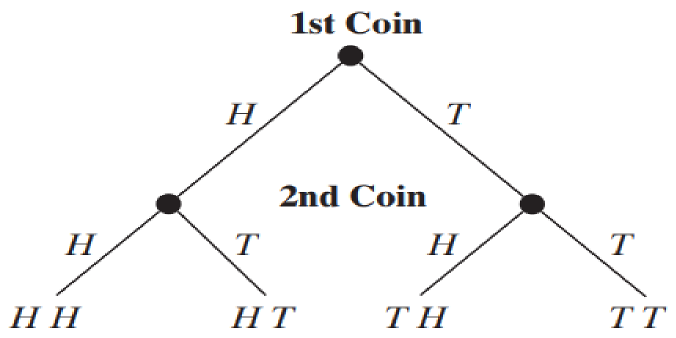
The tossing of a fair coin, the throwing of a balanced die, drawing of a card from a well-shuffled deck of 52 playing cards are examples of random experiments.

Example 1

The experiment of tossing a coin results in either of the two possible outcomes: a head (H) or a tail (T). The sample space for this experiment may be expressed in set notation as **S = {H, T}.** ‘H’ and ‘T’ are the two sample points.

Example 2

The sample space for tossing two coins once (or tossing a coin twice) will contain four possible outcomes denoted by S = {HH, HT, TH, TT}



Example 3

The sample space S for the random experiment of throwing two six-side dice can be described by the Cartesian product $A \times A$, where

$A = \{1, 2, 3, 4, 5,6\}$. In other words, $S = A \times A = \{(x, y) \mid x \in A \text{ and } y \in A\}$, Where x denotes the number of dots on the upper face of the first die, and y denotes the number of dots on the upper face of the second die. Hence, S contains 36 outcomes or sample points as shown below

FIGURE

Sample space for rolling two dice

$\mathcal{S}$

(1, 1)	(1, 2)	(1, 3)	(1, 4)	(1, 5)	(1, 6)
(2, 1)	(2, 2)	(2, 3)	(2, 4)	(2, 5)	(2, 6)
(3, 1)	(3, 2)	(3, 3)	(3, 4)	(3, 5)	(3, 6)
(4, 1)	(4, 2)	(4, 3)	(4, 4)	(4, 5)	(4, 6)
(5, 1)	(5, 2)	(5, 3)	(5, 4)	(5, 5)	(5, 6)
(6, 1)	(6, 2)	(6, 3)	(6, 4)	(6, 5)	(6, 6)

$\mathcal{S}$

A♡	2♡	3♡	4♡	5♡	6♡	7♡	8♡	9♡	10♡	J♡	Q♡	K♡
A♣	2♣	3♣	4♣	5♣	6♣	7♣	8♣	9♣	10♣	J♣	Q♣	K♣
A♢	2♢	3♢	4♢	5♢	6♢	7♢	8♢	9♢	10♢	J♢	Q♢	K♢
A♠	2♠	3♠	4♠	5♠	6♠	7♠	8♠	9♠	10♠	J♠	Q♠	K♠

FIGURE

Sample space for choosing one card

$\mathcal{S}$

$(A\heartsuit, A\heartsuit)$	$(A\heartsuit, 2\heartsuit)$	$(A\heartsuit, 3\heartsuit)$	$\cdots$	$(A\heartsuit, Q\spadesuit)$	$(A\heartsuit, K\spadesuit)$
$(2\heartsuit, A\heartsuit)$	$(2\heartsuit, 2\heartsuit)$	$(2\heartsuit, 3\heartsuit)$	$\cdots$	$(2\heartsuit, Q\spadesuit)$	$(2\heartsuit, K\spadesuit)$
$(3\heartsuit, A\heartsuit)$	$(3\heartsuit, 2\heartsuit)$	$(3\heartsuit, 3\heartsuit)$	$\cdots$	$(3\heartsuit, Q\spadesuit)$	$(3\heartsuit, K\spadesuit)$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
$(Q\spadesuit, A\heartsuit)$	$(Q\spadesuit, 2\heartsuit)$	$(Q\spadesuit, 3\heartsuit)$	$\cdots$	$(Q\spadesuit, Q\spadesuit)$	$(Q\spadesuit, K\spadesuit)$
$(K\spadesuit, A\heartsuit)$	$(K\spadesuit, 2\heartsuit)$	$(K\spadesuit, 3\heartsuit)$	$\cdots$	$(K\spadesuit, Q\spadesuit)$	$(K\spadesuit, K\spadesuit)$

FIGURE

Sample space for choosing two cards with replacement

$52 \times 52 = 2704$ elements of the sample space.

FIGURE

Sample space for choosing two cards without replacement

$\mathcal{S}$

	$(A\heartsuit, 2\heartsuit)$	$(A\heartsuit, 3\heartsuit)$	$\cdots$	$(A\heartsuit, Q\spadesuit)$	$(A\heartsuit, K\spadesuit)$
$(2\heartsuit, A\heartsuit)$		$(2\heartsuit, 3\heartsuit)$	$\cdots$	$(2\heartsuit, Q\spadesuit)$	$(2\heartsuit, K\spadesuit)$
$(3\heartsuit, A\heartsuit)$	$(3\heartsuit, 2\heartsuit)$		$\cdots$	$(3\heartsuit, Q\spadesuit)$	$(3\heartsuit, K\spadesuit)$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
$(Q\spadesuit, A\heartsuit)$	$(Q\spadesuit, 2\heartsuit)$	$(Q\spadesuit, 3\heartsuit)$	$\cdots$		$(Q\spadesuit, K\spadesuit)$
$(K\spadesuit, A\heartsuit)$	$(K\spadesuit, 2\heartsuit)$	$(K\spadesuit, 3\heartsuit)$	$\cdots$	$(K\spadesuit, Q\spadesuit)$	

If two cards are drawn *without replacement*, so that the second card is drawn from a reduced pack of 51 cards, then the sample space will be a subset of that above, as shown in Figure Specifically, events such as $(A\heartsuit, A\heartsuit)$, where a particular card is drawn twice, will not be in the sample space. The total number of elements in this new sample space will therefore be $2704 - 52 = 2652$.

Counting Techniques

The fundamental principle of counting is often referred to as the multiplication rule is defined as

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

Example 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.

The **generalized multiplication rule** covering *k* operations is stated in the following

> 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 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

$$n_l \times n_2 \times n_3 \times n_4 = 2 \times 4 \times 3 \times 5 = 120$$

different ways to order the parts.

Example 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

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

even four-digit numbers.

Since the above two cases are mutually exclusive, the total number of even four-digit numbers can be calculated as $60 + 96 = 156$. $\blacksquare$

Permutation

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

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

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

Consider the three letters *a*, *b*, and *c*. The possible permutations are

abc, *acb*, *bac*, *bca*, *cab*, and *cba*.

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

In general, n distinct objects can be arranged in

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

Factorial

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

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

with special case $0! = 1$.

The number of permutations of n objects is $n!$.

The number of permutations of the four letters a, b, c, and d will be $4! = 24$.

Theorem

The number of permutations of n objects is $n!$.

Example

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

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

Example

The number of ways to assemble chips in a controller

An electronic controlling mechanism requires 5 distinct, but interchangeable, memory chips. In how many ways can this mechanism be assembled

(a) by placing the 5 chips in the 5 positions within the controller?

(b) by placing 3 chips in the odd numbered positions within the controller?

Solution

(a) When all 5 chips must be placed, the answer is 5!. Alternatively, in the permutation notation with $n = 5$ and $r = 5$, the first formula yields

$$_5P_5 = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$$

and the second formula yields

$$_5P_5 = \frac{5!}{(5-5)!} = \frac{5!}{0!} = 5! = 120$$

(b) For $n = 5$ chips placed in $r = 3$ positions, the permutation is

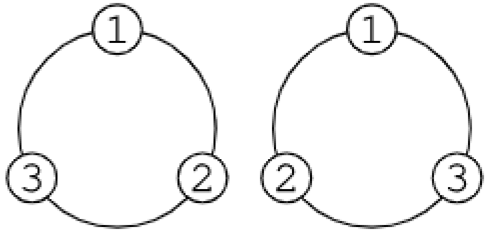
$$_5P_3 = \frac{5!}{2!} = \frac{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{2 \cdot 1} = 5 \cdot 4 \cdot 3 = 60$$

Circular Permutation

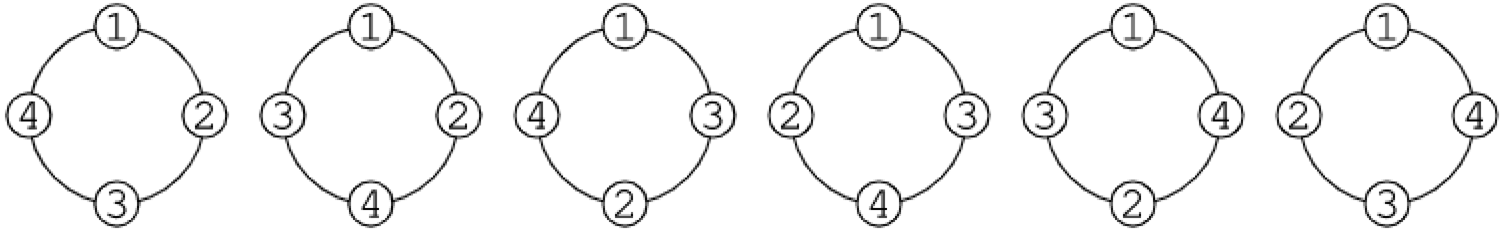
Permutations that occur by arranging objects in a circle are called **circular permutations**

Theorem

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



$(3-1)! = 2$ distinct circular permutations are $\{1, 2, 3\}$ and $\{1, 3, 2\}$.



$(4-1)! = 6$ distinct circular permutations are

$\{1, 2, 3, 4\}, \{1, 2, 4, 3\}, \{1, 3, 2, 4\}, \{1, 3, 4, 2\}, \{1, 4, 2, 3\}$, and $\{1, 4, 3, 2\}$

Theorem

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

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

Partitions Rule

Theorem

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

Solution The total number of possible partitions would be

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

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

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

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

Combination

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

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

Notice that

$$C_1^n = \binom{n}{1} = \frac{n!}{(n-1)!\,1!} = n$$

and

$$C_2^n = \binom{n}{2} = \frac{n!}{(n-2)!\,2!} = \frac{n(n-1)}{2}$$

$$C_{n-1}^n = \binom{n}{n-1} = \frac{n!}{1!\,(n-1)!} = n$$

and

$$C_n^n = \binom{n}{n} = \frac{n!}{0!\,n!} = 1$$

Example

Evaluating a combination

In how many different ways can 3 of 18 automotive engineers be chosen for a team to develop a new ceramic diesel engine.

Solution

For $n = 18$ and $r = 3$, the first formula for $\binom{n}{r}$ yields

$$\binom{18}{3} = \frac{18 \cdot 17 \cdot 16}{3!} = 816 \qquad \blacksquare$$

Example

Selection of machines for an experiment

A calibration study needs to be conducted to see if the readings on 15 test machines are giving similar results. In how many ways can 3 of the 15 be selected for the initial investigation?

Solution

$$\binom{15}{3} = \frac{15 \cdot 14 \cdot 13}{3 \cdot 2 \cdot 1} = 455 \text{ ways}$$

Note that selecting which 3 machines to use is the same as selecting which 12 not to include. That is, according to the second formula,

$$\binom{15}{12} = \frac{15!}{12!\,3!} = \frac{15!}{3!\,12!} = \binom{15}{3} \qquad \blacksquare$$

Example

The number of choices of new researchers

In how many different ways can the director of a research laboratory choose 2 chemists from among 7 applicants and 3 physicists from among 9 applicants?

Solution

The 2 chemists can be chosen in $\binom{7}{2} = 21$ ways and the 3 physicists can be chosen in $\binom{9}{3} = 84$ ways. By the multiplication rule, the whole selection can be made in $21 \cdot 84 = 1{,}764$ ways. ■

Example

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

Solution

The number of ways of selecting 3 cartridges from 10 is

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

The number of ways of selecting 2 cartridges from 5 is

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

Using the multiplication rule with $n_1 = 120$ and $n_2 = 10$,

$(120)(10) = 1200$ ways.

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

Solution

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

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.

	Repeats allowed	**No Repeats**
Permutations (order matters):	n^r	$\frac{n!}{(n-r)!}$
Combinations (order doesn't matter):	$\frac{(r+n-1)!}{r!(n-1)!}$	$\frac{n!}{r!(n-r)!}$

Summary of Counting Rules

1. *Multiplicative rule.* If you are drawing *one element from each of k sets of elements*, where the sizes of the sets are $n_1, n_2, \cdot \ldots \cdot, n_k$, then the number of different results is

$$n_1 n_2 n_3 \ldots n_k$$

2. *Permutations rule.* If you are drawing *n elements from a set of N elements and arranging the n elements in a distinct order*, then the number of different results is

$$P_n^N = \frac{N!}{(N-n)!}$$

3. *Partitions rule.* If you are partitioning the *elements of a set of N elements into k groups consisting* of $n_1, n_2, \ldots, n_k$ elements $(n_1 + \cdots + n_k = N)$, then the number of different results is

4. *Combinations rule.* If you are drawing *n elements from a set of N elements without regard to the order of the n elements*, then the number of different results is

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

[*Note:* The combinations rule is a special case of the partitions rule when $k = 2$.]

Definition

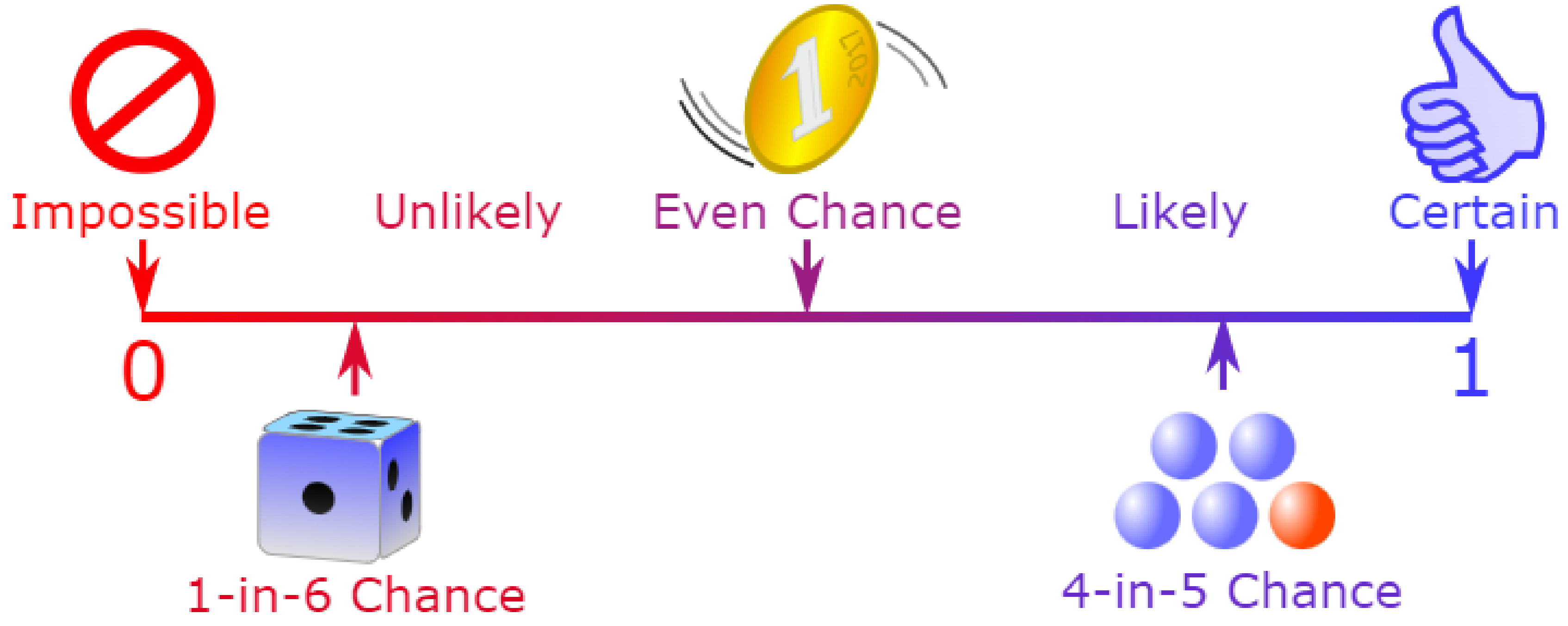
The **probability** of an event A is the sum of the weights of all sample points in A. Therefore,

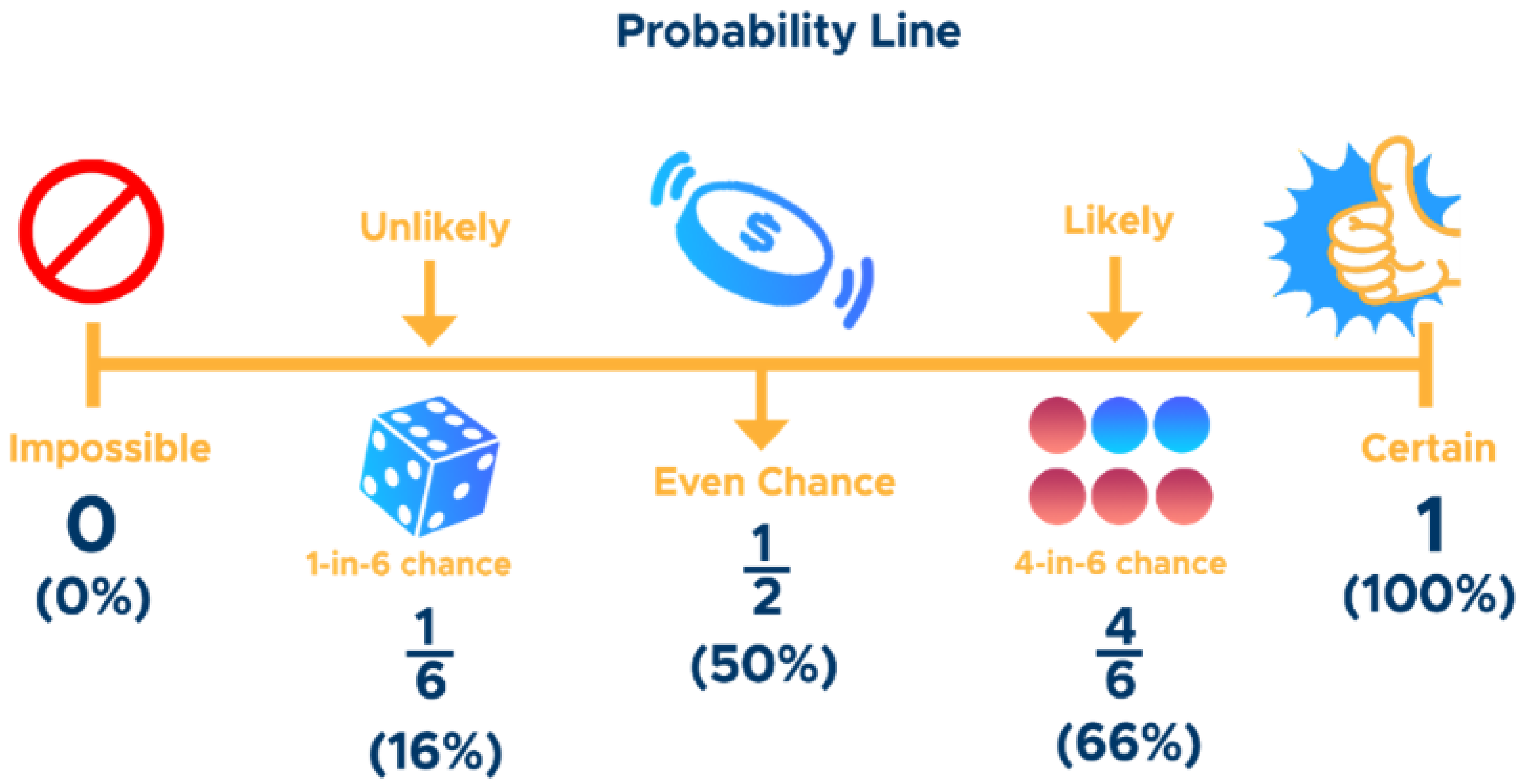
$$0 \leq P(A) \leq 1, \quad P(\phi) = 0, \quad \text{and} \quad P(S) = 1.$$

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

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

As a number, probability is from 0 (impossible) to 1 (certain).




Probability Line
Unlikely
Likely
Impossible
0
(0%)
1-in-6 chance
1
6
(16%)
Even Chance
1
2
(50%)
4-in-6 chance
4
6
(66%)
Certain
1
(100%)

Example

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

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

Example

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,

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

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

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

Probability values for rolling two dice

$\mathcal{S}$

(1, 1)	(1, 2)	(1, 3)	(1, 4)	(1, 5)	(1, 6)
1/36	1/36	1/36	1/36	1/36	1/36
(2, 1)	(2, 2)	(2, 3)	(2, 4)	(2, 5)	(2, 6)
1/36	1/36	1/36	1/36	1/36	1/36
(3, 1)	(3, 2)	(3, 3)	(3, 4)	(3, 5)	(3, 6)
1/36	1/36	1/36	1/36	1/36	1/36
(4, 1)	(4, 2)	(4, 3)	(4, 4)	(4, 5)	(4, 6)
1/36	1/36	1/36	1/36	1/36	1/36
(5, 1)	(5, 2)	(5, 3)	(5, 4)	(5, 5)	(5, 6)
1/36	1/36	1/36	1/36	1/36	1/36
(6, 1)	(6, 2)	(6, 3)	(6, 4)	(6, 5)	(6, 6)
1/36	1/36	1/36	1/36	1/36	1/36

$\mathcal{S}$

Head	Tail
0.6	0.4

FIGURE

Probability values for a biased coin

$\mathcal{S}$

1	2	3	4	5	6
1/6	1/6	1/6	1/6	1/6	1/6

FIGURE

Probability values for a fair die

$\mathcal{S}$

1	2	3	4	5	6
0.10	0.15	0.15	0.15	0.15	0.30

FIGURE

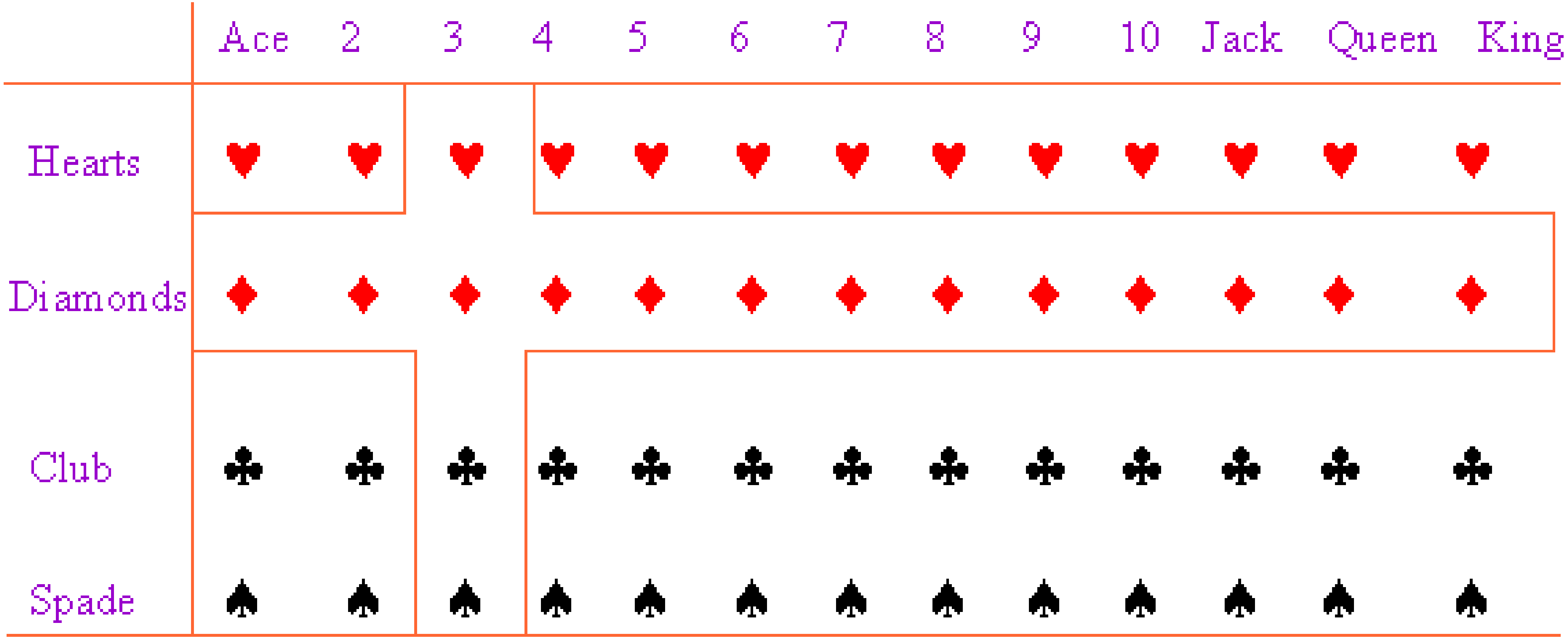
Probability values for a biased die

$\mathcal{S}$

A♡	2♡	3♡	4♡	5♡	6♡	7♡	8♡	9♡	10♡	J♡	Q♡	K♡
1/52	1/52	1/52	1/52	1/52	1/52	1/52	1/52	1/52	1/52	1/52	1/52	1/52
A♣	2♣	3♣	4♣	5♣	6♣	7♣	8♣	9♣	10♣	J♣	Q♣	K♣
1/52	1/52	1/52	1/52	1/52	1/52	1/52	1/52	1/52	1/52	1/52	1/52	1/52
A◇	2◇	3◇	4◇	5◇	6◇	7◇	8◇	9◇	10◇	J◇	Q◇	K◇
1/52	1/52	1/52	1/52	1/52	1/52	1/52	1/52	1/52	1/52	1/52	1/52	1/52
A♠	2♠	3♠	4♠	5♠	6♠	7♠	8♠	9♠	10♠	J♠	Q♠	K♠
1/52	1/52	1/52	1/52	1/52	1/52	1/52	1/52	1/52	1/52	1/52	1/52	1/52

FIGURE

Probability values for choosing one card


Ace 2 3 4 5 6 7 8 9 10 Jack Queen King
Hearts
Diamonds
Club
Spade

$\mathcal{S}$

(A♡, A♡)	(A♡, 2♡)	(A♡, 3♡)	⋯	(A♡, Q♠)	(A♡, K♠)
1/2704	1/2704	1/2704	⋯	1/2704	1/2704
(2♡, A♡)	(2♡, 2♡)	(2♡, 3♡)	⋯	(2♡, Q♠)	(2♡, K♠)
1/2704	1/2704	1/2704	⋯	1/2704	1/2704
(3♡, A♡)	(3♡, 2♡)	(3♡, 3♡)	⋯	(3♡, Q♠)	(3♡, K♠)
1/2704	1/2704	1/2704	⋯	1/2704	1/2704
⋮	⋮	⋮		⋮	⋮
(Q♠, A♡)	(Q♠, 2♡)	(Q♠, 3♡)	⋯	(Q♠, Q♠)	(Q♠, K♠)
1/2704	1/2704	1/2704	⋯	1/2704	1/2704
(K♠, A♡)	(K♠, 2♡)	(K♠, 3♡)	⋯	(K♠, Q♠)	(K♠, K♠)
1/2704	1/2704	1/2704	⋯	1/2704	1/2704

FIGURE

Probability values for choosing two cards with replacement

$\mathcal{S}$

	(A$\heartsuit$, 2$\heartsuit$) 1/2652	(A$\heartsuit$, 3$\heartsuit$) 1/2652	$\cdots$	(A$\heartsuit$, Q$\spadesuit$) 1/2652	(A$\heartsuit$, K$\spadesuit$) 1/2652
(2$\heartsuit$, A$\heartsuit$) 1/2652		(2$\heartsuit$, 3$\heartsuit$) 1/2652	$\cdots$	(2$\heartsuit$, Q$\spadesuit$) 1/2652	(2$\heartsuit$, K$\spadesuit$) 1/2652
(3$\heartsuit$, A$\heartsuit$) 1/2652	(3$\heartsuit$, 2$\heartsuit$) 1/2652		$\cdots$	(3$\heartsuit$, Q$\spadesuit$) 1/2652	(3$\heartsuit$, K$\spadesuit$) 1/2652
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
(Q$\spadesuit$, A$\heartsuit$) 1/2652	(Q$\spadesuit$, 2$\heartsuit$) 1/2652	(Q$\spadesuit$, 3$\heartsuit$) 1/2652	$\cdots$		(Q$\spadesuit$, K$\spadesuit$) 1/2652
(K$\spadesuit$, A$\heartsuit$) 1/2652	(K$\spadesuit$, 2$\heartsuit$) 1/2652	(K$\spadesuit$, 3$\heartsuit$) 1/2652	$\cdots$	(K$\spadesuit$, Q$\spadesuit$) 1/2652	

FIGURE

Probability values for choosing two cards without replacement

Example In a poker hand consisting of 5 cards, find the probability of holding 2 aces and 3 jacks.

Solution The number of ways of being dealt 2 aces from 4 cards is

$$\binom{4}{2} = \frac{4!}{2!\ 2!} = 6,$$

and the number of ways of being dealt 3 jacks from 4 cards is

$$\binom{4}{3} = \frac{4!}{3!\ 1!} = 4.$$

By the multiplication rule (Rule 2.1), there are $n = (6)(4) = 24$ hands with 2 aces and 3 jacks. The total number of 5-card poker hands, all of which are equally likely, is

$$N = \binom{52}{5} = \frac{52!}{5!\ 47!} = 2{,}598{,}960.$$

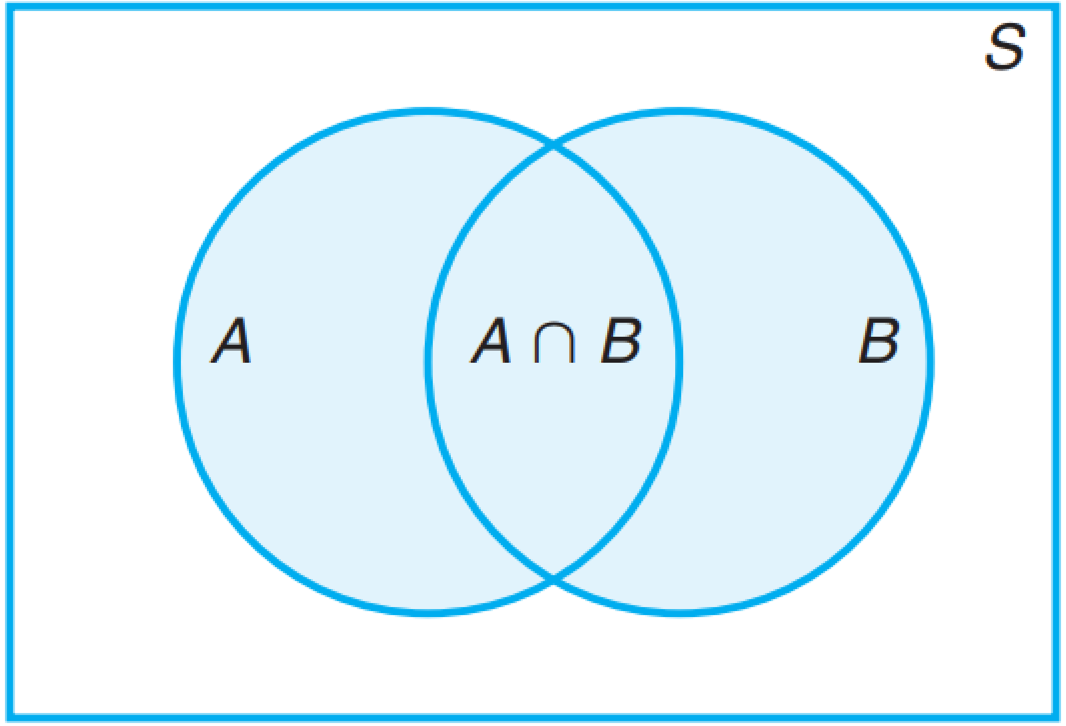
Therefore, the probability of getting 2 aces and 3 jacks in a 5-card poker hand is

$$P(C) = \frac{24}{2,598,960} = 0.9 \times 10^{-5}.$$

Additive Rules

If A and B are two events, then

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



If A and B are mutually exclusive, then

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

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

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

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

Additive Rule for Three Events

For three events A, B, and C,

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

Example

John is going to graduate from an industrial engineering department in a university by the end of the semester. After being interviewed at two companies he likes, he assesses that his probability of getting an offer from company A is 0.8, and his probability of getting an offer from company B is 0.6. If he believes that the probability that he will get offers from both companies is 0.5, what is the probability that he will get at least one offer from these two companies?

Solution

Using the additive rule, we have

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.8 + 0.6 - 0.5 = 0.9.$$

Example

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

Solution

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

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

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

Problem Hospital records show that 12% of all patients are admitted for surgical treatment, 16% are admitted for obstetrics, and 2% receive both obstetrics and surgical treatment. If a new patient is admitted to the hospital, what is the probability that the patient will be admitted for surgery, for obstetrics, or for both?

Solution Consider the following events:

A: {A patient admitted to the hospital receives surgical treatment.}

B: {A patient admitted to the hospital receives obstetrics treatment.}

Then, from the given information,

$$P(A) = .12$$
$$P(B) = .16$$

and the probability of the event that a patient receives both obstetrics and surgical treatment is

$$P(A \cap B) = .02$$

The event that a patient admitted to the hospital receives either surgical treatment, obstetrics treatment, or both is the union $A \cup B$, the probability of which is given by the additive rule of probability:

$$
\begin{aligned}
P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\
&= .12 + .16 - .02 = .26
\end{aligned}
$$

Thus, 26% of all patients admitted to the hospital receive either surgical treatment, obstetrics treatment, or both.

Example

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

$$\begin{aligned} P(G \cup W \cup R \cup B) &= P(G) + P(W) + P(R) + P(B) \\ &= 0.09 + 0.15 + 0.21 + 0.23 = 0.68. \end{aligned}$$

Theorem

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

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

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

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

Example

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

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

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

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

This also can be solved by using Theorem 2.9:

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

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

Conditional Probability

The probability of an event B occurring when it is known that some event A has occurred is called a **conditional probability** and is denoted by $P(B|A)$. The symbol $P(B|A)$ is usually read "the probability that B occurs given that A occurs" or simply "the probability of B, given A."

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

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

or $$P(A \cap B) \equiv P(A)\, P(B \mid A)$$

Example

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

(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

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

Example

Find the probability that a single toss of a die will result in a number less than 4 if (a) no other information is given and (b) it is given that the toss resulted in an odd number

Solution

(a) Let B denote the event {less than 4}. Since B is the union of the events 1, 2, or 3 turning up

$$P(B) = P(1) + P(2) + P(3) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$$

(b) Letting A be the event {odd number}, we see that $P(A) = \frac{3}{6} = \frac{1}{2}$. Also $P(A \cap B) = \frac{2}{6} = \frac{1}{3}$. Then

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} = \frac{1/3}{1/2} = \frac{2}{3}$$

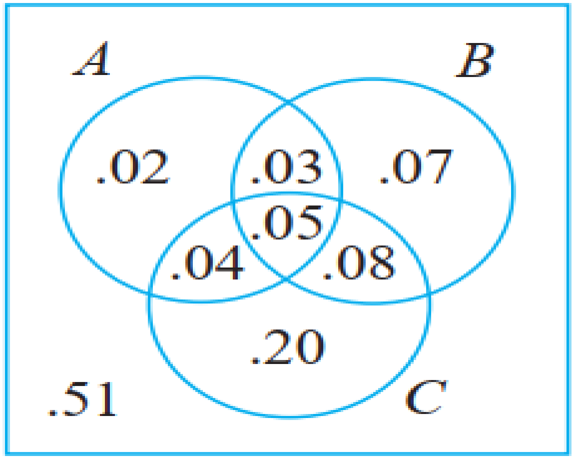
Example

A news magazine publishes three columns entitled "Art" (A), "Books" (B), and "Cinema" (C). Reading habits of a randomly selected reader with respect to these columns are

Read regularly	A	B	C	$A \cap B$	$A \cap C$	$B \cap C$	$A \cap B \cap C$
Probability	.14	.23	.37	.08	.09	.13	.05

We thus have

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{.08}{.23} = .348$$



$$P(A|B \cup C) = \frac{P(A \cap (B \cup C))}{P(B \cup C)} = \frac{.04 + .05 + .03}{.47} = \frac{.12}{.47} = .255$$

$$P(A|\text{reads at least one}) = P(A|A \cup B \cup C) = \frac{P(A \cap (A \cup B \cup C))}{P(A \cup B \cup C)}$$

$$= \frac{P(A)}{P(A \cup B \cup C)} = \frac{.14}{.49} = .286$$

and

$$P(A \cup B|C) = \frac{P((A \cup B) \cap C)}{P(C)} = \frac{.04 + .05 + .08}{.37} = .459 \quad \blacksquare$$

Problem Many medical researchers have conducted experiments to examine the relationship between cigarette smoking and cancer. Consider an individual randomly selected from the adult male population. Let A represent the event that the individual smokes, and let A^c denote the complement of A (the event that the individual does not smoke). Similarly, let B represent the event that the individual develops cancer, and let B^c be the complement of that event. Then the four sample points associated with the experiment are shown in Figure 3.15, and their probabilities for a certain section of the United States are given in Table 3.5. Use these sample point probabilities to examine the relationship between smoking and cancer.

Figure 3.15

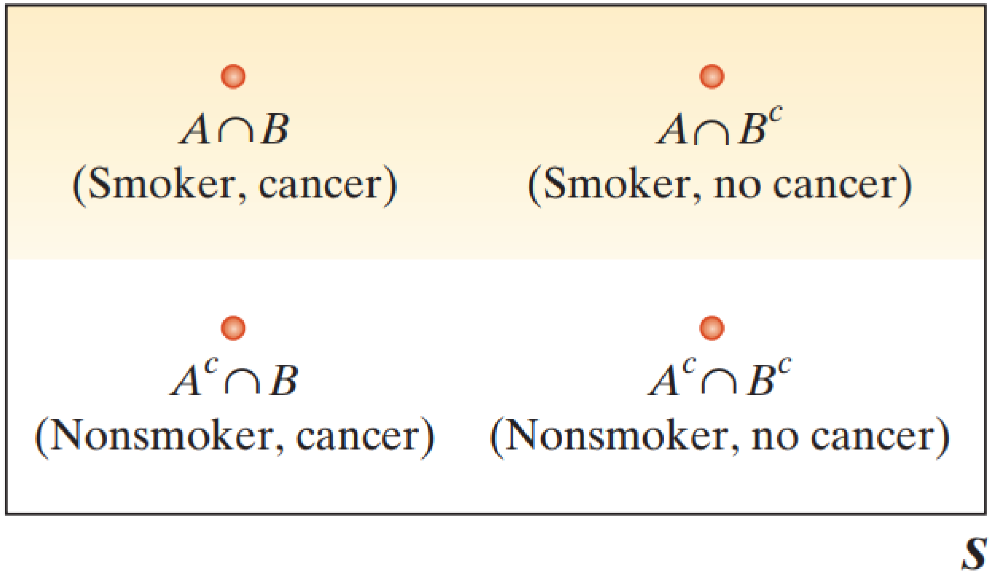


Table 3.5 Probabilities of Smoking and Developing Cancer

	Develops Cancer	
Smoker	Yes, B	No, B^c
Yes, A	.05	.20
No, A^c	.03	.72

Solution One method of determining whether the given probabilities indicate that smoking and cancer are related is to compare the *conditional probability* that an adult male acquires cancer given that he smokes with the conditional probability that an adult male acquires cancer given that he does not smoke [i.e., compare $P(B|A)$ with $P(B|A^c)$].

First, we consider the reduced sample space A corresponding to adult male smokers. This reduced sample space is shaded in Figure 3.15. The two sample points $A \cap B$ and $A \cap B^c$ are contained in this reduced sample space, and the adjusted probabilities of these two sample points are the two conditional probabilities

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \quad \text{and} \quad P(B^c|A) = \frac{P(A \cap B^c)}{P(A)}$$

The probability of event A is the sum of the probabilities of the sample points in A:

$$P(A) = P(A \cap B) + P(A \cap B^c) = .05 + .20 = .25$$

Then the values of the two conditional probabilities in the reduced sample space A are

$$P(B|A) = \frac{.05}{.25} = .20 \quad \text{and} \quad P(B^c|A) = \frac{.20}{.25} = .80$$

These two numbers represent the probabilities that an adult male smoker develops cancer and does not develop cancer, respectively.

In a like manner, the conditional probabilities of an adult male nonsmoker developing cancer and not developing cancer are

$$P(B|A^c) = \frac{P(A^c \cap B)}{P(A^c)} = \frac{.03}{.75} = .04$$

$$P(B^c|A^c) = \frac{P(A^c \cap B^c)}{P(A^c)} = \frac{.72}{.75} = .96$$

Two of the conditional probabilities give some insight into the relationship between cancer and smoking: the probability of developing cancer given that the adult male is a smoker, and the probability of developing cancer given that the adult male is not a smoker. The conditional probability that an adult male smoker develops cancer (.20) is five times the probability that a nonsmoker develops cancer (.04). This relationship does not imply that smoking *causes* cancer, but it does suggest a pronounced link between smoking and cancer.

Independent Events

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**.

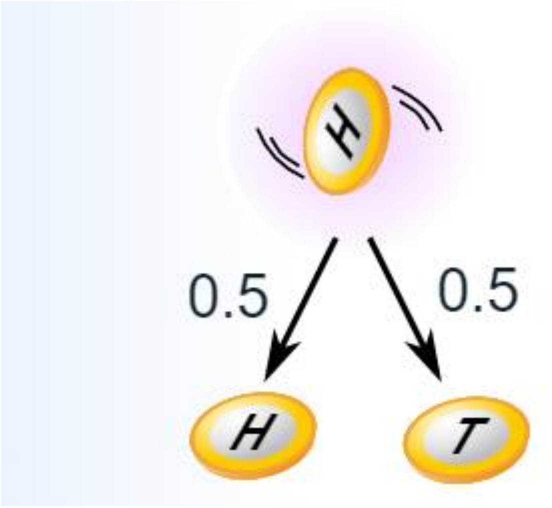
Two events A and B are independent if and only if

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

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

Independent Events

Independent Events are **not affected** by previous events.



Dependent Events

Events can be "dependent" ... which means they **can be affected by previous events** .

- **With** Replacement: the events are **Independent** (the chances don't change)
- **Without** Replacement: the events are **Dependent** (the chances change)

Problem Consider the experiment of tossing a fair die, and let

$$A = \{\text{Observe an even number.}\}$$
$$B = \{\text{Observe a number less than or equal to 4}$$

Are A and B independent events?

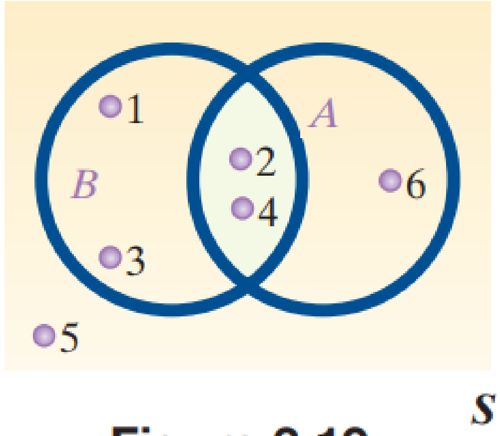


Figure 3.19

Solution The Venn diagram for this experiment is shown in Figure 3.19. We first calculate

$$P(A) = P(2) + P(4) + P(6) = \frac{1}{2}$$

$$P(B) = P(1) + P(2) + P(3) + P(4) = \frac{4}{6} = \frac{2}{3}$$

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

Now, assuming that B has occurred, we see that the conditional probability of A given B is

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{{}^1\!/_3}{{}^2\!/_3} = \frac{1}{2} = P(A)$$

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{{}^1\!/_3}{{}^1\!/_2} = \frac{2}{3} = P(B)$$

Thus, assuming that the occurrence of event B does not alter the probability of observing an even number, that probability remains $\frac{1}{2}$. Therefore, the events A and B are independent.

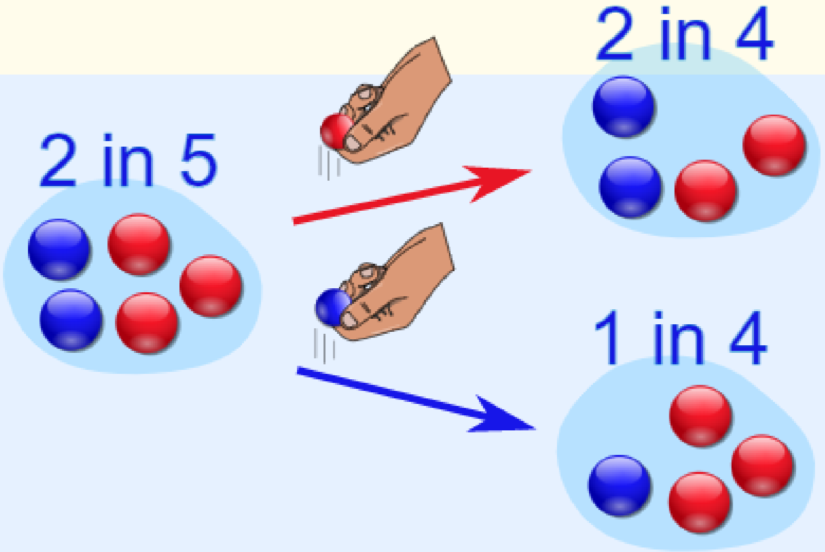
Example: Marbles in a Bag

2 blue and 3 red marbles are in a bag.

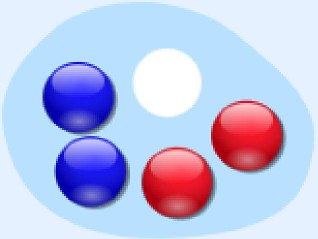
What are the chances of getting a blue marble?

The chance is **2 in 5**

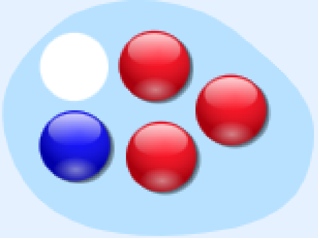
But after taking one out the chances change!



So the next time:



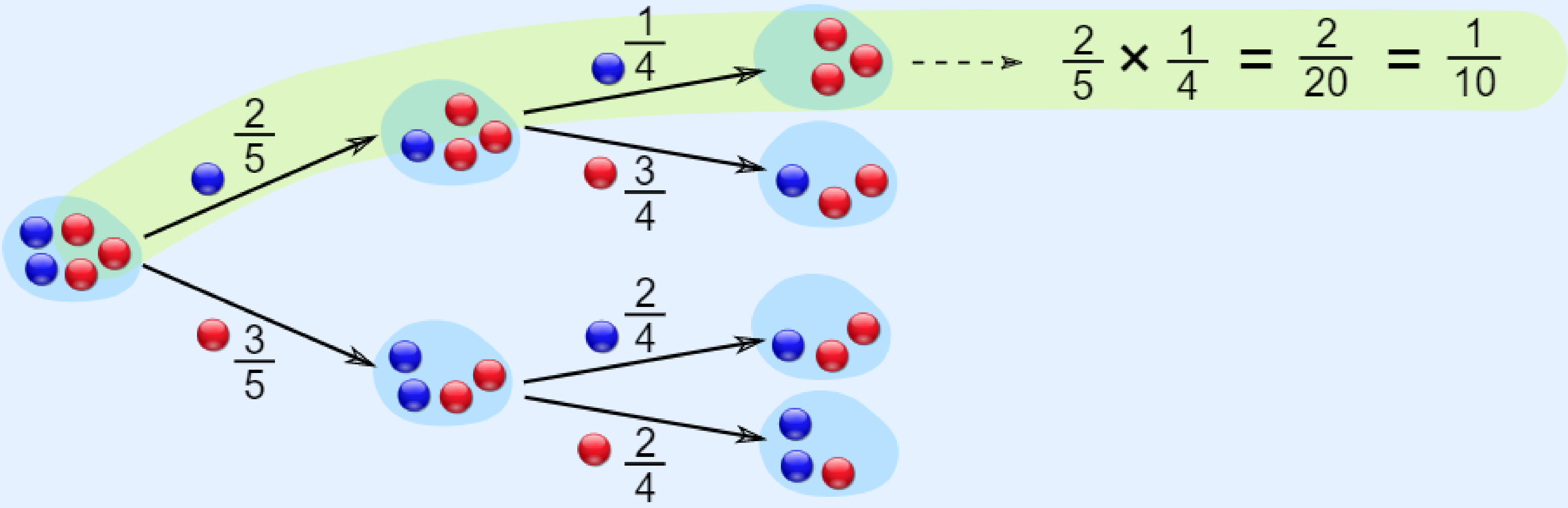
if we got a **red** marble before, then the chance of a blue marble next is **2 in 4**



if we got a **blue** marble before, then the chance of a blue marble next is **1 in 4**

Now we can answer questions like **"What are the chances of drawing 2 blue marbles?"**

Answer: it is a **2/5 chance** followed by a **1/4 chance**:



Did you see how we multiplied the chances? And got 1/10 as a result.

The chances of drawing 2 blue marbles is 1/10

Example

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

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

Example

Two cards are drawn at random from an ordinary deck of 52 playing cards. What is the probability of getting two aces if

(a) the first card is replaced before the second card is drawn;

(b) the first card is not replaced before the second card is drawn?

Solution

(a) Since there are four aces among the 52 cards, we get

$$\frac{4}{52} \cdot \frac{4}{52} = \frac{1}{169}$$

(b) Since there are only three aces among the 51 cards that remain after one ace has been removed from the deck, we get

$$\frac{4}{52} \cdot \frac{3}{51} = \frac{1}{221}$$

Note that

$$\frac{1}{221} \neq \frac{4}{52} \cdot \frac{4}{52}$$

so independence is violated when the sampling is without replacement.

Example (Pairwise Independent Events That Are Not Independent) Let a ball be drawn from an urn containing four balls, numbered 1, 2, 3, 4. Let $E = \{1, 2\}$, $F = \{1, 3\}$, $G = \{1, 4\}$. If all four outcomes are assumed equally likely, then

$$P(EF) = P(E)P(F) = \tfrac{1}{4},$$

$$P(EG) = P(E)P(G) = \tfrac{1}{4},$$

$$P(FG) = P(F)P(G) = \tfrac{1}{4}$$

However,

$$\tfrac{1}{4} = P(EFG) \neq P(E)P(F)P(G)$$

Hence, even though the events E, F, G are pairwise independent, they are not jointly independent. ■

The Product Rule, or the Multiplicative Rule

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

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

Thus, the probability that both A and B occur is equal to the probability that A occurs multiplied by the conditional probability that B occurs, given that A occurs. Since the events $A \cap B$ and $B \cap A$ are equivalent,

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

More generally, since

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

In other words, it does not matter which event is referred to as A and which event is referred to as B.

the probability of the intersection of three events can be calculated as

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

Example

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,

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

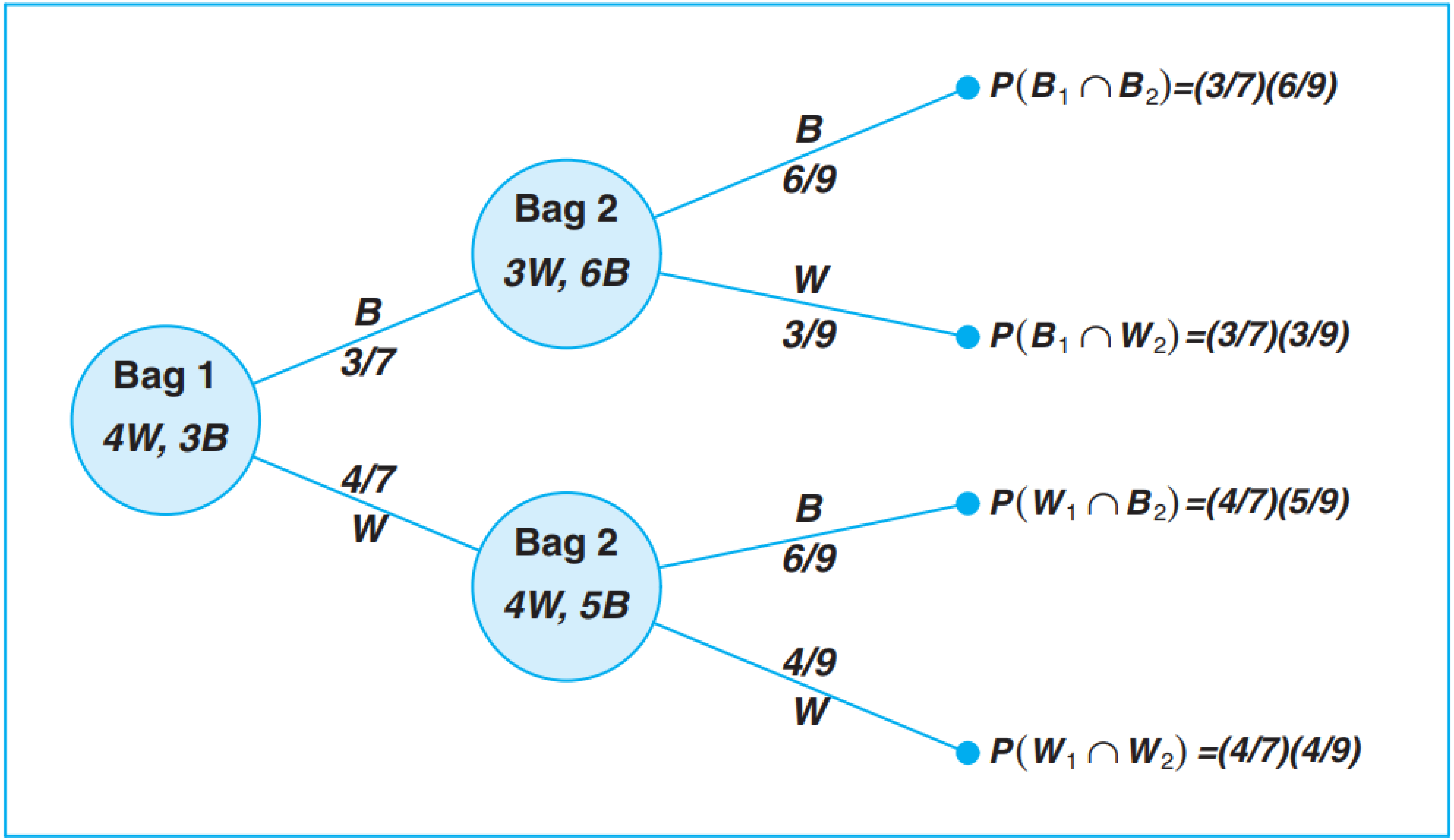
Example

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

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



Tree diagram

Example

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.

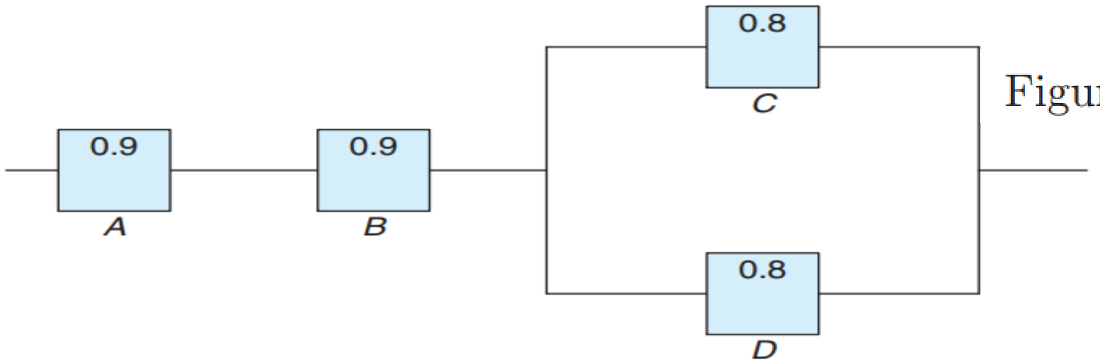


Figure 2.9

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 follows:

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

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

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

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

The multiplicative rule can be extended to more than two-event situations.

If, in an experiment, the events $A_1, A_2, \ldots, A_k$ 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 $A_1, A_2, \ldots, A_k$ 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

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,

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

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

Note: if two events are mutually exclusive, they cannot be independent

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 $C = \varphi$, the null set. Although $A \cap B \cap C = \varphi$, 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

A collection of events $\mathcal{A} = \{A_1, \ldots, A_n\}$ are mutually independent if for any subset of $\mathcal{A}$, $A_{i_1}, \ldots, A_{i_k}$, for $k \leq n$, we have

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

Rule of total probability or the **rule of elimination**

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,

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

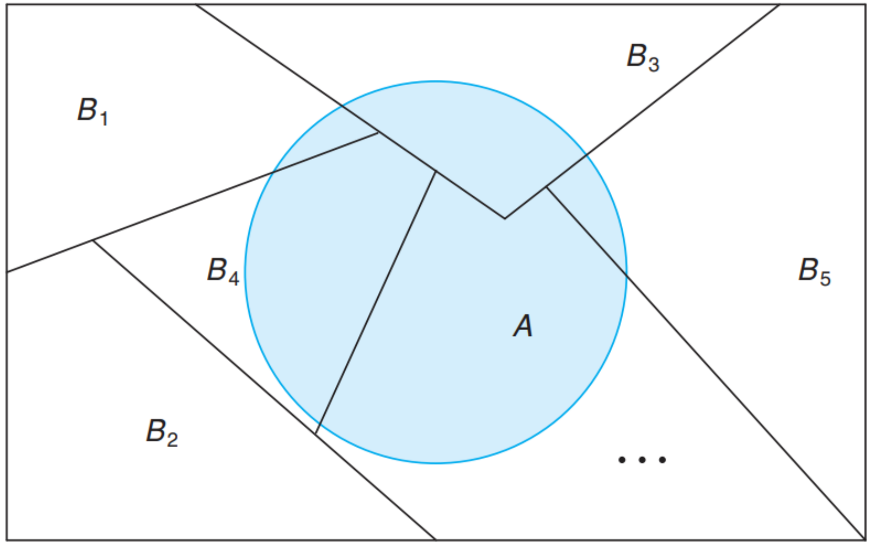


Figure 2.14: Partitioning the sample space S.

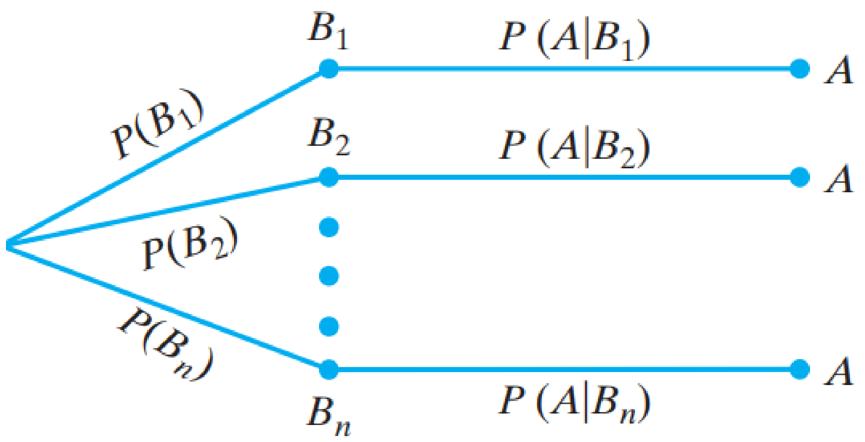
***Proof*:** Consider the Venn diagram of Figure 2.14. The event A is seen to be the union of the mutually exclusive events

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

that is,

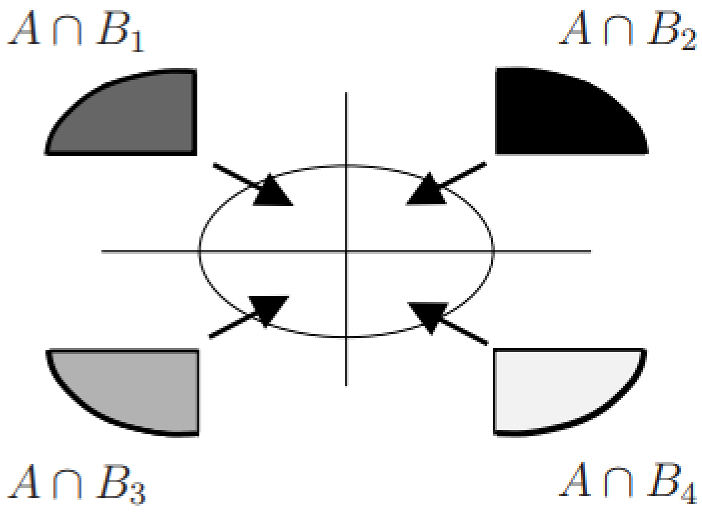
$$A = (B_1 \cap A) \cup (B_2 \cap A) \cup \cdots \cup (B_k \cap A).$$

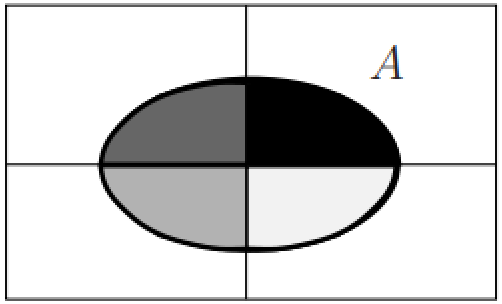
$$\begin{aligned} P(A) &= P[(B_1 \cap A) \cup (B_2 \cap A) \cup \cdots \cup (B_k \cap A)] \\ &= P(B_1 \cap A) + P(B_2 \cap A) + \cdots + P(B_k \cap A) \\ &= \sum_{i=1}^{k} P(B_i \cap A) \\ &= \sum_{i=1}^{k} P(B_i)P(A|B_i). \end{aligned}$$



Intuition behind the Partition Theorem (Law of Total Probability)

The Partition Theorem is easy to understand because it simply states that "the whole is the sum of its parts."





$$\mathbb{P}(A) = \mathbb{P}(A \cap B_1) + \mathbb{P}(A \cap B_2) + \mathbb{P}(A \cap B_3) + \mathbb{P}(A \cap B_4).$$

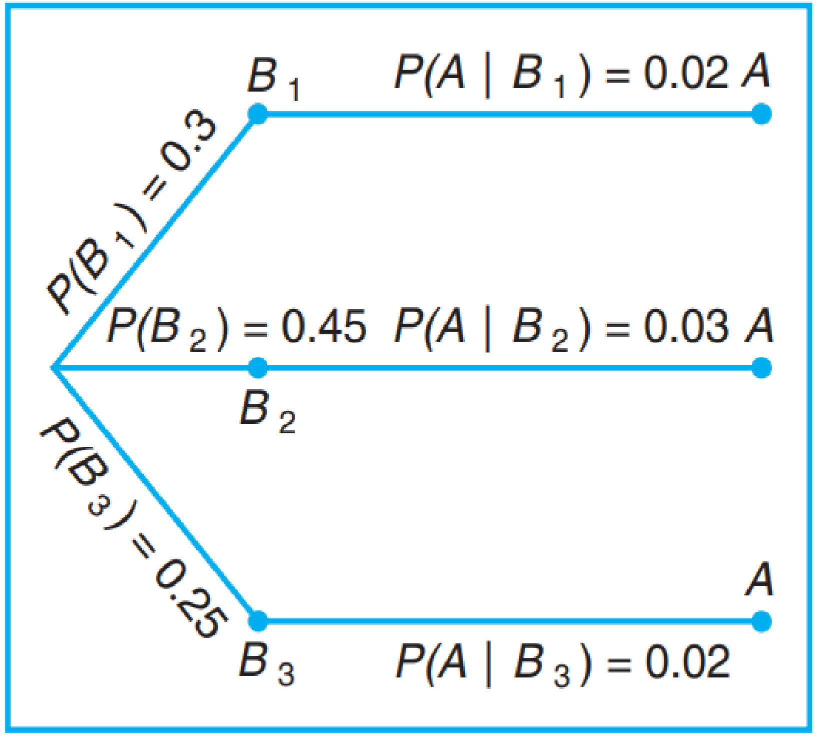
Problem

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

$$\begin{aligned} A&: \text{ the product is defective,} \\ B_1&: \text{ the product is made by machine } B_1, \\ B_2&: \text{ the product is made by machine } B_2, \\ B_3&: \text{ the product is made by machine } B_3. \end{aligned}$$

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



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

and hence

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

Example

An individual has 3 different email accounts. Most of her messages, in fact 70%, come into account #1, whereas 20% come into account #2 and the remaining 10% into account #3. Of the messages into account #1, only 1% are spam, whereas the corresponding percentages for accounts #2 and #3 are 2% and 5%, respectively. What is the probability that a randomly selected message is spam?

Solution

To answer this question, let's first establish some notation:

$$A_i = \{\text{message is from account } \# \ i\} \text{ for } i = 1, 2, 3, \quad B = \{\text{message is spam}\}$$

Then the given percentages imply that

$$P(A_1) = .70,\ P(A_2) = .20,\ P(A_3) = .10$$
$$P(B \mid A_1) = .01,\ P(B \mid A_2) = .02,\ P(B \mid A_3) = .05$$

Now it is simply a matter of substituting into the equation for the law of total probability:

$$P(B) = (.01)(.70) + (.02)(.20) + (.05)(.10) = .016$$

In the long run, 1.6% of this individual's messages will be spam. ■

Example

Suppose that in country S, 40% of the people support party A, 30% of the people support party B, 20% support party C, and 10% support party D. Let Q be a certain policy. We're given that 50% of the supporters of party A are in favor of Q, 40% of the supporters of party B are in favor of Q, 30% of the supporters of party C are in favor of Q, and 100% of the supporters of party D are in favor of Q. If we draw a citizen from this imaginary country at random, what is the probability that the citizen supports Q?

Solution

Let Q denote the event "is in favor of policy Q", A be the event "supports party A" (and so on for the rest of the parties). We can find P(Q) using the law of total probability:

$$\begin{aligned} P(Q) &= P(A)P(Q \mid A) + P(B)P(Q \mid B) + P(C)P(Q \mid C) + P(D)P(Q \mid D) \\ &= 0.4 \cdot 0.5 + 0.3 \cdot 0.4 + 0.2 \cdot 0.3 + 0.1 \cdot 1 = 0.48. \end{aligned}$$

Bayes' Rule

Bayesian statistics is a collection of tools that is used in a special form of statistical inference which applies in the analysis of experimental data in many practical situations in science and engineering. Bayes' rule is one of the most important rules in probability theory.

Theorem

For any events A and B:
$$\boxed{\mathbb{P}(B \mid A) = \frac{\mathbb{P}(A \mid B)\mathbb{P}(B)}{\mathbb{P}(A)}.}$$

Proof:

$$\begin{aligned}
\mathbb{P}(B \cap A) &= \mathbb{P}(A \cap B) \\
\mathbb{P}(B \mid A)\mathbb{P}(A) &= \mathbb{P}(A \mid B)\mathbb{P}(B) \qquad \text{(multiplication rule)} \\
\therefore \qquad \mathbb{P}(B \mid A) &= \frac{\mathbb{P}(A \mid B)\mathbb{P}(B)}{\mathbb{P}(A)}. \qquad \square
\end{aligned}$$

(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,

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

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

Example

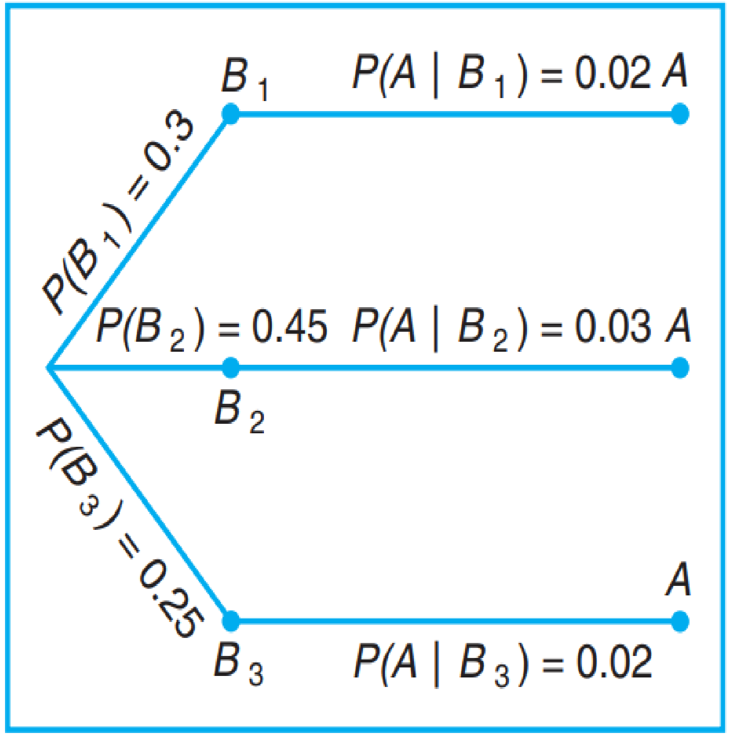
In a certain assembly plant, three machines, *B1*, *B2*, and *B3*, make 30%, 45%, and 25%, respectively, of the products. It is known from past experience that 2%, 3%, and 2% of the products made by each machine, respectively, are defective. if a product was chosen randomly and found to be defective, what is the probability that it was made by machine B3?

Solution

Using Bayes' rule to write

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

$$P(A) = P(B_1)P(A|B_1) + P(B_2)P(A|B_2) + P(B_3)P(A|B_3)$$
$$= 0.006 + 0.0135 + 0.005 = 0.0245$$



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

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

Example

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:

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

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

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

Similarly,

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

Let us say P(Fire) means how often there is fire, and P(Smoke) means how often we see smoke, then:

P(Fire|Smoke) means how often there is fire when we can see smoke
P(Smoke|Fire) means how often we can see smoke when there is fire

Example:

- dangerous fires are rare (1%)
- but smoke is fairly common (10%) due to barbecues,
- and 90% of dangerous fires make smoke

We can then discover the **probability of dangerous Fire when there is Smoke**:

$$\begin{aligned} P(\text{Fire}|\text{Smoke}) &= \frac{P(\text{Fire})\ P(\text{Smoke}|\text{Fire})}{P(\text{Smoke})} \\ &= \frac{1\% \times 90\%}{10\%} \\ &= 9\% \end{aligned}$$

So it is still worth checking out any smoke to be sure.

Example: Picnic Day

You are planning a picnic today, but the morning is cloudy

- Oh no! 50% of all rainy days start off cloudy!
- But cloudy mornings are common (about 40% of days start cloudy)
- And this is usually a dry month (only 3 of 30 days tend to be rainy, or 10%)

What is the chance of rain during the day?

We will use Rain to mean rain during the day, and Cloud to mean cloudy morning.

The chance of Rain given Cloud is written P(Rain|Cloud)

So let's put that in the formula:

$$P(\text{Rain}|\text{Cloud}) = \frac{P(\text{Rain})\ P(\text{Cloud}|\text{Rain})}{P(\text{Cloud})}$$

$$P(\text{Rain}|\text{Cloud}) = \frac{P(\text{Rain})\, P(\text{Cloud}|\text{Rain})}{P(\text{Cloud})}$$

- P(Rain) is Probability of Rain = 10%
- P(Cloud|Rain) is Probability of Cloud, given that Rain happens = 50%
- P(Cloud) is Probability of Cloud = 40%

$$P(\text{Rain}|\text{Cloud}) = \frac{0.1 \times 0.5}{0.4} = .125$$

Or a 12.5% chance of rain. Not too bad, let's have a picnic!

Example

Identifying spam using Bayes' Theorem

A first step towards identifying spam is to create a list of words that are more likely to appear in spam than in normal messages. For instance, words like buy or the brand name of an enhancement drug are more likely to occur in spam messages than in normal messages. Suppose a specified list of words is available and that your data base of 5000 messages contains 1700 that are spam. Among the spam messages, 1343 contain words in the list. Of the 3300 normal messages, only 297 contain words in the list.

Obtain the probability that a message is spam given that the message contains words in the list.

Solution

Let $A =$ [message contains words in list] be the event a message is identified as spam and let $B_1 =$ [message is spam] and $B_2 =$ [message is normal]. We use the observed relative frequencies from the data base as approximations to the probabilities.

$$P(B_1) = \frac{1700}{5000} = .34 \qquad P(B_2) = \frac{3300}{5000} = .66$$

$$P(A \mid B_1) = \frac{1343}{1700} = .79 \qquad P(A \mid B_2) = \frac{297}{3300} = .09$$

Bayes' Theorem expresses the probability of being spam, given that a message is identified as spam, as

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

The updated, or posterior probability, is

$$P(B_1 \mid A) = \frac{.79 \times .34}{.79 \times .34 + .09 \times .66} = \frac{.2686}{.328} = .819$$

Because this posterior probability of being spam is quite large, we suspect that this message really is spam. Since $P(B_1) = .34$, or 34% of the incoming messages are spam, we likely would want the spam filter to remove this message. Existing spam filer programs learn and improve as you mark your incoming messages spam. ■

Example

Incidence of a rare disease. Only 1 in 1000 adults is afflicted with a rare disease for which a diagnostic test has been developed. The test is such that when an individual actually has the disease, a positive result will occur 99% of the time, whereas an individual without the disease will show a positive test result only 2% of the time. If a randomly selected individual is tested and the result is positive, what is the probability that the individual has the disease?

Solution

To use Bayes' theorem, let A_1 = individual has the disease, A_2 = individual does not have the disease, and B = positive test result. Then $P(A_1) = .001$, $P(A_2) = .999$, $P(B|A_1) = .99$, and $P(B|A_2) = .02$. The tree diagram for this problem is in Figure 2.12.

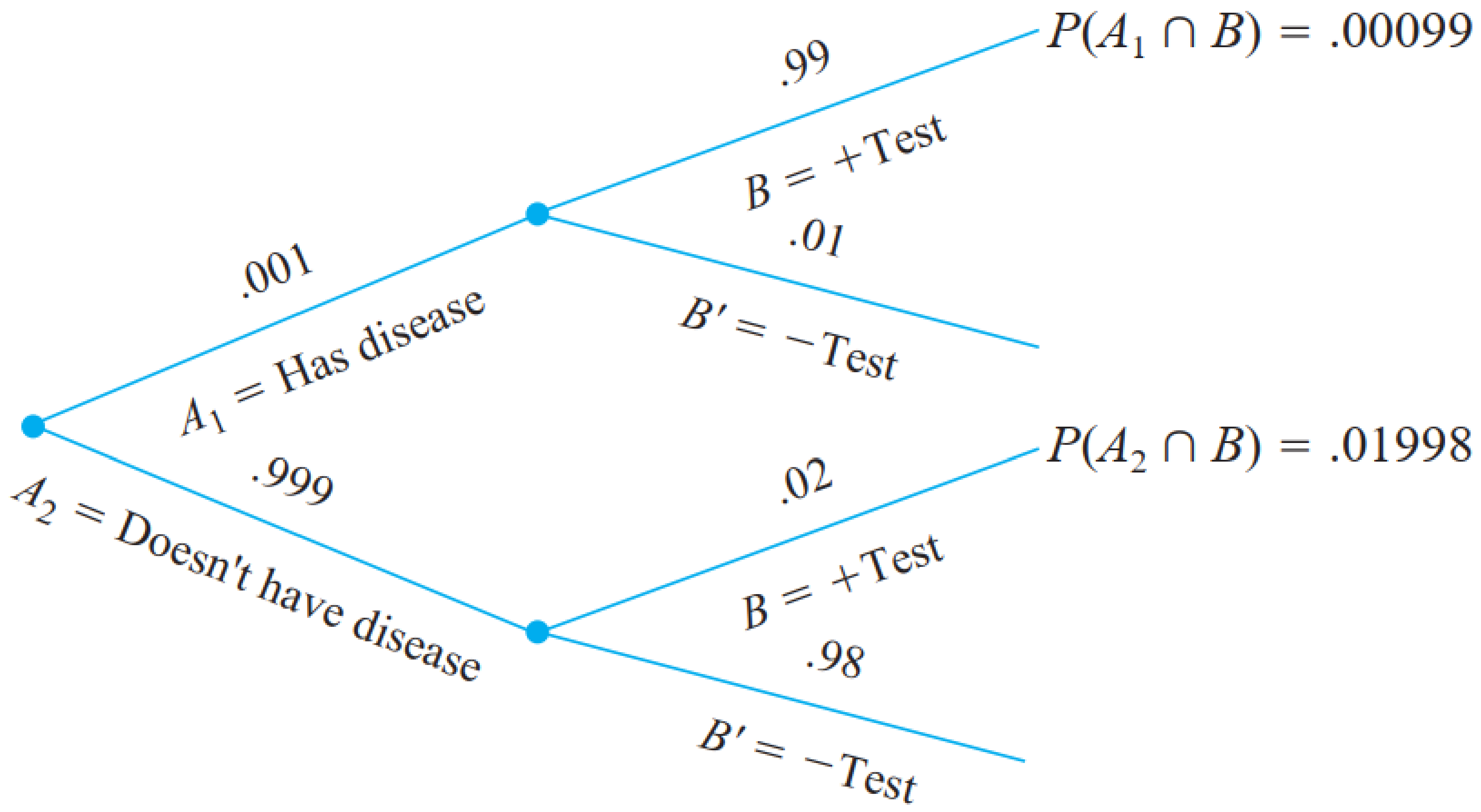


Figure 2.12 Tree diagram for the rare-disease problem

Next to each branch corresponding to a positive test result, the multiplication rule yields the recorded probabilities. Therefore, $P(B) = .00099 + .01998 = .02097$, from which we have

$$P(A_1 | B) = \frac{P(A_1 \cap B)}{P(B)} = \frac{.00099}{.02097} = .047$$

This result seems counterintuitive; the diagnostic test appears so accurate that we expect someone with a positive test result to be highly likely to have the disease, whereas the computed conditional probability is only .047. However, the rarity of the disease implies that most positive test results arise from errors rather than from diseased individuals. The probability of having the disease has increased by a multiplicative factor of 47 (from prior .001 to posterior .047); but to get a further increase in the posterior probability, a diagnostic test with much smaller error rates is needed. ■

Problem Electric wheelchairs are difficult to maneuver for many disabled people. In a paper presented at the *1st International Workshop on Advances in Service Robotics* (Mar. 2003), researchers applied Bayes's rule to evaluate an "intelligent" robotic controller that aims to capture the intent of a wheelchair user and aid in navigation. Consider the following scenario. From a certain location in a room, a wheelchair user will either (1) turn sharply to the left and navigate through a door, (2) proceed straight to the other side of the room, or (3) turn slightly right and stop at a table. Denote these three events as D (for door), S (Straight), and T (for table). Based on previous trips, $P(D) = .5$, $P(S) = .2$, and $P(T) = .3$. The wheelchair is installed with a robot-controlled joystick. When the user intends to go through the door, she points the joystick straight 30% of the time; when the user intends to go straight, she points the joystick straight 40% of the time; and, when the user intends to go to the table, she points the joystick straight 5% of the time. If the wheelchair user points the joystick straight, what is her most likely destination?

Example

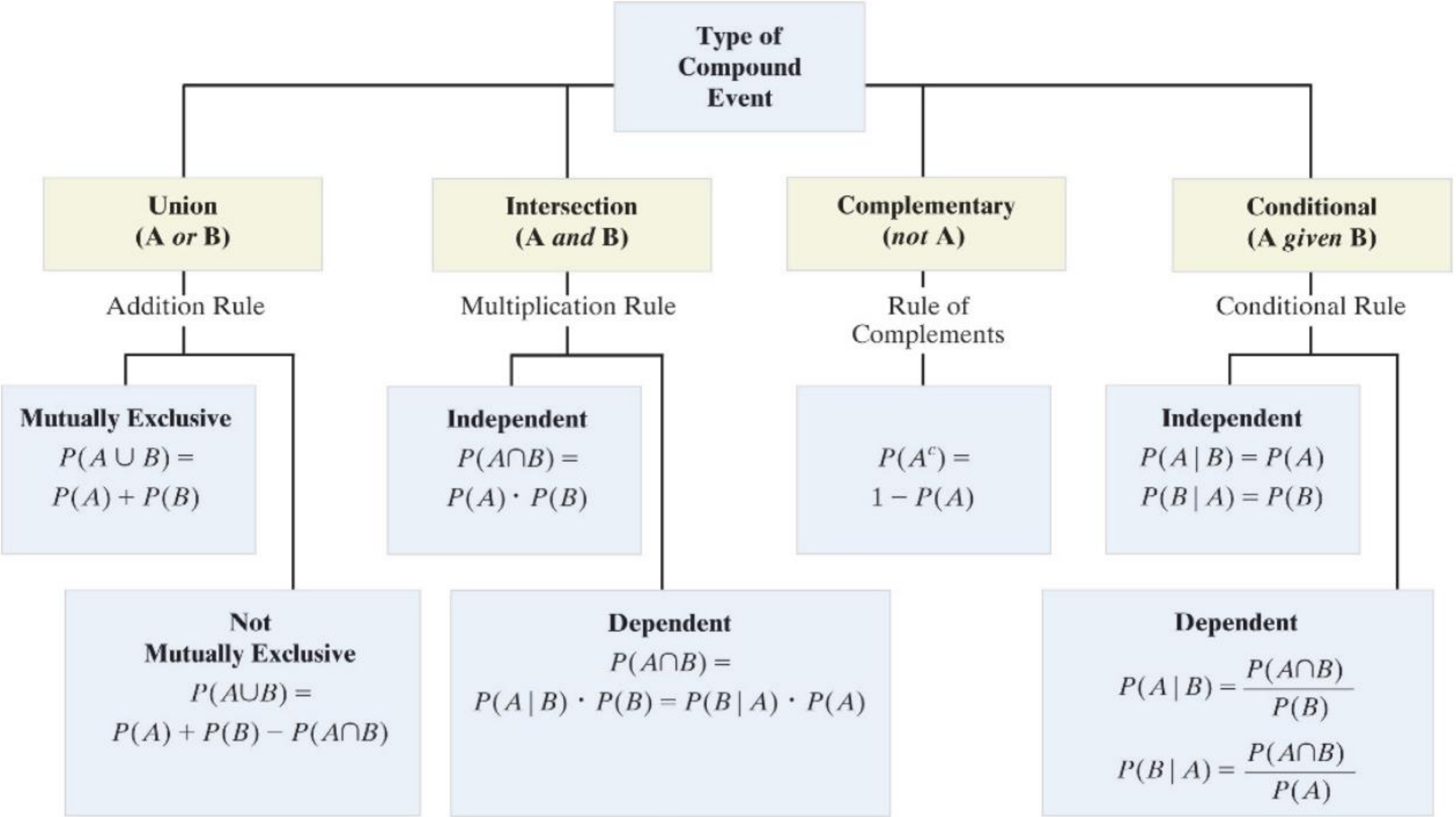
Bayes's Rule Application—Wheelchair Control

Solution Let $J = \{\text{joystick is pointed straight}\}$. The user intention percentages can be restated as the following conditional probabilities: $P(J|D) = .3$, $P(J|S) = .4$, and $P(J|T) = .05$. Since the user has pointed the joystick straight, we want to find the following probabilities: $P(D|J)$, $P(S|J)$, and $P(T|J)$. Now, the three events, D, S, and T, represent mutually exclusive and exhaustive events, where $P(D) = .5$, $P(S) = .2$, and $P(T) = .3$. Consequently, we can apply Bayes's rule as follows:

$$\begin{aligned}
P(D|J) &= P(J|D) \cdot P(D)/[P(J|D) \cdot P(D) + P(J|S) \cdot P(S) + P(J|T) \cdot P(T)] \\
&= (.3)(.5)/[(.3)(.5) + (.4)(.2) + (.05)(.3)] = .15/.245 = .612 \\
P(S|J) &= P(J|S) \cdot P(S)/[P(J|D) \cdot P(D) + P(J|S) \cdot P(S) + P(J|T) \cdot P(T)] \\
&= (.4)(.2)/[(.3)(.5) + (.4)(.2) + (.05)(.3)] = .08/.245 = .327 \\
P(T|J) &= P(J|T) \cdot P(T)/[P(J|D) \cdot P(D) + P(J|S) \cdot P(S) + P(J|T) \cdot P(T)] \\
&= (.05)(.3)/[(.3)(.5) + (.4)(.2) + (.05)(.3)] = .015/.245 = .061
\end{aligned}$$

Note that the largest conditional probability is $P(D|J) = .612$. Thus, if the joystick is pointed straight, the wheelchair user is most likely headed through the door.

Guide to Selecting Probability Rules



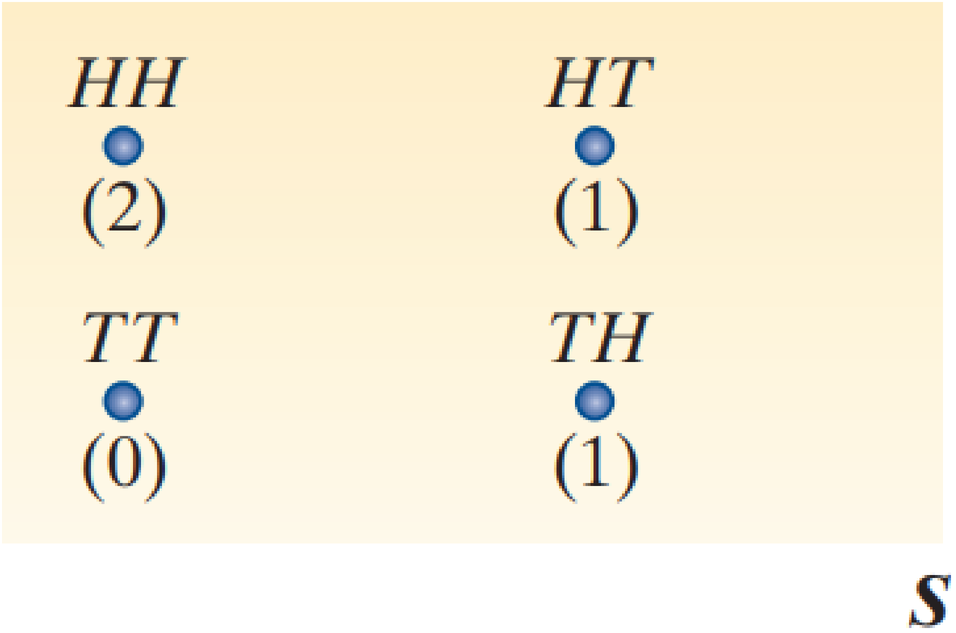
Chapter 3 Random Variables and Probability Distributions

Random Variable

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

A **random variable** is obtained by assigning a numerical value to each outcome of a particular experiment.

A random variable is therefore a special kind of experiment in which the outcomes are numerical values, either positive or negative or possibly zero. Sometimes the experimental outcomes are already numbers, and then these may just be used to define the random variable.



Random variables are typically denoted by uppercase letters, such as X, Y, and Z. The actual numerical values that a random variable can assume are denoted by lowercase letters, such as x, y, and z.

Mathematically,

$$X : \Omega \longrightarrow \mathbb{R}$$

where Ω represents sample space

Example

Suppose that a coin is tossed twice so that the sample space is $S = \{HH, HT, TH, TT\}$. Let X represent the number of heads that can come up. With each sample point we can associate a number for X as shown in Table . Thus, for example, in the case of HH (i.e., 2 heads), X =2 while for TH (1 head), X =1. It follows that X is a random variable

Sample Point	HH	HT	TH	TT
X	2	1	1	0

Example

	1	2	3	4	5	6
1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

Examples :

Let X denote the random variable that is defined as the sum of two fair dice, then

$P\{X = 2\} = P\{(1, 1)\} = 1/36$

$P\{X = 3\} = P\{(1, 2), (2, 1)\} = 2/36$

$P\{X = 4\} = P\{(1, 3), (2, 2), (3, 1)\} = 3/36$

$P\{X = 5\} = P\{(1, 4), (2, 3), (3, 2), (4, 1)\} = 4/36$

$P\{X = 6\} = P\{(1, 5), (2, 4), (3, 3), (4, 2), (5, 1)\} = 5/36$

$P\{X = 7\} = P\{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\} = 6/36$

$P\{X = 8\} = P\{(2, 6), (3, 5), (4, 4), (5, 3), (6, 2)\} = 5/36$

$P\{X = 9\} = P\{(3, 6), (4, 5), (5, 4), (6, 3)\} = 4/36$

P{X = 10} = P{(4, 6), (5, 5), (6, 4)} = 3/36

P{X = 11} = P{(5, 6), (6, 5)} = 2/36

P{X = 12} = P{(6, 6)} = 1/36

In other words, the random variable X can take on any integral value between 2 and 12.

Note:

$$1 = P(S) = P\left(\bigcup_{i=2}^{12}\{X = i\}\right) = \sum_{i=2}^{12} P\{X = i\}$$

Another random variable of possible interest in this experiment is the value of the first die. Letting Y denote this random variable, then Y is equally likely to take on any of the values 1 through 6. That is,

P{Y = i} = 1/6, i = 1, 2, 3, 4, 5, 6.

Next Example :

Suppose that an individual purchases two electronic components, each of which may be either defective or acceptable. In addition, suppose that the four possible results — (d, d), (d, a), (a, d), (a, a) — have respective probabilities .09, .21, .21, .49 [where (d, d) means that both components are defective, (d, a) that the first component is defective and the second acceptable, and so on]. If we let X denote the number of acceptable components obtained in the purchase, then X is a random variable taking on one of the values 0, 1, 2 with respective probabilities

$$P\{X = 0\} = .09$$
$$P\{X = 1\} = .42$$
$$P\{X = 2\} = .49$$

If we were mainly concerned with whether there was at least one acceptable component, we could define the random variable I by

$$I = \begin{cases} 1 & \text{if } X = 1 \text{ or } 2 \\ 0 & \text{if } X = 0 \end{cases}$$

If A denotes the event that at least one acceptable component is obtained, then the random variable I is called the indicator random variable for the event A, since I will equal 1 or 0 depending upon whether A occurs. The probabilities attached to the possible values of I are

$$P\{I = 1\} = .91$$
$$P\{I = 0\} = .09$$

Example

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	$\boldsymbol{y}$
RR	2
RB	1
BR	1
BB	0

Example

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.

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

$M = \{0,1,3\}$

Example

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

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

DEFINITION

Any random variable whose only possible values are 0 and 1 is called a **Bernoulli random variable.**

Example

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 $0 \leq x \leq 1$.

Example

Let X be the random variable defined by the waiting time, in hours, between successive speeders spotted by a radar unit. The random variable X takes on all values x for which $x \geq 0$.

Definition

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

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**.

In most practical problems, continuous random variables represent *measured* data, such as all possible heights, weights, temperatures, distance, or life periods, whereas discrete random variables represent *count* data, such as the number of defectives in a sample of k items or the number of highway fatalities per year in a given state.

To study basic properties of discrete rv's, only the tools of discrete mathematics—summation and differences—are required. The study of continuous variables requires the continuous mathematics of the calculus—integrals and derivatives.

Distribution of a randam variable X

Collection of all the probabilities related to X is the distribution of X.

Let X is the number of 1's in a random binary string of 3 characters then the distribution of X is

x	$P\{X = x\}$
0	1/8
1	3/8
2	3/8
3	1/8
Total	1

Because

$$\Omega = \{000, 001, 010, 100, 111, 110, 101, 011\}$$

Discrete Probability Distributions

Definition

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. $P(X = x) = f(x)$.

Example

Checking for nonnegativity and total probability equals one

Check whether the following can serve as probability distributions:

(a) $f(x) = \dfrac{x-2}{2}$ for $x = 1, 2, 3, 4$

(b) $h(x) = \dfrac{x^2}{25}$ for $x = 0, 1, 2, 3, 4$

Solution

(a) This function cannot serve as a probability distribution because $f(1)$ is negative.

(b) The function cannot serve as a probability distribution because the sum of the five probabilities is $\frac{6}{5}$ and not 1. ■

Example

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

Example

Suppose that a coin is tossed twice so that the sample space is S = {HH, HT, TH, TT}. Let X represent the number of heads that can come up. With each sample point we can associate a number for X as shown in Table . Thus, for example, in the case of HH (i.e., 2 heads), X =2 while for TH (1 head), X =1. Find the probability function corresponding to the random variable X.

Solution

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

$$P(HT) = \frac{1}{4}$$

$$P(TT) = \frac{1}{4}$$

Then

$$P(X = 0) = P(TT) = \frac{1}{4}$$

$$P(X = 1) = P(HT \cup TH) = P(HT) + P(TH) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

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

The probability function is thus given by Table

x	0	1	2
$f(x)$	1/4	1/2	1/4

Definition

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

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

If X takes on only a finite number of values $x_1, x_2, \ldots, x_n$, then the distribution function is given by

$$F(x) = \begin{cases} 0 & -\infty < x < x_1 \\ f(x_1) & x_1 \leq x < x_2 \\ f(x_1) + f(x_2) & x_2 \leq x < x_3 \\ \vdots & \vdots \\ f(x_1) + \cdots + f(x_n) & x_n \leq x < \infty \end{cases}$$

Note: CDF is defined for both Discrete & Continuous r.v.s.

The distribution function $F(x)$ has the following properties:

1. $F(x)$ is nondecreasing [i.e., $F(x) \le F(y)$ if $x \le y$].
2. $\lim_{x \to -\infty} F(x) = 0$; $\lim_{x \to \infty} F(x) = 1$.
3. $F(x)$ is continuous from the right [i.e., $\lim_{h \to 0^+} F(x + h) = F(x)$ for all x].

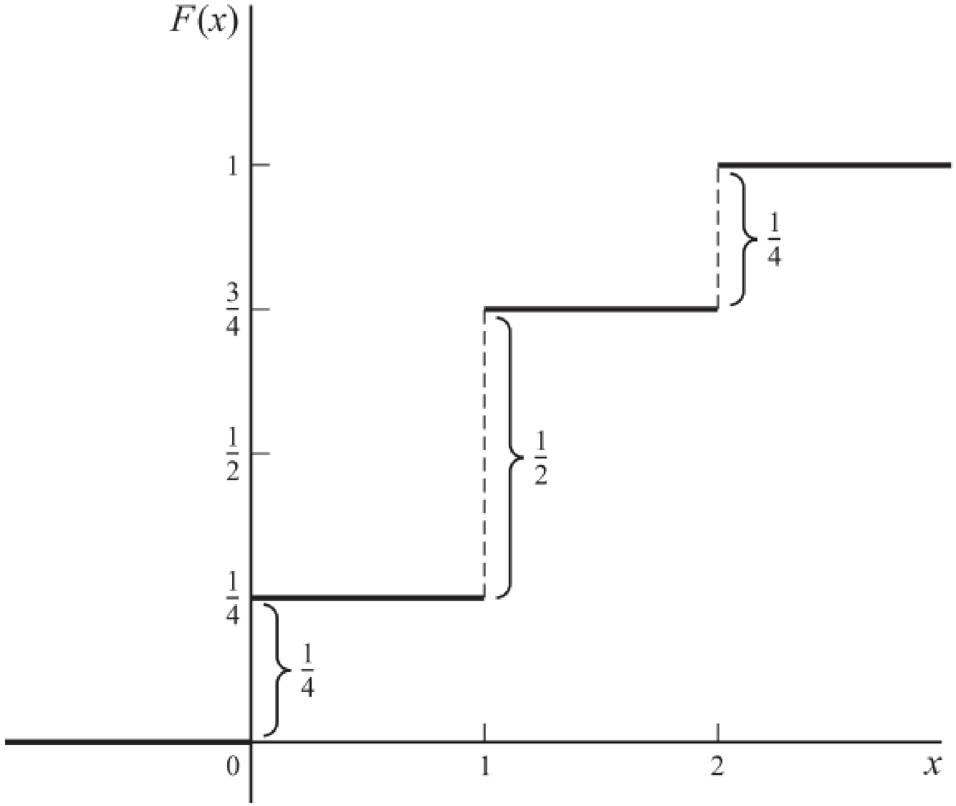
Example

Suppose that a coin is tossed twice so that the sample space is $S = \{HH, HT, TH, TT\}$. Let X represent the number of heads that can come up. With each sample point we can associate a number for X as shown in Table . Thus, for example, in the case of HH (i.e., 2 heads), X =2 while for TH (1 head), X =1. Find the distribution function for the random variable X.

Solution

The distribution function is

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



EXAMPLE

a) Find the probability distribution of the sum of the dots when two fair dice are thrown
b) Use the probability distribution to find the probabilities of obtaining (i) a sum that is greater than 8, and (ii) a sum that is greater than 5 but less than or equal to 10.

SOLUTION

a) The sample space S is represented by the following 36 outcomes:

$$
\begin{aligned}
S = \{&(1, 1), (1, 2), (1, 3), (1, 5), (1, 6);\\
&(2, 1), (2, 2), (2, 3), (2, 5), (2, 6);\\
&(3, 1), (3, 2), (3, 3), (3, 5), (3, 6);\\
&(4, 1), (4, 2), (4, 3), (4, 5), (4, 6);\\
&(5, 1), (5, 2), (5, 3), (5, 5), (5, 6);\\
&(6, 1), (6, 2), (6, 3), (6, 5), (6, 6)\ \}
\end{aligned}
$$

Since each of the 36 outcomes is equally likely to occur, therefore each outcome has probability 1/36.
Let X be the random variable representing the sum of dots which appear on the dice. Then the values of the r.v. are 2, 3, 4… 12.
The probabilities of these values are computed as below:

$f(2) = P(X = 2) = P[\{1, 1\}] = \frac{1}{36}$, as there is only one outcome resulting in a sum of 2,

$$f(3) = P(X = 3) = P[\{(1, 2), (2,1)\}] = \frac{2}{36},$$

$$f(4) = P(X = 4) = P[\{(1, 3), (2,2), (3,1)\}] = \frac{3}{36},$$

Similarly

$$f(5) = \frac{4}{36}, f(6) = \frac{5}{36}, f(7) = \frac{6}{36}, f(8) = \frac{5}{36}, f(9) = \frac{4}{36},$$

$$f(10)=\frac{3}{36},\ f(11)=\frac{2}{36} \text{ and } f(12)=\frac{1}{36}.$$

Therefore the desired probability distribution of the r.v X is

x_i	2	3	4	5	6	7	8	9	10	11	12
$f(x_i)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

b) Using the probability distribution, we get the required probabilities as follows:

i) P(a sum that is greater than 8)

$$= P(X > 8)$$

$$= P(X=9) + P(X=10) + P(X=11) + P(X=12)$$

$$= f(9) + f(10) + f(11) + f(12)$$

$$= \frac{4}{36} + \frac{3}{36} + \frac{2}{36} + \frac{1}{36} = \frac{10}{36}$$

ii) P(a sum that is greater than 5 but less than or equal to 10)

$$
\begin{aligned}
&= P(5 < X \leq 10) \\
&= P(X = 6) + P(X = 7) + P(X = 8) \\
&\qquad + P(X = 9) + P(X = 10) \\
&= f(6) + f(7) + f(8) + f(9) + f(10) \\
&= \frac{5}{36} + \frac{6}{36} + \frac{5}{36} + \frac{4}{36} + \frac{3}{36} = \frac{23}{36}.
\end{aligned}
$$

DISTRIBUTION FUNCTION

The distribution function of a random variable X, denoted by F(x), is defined by $F(x) = P(X < x)$.

The term 'distribution function' implies the cumulating of the probabilities

x_i	2	3	4	5	6	7	8	9	10	11	12
$f(x_i)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$
$F(x_i)$	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{6}{36}$	$\frac{10}{36}$	$\frac{15}{36}$	$\frac{21}{36}$	$\frac{26}{36}$	$\frac{30}{36}$	$\frac{33}{36}$	$\frac{35}{36}$	$\frac{36}{36}$

EXAMPLE

Find the probability distribution and distribution function for the number of heads when 3 balanced coins are tossed. Depict both the probability distribution and the distribution function graphically. Since the coins are balanced, therefore the equally probable sample space for this experiment is

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

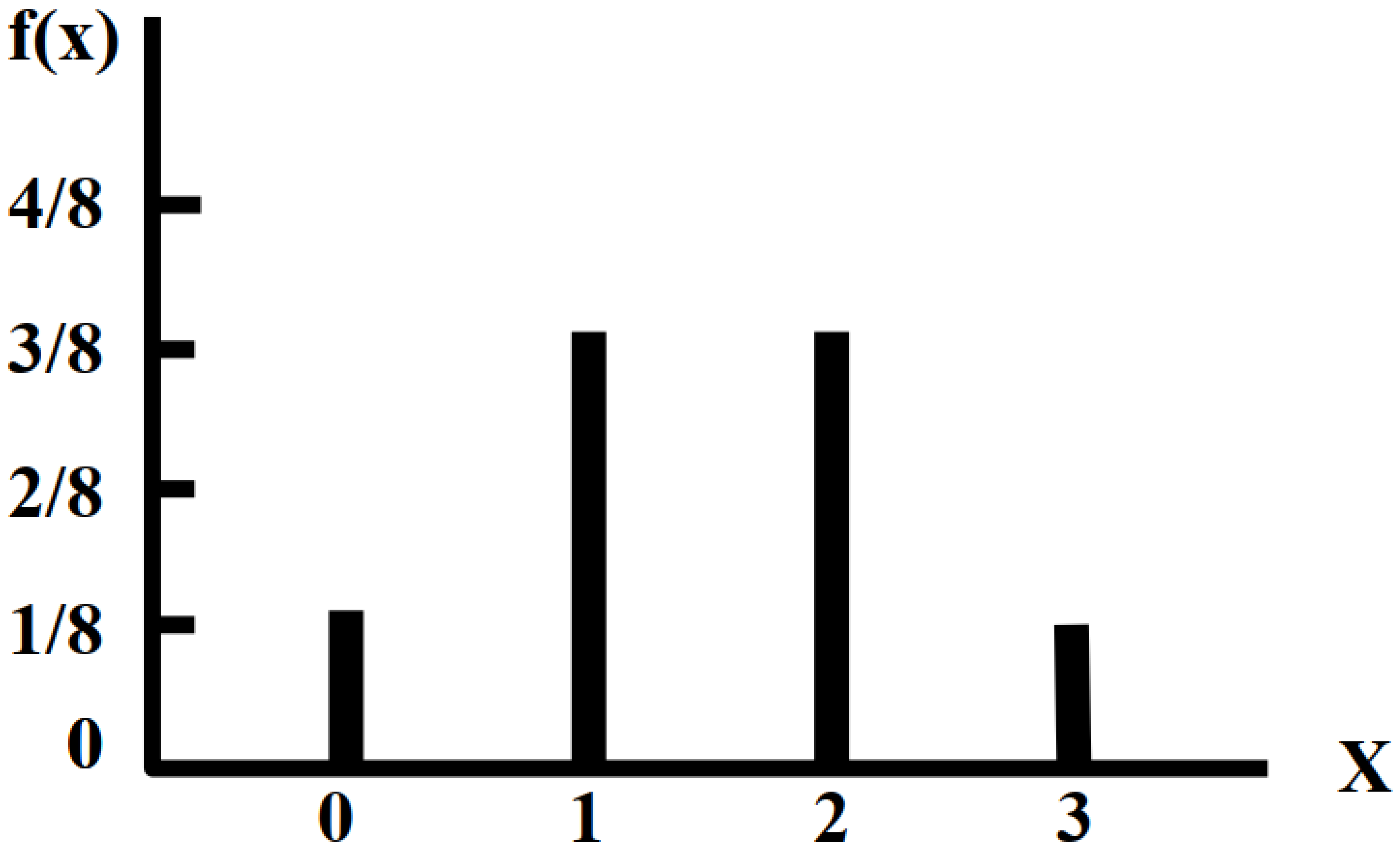
Let X be the random variable that denotes the number of heads.
Then the values of X are 0, 1, 2 and 3.
And their probabilities are:

$$
\begin{aligned}
f(0) &= P(X = 0) \\
&= P[\{TTT\}] = 1/8 \\
f(1) &= P(X = 1) \\
&= P[\{HTT, THT, TTH\}] = 3/8 \\
f(2) &= P(X = 2) \\
&= P[\{HHT, HTH, THH\}] = 3/8 \\
f(2) &= P(X = 3) \\
&= P[\{HHH\}] = 1/8
\end{aligned}
$$

Expressing the above information in the tabular form, we obtain the desired probability distribution of X as follows:

Number of Heads (x_i)	Probability $f(x_i)$
0	$\frac{1}{8}$
1	$\frac{3}{8}$
2	$\frac{3}{8}$
3	$\frac{1}{8}$
Total	1

The line chart of the above probability distribution is as follows:



In order to obtain the distribution function of this random variable, we compute the cumulative probabilities as follows:

Number of Heads (x_i)	Probability $f(x_i)$	Cumulative Probability $F(x_i)$
0	$\frac{1}{8}$	$\frac{1}{8}$
1	$\frac{3}{8}$	$\frac{1}{8}+\frac{3}{8}=\frac{4}{8}$
2	$\frac{3}{8}$	$\frac{4}{8}+\frac{3}{8}=\frac{7}{8}$
3	$\frac{1}{8}$	$\frac{7}{8}+\frac{1}{8}=1$

Hence the desired distribution function is

$$F(x)=\begin{cases} 0, & for\ \ x<0 \\ \dfrac{1}{8}, & for\ 0\le x<1 \\ \dfrac{4}{8}, & for\ 1\le x<2 \\ \dfrac{7}{8}, & for\ 2\le x<3 \\ 1, & for\ \ x\ge 3 \end{cases}$$

Example Suppose the random variable X has distribution function

$$F(x) = \begin{cases} 0 & x \leq 0 \\ 1 - \exp\{-x^2\} & x > 0 \end{cases}$$

What is the probability that X exceeds 1?

Solution. The desired probability is computed as follows:

$$\begin{aligned} P\{X > 1\} &= 1 - P\{X \leq 1\} \\ &= 1 - F(1) \\ &= e^{-1} \\ &= .368 \quad \blacksquare \end{aligned}$$

Note

$$P\{X \leq b\} = P\{X \leq a\} + P\{a < X \leq b\}$$

or

$$P\{a < X \leq b\} = F(b) - F(a)$$

Example Consider a random variable X that is equal to 1, 2, or 3. If we know that

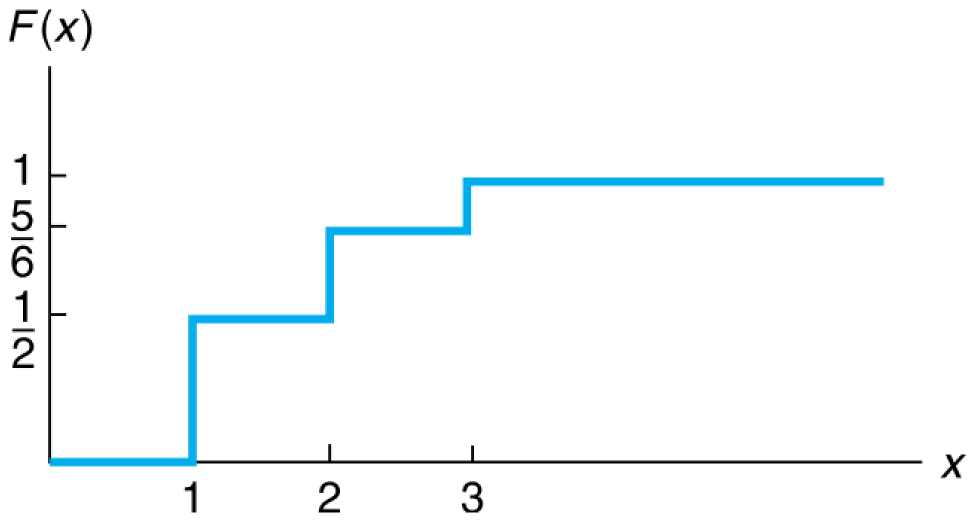
$$p(1) = \tfrac{1}{2} \qquad \text{and} \qquad p(2) = \tfrac{1}{3}$$

then it follows (since $p(1) + p(2) + p(3) = 1$) that

$$p(3) = \tfrac{1}{6}$$

Then the cumulative distribution function F of X is given by

$$F(a) = \begin{cases} 0 & a < 1 \\ \frac{1}{2} & 1 \leq a < 2 \\ \frac{5}{6} & 2 \leq a < 3 \\ 1 & 3 \leq a \end{cases}$$



EXAMPLE 2.1 If X is a discrete random variable having the probability distribution

$X = x$	1	2	3
$P(X = x)$	k	$2k$	k

Find $P(X \leq 2)$.

Solution We know that

$$\sum P(X = x) = 1 \quad \Rightarrow \quad 4k = 1$$

$$\therefore \qquad k = \frac{1}{4}$$

$$P(X \leq 2) = P(X = 1) + P(X = 2) = 3k \quad \Rightarrow \quad P(X \leq 2) = \frac{3}{4}$$

EXAMPLE 2.2 If X is a discrete random variable having the PMF

x	-1	0	1
$P(x)$	k	$2k$	$3k$

Find $P(X \geq 0)$.

Solution We know that

$$\sum P(X = x) = P(X = -1) + P(X = 0) + P(X = 1)$$

$$k + 2k + 3k = 1 \Rightarrow 6k = 1$$

$$k = \frac{1}{6}$$

Example

Let X be a discrete random variable with the following PMF

$$P_X(x) = \begin{cases} \frac{1}{2} & \text{for } x = 0 \\ \frac{1}{3} & \text{for } x = 1 \\ \frac{1}{6} & \text{for } x = 2 \\ 0 & \text{otherwise} \end{cases}$$

(a) Find R_X, the range of the random variable X.

(b) Find $P(X \geq 1.5)$.

(c) Find $P(0 < X < 2)$.

(d) Find $P(X = 0|X < 2)$

Solution

(a) The range of X can be found from the PMF. The range of X consists of possible values for X. Here we have

$$R_X = \{0, 1, 2\}.$$

(b) The event $X \geq 1.5$ can happen only if X is 2. Thus,

$$\begin{aligned} P(X \geq 1.5) &= P(X = 2) \\ &= P_X(2) = \frac{1}{6}. \end{aligned}$$

(c) Similarly, we have

$$\begin{aligned} P(0 < X < 2) &= P(X = 1) \\ &= P_X(1) = \frac{1}{3}. \end{aligned}$$

(d) This is a conditional probability problem, so we can use our famous formula $P(A|B) = \frac{P(A \cap B)}{P(B)}$.

We have

$$
\begin{aligned}
P(X = 0 | X < 2) &= \frac{P\big(X = 0, X < 2\big)}{P(X < 2)} \\
&= \frac{P(X = 0)}{P(X < 2)} \\
&= \frac{P_X(0)}{P_X(0) + P_X(1)} \\
&= \frac{\frac{1}{2}}{\frac{1}{2} + \frac{1}{3}} = \frac{3}{5}.
\end{aligned}
$$

Example

Let X be a discrete random variable with the following CDF:

$$F_X(x) = \begin{cases} 0 & \text{for } x < 0 \\ \frac{1}{6} & \text{for } 0 \leq x < 1 \\ \frac{1}{2} & \text{for } 1 \leq x < 2 \\ \frac{3}{4} & \text{for } 2 \leq x < 3 \\ 1 & \text{for } x \geq 3 \end{cases}$$

Find the range and PMF of X.

Solution

$$R_X = \{0, 1, 2, 3\}.$$

$$P_X(x) = F_X(x) - F_X(x - \epsilon).$$

$$P_X(0) = F_X(0) - F_X(0 - \epsilon) = \frac{1}{6} - 0 = \frac{1}{6}$$

$$P_X(1) = F_X(1) - F_X(1 - \epsilon) = \frac{1}{2} - \frac{1}{6} = \frac{1}{3}$$

$$P_X(2) = F_X(2) - F_X(2 - \epsilon) = \frac{3}{4} - \frac{1}{2} = \frac{1}{4}$$

$$P_X(3) = F_X(3) - F_X(3 - \epsilon) = 1 - \frac{3}{4} = \frac{1}{4}.$$

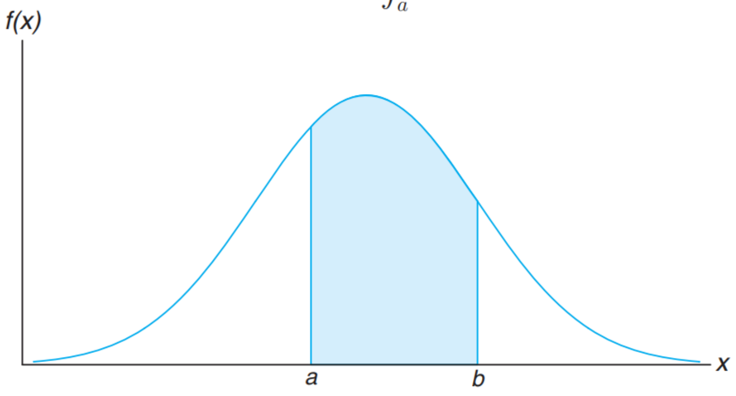
$$P_X(x) = \begin{cases} \frac{1}{6} & \text{for } x = 0 \\ \frac{1}{3} & \text{for } x = 1 \\ \frac{1}{4} & \text{for } x = 2 \\ \frac{1}{4} & \text{for } x = 3 \\ 0 & \text{otherwise} \end{cases}$$

Continuous Probability Distributions

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

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



$P(a < X < b)$

Example

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

(a) Obviously, $f(x) \geq 0$. To verify condition 2, 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

$$P(0 < X \leq 1) = \int_{0}^{1} \frac{x^2}{3} dx = \left.\frac{x^3}{9}\right|_{0}^{1} = \frac{1}{9}.$$

Definition

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 , 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

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

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

evaluate $P(0 < X \leq 1)$.

Solution

For $-1 < x < 2$,

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

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

Example

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

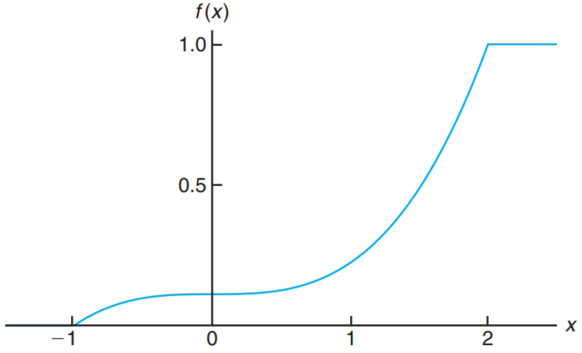
$$f(y) = \begin{cases} \frac{5}{8b}, & \frac{2}{5}b \leq y \leq 2b, \\ 0, & \text{elsewhere.} \end{cases}$$

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 $2b/5 \leq y \leq 2b$,

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



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

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

Example Suppose that X is a continuous random variable whose probability density function is given by

$$f(x) = \begin{cases} C(4x - 2x^2) & 0 < x < 2 \\ 0 & \text{otherwise} \end{cases}$$

(a) What is the value of C?
(b) Find $P\{X > 1\}$.

Solution. **(a)** Since f is a probability density function, we must have that $\int_{-\infty}^{\infty} f(x)\,dx = 1$, implying that

$$C \int_0^2 (4x - 2x^2)\,dx = 1$$

or

or

$$C\left[2x^2 - \frac{2x^3}{3}\right]\Bigg|_{x=0}^{x=2} = 1$$

or

$$C = \tfrac{3}{8}$$

(b) Hence

$$P\{X > 1\} = \int_1^{\infty} f(x)\,dx = \tfrac{3}{8}\int_1^{2}(4x - 2x^2)\,dx = \tfrac{1}{2} \quad \blacksquare$$

Relation b/w CDF and PDF

The relationship between the cumulative distribution $F(\cdot)$ and the probability density $f(\cdot)$ is expressed by

$$F(a) = P\{X \in (-\infty, a]\} = \int_{-\infty}^{a} f(x)\, dx$$

Differentiating both sides yields

$$\frac{d}{da}F(a) = f(a)$$ (By Fundamental Theorem of Calculus)

Example

Let X be a continuous random variable with PDF

$$f_X(x) = \begin{cases} \frac{5}{32}x^4 & 0 \leq x \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

and let $Y = X^2$.

(a) Find CDF of Y.

(b) Find PDF of Y.

Solution

(a) First, we note that $R_Y = [0, 4]$. As usual, we start with the CDF. For $y \in [0, 4]$, we have

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(X^2 \leq y) \\ &= P(0 \leq X \leq \sqrt{y}) \quad \text{since } x \text{ is not negative} \end{aligned}$$

$$= \int_0^{\sqrt{y}} \frac{5}{32} x^4 dx$$
$$= \frac{1}{32} (\sqrt{y})^5$$
$$= \frac{1}{32} y^2 \sqrt{y}$$

Thus, the CDF of Y is given by

$$F_Y(y) = \begin{cases} 0 & \text{for } y < 0 \\ \frac{1}{32} y^2 \sqrt{y} & \text{for } 0 \le y \le 4 \\ 1 & \text{for } y > 4. \end{cases}$$

(b)

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \begin{cases} \frac{5}{64} y \sqrt{y} & \text{for } 0 \le y \le 4 \\ 0 & \text{otherwise} \end{cases}$$

Joint Probability Distributions

For a given experiment, we are often interested not only in probability distribution functions of individual random variables but also in the relationships between two or more random variables. For instance, in an experiment into the possible causes of cancer, we might be interested in the relationship between the average number of cigarettes smoked daily and the age at which an individual contracts cancer. Similarly, an engineer might be interested in the relationship between the shear strength and the diameter of a spot weld in a fabricated sheet steel specimen.

Definition

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. $\sum\limits_x \sum\limits_y f(x,y) = 1$,
3. $P(X = x, Y = y) = f(x,y)$.

For any region A in the xy plane, $P[(X,Y) \in A] = \sum\sum\limits_A f(x,y)$.

Definition

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.

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

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.

$X \backslash Y$	y_1	y_2	$\dots$	y_n	Totals $\downarrow$
x_1	$f(x_1, y_1)$	$f(x_1, y_2)$	$\dots$	$f(x_1, y_n)$	$f_1(x_1)$
x_2	$f(x_2, y_1)$	$f(x_2, y_2)$	$\dots$	$f(x_2, y_n)$	$f_1(x_2)$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
x_m	$f(x_m, y_1)$	$f(x_m, y_2)$	$\dots$	$f(x_m, y_n)$	$f_1(x_m)$
Totals $\rightarrow$	$f_2(y_1)$	$f_2(y_2)$	$\dots$	$f_2(y_n)$	1 $\leftarrow$ Grand Total

The row totals give the marginal pmf of X and the column totals give the marginal pmf of Y. Once these marginal pmf's are available, the probability of any event involving only X or only Y can be calculated.

Joint Probability Distribution of X and Y

X\Y	y_1	y_2	…	y_j	…	y_n	$P(X = x_i)$
x_1	$f(x_1,y_1)$	$f(x_1,y_2)$	…	$f(x_1,y_j)$	…	$f(x_1,y_n)$	$g(x_1)$
x_2	$f(x_2,y_1)$	$f(x_2,y_2)$	…	$f(x_2,y_j)$	…	$f(x_2,y_n)$	$g(x_2)$
$\vdots$	$\vdots$					$\vdots$	$\vdots$
x_i	$f(x_i,y_1)$	$f(x_i,y_2)$	…	$f(x_i,y_j)$	…	$f(x_i,y_n)$	$g(x_i)$
$\vdots$	$\vdots$					$\vdots$	$\vdots$
x_m	$f(x_m,y_1)$	$f(x_m,y_2)$	…	$f(x_m,y_j)$	…	$f(x_m,y_n)$	$g(x_m)$
$P(Y=y_j)$	$h(y_1)$	$h(y_2)$	…	$h(y_j)$	…	$h(y_n)$	1

Definition

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

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

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

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

Definition

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.

If you can find any point (x, y) for which $f(x, y)$ is defined such that $f(x, y) \neq g(x)h(y)$, the discrete variables X and Y are not statistically independent

Example

The joint probability function of two discrete random variables X and Y is given by $f(x, y) = c(2x + y)$, where x and y can assume all integers such that $0 \le x \le 2$, $0 \le y \le 3$, and $f(x, y) = 0$ otherwise.

(a) Find the value of the constant c.

(b) Find $P(X = 2, Y = 1)$.

(c) Find $P(X \ge 1, Y \le 2)$.

(a) The sample points (x, y) for which probabilities are different from zero are indicated in Fig. 2-8. The probabilities associated with these points, given by $c(2x + y)$, are shown in Table 2-6. Since the grand total, $42c$, must equal 1, we have $c = 1/42$.

X \ Y	0	1	2	3	Totals ↓
0	0	c	$2c$	$3c$	$6c$
1	$2c$	$3c$	$4c$	$5c$	$14c$
2	$4c$	$5c$	$6c$	$7c$	$22c$
Totals →	$6c$	$9c$	$12c$	$15c$	$42c$

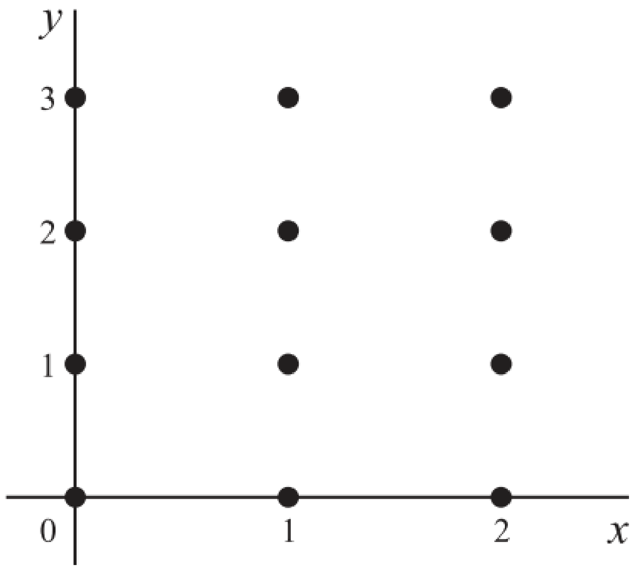


Fig. 2-8

(c) From Table we see that

$$
\begin{aligned}
P(X \geq 1, Y \leq 2) &= \sum_{x \geq 1} \sum_{y \leq 2} f(x, y) \\
&= (2c + 3c + 4c)(4c + 5c + 6c) \\
&= 24c = \frac{24}{42} = \frac{4}{7}
\end{aligned}
$$

as indicated by the entries shown shaded in the table.

Example

Find the marginal probability functions of X and of Y for the random variables

Solution

The marginal probability function for X is given by $P(X = x) = f_1(x)$ and can be obtained from the margin totals in the right-hand column of Table 2-6. From these we see that

$$P(X = x) = f_1(x) = \begin{cases} 6c = 1/7 & x = 0 \\ 14c = 1/3 & x = 1 \\ 22c = 11/21 & x = 2 \end{cases}$$

Check: $\frac{1}{7} + \frac{1}{3} + \frac{11}{21} = 1$

(b) The marginal probability function for Y is given by $P(Y = y) = f_2(y)$ and can be obtained from the margin totals in the last row of Table 2-6. From these we see that

$$P(Y = y) = f_2(y) = \begin{cases} 6c = 1/7 & y = 0 \\ 9c = 3/14 & y = 1 \\ 12c = 2/7 & y = 2 \\ 15c = 5/14 & y = 3 \end{cases}$$

Check: $\frac{1}{7} + \frac{3}{14} + \frac{2}{7} + \frac{5}{14} = 1$

Example

. Show that the random variables X and Y are dependent.

Solution

If the random variables X and Y are independent, then we must have, for all x and y,

$$P(X = x, Y = y) = P(X = x)P(Y = y)$$

But, as seen from Problems 2.8(b) and 2.9,

$$P(X = 2, Y = 1) = \frac{5}{42} \qquad P(X = 2) = \frac{11}{21} \qquad P(Y = 1) = \frac{3}{14}$$

so that $$P(X = 2, Y = 1) \neq P(X = 2)P(Y = 1)$$

The result also follows from the fact that the joint probability function $(2x + y)/42$ cannot be expressed as a function of x alone times a function of y alone.

EXAMPLE

An urn contains 3 black, 2 red and 3 green balls and 2 balls are selected at random from it. If X is the number of black balls and Y is the number of red balls selected, then find

i) the joint probability function $f(x, y)$;
ii) $P(X + Y \leq 1)$;
iii) the marginal p.d. $g(x)$ and $h(y)$;
iv) the conditional p.d. $f(x \mid 1)$,
v) $P(X = 0 \mid Y = 1)$; and
vi) Are x and Y independent?

i) The sample space S for this experiment contains sample points. The possible values of X are 0, 1, and 2, and those for Y are 0, 1, and 2. The values that (X, Y) can take on are (0, 0), (0, 1), (1, 0), (1, 1), (0, 2) and (2, 0). We desire to find f(x, y) for each value (x, y).

The total number of ways in which 2 balls can be drawn out of a total of 8 balls is

$$\binom{8}{2} = \frac{8x7}{2} = 28.$$

Now f(0, 0) = P(X = 0 and Y = 0), where the event (X = 0 and Y = 0) represents that neither black nor red ball is selected, implying that the 2 selected are green balls. This event therefore contains $\binom{3}{0}\binom{2}{0}\binom{3}{2} = 3$ sample points,

and

f(0, 0) = P(X = 0 and Y = 0) = 3/28

Again f(0, 1) = P(X = 0 and Y = 1)

= P (none is black, 1 is red and 1 is green)

$$= \frac{\binom{3}{0}\binom{2}{1}\binom{3}{1}}{28} = \frac{6}{28}$$

Similarly,

$$f(1,1) = P(X = 1 \text{ and } Y = 1)$$
$$= P(1 \text{ is black } 1 \text{ is red and none is green})$$
$$= \frac{\binom{3}{1}\binom{2}{1}\binom{3}{0}}{28} = \frac{6}{28}$$

Similar calculations give the probabilities of other values and the joint probability function of X and Y is given as

X \ Y	0	1	2	$P(X = x_i)$ $g(x)$
0	3/28	6/28	1/28	10/28
1	9/28	6/28	0	15/28
2	3/28	0	0	3/28
$P(Y = y_j)$ $h(y)$	15/28	12/28	1/28	1

This joint p.d. of the two r.v.'s (X, Y) can be represented by the formula

$$f(x,y)=\frac{\binom{3}{x}\binom{2}{y}\binom{3}{2-x-y}}{28}=\begin{matrix} x=0,1,2 \\ y=0,1,2 \\ 0\le x+y\le 2. \end{matrix}$$

ii) To compute $P(X + Y < 1)$, we see that $x + y < 1$ for the cells (0, 0), (0, 1) and (1, 0).

Therefore

$$\begin{aligned} &P(X + Y < 1) \\ &= f(0, 0) + f(0, 1) + f(1, 0) \\ &= 3/28 + 6/28 + 9/28 \\ &= 18/28 = 9/14 \end{aligned}$$

iii) The marginal p.d.'s are:

x	0	1	2
g(x)	10/28	15/28	3/28

y	0	1	2
h(x)	15/28	12/28	1/28

iv) By definition, the conditional p.d. f(x | 1) is

$$f(x \mid 1) = P(X = x \mid Y = 1)$$

$$= \frac{P(X = x \, and \, Y = 1)}{P(Y = 1)} = \frac{f(x,1)}{h(1)}$$

Now

$$h(1) = \sum_{x=0}^{2} f(x,1)$$

$$= \frac{6}{28} + \frac{6}{28} + 0$$

$$= \frac{12}{28} = \frac{3}{7}$$

Therefore

$$f(x \mid 1) = \frac{f(x,1)}{h(1)}$$

That is, $$= \frac{3}{7} f(x,1), \qquad x = 0,1,2$$

$$f(0 \mid 1) = \frac{7}{3} f(0,1) = \left(\frac{7}{3}\right)\left(\frac{6}{28}\right) = \frac{1}{2}$$

$$f(1 \mid 1) = \frac{7}{3} f(1,1) = \left(\frac{7}{3}\right)\left(\frac{6}{28}\right) = \frac{1}{2}$$

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

Hence the conditional p.d. of X given that Y = 1, is

x	0	1	2
$f(x\|1)$	1/2	1/2	0

vi) We find that f(0, 1) = 6/28,

$$g(0) = \sum_{y=0}^{2} f(0, y)$$

$$= \frac{3}{28} + \frac{6}{28} + \frac{1}{28} = \frac{10}{28}$$

$$h(1) = \sum_{x=0}^{2} f(x,1)$$

$$= \frac{6}{28} + \frac{6}{28} + 0 = \frac{12}{28}$$

v) Finally,

$$P(X = 0 \mid Y = 1)$$
$$= f(0 \mid 1) = 1/2$$

Now $$\frac{6}{28} \neq \frac{10}{28} \times \frac{12}{28},$$

i.e. $$f(0,1) \neq g(0)h(1),$$

therefore X and Y are ***NOT***

Satistically independent.

Problem

A program consists of two modules. The number of errors, X, in the first module and the number of errors, Y , in the second module have the joint distribution, $p(0,0) = p(0,1) = p(1,0) = 0.2$, $p(1,1) = p(1,2) = p(1,3) = 0.1$, $p(0,2) = p(0,3) = 0.05$. Find (a) the marginal distributions of X and Y , (b) the probability of no errors in the first module, and (c) the distribution of the total number of errors in the program.

[Remember: Collection of all the probabilities related to some random variable is sometime called the distribution of the random variable. i.e. joint distribution mean joint PMF etc.]

Solution

It is convenient to organize the joint pmf of X and Y in a table. Adding row wise and column wise, we get the marginal pmfs,

$p(x,y)$		$y=0$	$y=1$	$y=2$	$y=3$	$p_X(x)$
x	0	0.20	0.20	0.05	0.05	0.50
	1	0.20	0.10	0.10	0.10	0.50
$p_Y(y)$		0.40	0.30	0.15	0.15	1.00

This solves (a).

(b) $P_X(0) = 0.50$. (Because the probability of no errors in the first module mean $P\{X = 0\}$.

(c) Let Z = X + Y be the total number of errors. To find the distribution of Z, we first identify its possible values, then find the probability of each value. We see that Z can be as small as 0 and as large as 4. Then,

$$p_Z(0) = P\{X + Y = 0\} = P\{X = 0, Y = 0\} = p(0,0) = 0.20,$$

$$p_Z(1) = P\{X = 0, Y = 1\} + P\{X = 1, Y = 0\}$$

$$= p(0,1) + p(1,0) = 0.20 + 0.20 = 0.40,$$

$$p_Z(2) = p(0,2) + p(1,1) = 0.05 + 0.10 = 0.15,$$

$$p_Z(3) = p(0,3) + p(1,2) = 0.05 + 0.10 = 0.15,$$

$$p_Z(4) = p(1,3) = 0.10.$$

It is a good check to verify that

$$\sum_z p_Z(\mathrm{z}) = 1$$

Example The joint density function of X and Y is given by

$$f(x, y) = \begin{cases} 2e^{-x}e^{-2y} & 0 < x < \infty, 0 < y < \infty \\ 0 & \text{otherwise} \end{cases}$$

Compute **(a)** $P\{X > 1, Y < 1\}$; **(b)** $P\{X < Y\}$; and **(c)** $P\{X < a\}$.

Solution. (a)

$$\begin{aligned} P\{X > 1, Y < 1\} &= \int_0^1 \int_1^\infty 2e^{-x}e^{-2y}\,dx\,dy \\ &= \int_0^1 2e^{-2y}(-e^{-x}|_1^\infty)\,dy \end{aligned}$$

$$= e^{-1} \int_0^1 2e^{-2y}\, dy$$

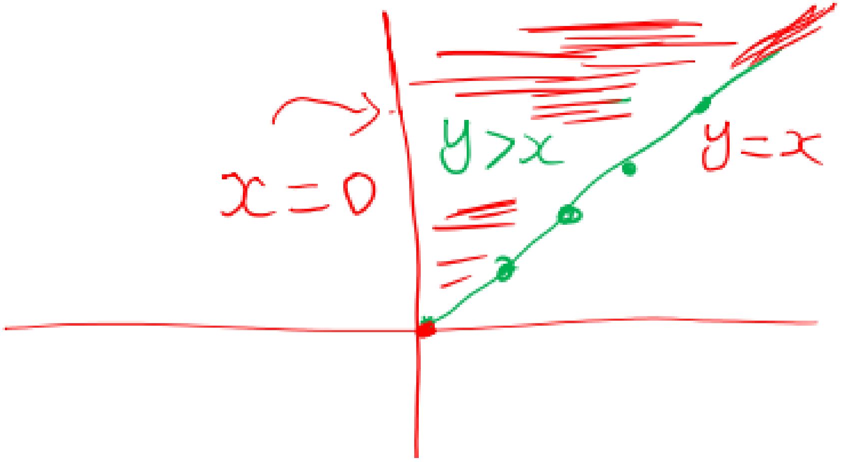
$$= e^{-1}(1 - e^{-2})$$

(b)

$$P\{X < Y\} = \iint\limits_{(x,y):x<y} 2e^{-x}e^{-2y}\, dx\, dy$$

$$= \int_0^\infty \int_0^y 2e^{-x}e^{-2y}\, dx\, dy$$

$$= \int_0^\infty 2e^{-2y}(1 - e^{-y})\, dy$$



$$= \int_0^\infty 2e^{-2y}\,dy - \int_0^\infty 2e^{-3y}\,dy$$

$$= 1 - \tfrac{2}{3}$$

$$= \tfrac{1}{3}$$

(c)

$$P\{X < a\} = \int_0^a \int_0^\infty 2e^{-2y} e^{-x}\,dy\,dx$$

$$= \int_0^a e^{-x}\,dx$$

$$= 1 - e^{-a} \quad \blacksquare$$

Example

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

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

(a) Verify condition 2 of Definition of **joint density function**

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

(c) Find $g(x)$ and $h(y)$ for the joint density function

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

$$P[(X,Y)\in A] = P\left(0 < X < \frac{1}{2}, \frac{1}{4} < Y < \frac{1}{2}\right)$$
$$= \int_{1/4}^{1/2}\int_0^{1/2} \frac{2}{5}(2x+3y)\ dx\ dy$$

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

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

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

(c)

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

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

Example

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:

$$g(x) = \int_{-\infty}^{\infty} f(x, y)\, dy = \int_{x}^{1} 10xy^2\, dy$$

$$= \frac{10}{3} x y^3 \Big|_{y=x}^{y=1} = \frac{10}{3} x(1 - x^3), \ 0 < x < 1,$$

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

Now

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

(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

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 $0 < x < 2$,

$$g(x) = \int_{-\infty}^{\infty} f(x,y)\, dy = \int_0^1 \frac{x(1+3y^2)}{4} dy$$

$$= \left(\frac{xy}{4} + \frac{xy^3}{4} \right) \bigg|_{y=0}^{y=1} = \frac{x}{2},$$

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

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

Chapter 4 Mathematical Expectation

Definition

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.

The reader should note that the way to calculate the expected value, or mean, shown here is different from the way to calculate the sample mean described in Chapter 1, where the sample mean is obtained by using data. In mathematical expectation, the expected value is calculated by using the probability distribution.

Example

Find $E[X]$ where X is the outcome when we roll a fair die.

Solution Since $p(1) = p(2) = p(3) = p(4) = p(5) = p(6) = \frac{1}{6}$, we obtain that

$$E[X] = 1\left(\frac{1}{6}\right) + 2\left(\frac{1}{6}\right) + 3\left(\frac{1}{6}\right) + 4\left(\frac{1}{6}\right) + 5\left(\frac{1}{6}\right) + 6\left(\frac{1}{6}\right) = \frac{7}{2}$$

The reader should note that, for this example, the expected value of X is not a value that X could possibly assume. (That is, rolling a die cannot possibly lead to an outcome of 7/2.) Thus, even though we call $E[X]$ the *expectation* of X, it should not be interpreted as the value that we *expect* X to have but rather as the average value of X in a large number of repetitions of the experiment. That is, if we continually roll a fair die, then after a large number of rolls the average of all the outcomes will be approximately 7/2. (The interested reader should try this as an experiment.) ■

Example If I is an indicator random variable for the event A, that is, if

$$I = \begin{cases} 1 & \text{if } A \text{ occurs} \\ 0 & \text{if } A \text{ does not occur} \end{cases}$$

then

$$E[I] = 1P(A) + 0P(A^c) = P(A)$$

Hence, the expectation of the indicator random variable for the event A is just the probability that A occurs. ■

Example Suppose that you are expecting a message at some time past 5 P.M. From experience you know that X, the number of hours after 5 P.M. until the message arrives, is a random variable with the following probability density function:

$$f(x) = \begin{cases} \dfrac{1}{1.5} & \text{if } 0 < x < 1.5 \\ 0 & \text{otherwise} \end{cases}$$

The expected amount of time past 5 P.M. until the message arrives is given by

$$E[X] = \int_0^{1.5} \frac{x}{1.5}\, dx = .75$$

Hence, on average, you would have to wait three-fourths of an hour. ■

Example

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

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

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

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

Suppose X has the following probability mass function

$$p(0) = .2, \quad p(1) = .5, \quad p(2) = .3$$

Calculate $E[X^2]$.

Solution

Letting $Y = X^2$, we have that Y is a random variable that can take on one of the values 0^2, 1^2, 2^2 with respective probabilities

$$p_Y(0) = P\{Y = 0^2\} = .2$$

$$p_Y(1) = P\{Y = 1^2\} = .5$$

$$p_Y(4) = P\{Y = 2^2\} = .3$$

$$E[X^2] = 0^2(0.2) + (1^2)(0.5) + (2^2)(0.3) = 1.7$$

Example The time, in hours, it takes to locate and repair an electrical breakdown in a certain factory is a random variable — call it X — whose density function is given by

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

If the cost involved in a breakdown of duration x is x^3, what is the expected cost of such a breakdown?

Solution.

$$E[X^3] = \int_0^1 x^3 dx \quad (\text{since } f(x) = 1, 0 < x < 1)$$

$$= \frac{1}{4} \quad \blacksquare$$

Example

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*:**

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

Corollary If a and b are constants, then

$$E[aX + b] = aE[X] + b$$

$$E[b] = b$$

$$E[aX] = aE[X]$$

$$E[X^n] = \begin{cases} \sum_x x^n p(x) & \text{if } X \text{ is discrete} \\ \int_{-\infty}^{\infty} x^n f(x)\, dx & \text{if } X \text{ is continuous} \end{cases}$$

In general, for any n,

$$E[X_1 + X_2 \cdots + X_n] = E[X_1] + E[X_2] + \cdots + E[X_n]$$

Variance and Covariance of Random Variables

Another important summary measure of the distribution of a random variable is the **variance**, which measures the *spread* or *variability* in the values taken by the random variable. Whereas the mean or expectation measures the central or average value of the random variable, the variance measures the spread or deviation of the random variable about its mean value.

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.

Variance

The **variance** of a random variable X is defined to be

$$\mathrm{Var}(X) = E((X - E(X))^2)$$

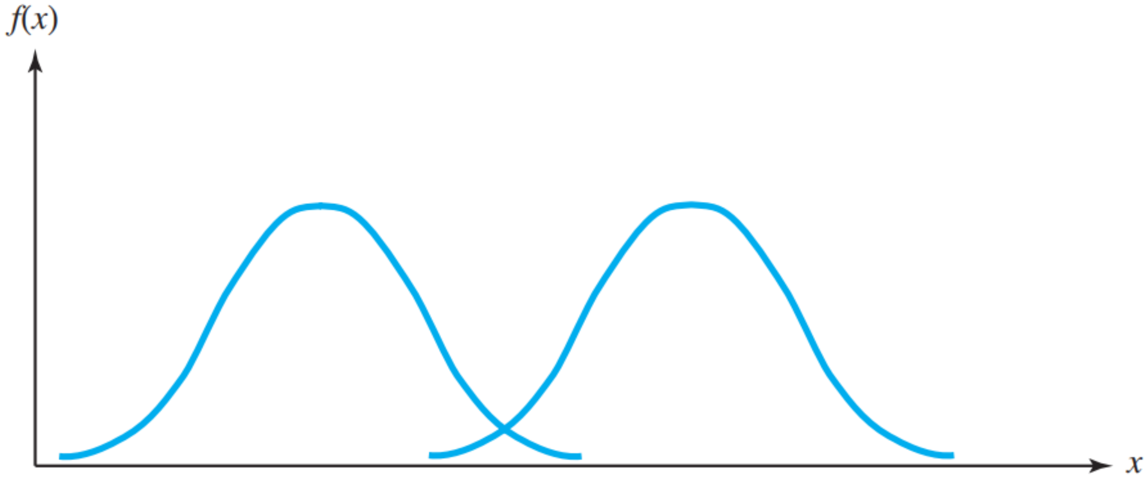
or equivalently

$$\mathrm{Var}(X) = E(X^2) - (E(X))^2$$

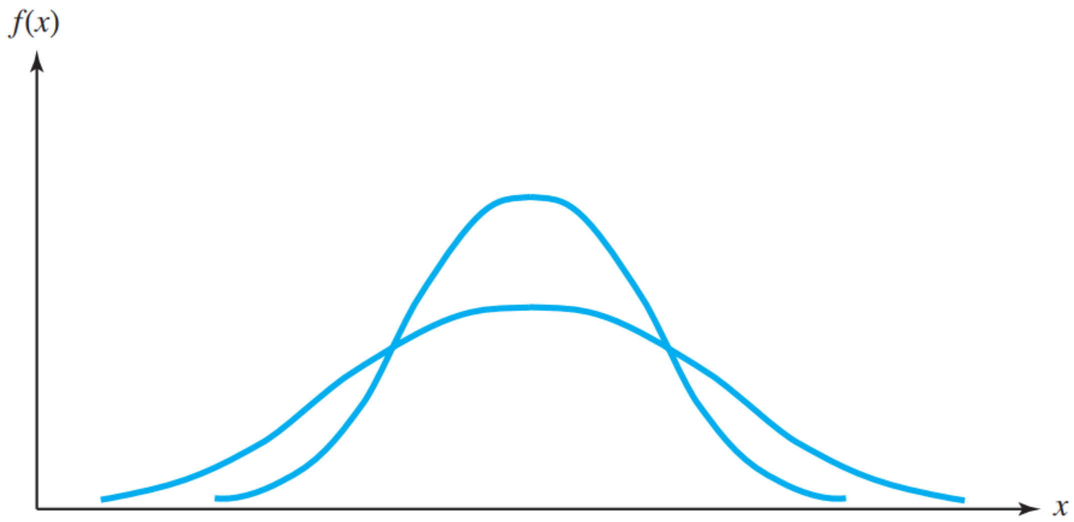
The variance is a positive quantity that measures the spread of the distribution of the random variable about its mean value. Larger values of the variance indicate that the distribution is more spread out.

The variance of a random variable X is

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



Two distributions with different mean values but identical variances



Two distributions with identical mean values but different variances

Example

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

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

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.

Solution: Calculating $E(X)$ and $E(X^2$, we have

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

Therefore,

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

Theorem

:

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.

Example

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

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

Solution:

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

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

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 variance of the random variable $g(X) = 4X + 3$.

Solution: we found that $\mu_{4X+3} = 8$.

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

Definition

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.

The covariance between two random variables is a measure of the nature of the association between the two. If large values of X often result in large values of Y or small values of X result in small values of Y, positive $X - \mu_X$ will often result in positive $Y - \mu_Y$ and negative $X - \mu_X$ will often result in negative $Y - \mu_Y$. Thus, the product $(X - \mu_X)(Y - \mu_Y)$ will tend to be positive. On the other hand, if large X values often result in small Y values, the product $(X - \mu_X)(Y - \mu_Y)$ will tend to be negative. The *sign* of the covariance indicates whether the relationship between two dependent random variables is positive or negative. When X and Y are statistically independent, it can be shown that the covariance is zero (see Corollary 4.5). The converse, however, is not generally true. Two variables may have zero covariance and still not be statistically independent. Note that the covariance only describes the *linear* relationship between two random variables. Therefore, if a covariance between X and Y is zero, X and Y may have a nonlinear relationship, which means that they are not necessarily independent.

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

$$\sigma_{XY} = E(XY) - \mu_X \mu_Y = E[XY] - E[X]E[Y]$$

In practice, the most convenient way to assess the strength of the dependence between two random variables is through their **correlation**.

Correlation

The **correlation** between two random variables X and Y is defined to be

$$\mathrm{Corr}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}.$$

The correlation takes values between -1 and 1, and independent random variables have a correlation of 0.

Example 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.

Find the correlation coefficient between X and Y.

Solution:

$$E(XY) = \sum_{x=0}^{2}\sum_{y=0}^{2} xyf(x,y)$$

$$= (0)(0)f(0,0) + (0)(1)f(0,1)$$

$$+ (1)(0)f(1,0) + (1)(1)f(1,1) + (2)(0)f(2,0)$$

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

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

and

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

Therefore,

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

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

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

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

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

Find the covariance of X and Y.

Find the correlation coefficient of X and Y

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

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

and

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

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

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

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

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

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

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

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

Theorem

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,

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

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

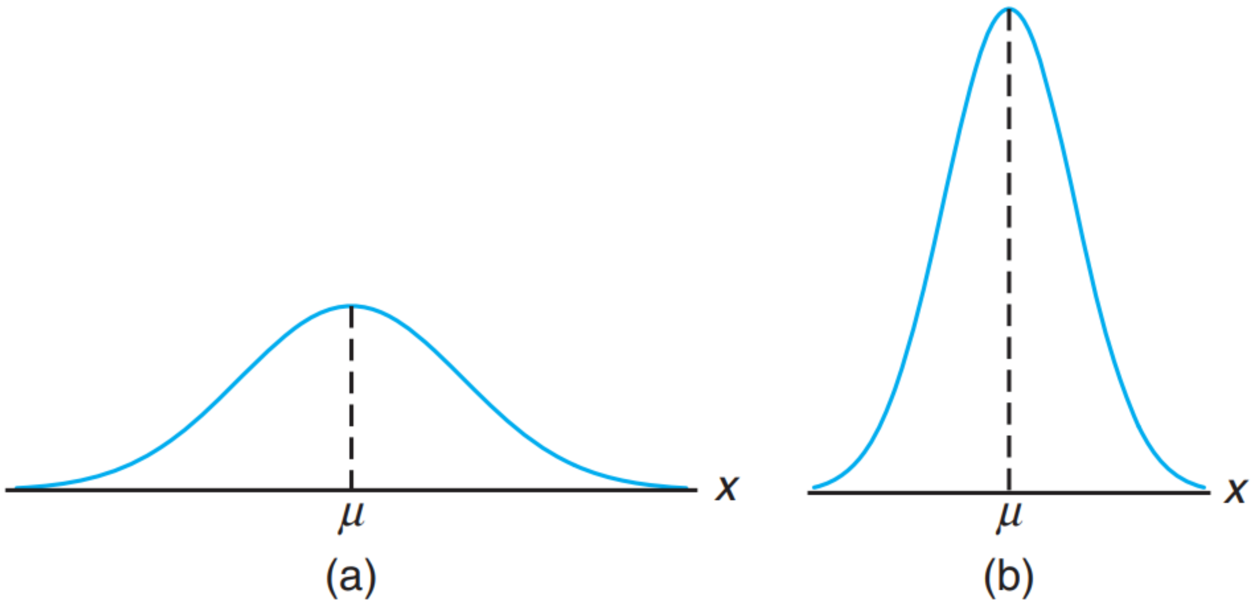
$$E[X \pm Y] = E[X] \pm E[Y].$$

Let X and Y be two independent random variables. Then

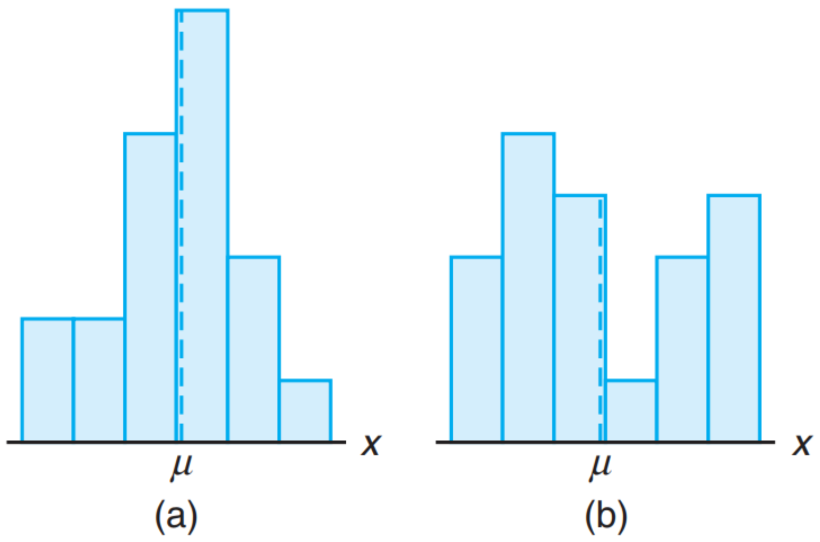
$$E(XY) = E(X)E(Y).$$

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

Chebyshev's Theorem



Variability of continuous observations about the mean.



Variability of discrete observations about the mean

Theorem

(Chebyshev's Theorem) The probability that any random variable X will assume a value within k standard deviations of the mean is at least $1 - 1/k^2$. That is,

$$P(\mu - k\sigma < X < \mu + k\sigma) \geq 1 - \frac{1}{k^2}.$$

For $k = 2$, the theorem states that the random variable X has a probability of at least $1 - 1/2^2 = 3/4$ of falling within two standard deviations of the mean. That is, three-fourths or more of the observations of any distribution lie in the interval $\mu \pm 2\sigma$. Similarly, the theorem says that at least eight-ninths of the observations of any distribution fall in the interval $\mu \pm 3\sigma$.

Example

A random variable X has a mean $\mu = 8$, a variance $\sigma^2 = 9$, and an unknown probability distribution. Find

(a) $P(-4 < X < 20)$,

(b) $P(|X - 8| \geq 6)$.

Solution:

(a) $P(-4 < X < 20) = P[8 - (4)(3) < X < 8 + (4)(3)] \geq \frac{15}{16}$.

(b) $P(|X - 8| \geq 6) = 1 - P(|X - 8| < 6) = 1 - P(-6 < X - 8 < 6)$

$$= 1 - P[8 - (2)(3) < X < 8 + (2)(3)] \leq \frac{1}{4}.$$

Chebyshev's theorem holds for any distribution of observations, and for this reason the results are usually weak. The value given by the theorem is a lower bound only. That is, we know that the probability of a random variable falling within two standard deviations of the mean can be *no less* than 3/4, but we never know how much more it might actually be. Only when the probability distribution is known can we determine exact probabilities. For this reason we call the theorem a *distribution-free* result. When specific distributions are assumed, as in future chapters, the results will be less conservative. The use of Chebyshev's theorem is relegated to situations where the form of the distribution is unknown.

Chapter 5

Some Discrete Probability Distributions

Bernoulli Random Variables

A **Bernoulli** random variable with parameter p, $0 \leq p \leq 1$, takes the values 0 and 1 with $P(X = 0) = 1 - p$ and $P(X = 1) = p$. The expectation and variance of the random variable are

$$E(X) = p \qquad \text{and} \qquad \mathrm{Var}(X) = p(1 - p)$$

An experiment that has only two outcomes is often referred to as a *Bernoulli trial*.

$$E(X) = (0 \times P(X = 0)) + (1 \times P(X = 1)) = (0 \times (1 - p)) + (1 \times p) = p$$

$$E(X^2) = (0^2 \times P(X = 0)) + (1^2 \times P(X = 1)) = p$$

their variance is

$$\mathrm{Var}(X) = E(X^2) - (E(X))^2 = p - p^2 = p(1 - p)$$

Definition of the Binomial Distribution

Many processes can be thought of as consisting of a *sequence* of Bernoulli trials, such as, for example, the repeated tossing of a coin or the repeated examination of objects to determine whether or not they are defective. In such cases, a random variable of interest is the *number of successes* obtained within a fixed number of trials n, where a "success" is defined in an appropriate manner. Such a random variable is called a **binomial** random variable and it is probably the most important of all discrete probability distributions.

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

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

The Binomial Distribution

Consider an experiment consisting of

- n Bernoulli trials
- that are independent and
- that each have a constant probability p of success.

Then the total number of successes X is a random variable that has a **binomial** distribution with parameters n and p, which is written

$$X \sim B(n, p)$$

The probability mass function of a $B(n, p)$ random variable is

$$P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}$$

$$q = 1 - p.$$

for $x = 0, 1, \ldots, n$, with

$$E(X) = np \qquad \text{and} \quad \sigma^2 = \text{Var}(X) = np(1 - p) = npq.$$

There are many experiments that conform either exactly or approximately to the following list of requirements:

1. The experiment consists of a sequence of n smaller experiments called *trials,* where n is fixed in advance of the experiment.
2. Each trial can result in one of the same two possible outcomes (dichotomous trials), which we generically denote by success (S) and failure (F).
3. The trials are independent, so that the outcome on any particular trial does not influence the outcome on any other trial.
4. The probability of success $P(S)$ is constant from trial to trial; we denote this probability by p.

DEFINITION

An experiment for which Conditions 1–4 are satisfied is called a **binomial experiment.**

Example

The probability that a certain kind of component will survive a shock test is $3/4$. Find the probability that exactly 2 of the next 4 components tested survive.

***Solution*:**

Assuming that the tests are independent and $p = 3/4$ for each of the 4 tests, we obtain

$$b\left(2; 4, \frac{3}{4}\right) = \binom{4}{2}\left(\frac{3}{4}\right)^2\left(\frac{1}{4}\right)^2 = \left(\frac{4!}{2!\,2!}\right)\left(\frac{3^2}{4^4}\right) = \frac{27}{128}.$$

Example 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.

$$\text{(a)} \qquad P(X \geq 10) = 1 - P(X < 10) = 1 - \sum_{x=0}^{9} b(x; 15, 0.4) = 1 - 0.9662$$
$$= 0.0338$$

$$\text{(b)} \quad P(3 \leq X \leq 8) = \sum_{x=3}^{8} b(x; 15, 0.4) = \sum_{x=0}^{8} b(x; 15, 0.4) - \sum_{x=0}^{2} b(x; 15, 0.4)$$
$$= 0.9050 - 0.0271 = 0.8779$$

$$\text{(c)} \qquad P(X = 5) = b(5; 15, 0.4) = \sum_{x=0}^{5} b(x; 15, 0.4) - \sum_{x=0}^{4} b(x; 15, 0.4)$$
$$= 0.4032 - 0.2173 = 0.1859$$

Example Find the mean and variance of the binomial random variable of Example and then use Chebyshev's theorem (on page 157) 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.

EXAMPLE

At any given time, a professional basketball player can make a free throw with probability equal to .8. Suppose he shoots four free throws.

1. What is the probability that he will make exactly two free throws?
2. What is the probability that he will make at least one free throw?

Solution A "trial" is a single free throw. Define a "success" as a basket and a "failure" as a miss, so that $n = 4$ and $p = .8$. If you assume that the player's chance of making the free throw does not change from shot to shot, then x, the number of free throws that he makes is a *binomial random variable*.

1. $P(x = 2) = C_2^4(.8)^2(.2)^2$

$$= \frac{4!}{2!2!}(.64)(.04) = \frac{4(3)(2)(1)}{2(1)(2)(1)}(.64)(.04) = .1536$$

The probability is .1536 that he will make exactly two free throws.

2. $$\begin{aligned} P(\text{at least one}) = P(x \geq 1) &= p(1) + p(2) + p(3) + p(4) \\ &= 1 - p(0) \\ &= 1 - C_0^4 (.8)^0 (.2)^4 \\ &= 1 - .0016 = .9984. \end{aligned}$$

Although you could have calculated $P(x = 1)$, $P(x = 2)$, $P(x = 3)$, and $P(x = 4)$ to find this probability, using the complement of the event made your job easier; that is,

$$P(x \geq 1) = 1 - P(x < 1) = 1 - P(x = 0).$$

Can you think of any reason your assumption of independent trials might be wrong? If the player learns from his previous attempt (that is, he adjusts his shooting according to his last attempt), then his probability p of making the free throw may change, possibly increase, from shot to shot. The trials would *not* be independent and the experiment would *not* be binomial.

Example 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,

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

(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,

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

Example | 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, $P(X > 3) = 1 - 0.6496 = 0.3504$.

Poisson Distribution and the Poisson Process

Experiments yielding numerical values of a random variable X, the number of outcomes occurring during a given time interval or in a specified region, are called **Poisson experiments**. The given time interval may be of any length, such as a minute, a day, a week, a month, or even a year. For example, a Poisson experiment can generate observations for the random variable X representing the number of telephone calls received per hour by an office, the number of days school is closed due to snow during the winter, or the number of games postponed due to rain during a baseball season. The specified region could be a line segment, an area, a volume, or perhaps a piece of material. In such instances, X might represent the number of field mice per acre, the number of bacteria in a given culture, or the number of typing errors per page. A Poisson experiment is derived from the **Poisson process** and possesses the following properties.

Properties of the Poisson Process

1. The number of outcomes occurring in one time interval or specified region of space is independent of the number that occur in any other disjoint time interval or region. In this sense we say that the Poisson process has no memory.

2. The probability that a single outcome will occur during a very short time interval or in a small region is proportional to the length of the time interval or the size of the region and does not depend on the number of outcomes occurring outside this time interval or region.

3. The probability that more than one outcome will occur in such a short time interval or fall in such a small region is negligible.

The number X of outcomes occurring during a Poisson experiment is called a **Poisson random variable**, and its probability distribution is called the **Poisson distribution**. The mean number of outcomes is computed from $\mu = \lambda t$, where t is the specific "time," "distance," "area," or "volume" of interest. Since the probabilities depend on λ, the rate of occurrence of outcomes, we shall denote them by $p(x; \lambda t)$.

Some examples of random variables that usually obey, to a good approximation, the Poisson probability law (that is, they usually obey equation for some value of λ) are:

1. The number of misprints on a page (or a group of pages) of a book.
2. The number of people in a community living to 100 years of age.
3. The number of wrong telephone numbers that are dialed in a day.
4. The number of transistors that fail on their first day of use.

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 $e = 2.71828\ldots$.

Theorem

$\boxed{\text{Both the mean and the variance of the Poisson distribution } p(x; \lambda t) \text{ are } \lambda t.}$

Example

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

$$p(6;4) = \frac{e^{-4}4^6}{6!} = \sum_{x=0}^{6} p(x;4) - \sum_{x=0}^{5} p(x;4) = 0.8893 - 0.7851 = 0.1042.$$

Example

Ten is the average number of oil tankers arriving each day at a certain port. The facilities at the port can handle at most 15 tankers per day. What is the probability that on a given day tankers have to be turned away?

***Solution*:**

Let X be the number of tankers arriving each day. Then, using Table A.2, we have

$$P(X > 15) = 1 - P(X \leq 15) = 1 - \sum_{x=0}^{15} p(x; 10) = 1 - 0.9513 = 0.0487.$$

Like the binomial distribution, the Poisson distribution is used for quality control, quality assurance, and acceptance sampling.

Approximation of Binomial Distribution by a Poisson Distribution

In the case of the binomial, if n is quite large and p is small, the conditions begin to simulate the *continuous space or time* implications of the Poisson process.

Let X be a binomial random variable with probability distribution $b(x; n, p)$. When $n \to \infty$, $p \to 0$, and $np \overset{n\to\infty}{\longrightarrow} \mu$ remains constant,

$$b(x; n, p) \overset{n\to\infty}{\longrightarrow} p(x; \mu).$$

Example

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

Example

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

$$P(X < 7) = \sum_{x=0}^{6} b(x; 8000, 0.001) \approx p(x; 8) = 0.3134.$$

Independent Bernoulli Trials

Binomial

e.g. If a coin is tossed 20 times, what is the probability heads comes up exactly 14 times?

The number of trials is fixed.
The number of successes is a random variable.

Negative Binomial

e.g. If a coin is repeatedly tossed, what is the probability the third time heads appears occurs on the 9th toss?

The number of successes is fixed.
The number of trials is a random variable.

Geometric

e.g. If a coin is repeatedly tossed, what is the probability the *first* time heads appears occurs on the 8th toss?

The *geometric* distribution is the distribution of the number of trials needed to get the first success in repeated independent Bernoulli trials.

The *negative binomial* distribution is the distribution of the number of trials needed to get the rth success.

The Negative Binomial Distribution

The number of trials up to and including the rth success in a sequence of independent Bernoulli trials with a constant success probability p has a **negative binomial** distribution with parameters p and r. The probability mass function is

$$P(X = x) = \binom{x-1}{r-1}(1-p)^{x-r}p^r = b^*(x; r, p)$$

for $x = r, r+1, r+2, r+3, \ldots,$ with an expected value and a variance of

$$E(X) = \frac{r}{p} \qquad \text{and} \qquad \mathrm{Var}(X) = \frac{r(1-p)}{p^2}$$

The Geometric Distribution

The number of trials up to and including the *first* success in a sequence of independent Bernoulli trials with a constant success probability p has a **geometric** distribution with parameter p. The probability mass function is

$$P(X = x) = (1 - p)^{x-1}\ p$$

for $x = 1, 2, 3, 4 \ldots,$ and the cumulative distribution function is

$$P(X \leq x) = 1 - (1 - p)^{x}$$

The geometric distribution with parameter p has an expected value and a variance of

$$E(X) = \frac{1}{p} \qquad \text{and} \qquad \text{Var}(X) = \frac{1 - p}{p^2}$$

Example

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) $P(\text{team } A \text{ wins the championship series})$ is

$$\begin{aligned} &b^*(4;4,0.55) + b^*(5;4,0.55) + b^*(6;4,0.55) + b^*(7;4,0.55) \\ &= 0.0915 + 0.1647 + 0.1853 + 0.1668 = 0.6083. \end{aligned}$$

(c) $P(\text{team } A \text{ wins the playoff})$ is

$$\begin{aligned} &b^*(3;3,0.55) + b^*(4;3,0.55) + b^*(5;3,0.55) \\ &= 0.1664 + 0.2246 + 0.2021 = 0.5931. \end{aligned}$$

Problem

Let X be a geometric random variable with $p = 0.25$. What is the probability that $X = 4$ (i.e. that the first success occurs on the 4th trial)?

Note: For X to be equal to 4, we must have had 3 failures, and then a success.

Solution

$$P\{X = 4\} = (0.25)(1 - 0.25)^{4-1} = 0.1055$$

Problem

For a certain manufacturing process, it is known that 1 in every 100 items is defective. What is the probability that the fifth item inspected is the first defective item found?

Solution

Let X is the number of items inspected before we found a defective item.

p =probability of being defective item= $1/100$, then X is a geometric R.V. with PMF given by

$$P\{X = i\} = (1/100)(1 - 1/100)^{i-1}, i = 1,2,3, \dots$$

Now

$$P\{X = 5\} = (1/100)(1 - 1/100)^{5-1} = 0.0096$$

Example

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

$$g(5; 0.01) = (0.01)(0.99)^4 = 0.0096.$$

GAMES OF CHANCE

If a fair die is repeatedly thrown, the number of throws made until a 6 is obtained has a geometric distribution with parameter $p = 1/6$. The expected number of throws made until a 6 is obtained is therefore

$$E(X) = \frac{1}{p} = 6$$

The probability that a 6 is *not* obtained in the first six throws is

$$P(X \geq 7) = 1 - P(X \leq 6) = 1 - \left(1 - \left(\frac{5}{6}\right)^6\right) = \left(\frac{5}{6}\right)^6 = 0.335$$

The distribution of the number of throws made until a 6 is obtained for the *second* time has a negative binomial distribution with parameters $p = 1/6$ and $r = 2$. The expected number of throws required is

$$E(X) = \frac{r}{p} = 12$$

Hypergeometric

e.g. If 5 cards are drawn without replacement, what is the probability 3 hearts are drawn?

Binomial

If the cards are drawn *with replacement*, this would be a binomial problem.

The Hypergeometric Distribution

The **hypergeometric distribution** has a probability mass function given by

$$P(X = x) = \frac{\binom{r}{x} \times \binom{N-r}{n-x}}{\binom{N}{n}}$$

for $\max\{0, n + r - N\} \leq x \leq \min\{n, r\}$, with an expected value of

$$E(X) = \frac{nr}{N}$$

and a variance of

$$\mathrm{Var}(X) = \left(\frac{N-n}{N-1}\right) \times n \times \frac{r}{N} \times \left(1 - \frac{r}{N}\right)$$

It represents the distribution of the number of items of a certain kind in a random sample of size n drawn *without replacement* from a population of size N that contains r items of this kind.

Like the binomial distribution, the hypergeometric distribution finds applications in acceptance sampling, where lots of materials or parts are sampled in order to determine whether or not the entire lot is accepted.

Example

A particular part that is used as an injection device is sold in lots of 10. The producer deems a lot acceptable if no more than one defective is in the lot. A sampling plan involves random sampling and testing 3 of the parts out of 10. If none of the 3 is defective, the lot is accepted. Comment on the utility of this plan.

***Solution*:**

Let us assume that the lot is truly **unacceptable** (i.e., that 2 out of 10 parts are defective). The probability that the sampling plan finds the lot acceptable is

$$P(X=0) = \frac{\binom{2}{0}\binom{8}{3}}{\binom{10}{3}} = 0.467.$$

Example

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.

EXAMPLE

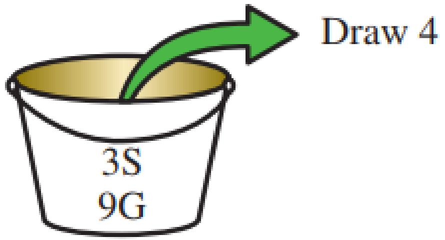
A case of wine has 12 bottles, 3 of which contain spoiled wine. A sample of 4 bottles is randomly selected from the case.

1. Find the probability distribution for x, the number of bottles of spoiled wine in the sample.
2. What are the mean and variance of x?

Solution For this example, $N = 12$, $n = 4$, $M = 3$, and $(N - M) = 9$. Then

$$p(x) = \frac{C_x^3 C_{4-x}^9}{C_4^{12}}$$

● **Need a Tip?**



1. The possible values for x are 0, 1, 2, and 3, with probabilities

$$p(0) = \frac{C_0^3 C_4^9}{C_4^{12}} = \frac{1(126)}{495} = .25 \qquad p(2) = \frac{C_2^3 C_2^9}{C_4^{12}} = \frac{3(36)}{495} = .22$$

$$p(1) = \frac{C_1^3 C_3^9}{C_4^{12}} = \frac{3(84)}{495} = .51 \qquad p(3) = \frac{C_3^3 C_1^9}{C_4^{12}} = \frac{1(9)}{495} = .02$$

2. The mean is given by

$$\mu = 4\left(\frac{3}{12}\right) = 1$$

and the variance is

$$\sigma^2 = 4\left(\frac{3}{12}\right)\left(\frac{9}{12}\right)\left(\frac{12-4}{11}\right) = .5455$$

EXAMPLE

An industrial product is shipped in lots of 20. Testing to determine whether an item is defective is costly; hence, the manufacturer samples production rather than using a 100% inspection plan. The sampling plan calls for sampling five items from each lot and rejecting the lot if more than one defective is observed. (If the lot is rejected, each item in the lot is then tested to isolate the defectives.) If a lot contains four defectives, what is the probability that it will be accepted?

Solution Let x be the number of defectives in the sample. Then $N = 20$, $M = 4$, $(N - M) = 16$, and $n = 5$. The lot will be rejected if $x = 2, 3,$ or 4. Then

$$P(\text{accept the lot}) = P(x \le 1) = p(0) + p(1) = \frac{C_0^4 C_5^{16}}{C_5^{20}} + \frac{C_1^4 C_4^{16}}{C_5^{20}}$$

$$= \frac{\left(\frac{4!}{0!4!}\right)\left(\frac{16!}{5!11!}\right)}{\frac{20!}{5!15!}} + \frac{\left(\frac{4!}{1!3!}\right)\left(\frac{16!}{4!12!}\right)}{\frac{20!}{5!15!}}$$

$$= \frac{91}{323} + \frac{455}{969} = .2817 + .4696 = .7513$$

Multinomial Distribution

If a given trial can result in the k outcomes $E_1, E_2, \ldots, E_k$ with probabilities $p_1, p_2, \ldots, p_k$, then the probability distribution of the random variables $X_1, X_2, \ldots, X_k$, representing the number of occurrences for $E_1, E_2, \ldots, E_k$ in n independent trials, is

$$f(x_1, x_2, \ldots, x_k; p_1, p_2, \ldots, p_k, n) = \binom{n}{x_1, x_2, \ldots, x_k} p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k},$$

with

$$\sum_{i=1}^{k} x_i = n \text{ and } \sum_{i=1}^{k} p_i = 1.$$

$$\binom{n}{x_1, x_2, \ldots, x_k} = \frac{n!}{x_1! \, x_2! \cdots x_k!}$$

Example

The complexity of arrivals and departures of planes at an airport is such that computer simulation is often used to model the "ideal" conditions. For a certain airport with three runways, it is known that in the ideal setting the following are the probabilities that the individual runways are accessed by a randomly arriving commercial jet:

$$\begin{aligned} &\text{Runway 1:} \quad p_1 = 2/9, \\ &\text{Runway 2:} \quad p_2 = 1/6, \\ &\text{Runway 3:} \quad p_3 = 11/18. \end{aligned}$$

What is the probability that 6 randomly arriving airplanes are distributed in the following fashion?

$$\begin{aligned} &\text{Runway 1:} \quad \text{2 airplanes}, \\ &\text{Runway 2:} \quad \text{1 airplane}, \\ &\text{Runway 3:} \quad \text{3 airplanes} \end{aligned}$$

***Solution*:**

Using the multinomial distribution, we have

$$f\left(2,1,3;\frac{2}{9},\frac{1}{6},\frac{11}{18},6\right)=\binom{6}{2,1,3}\left(\frac{2}{9}\right)^2\left(\frac{1}{6}\right)^1\left(\frac{11}{18}\right)^3$$
$$=\frac{6!}{2!\,1!\,3!}\cdot\frac{2^2}{9^2}\cdot\frac{1}{6}\cdot\frac{11^3}{18^3}=0.1127.$$

Distribution	$f(x)$	Support	Mean	Variance
Bernoulli (p)	$p^x(1-p)^{1-x}$	$0, 1$	p	$p(1-p)$
binomial(n, p)	$\binom{n}{x} p^x(1-p)^{n-x}$	$0, 1, \ldots, n$	np	$np(1-p)$
geometric(p)	$p(1-p)^x$	$0, 1, 2, \ldots$	$\frac{1-p}{p}$	$\frac{1-p}{p^2}$
neg. binom.(r, p)	$\binom{x+r-1}{r-1} p^r(1-p)^x$	$0, 1, 2, \ldots$	$\frac{r(1-p)}{p}$	$\frac{r(1-p)}{p^2}$
Poisson(λ)	$\frac{\lambda^x e^{-\lambda}}{x!}$	$0, 1, 2, \ldots$	λ	λ
hypergeom.(m, N, n)	$\frac{\binom{m}{x}\binom{N-m}{n-x}}{\binom{N}{n}}$	$0, 1, \ldots, n$	$\frac{nm}{N}$	$\frac{nm(N-n)(N-m)}{N^2(N-1)}$

Chapter 6

Some Continuous Probability Distributions

Continuous Uniform Distribution

One of the simplest continuous distributions in all of statistics is the **continuous uniform distribution**. This distribution is characterized by a density function that is "flat," and thus the probability is uniform in a closed interval, say $[A, B]$.

The density function of the continuous uniform random variable X on the interval $[A, B]$ is

$$f(x; A, B) = \begin{cases} \frac{1}{B-A}, & A \leq x \leq B, \\ 0, & \text{elsewhere.} \end{cases}$$

The density function forms a rectangle with base $B-A$ and **constant height** $\frac{1}{B-A}$. As a result, the uniform distribution is often called the **rectangular distribution**.

The mean and variance of the uniform distribution are

$$\mu = \frac{A+B}{2} \text{ and } \sigma^2 = \frac{(B-A)^2}{12}.$$

Example

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 $[0, 4]$.

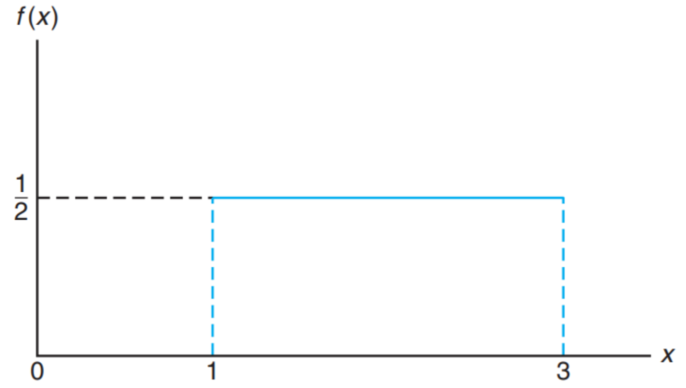
(a) What is the probability density function?

(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

$$f(x) = \begin{cases} \frac{1}{4}, & 0 \leq x \leq 4, \\ 0, & \text{elsewhere.} \end{cases}$$



: The density function for a random variable on the interval $[1, 3]$.

(b) $P[X \geq 3] = \int_3^4 \frac{1}{4}\, dx = \frac{1}{4}$.

Normal Distribution

The most important continuous probability distribution in the entire field of statistics is the **normal distribution**. Its graph, called the **normal curve**, is the bell-shaped curve of Figure 6.2, which approximately describes many phenomena that occur in nature, industry, and research. For example, physical measurements in areas such as meteorological experiments, rainfall studies, and measurements of manufactured parts are often more than adequately explained with a normal distribution. In addition, errors in scientific measurements are extremely well approximated by a normal distribution. In 1733, Abraham DeMoivre developed the mathematical equation of the normal curve. It provided a basis from which much of the theory of inductive statistics is founded. The normal distribution is often referred to as the **Gaussian distribution**, in honor of Karl Friedrich Gauss (1777–1855), who also derived its equation from a study of errors in repeated measurements of the same quantity.

A continuous random variable X having the bell-shaped distribution of Figure 6.2 is called a **normal random variable**. The mathematical equation for the probability distribution of the normal variable depends on the two parameters μ and σ, its mean and standard deviation, respectively. Hence, we denote the values of the density of X by $n(x; \mu, \sigma)$.

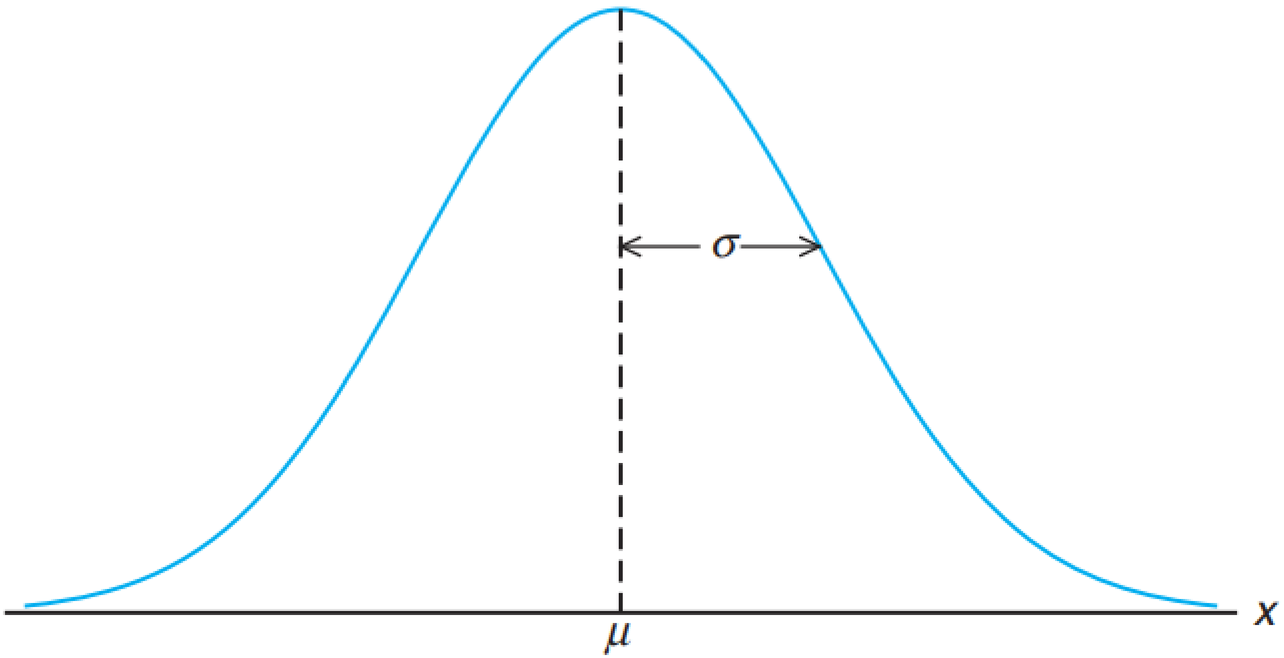


Figure 6.2: The normal curve.

Normal Distribution

The density of the normal random variable X, with mean μ and variance σ^2, is

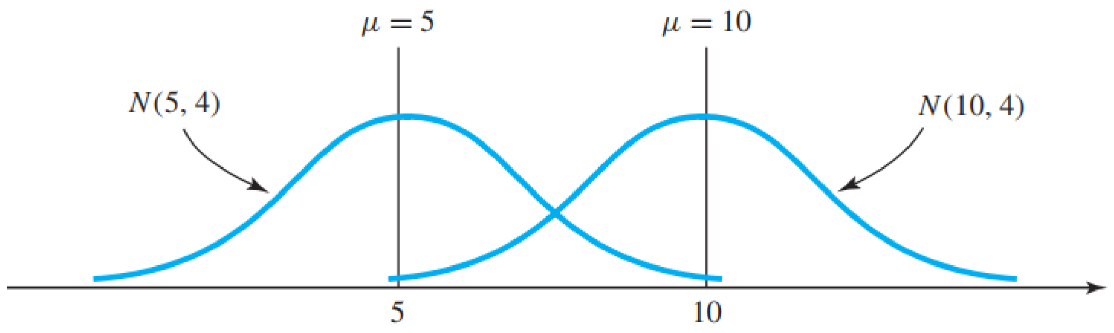
$$n(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}, \qquad -\infty < x < \infty,$$

where $\pi = 3.14159\ldots$ and $e = 2.71828\ldots$.

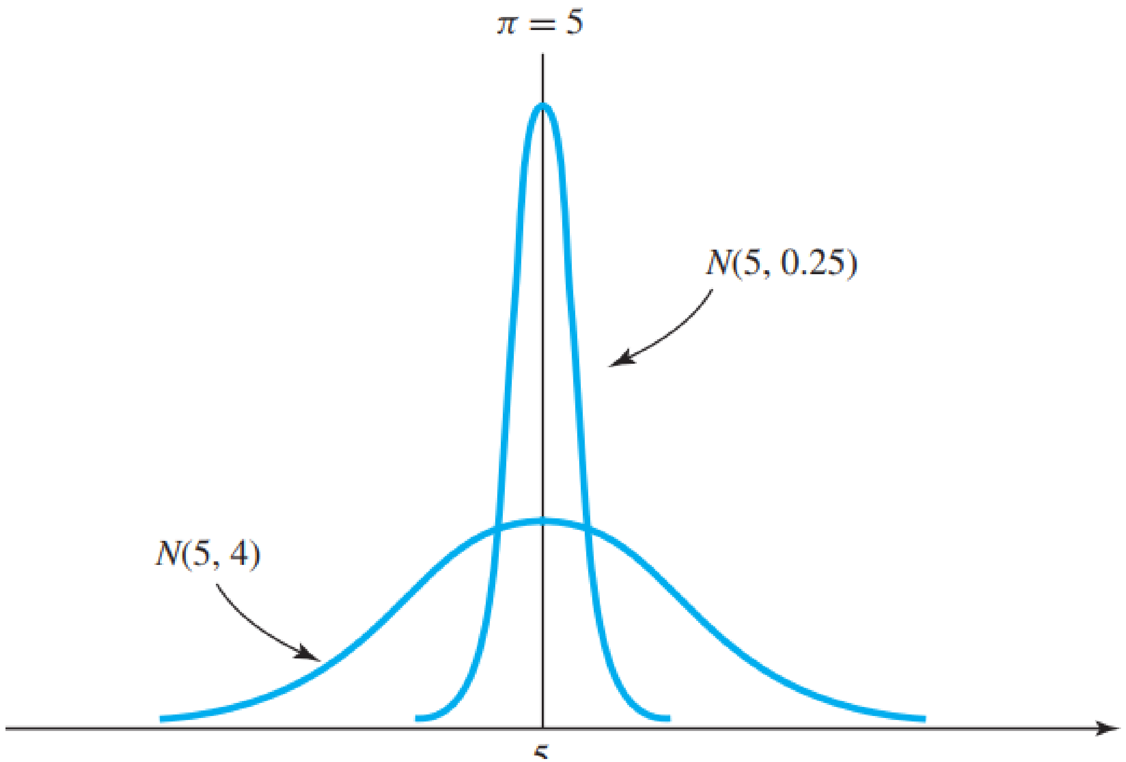
Theorem

The mean and variance of $n(x;\mu,\sigma)$ are μ and σ^2, respectively. Hence, the standard deviation is σ.

$$E(X) = \mu \qquad \text{and} \qquad \mathrm{Var}(X) = \sigma^2$$



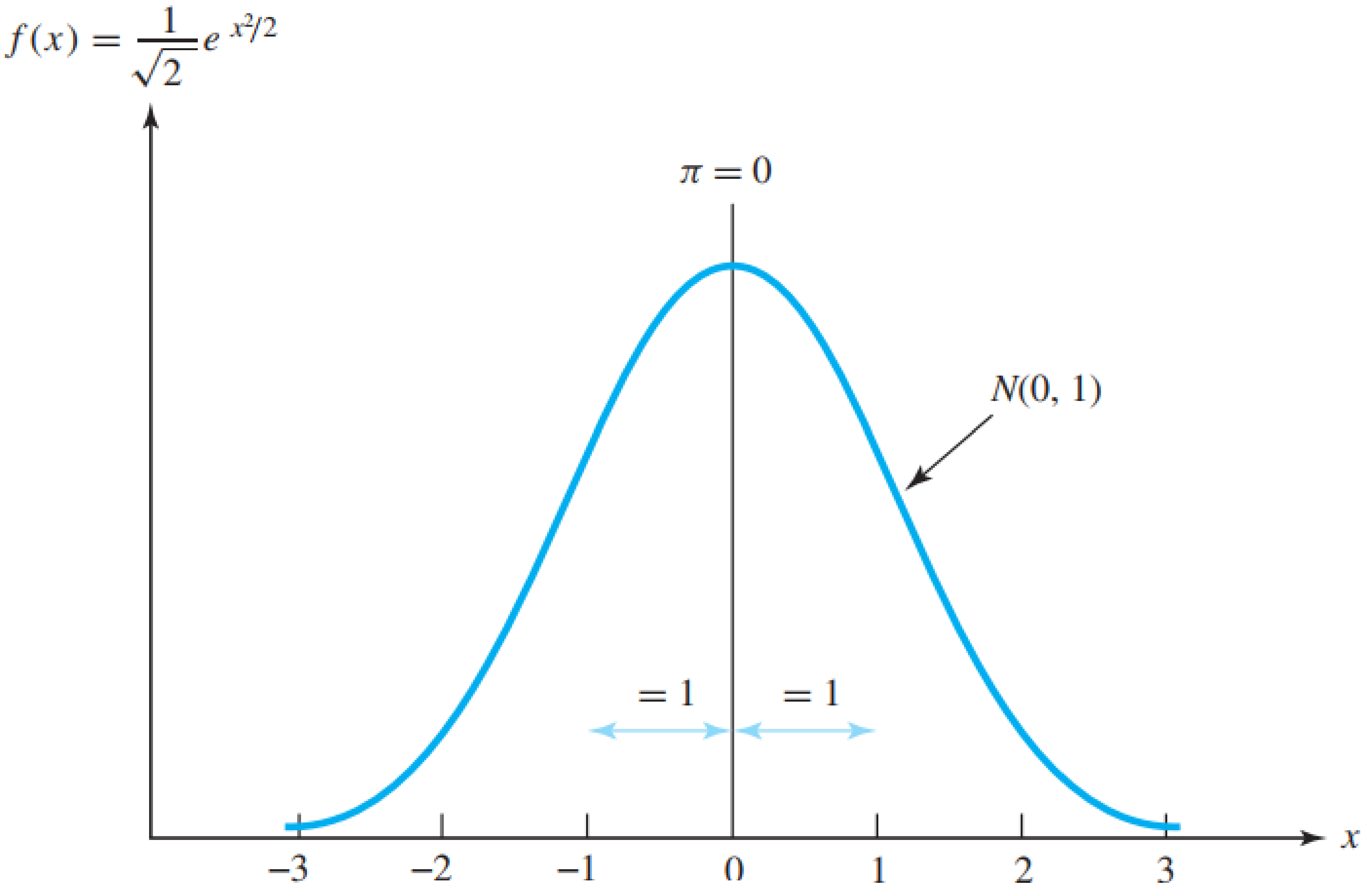
The effect of changing the mean of
a normal distribution



The effect of changing the variance
of a normal distribution

Definition

> The distribution of a normal random variable with mean 0 and variance 1 is called a **standard normal distribution**.



The standard normal distribution

$$Z = \frac{X - \mu}{\sigma}.$$

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2} dy, \qquad -\infty < x < \infty$$

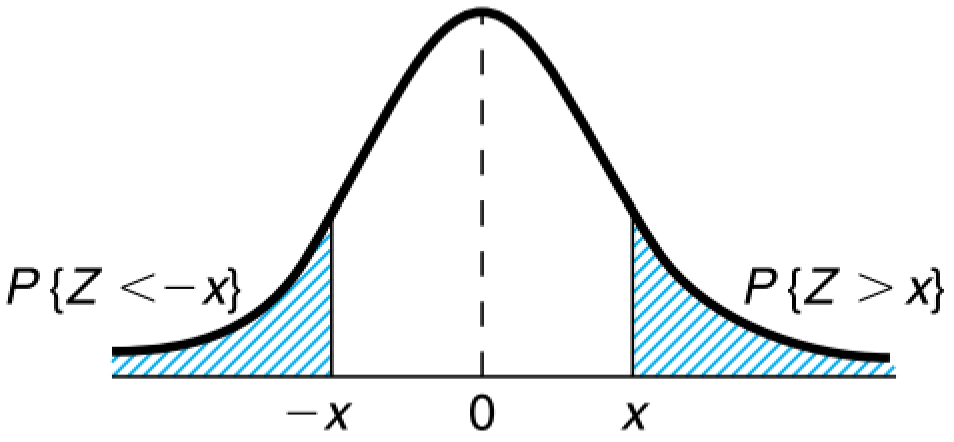
$$P\{X < b\} = P\left\{\frac{X-\mu}{\sigma} < \frac{b-\mu}{\sigma}\right\} = \Phi\left(\frac{b-\mu}{\sigma}\right)$$

$$P\{a < X < b\} = P\left\{\frac{a-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{b-\mu}{\sigma}\right\} = P\left\{\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}\right\}$$

$$= P\left\{Z < \frac{b-\mu}{\sigma}\right\} - P\left\{Z < \frac{a-\mu}{\sigma}\right\}$$

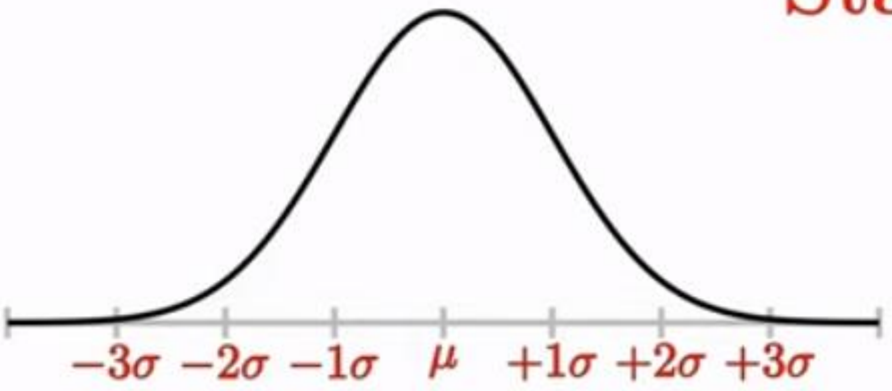
$$= \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

$$P\{Z > x\} = 1 - \Phi(x)$$

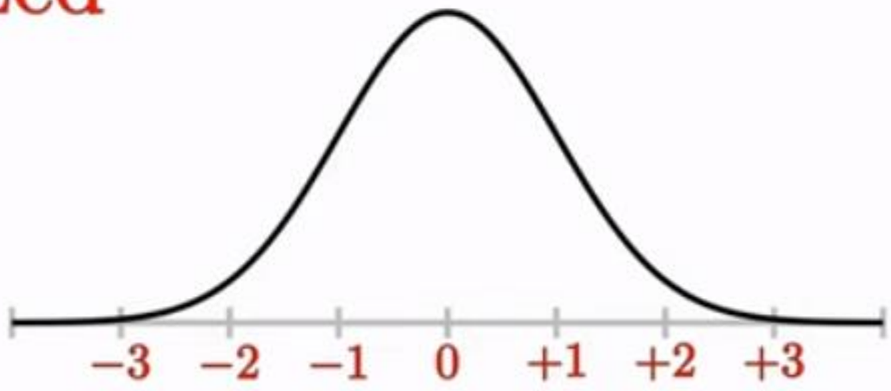


Normal Distribution

Standardized



$\Rightarrow$



Z-score

$$z = \frac{x - \mu}{\sigma}$$

Heights:

$\mu = 66$ in.
$\sigma = 2$ in.

$x = 67$ in. $\Rightarrow$ $z = \frac{67 - 66}{2} = \frac{1}{2} = 0.5$

Example If X is a normal random variable with mean $\mu = 3$ and variance $\sigma^2 = 16$, find

(a) $P\{X < 11\}$;
(b) $P\{X > -1\}$;
(c) $P\{2 < X < 7\}$.

Solution.

(a)

$$P\{X < 11\} = P\left\{\frac{X-3}{4} < \frac{11-3}{4}\right\}$$
$$= \Phi(2)$$
$$= .9772$$

(b)

$$P\{X > -1\} = P\left\{\frac{X-3}{4} > \frac{-1-3}{4}\right\}$$

$$= P\{Z > -1\}$$

$$= P\{Z < 1\}$$

$$= .8413$$

$$\begin{aligned}\Phi(-x) &= P\{Z < -x\}\\ &= P\{Z > x\} \quad \text{by symmetry}\\ &= 1 - \Phi(x)\end{aligned}$$

(c)

$$P\{2 < X < 7\} = P\left\{\frac{2-3}{4} < \frac{X-3}{4} < \frac{7-3}{4}\right\}$$

$$= \Phi(1) - \Phi(-1/4)$$

$$= \Phi(1) - (1 - \Phi(1/4))$$

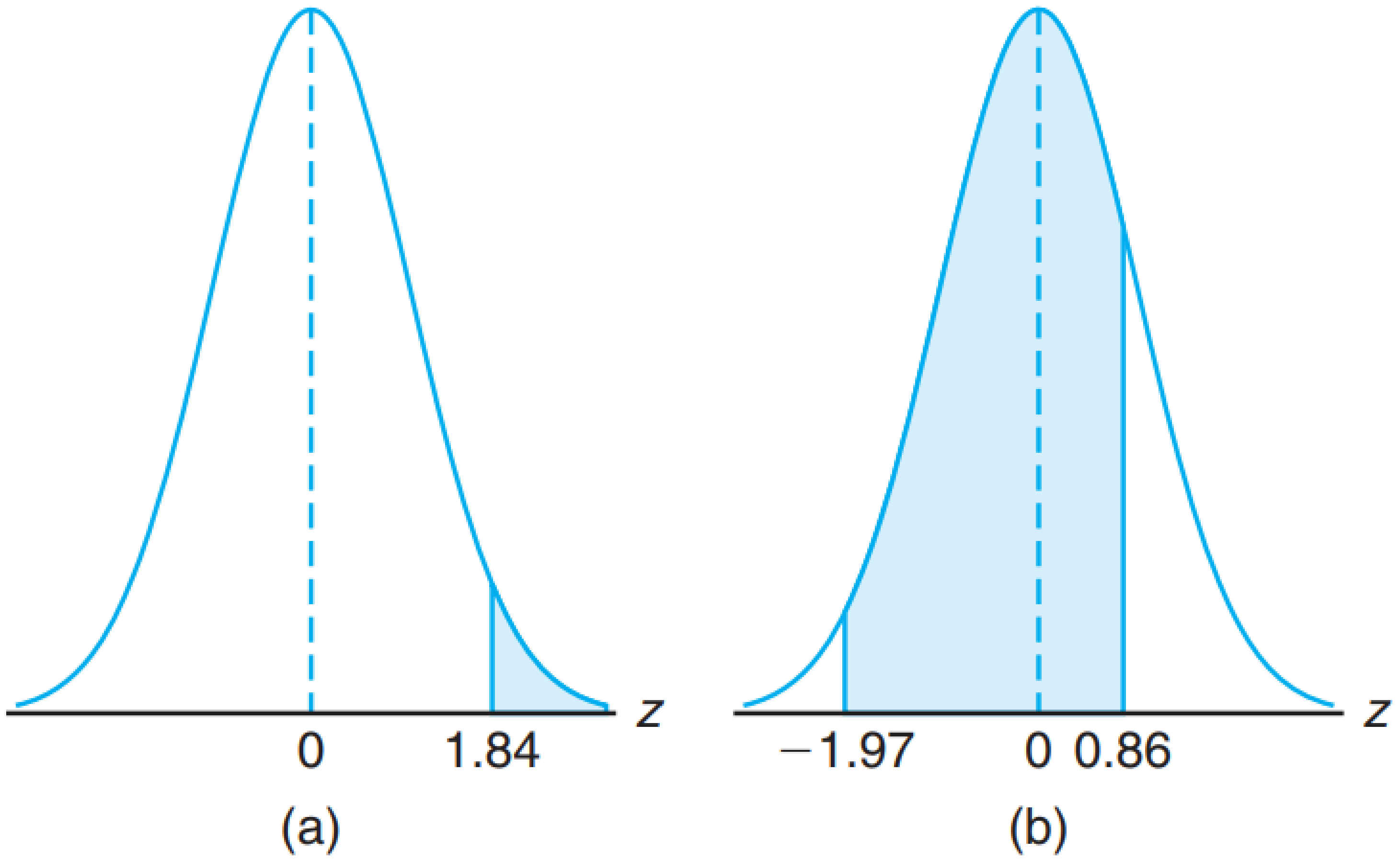
$$= .8413 + .5987 - 1 = .4400 \quad \blacksquare$$

Example

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 $z = 0.86$.



***Solution*:** See Figure

(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, $1 - 0.9671 = 0.0329$.

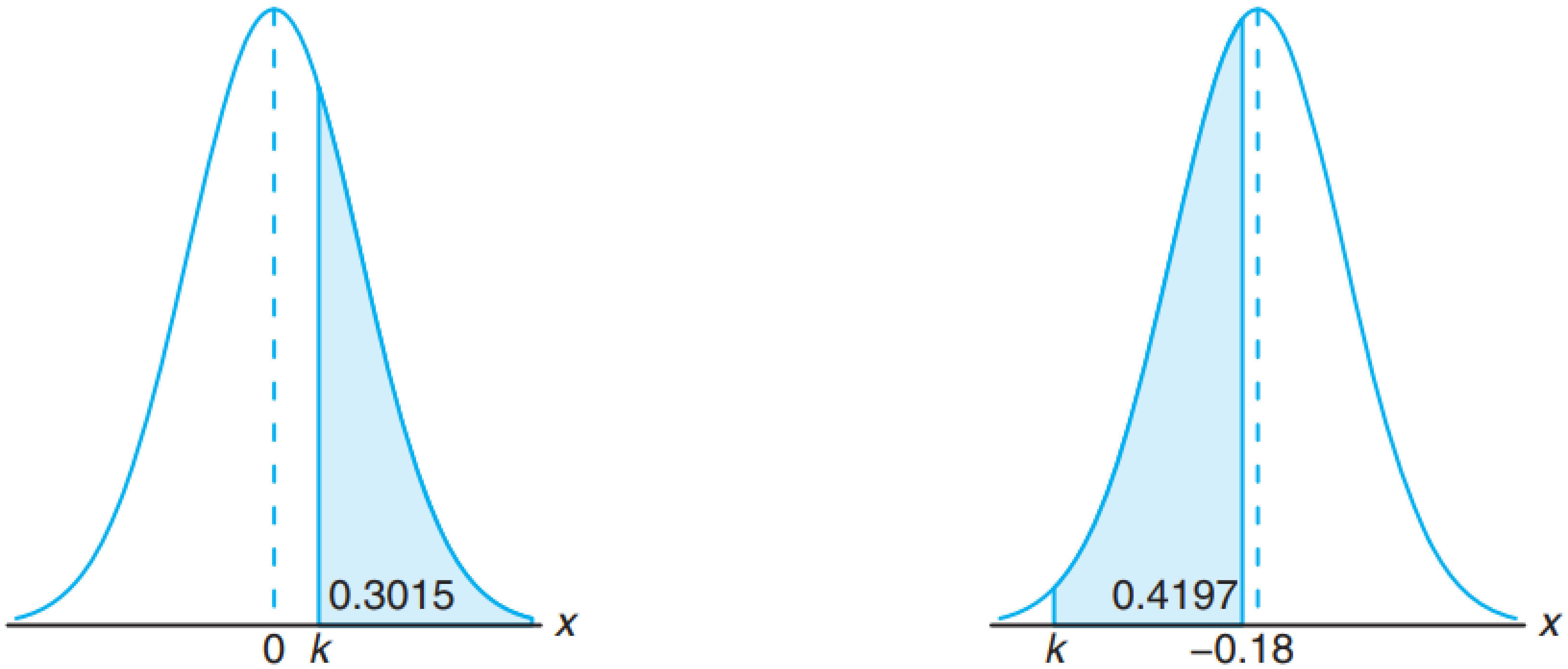
(b) The area in Figure 6.9(b) between $z = -1.97$ and $z = 0.86$ is equal to the area to the left of $z = 0.86$ minus the area to the left of $z = -1.97$. From Table A.3 we find the desired area to be $0.8051 - 0.0244 = 0.7807$.

Example

Given a standard normal distribution, find the value of k such that

(a) $P(Z > k) = 0.3015$ and

(b) $P(k < Z < -0.18) = 0.4197$.



(a) (b)

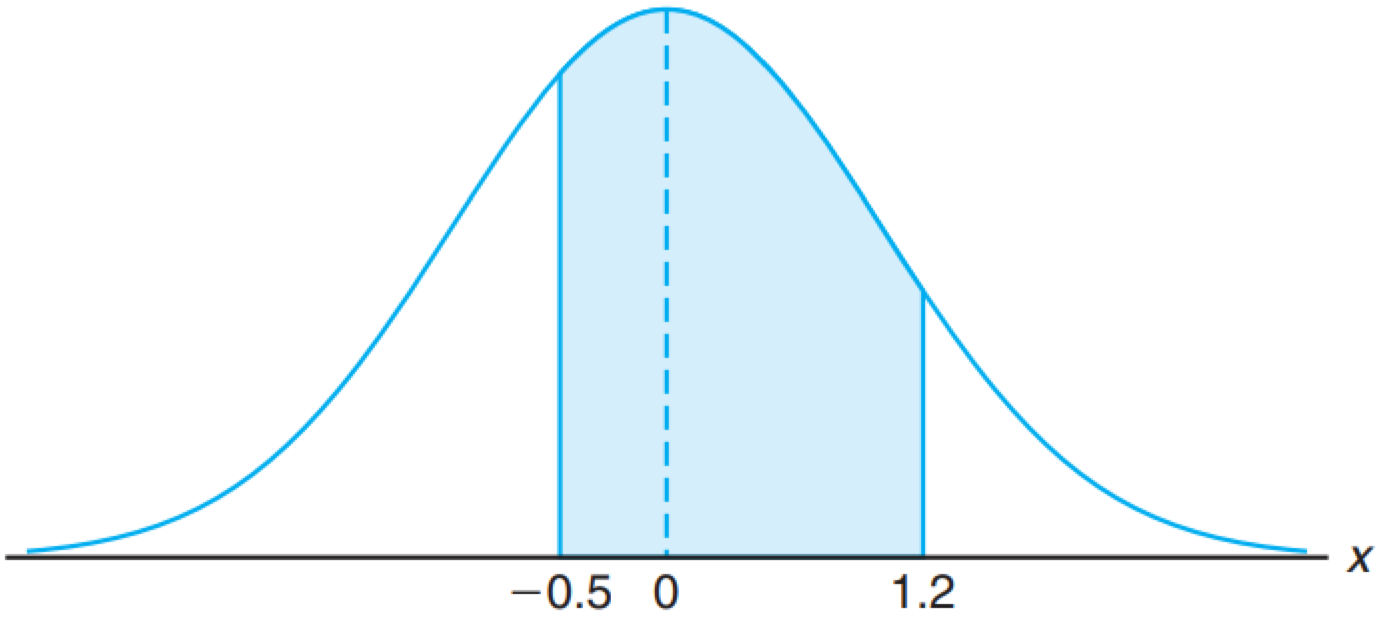
***Solution*:** Distributions and the desired areas are shown in Figure

(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 $k = -2.37$.

Example

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.



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

$$\begin{aligned} P(45 < X < 62) &= P(-0.5 < Z < 1.2) = P(Z < 1.2) - P(Z < -0.5) \\ &= 0.8849 - 0.3085 = 0.5764. \end{aligned}$$

Sometimes, we are required to find the value of z corresponding to a specified probability that falls between values listed in Table A.3 (see Example 6.6). For convenience, we shall always choose the z value corresponding to the tabular probability that comes closest to the specified probability.

The preceding two examples were solved by going first from a value of x to a z value and then computing the desired area. In Example 6.6, we reverse the process and begin with a known area or probability, find the z value, and then determine x by rearranging the formula

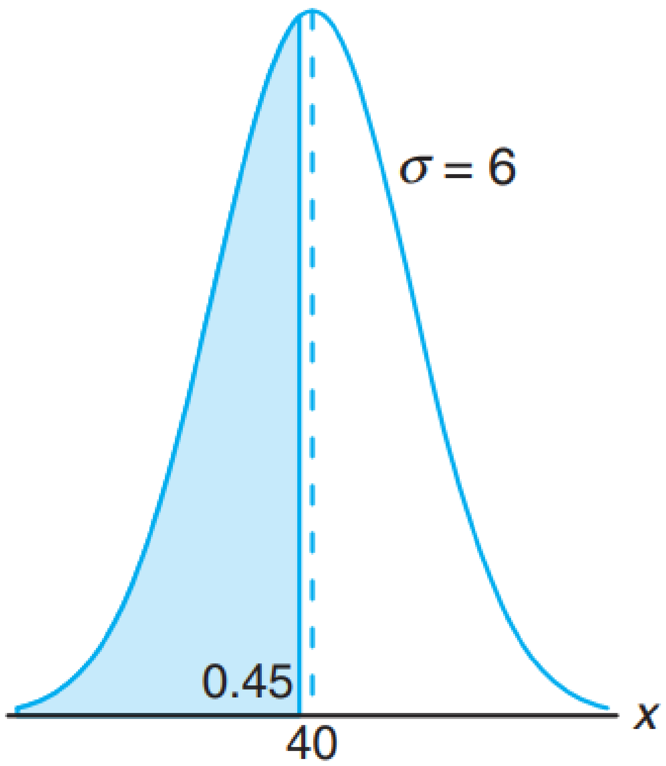
$$z = \frac{x - \mu}{\sigma} \quad \text{to give} \quad x = \sigma z + \mu.$$

Example 6.6:

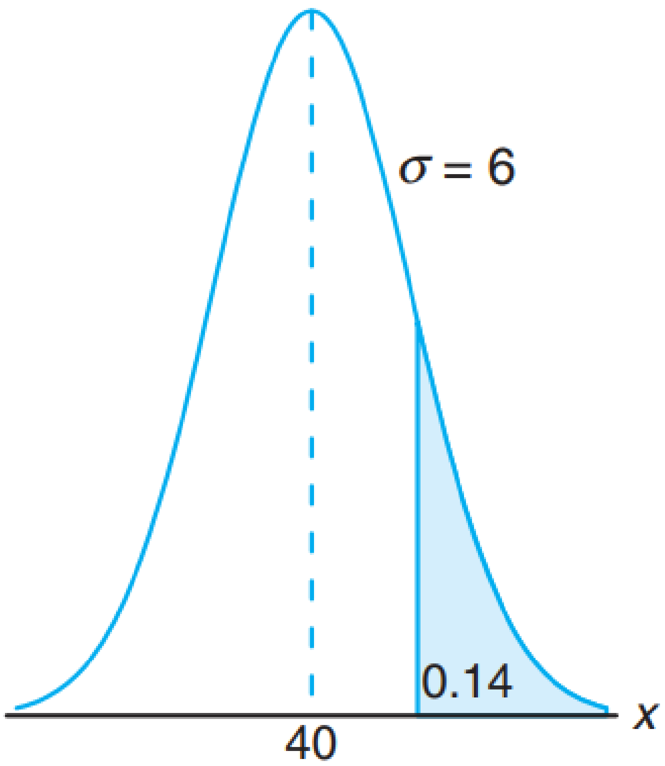
Given a normal distribution with $\mu = 40$ and $\sigma = 6$, find the value of x that has

(a) 45% of the area to the left and

(b) 14% of the area to the right.



(a)



(b)

***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

$$x = (6)(1.08) + 40 = 46.48.$$

Applications of the Normal Distribution

Example

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.

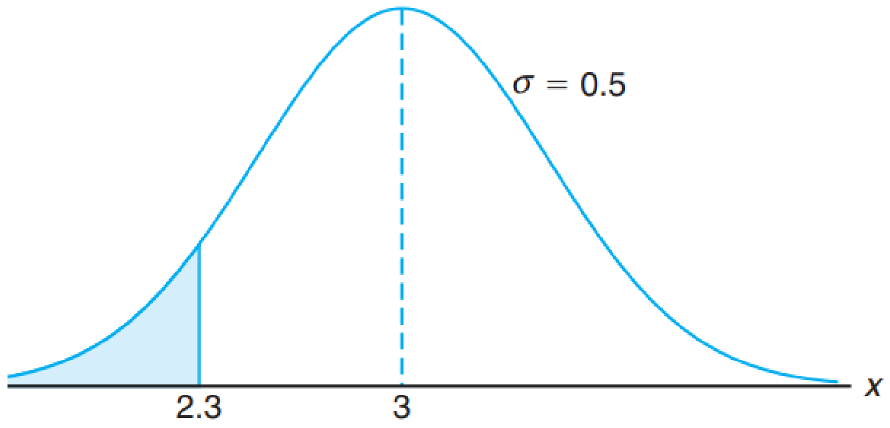
***Solution*:**

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

$$P(X < 2.3) = P(Z < -1.4) = 0.0808.$$



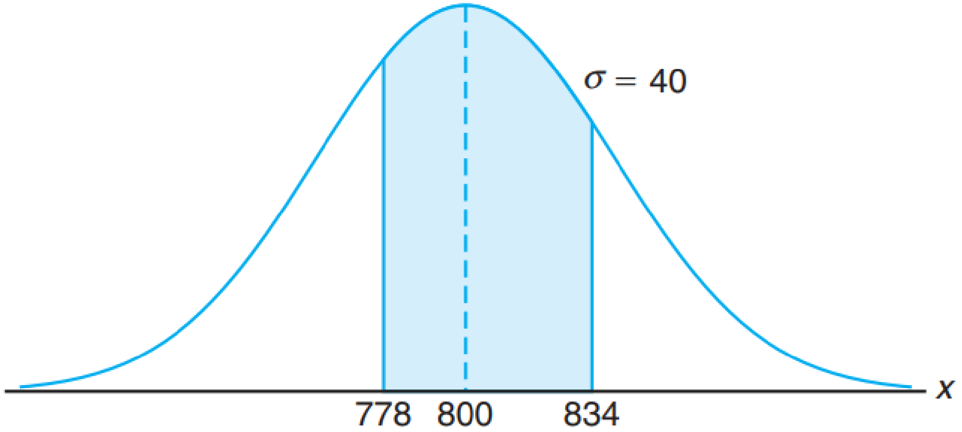
Example An electrical firm manufactures light bulbs that have a life, before burn-out, that is normally distributed with mean equal to 800 hours and a standard deviation of 40 hours. Find the probability that a bulb burns between 778 and 834 hours.

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

$$z_1 = \frac{778 - 800}{40} = -0.55 \text{ and } z_2 = \frac{834 - 800}{40} = 0.85.$$

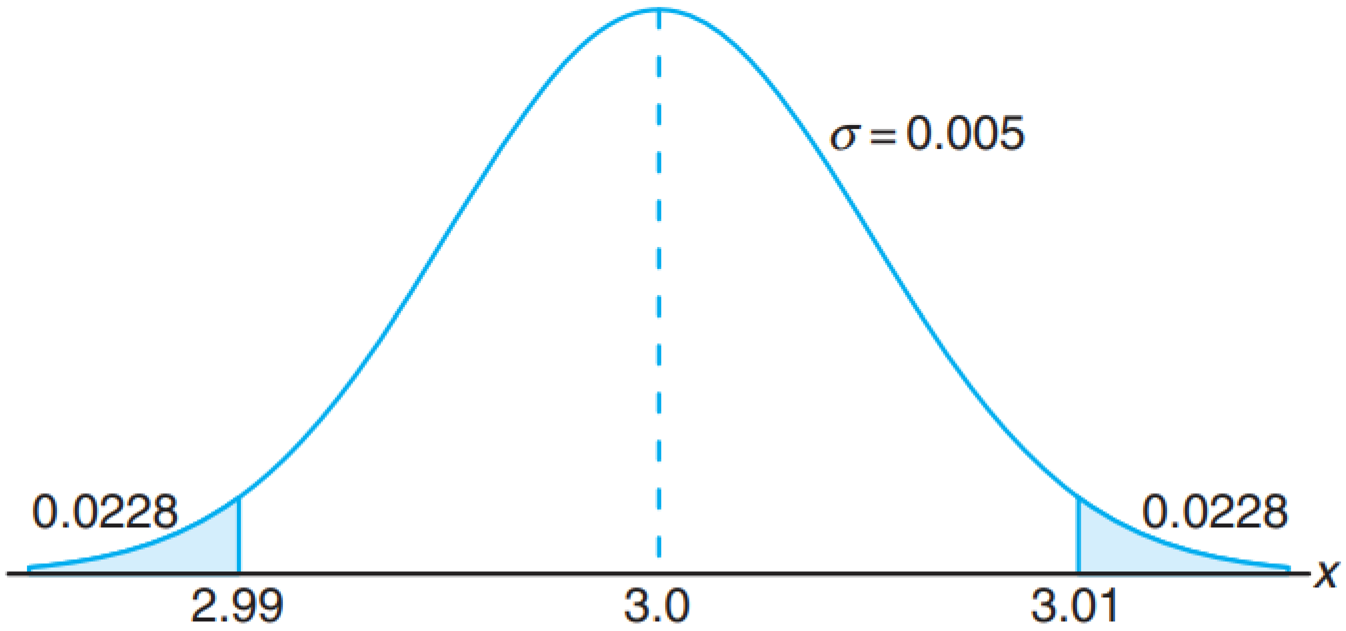
Hence,

$$\begin{aligned} P(778 < X < 834) &= P(-0.55 < Z < 0.85) = P(Z < 0.85) - P(Z < -0.55) \\ &= 0.8023 - 0.2912 = 0.5111. \end{aligned}$$



Example

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. ◾

Example 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.

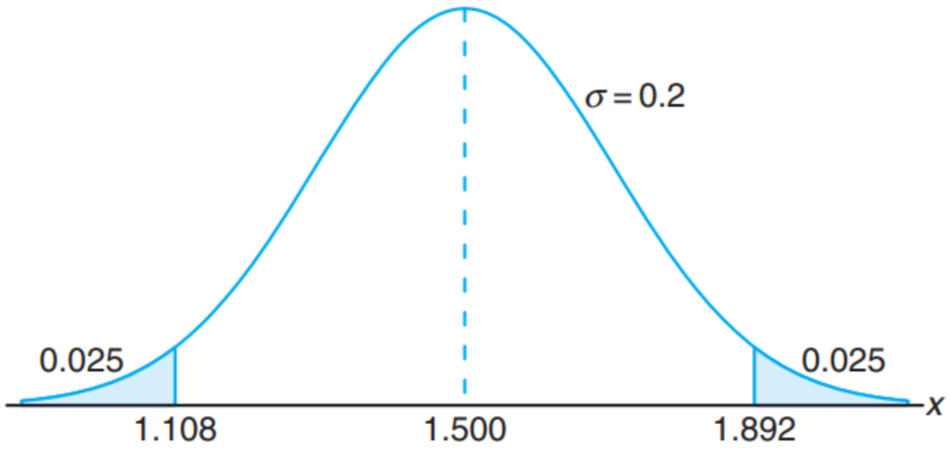
Solution: From Table A.3 we know that

$$P(-1.96 < Z < 1.96) = 0.95.$$

Therefore,

$$1.96 = \frac{(1.50 + d) - 1.50}{0.2},$$

from which we obtain



$$d = (0.2)(1.96) = 0.392.$$

The Exponential Distribution

An **exponential distribution** with parameter $\lambda > 0$ has a probability density function

$$f(x) = \lambda e^{-\lambda x}$$

for $x \geq 0$ and $f(x) = 0$ for $x < 0$, and a cumulative distribution function

$$F(x) = 1 - e^{-\lambda x}$$

for $x \geq 0$. It is useful for modeling failure times and waiting times. Its expectation and variance are

$$E(X) = \frac{1}{\lambda} \qquad \text{and} \qquad \mathrm{Var}(X) = \frac{1}{\lambda^2}$$

The Gamma Function, continued

for $k > 1$. If n is a positive integer, then

$$\Gamma(n) = (n-1)!$$

but except for these special cases there is in general no closed-form expression for the gamma function.

The Gamma Distribution

A **gamma distribution** with parameters $k > 0$ and $\lambda > 0$ has a probability density function

$$f(x) = \frac{\lambda^k x^{k-1} e^{-\lambda x}}{\Gamma(k)}$$

for $x \geq 0$ and $f(x) = 0$ for $x < 0$, where $\Gamma(k)$ is the gamma function. It has an expectation and variance of

$$E(X) = \frac{k}{\lambda} \qquad \text{and} \qquad \mathrm{Var}(X) = \frac{k}{\lambda^2}$$

Example

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

$$P(T > 8) = \frac{1}{5}\int_8^\infty e^{-t/5}\,dt = e^{-8/5} \approx 0.2.$$

Let X represent the number of components functioning after 8 years. Then using the binomial distribution, we have

$$P(X \geq 2) = \sum_{x=2}^{5} b(x; 5, 0.2) = 1 - \sum_{x=0}^{1} b(x; 5, 0.2) = 1 - 0.7373 = 0.2627.$$

Type of Random Variable

DISCRETE (Countable)

Binomial

x = # of S's in n trials

1. n identical trials
2. 2 outcomes: S, F
3. $P(S)$ & $P(F)$ same across trials
4. Trials independent

Poisson

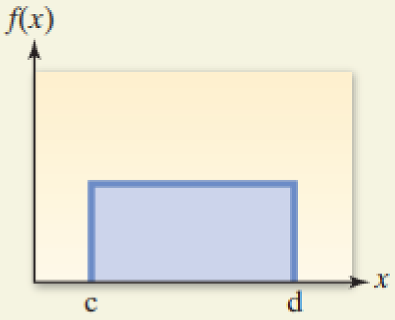
x = # times a rare event (S) occurs in a unit

1. $P(S)$ remains constant across units
2. Unit x-values are independent

CONTINUOUS (Uncountable)

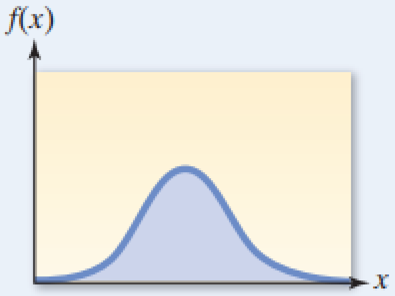
Uniform

Randomness (evenness) distribution



Normal

Bell-shaped curve



Exponential

Waiting time distribution

Moments and Moment-Generating Functions

Definition 7.1:

In this section, we concentrate on applications of moment-generating functions. The obvious purpose of the moment-generating function is in determining moments of random variables. However, the most important contribution is to establish distributions of functions of random variables.

If $g(X) = X^r$ for $r = 0, 1, 2, 3, \ldots,$ Definition 7.1 yields an expected value called the rth **moment about the origin** of the random variable X, which we denote by μ'_r.

The rth **moment about the origin** of the random variable X is given by

$$\mu'_r = E(X^r) = \begin{cases} \sum_x x^r f(x), & \text{if } X \text{ is discrete,} \\ \int_{-\infty}^{\infty} x^r f(x)\, dx, & \text{if } X \text{ is continuous.} \end{cases}$$

Since the first and second moments about the origin are given by $\mu_1' = E(X)$ and $\mu_2' = E(X^2)$, we can write the mean and variance of a random variable as

$$\mu = \mu_1' \qquad \text{and} \qquad \sigma^2 = \mu_2' - \mu^2.$$

Although the moments of a random variable can be determined directly from Definition 7.1, an alternative procedure exists. This procedure requires us to utilize a **moment-generating function**.

Definition 7.2:

The **moment-generating function** of the random variable X is given by $E(e^{tX})$ and is denoted by $M_X(t)$. Hence,

$$M_X(t) = E(e^{tX}) = \begin{cases} \sum\limits_x e^{tx} f(x), & \text{if } X \text{ is discrete,} \\ \int_{-\infty}^{\infty} e^{tx} f(x)\, dx, & \text{if } X \text{ is continuous.} \end{cases}$$

Moment-generating functions will exist only if the sum or integral of Definition 7.2 converges. If a moment-generating function of a random variable X does exist, it can be used to generate all the moments of that variable. The method is described in Theorem 7.6 without proof.

Theorem

Let X be a random variable with moment-generating function $M_X(t)$. Then

$$\left.\frac{d^r M_X(t)}{dt^r}\right|_{t=0} = \mu_r'.$$

Example

Find the moment-generating function of the binomial random variable X and then use it to verify that $\mu = np$ and $\sigma^2 = npq$.

***Solution*:** From Definition

$$M_X(t) = \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x q^{n-x} = \sum_{x=0}^{n} \binom{n}{x} (pe^t)^x q^{n-x}.$$

Recognizing this last sum as the binomial expansion of $(pe^t + q)^n$, we obtain

$$M_X(t) = (pe^t + q)^n.$$

Now

$$\frac{dM_X(t)}{dt} = n(pe^t + q)^{n-1} pe^t$$

and

$$\frac{d^2 M_X(t)}{dt^2} = np[e^t(n-1)(pe^t+q)^{n-2}pe^t + (pe^t+q)^{n-1}e^t].$$

Setting $t = 0$, we get

$$\mu_1' = np \text{ and } \mu_2' = np[(n-1)p+1].$$

Therefore,

$$\mu = \mu_1' = np \text{ and } \sigma^2 = \mu_2' - \mu^2 = np(1-p) = npq,$$

Chapter 8

Fundamental Sampling Distributions and Data Descriptions

Definition

A **population** consists of the totality of the observations with which we are concerned.

A **sample** is a subset of a population.

If our inferences from the sample to the population are to be valid, we must obtain samples that are representative of the population. All too often we are tempted to choose a sample by selecting the most convenient members of the population. Such a procedure may lead to erroneous inferences concerning the population. Any sampling procedure that produces inferences that consistently overestimate or consistently underestimate some characteristic of the population is said to be **biased**. To eliminate any possibility of bias in the sampling procedure, it is desirable to choose a **random sample** in the sense that the observations are made independently and at random.

Any function of the random variables constituting a random sample is called a **statistic**.

Location Measures of a Sample: The Sample Mean, Median, and Mode

(a) Sample mean:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

(b) Sample median:

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

(c) The sample mode is the value of the sample that occurs most often.

Variability Measures of a Sample: The Sample Variance, Standard Deviation, and Range

(a) Sample variance:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

Note that S^2 is essentially defined to be the average of the squares of the deviations of the observations from their mean

If S^2 is the variance of a random sample of size n, we may write

$$S^2 = \frac{1}{n(n-1)}\left[n\sum_{i=1}^{n}X_i^2 - \left(\sum_{i=1}^{n}X_i\right)^2\right].$$

(b) Sample standard deviation: $S = \sqrt{S^2}$,

(c) Sample range: $R = X_{\max} - X_{\min}$.

Sampling Distributions

Inference about the Population from Sample Information

In each of the examples above, we computed a statistic from a sample selected from the population, and from this statistic we made various statements concerning the values of population parameters that may or may not be true.

Since a statistic is a random variable that depends only on the observed sample, it must have a probability distribution.

:

The probability distribution of a statistic is called a **sampling distribution**.

The sampling distribution of a statistic depends on the distribution of the population, the size of the samples, and the method of choosing the samples. In the remainder of this chapter we study several of the important sampling distributions of frequently used statistics. Applications of these sampling distributions to problems of statistical inference are considered throughout most of the remaining chapters. The probability distribution of $\bar{X}$ is called the **sampling distribution of the mean**.

What Is the Sampling Distribution of $\bar{X}$?

We should view the sampling distributions of $\bar{X}$ and S^2 as the mechanisms from which we will be able to make inferences on the parameters μ and σ^2. The sampling distribution of $\bar{X}$ with sample size n is the distribution that results when an **experiment is conducted over and over** (always with sample size n) **and the many values of $\bar{X}$ result**. This sampling distribution, then, describes the variability of sample averages around the population mean μ. In the case of the soft-drink machine, knowledge of the sampling distribution of $\bar{X}$ arms the analyst with the knowledge of a "typical" discrepancy between an observed $\bar{x}$ value and true μ. The same principle applies in the case of the distribution of S^2. The sampling distribution produces information about the variability of s^2 values around σ^2 in repeated experiments.

The central limit theorem

In this section, we will consider one of the most remarkable results in probability — namely, the *central limit theorem*. Loosely speaking, this theorem asserts that the sum of a large number of independent random variables has a distribution that is approximately normal. Hence, it not only provides a simple method for computing approximate probabilities for sums of independent random variables, but it also helps explain the remarkable fact that the empirical frequencies of so many natural populations exhibit a bell-shaped (that is, a normal) curve.

the reproductive property of the normal distribution, we conclude that

$$\bar{X} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$$

has a normal distribution with mean

$$\mu_{\bar{X}} = \frac{1}{n}\underbrace{(\mu + \mu + \cdots + \mu)}_{n \text{ terms}} = \mu \text{ and variance } \sigma^2_{\bar{X}} = \frac{1}{n^2}\underbrace{(\sigma^2 + \sigma^2 + \cdots + \sigma^2)}_{n \text{ terms}} = \frac{\sigma^2}{n}.$$

Central Limit Theorem: If $\bar{X}$ is the mean of a random sample of size n taken from a population with mean μ and finite variance σ^2, then the limiting form of the distribution of

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}},$$

as $n \to \infty$, is the standard normal distribution $n(z; 0, 1)$.

? Need to Know...

When the Sample Size Is Large Enough to Use the Central Limit Theorem

- If the sampled population is **normal**, then the sampling distribution of $\overline{x}$ will also be normal, no matter what sample size you choose. This result can be proven theoretically, but it should not be too difficult for you to accept without proof.
- When the sampled population is approximately **symmetric**, the sampling distribution of $\overline{x}$ becomes approximately normal for relatively small values of n. Remember how rapidly the discrete uniform distribution in the dice example became mound-shaped ($n = 3$).
- When the sampled population is **skewed**, the sample size n must be larger, with n at least 30 before the sampling distribution of $\overline{x}$ becomes approximately normal.

Example

An electrical firm manufactures light bulbs that have a length of life that is approximately normally distributed, with mean equal to 800 hours and a standard deviation of 40 hours. Find the probability that a random sample of 16 bulbs will have an average life of less than 775 hours.

***Solution*:**

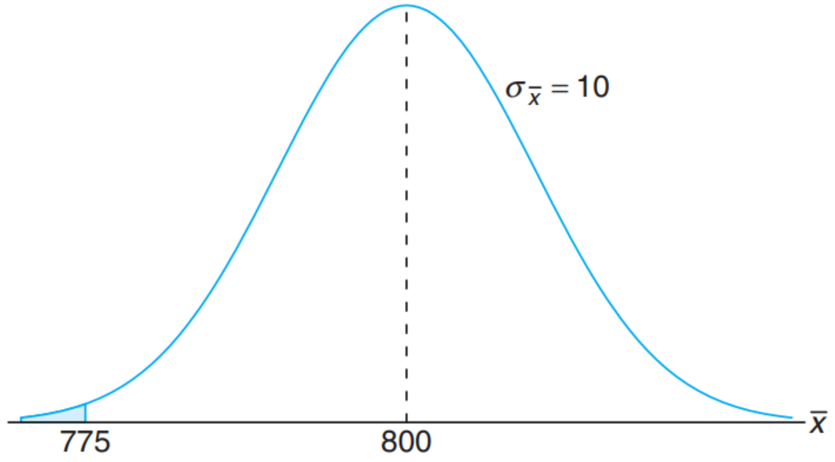
The sampling distribution of $\bar{X}$ will be approximately normal, with $\mu_{\bar{X}} = 800$ and $\sigma_{\bar{X}} = 40/\sqrt{16} = 10$. The desired probability is given by the area of the shaded

Corresponding to $\bar{x} = 775$, we find that

$$z = \frac{775 - 800}{10} = -2.5,$$

and therefore

$$P(\bar{X} < 775) = P(Z < -2.5) = 0.0062.$$



Sampling Distribution of the Difference between Two Means

The Central Limit Theorem can be easily extended to the two-sample, two-population case.

If independent samples of size n_1 and n_2 are drawn at random from two populations, discrete or continuous, with means μ_1 and μ_2 and variances σ_1^2 and σ_2^2, respectively, then the sampling distribution of the differences of means, $\bar{X}_1 - \bar{X}_2$, is approximately normally distributed with mean and variance given by

$$\mu_{\bar{X}_1 - \bar{X}_2} = \mu_1 - \mu_2 \text{ and } \sigma^2_{\bar{X}_1 - \bar{X}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}.$$

Hence,

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{(\sigma_1^2/n_1) + (\sigma_2^2/n_2)}}$$

is approximately a standard normal variable.

Theorem

If S^2 is the variance of a random sample of size n taken from a normal population having the variance σ^2, then the statistic

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} = \sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{\sigma^2}$$

has a chi-squared distribution with $v = n - 1$ degrees of freedom.

Topics dealing with sampling distributions, analysis of variance, and nonparametric statistics involve extensive use of the chi-squared distribution.

Chi-Squared Distribution

Another very important special case of the gamma distribution is obtained by letting $\alpha = v/2$ and $\beta = 2$, where v is a positive integer. The result is called **the chi-squared distribution**. The distribution has a single parameter, v, called the **degrees of freedom**.

The continuous random variable X has a **chi-squared distribution**, with v **degrees of freedom**, if its density function is given by

$$f(x; v) = \begin{cases} \frac{1}{2^{v/2}\Gamma(v/2)} x^{v/2-1} e^{-x/2}, & x > 0, \\ 0, & \text{elsewhere}, \end{cases}$$

where v is a positive integer.

The mean and variance of the chi-squared distribution are

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

Example

A manufacturer of car batteries guarantees that the batteries will last, on average, 3 years with a standard deviation of 1 year. If five of these batteries have lifetimes of 1.9, 2.4, 3.0, 3.5, and 4.2 years, should the manufacturer still be convinced that the batteries have a standard deviation of 1 year? Assume that the battery lifetime follows a normal distribution.

Solution: We first find the sample variance

$$s^2 = \frac{(5)(48.26) - (15)^2}{(5)(4)} = 0.815.$$

Then

$$\chi^2 = \frac{(4)(0.815)}{1} = 3.26$$

is a value from a chi-squared distribution with 4 degrees of freedom. Since 95% of the χ^2 values with 4 degrees of freedom fall between 0.484 and 11.143, the computed value with $\sigma^2 = 1$ is reasonable, and therefore the manufacturer has no reason to suspect that the standard deviation is other than 1 year.

t-Distribution

Let Z be a standard normal random variable and V a chi-squared random variable with v degrees of freedom. If Z and V are independent, then the distribution of the random variable T, where

$$T = \frac{Z}{\sqrt{V/v}},$$

is given by the density function

$$h(t) = \frac{\Gamma[(v+1)/2]}{\Gamma(v/2)\sqrt{\pi v}} \left(1 + \frac{t^2}{v}\right)^{-(v+1)/2}, \quad -\infty < t < \infty.$$

This is known as the ***t*-distribution** with v degrees of freedom.

Example

The t-value with $v = 14$ degrees of freedom that leaves an area of 0.025 to the left, and therefore an area of 0.975 to the right, is

$$t_{0.975} = -t_{0.025} = -2.145.$$

Example

Find $P(-t_{0.025} < T < t_{0.05})$.

Since $t_{0.05}$ leaves an area of 0.05 to the right, and $-t_{0.025}$ leaves an area of 0.025 to the left, we find a total area of

$$1 - 0.05 - 0.025 = 0.925$$

between $-t_{0.025}$ and $t_{0.05}$. Hence

$$P(-t_{0.025} < T < t_{0.05}) = 0.925.$$

What Is the t-Distribution Used For?

The t-distribution is used extensively in problems that deal with inference about the population mean (as illustrated in Example 8.11) or in problems that involve comparative samples (i.e., in cases where one is trying to determine if means from two samples are significantly different).

F-Distribution

We have motivated the t-distribution in part by its application to problems in which there is comparative sampling (i.e., a comparison between two sample means).

The F-distribution finds enormous application in comparing sample variances.

Let U and V be two independent random variables having chi-squared distributions with v_1 and v_2 degrees of freedom, respectively. Then the distribution of the random variable $F = \frac{U/v_1}{V/v_2}$ is given by the density function

$$h(f) = \begin{cases} \frac{\Gamma[(v_1+v_2)/2](v_1/v_2)^{v_1/2}}{\Gamma(v_1/2)\Gamma(v_2/2)} \frac{f^{(v_1/2)-1}}{(1+v_1 f/v_2)^{(v_1+v_2)/2}}, & f > 0, \\ 0, & f \le 0. \end{cases}$$

This is known as the ***F*-distribution** with v_1 and v_2 degrees of freedom (d.f.).

Theorem

Writing $f_\alpha(v_1, v_2)$ for f_α with v_1 and v_2 degrees of freedom, we obtain

$$f_{1-\alpha}(v_1, v_2) = \frac{1}{f_\alpha(v_2, v_1)}.$$

Thus, the f-value with 6 and 10 degrees of freedom, leaving an area of 0.95 to the right, is

$$f_{0.95}(6, 10) = \frac{1}{f_{0.05}(10, 6)} = \frac{1}{4.06} = 0.246.$$

Chapter 9

One- and Two-Sample Estimation Problems

Inferential statistics

The process of drawing inferences about a population on the basis of information Contained in a sample taken from the population is called statistical inferences.

Or

Statistical inference is the process of drawing conclusions or making predictions about a population based on sample data.

Types of statistical inference include:

1. Estimation
2. Hypothesis Testing

Estimation

Estimation aims to estimate unknown population parameters based on sample data

a. Point estimation:

Providing a single value as the best estimate of the parameter.

Sample mean: Calculating the sample mean as an estimate of the population mean.

Sample proportion: Estimating the population proportion based on the sample proportion.

b. Interval estimation:

Providing a range of values (confidence interval) likely to contain the true population parameter.

Confidence interval: A range of values with a specified level of confidence.

Confidence level: The level of confidence associated with the interval estimate

Hypothesis testing

A hypothesis test is a procedure that enables us to decide on the basis of information obtained by sampling whether to accept or reject any specified statement or hypothesis regarding the values of parameter in a statistical problem.

It involves formulating a null hypothesis (H_0) and an alternative hypothesis (H_1),

Definition

A statistic $\hat{\Theta}$ is said to be an **unbiased estimator** of the parameter θ if

$$\mu_{\hat{\Theta}} = E(\hat{\Theta}) = \theta.$$

Variance of a Point Estimator

If $\hat{\Theta}_1$ and $\hat{\Theta}_2$ are two unbiased estimators of the same population parameter θ, we want to choose the estimator whose sampling distribution has the smaller variance. Hence, if $\sigma^2_{\hat{\theta}_1} < \sigma^2_{\hat{\theta}_2}$, we say that $\hat{\Theta}_1$ is a **more efficient estimator** of θ than $\hat{\Theta}_2$.

Definition

If we consider all possible unbiased estimators of some parameter θ, the one with the smallest variance is called the **most efficient estimator** of θ.

Maximum Likelihood Estimation

Maximum likelihood estimation is one of the most important approaches to estimation in all statistical inference

Given independent observations $x_1, x_2, \ldots, x_n$ from a probability density function (continuous case) or probability mass function (discrete case) $f(\mathbf{x}; \theta)$, the maximum likelihood estimator $\hat{\theta}$ is that which maximizes the likelihood function

$$L(x_1, x_2, \ldots, x_n; \theta) = f(\mathbf{x}; \theta) = f(x_1, \theta) f(x_2, \theta) \cdots f(x_n, \theta).$$

Quite often it is convenient to work with the natural log of the likelihood function in finding the maximum of that function. Consider the following example dealing with the parameter μ of a Poisson distribution.

Example

Consider a Poisson distribution with probability mass function

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

Suppose that a random sample $x_1, x_2, \ldots, x_n$ is taken from the distribution. What is the maximum likelihood estimate of u?

Solution:

The likelihood function is

$$L(x_1, x_2, \ldots, x_n; \mu) = \prod_{i=1}^{n} f(x_i|\mu) = \frac{e^{-n\mu}\mu^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!}.$$

Now consider

Now consider

$$\ln L(x_1, x_2, \ldots, x_n; \mu) = -n\mu + \sum_{i=1}^{n} x_i \ln \mu - \ln \prod_{i=1}^{n} x_i!$$

$$\frac{\partial \ln L(x_1, x_2, \ldots, x_n; \mu)}{\partial \mu} = -n + \sum_{i=1}^{n} \frac{x_i}{\mu}.$$

Solving for $\hat{\mu}$, the maximum likelihood estimator, involves setting the derivative to zero and solving for the parameter. Thus,

$$\hat{\mu} = \sum_{i=1}^{n} \frac{x_i}{n} = \bar{x}.$$

The second derivative of the log-likelihood function is negative, which implies that the solution above indeed is a maximum. Since μ is the mean of the Poisson distribution (Chapter 5), the sample average would certainly seem like a reasonable estimator. ∎

Example

Suppose 10 rats are used in a biomedical study where they are injected with cancer cells and then given a cancer drug that is designed to increase their survival rate. The survival times, in months, are 14, 17, 27, 18, 12, 8, 22, 13, 19, and 12. Assume that the exponential distribution applies. Give a maximum likelihood estimate of the mean survival time.

***Solution*:**

; the probability density function for the exponential random variable X is

$$f(x, \beta) = \begin{cases} \frac{1}{\beta} e^{-x/\beta}, & x > 0, \\ 0, & \text{elsewhere.} \end{cases}$$

Thus, the log-likelihood function for the data, given $n = 10$, is

$$\ln L(x_1, x_2, \ldots, x_{10}; \beta) = -10 \ln \beta - \frac{1}{\beta} \sum_{i=1}^{10} x_i .$$

Setting

$$\frac{\partial \ln L}{\partial \beta} = -\frac{10}{\beta} + \frac{1}{\beta^2} \sum_{i=1}^{10} x_i = 0$$

implies that

$$\hat{\beta} = \frac{1}{10} \sum_{i=1}^{10} x_i = \bar{x} = 16.2.$$

Evaluating the second derivative of the log-likelihood function at the value $\hat{\beta}$ above yields a negative value. As a result, the estimator of the parameter β, the population mean, is the sample average $\bar{x}$.

Interval Estimation

If, for instance, we find $\hat{\Theta}_L$ and $\hat{\Theta}_U$ such that

$$P(\hat{\Theta}_L < \theta < \hat{\Theta}_U) = 1 - \alpha,$$

for $0 < \alpha < 1$, then we have a probability of $1-\alpha$ of selecting a random sample that will produce an interval containing θ. The interval $\hat{\theta}_L < \theta < \hat{\theta}_U$, computed from the selected sample, is called a $100(1-\alpha)\%$ **confidence interval**, the fraction $1-\alpha$ is called the **confidence coefficient** or the **degree of confidence**, and the endpoints, $\hat{\theta}_L$ and $\hat{\theta}_U$, are called the lower and upper **confidence limits**.

Single Sample: Estimating the Mean

Confidence Interval on μ, σ^2 Known

If $\bar{x}$ is the mean of a random sample of size n from a population with known variance σ^2, a $100(1-\alpha)\%$ confidence interval for μ is given by

$$\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}},$$

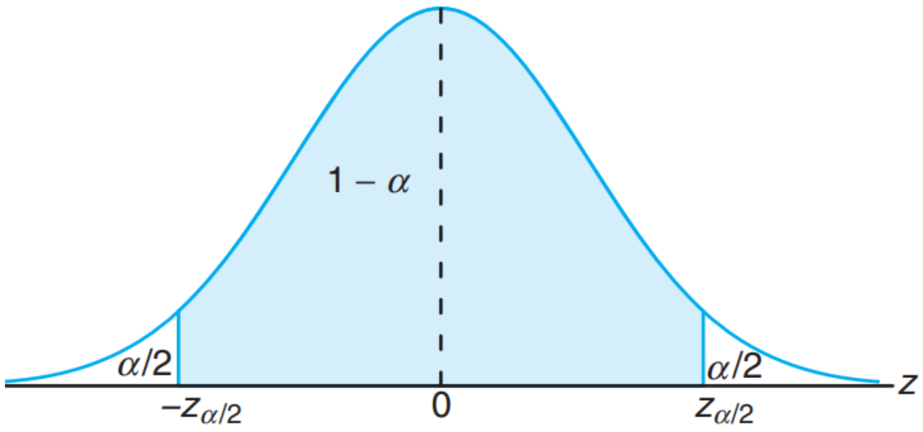
where $z_{\alpha/2}$ is the z-value leaving an area of $\alpha/2$ to the right.

$$P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha,$$

where

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}.$$

Hence,



$$P\left(-z_{\alpha/2} < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < z_{\alpha/2}\right) = 1-\alpha.$$

$$P\left(\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha.$$

Clearly, the values of the random variables $\hat{\Theta}_L$ and $\hat{\Theta}_U$ are the confidence limits

$$\hat{\theta}_L = \bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \quad \text{and} \quad \hat{\theta}_U = \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}.$$

Example

The average zinc concentration recovered from a sample of measurements taken in 36 different locations in a river is found to be 2.6 grams per milliliter. Find the 95% and 99% confidence intervals for the mean zinc concentration in the river. Assume that the population standard deviation is 0.3 gram per milliliter.

***Solution*:**

The point estimate of μ is $\bar{x} = 2.6$. The z-value leaving an area of 0.025 to the right, and therefore an area of 0.975 to the left, is $z_{0.025} = 1.96$ (Table A.3). Hence, the 95% confidence interval is

$$2.6 - (1.96)\left(\frac{0.3}{\sqrt{36}}\right) < \mu < 2.6 + (1.96)\left(\frac{0.3}{\sqrt{36}}\right),$$

which reduces to $2.50 < \mu < 2.70$. To find a 99% confidence interval, we find the z-value leaving an area of 0.005 to the right and 0.995 to the left. From Table A.3 again, $z_{0.005} = 2.575$, and the 99% confidence interval is

$$2.6 - (2.575)\left(\frac{0.3}{\sqrt{36}}\right) < \mu < 2.6 + (2.575)\left(\frac{0.3}{\sqrt{36}}\right),$$

or simply

$$2.47 < \mu < 2.73.$$

We now see that a longer interval is required to estimate μ with a higher degree of confidence.

Theorem

If $\bar{x}$ is used as an estimate of μ, we can be $100(1-\alpha)\%$ confident that the error will not exceed $z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$.

Theorem

If $\bar{x}$ is used as an estimate of μ, we can be $100(1-\alpha)\%$ confident that the error will not exceed a specified amount e when the sample size is

$$n = \left(\frac{z_{\alpha/2}\sigma}{e}\right)^2.$$

When solving for the sample size, n, we round all fractional values up to the next whole number. By adhering to this principle, we can be sure that our degree of confidence never falls below $100(1-\alpha)\%$.

Example

How large a sample is required if we want to be 95% confident that our estimate of μ in Example 9.2 is off by less than 0.05?

***Solution*:**

The population standard deviation is $\sigma = 0.3$. Then, by Theorem 9.2,

$$n = \left[\frac{(1.96)(0.3)}{0.05}\right]^2 = 138.3.$$

Therefore, we can be 95% confident that a random sample of size 139 will provide an estimate $\bar{x}$ differing from μ by an amount less than 0.05.

One-Sided Confidence Bounds on μ, σ^2 Known

If $\bar{X}$ is the mean of a random sample of size n from a population with σ^2, the one-sided $100(1-\alpha)\%$ confidence bounds for μ are given by

upper one-sided bound: $\bar{x} + z_\alpha \sigma/\sqrt{n}$;

lower one-sided bound: $\bar{x} - z_\alpha \sigma/\sqrt{n}$.

Confidence Interval on μ, σ^2 Unknown

If $\bar{x}$ and s are the mean and standard deviation of a random sample from a normal population with unknown variance σ^2, a $100(1-\alpha)\%$ confidence interval for μ is

$$\bar{x} - t_{\alpha/2}\frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{\alpha/2}\frac{s}{\sqrt{n}},$$

where $t_{\alpha/2}$ is the t-value with $v = n - 1$ degrees of freedom, leaving an area of $\alpha/2$ to the right.

Computed one-sided confidence bounds for μ with σ unknown are as the reader would expect, namely

$$\bar{x} + t_{\alpha}\frac{s}{\sqrt{n}} \qquad \text{and} \qquad \bar{x} - t_{\alpha}\frac{s}{\sqrt{n}}.$$

Example

The contents of seven similar containers of sulfuric acid are 9.8, 10.2, 10.4, 9.8, 10.0, 10.2, and 9.6 liters. Find a 95% confidence interval for the mean contents of all such containers, assuming an approximately normal distribution.

Solution: The sample mean and standard deviation for the given data are

$$\bar{x} = 10.0 \qquad \text{and} \qquad s = 0.283.$$

we find $t_{0.025} = 2.447$ for $v = 6$ degrees of freedom. Hence, the

95% confidence interval for μ is

$$10.0 - (2.447)\left(\frac{0.283}{\sqrt{7}}\right) < \mu < 10.0 + (2.447)\left(\frac{0.283}{\sqrt{7}}\right),$$

which reduces to $9.74 < \mu < 10.26$.

Single Sample: Estimating a Proportion

Large-Sample Confidence Intervals for p

If $\hat{p}$ is the proportion of successes in a random sample of size n and $\hat{q} = 1 - \hat{p}$, an approximate $100(1-\alpha)\%$ confidence interval, for the binomial parameter p is given by (method 1)

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} < p < \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

or by (method 2)

$$\frac{\hat{p} + \frac{z_{\alpha/2}^2}{2n}}{1 + \frac{z_{\alpha/2}^2}{n}} - \frac{z_{\alpha/2}}{1 + \frac{z_{\alpha/2}^2}{n}}\sqrt{\frac{\hat{p}\hat{q}}{n} + \frac{z_{\alpha/2}^2}{4n^2}} < p < \frac{\hat{p} + \frac{z_{\alpha/2}^2}{2n}}{1 + \frac{z_{\alpha/2}^2}{n}} + \frac{z_{\alpha/2}}{1 + \frac{z_{\alpha/2}^2}{n}}\sqrt{\frac{\hat{p}\hat{q}}{n} + \frac{z_{\alpha/2}^2}{4n^2}},$$

where $z_{\alpha/2}$ is the z-value leaving an area of $\alpha/2$ to the right.

Example

In a random sample of $n = 500$ families owning televisions in the city of Hamilton, Canada, it is found that $x = 340$ subscribe to HBO. Find a 95% confidence interval for the actual proportion of families with televisions in this city that subscribe to HBO.

***Solution*:**

The point estimate of p is $\hat{p} = 340/500 = 0.68$. Using Table A.3, we find that $z_{0.025} = 1.96$. Therefore, using method 1, the 95% confidence interval for p is

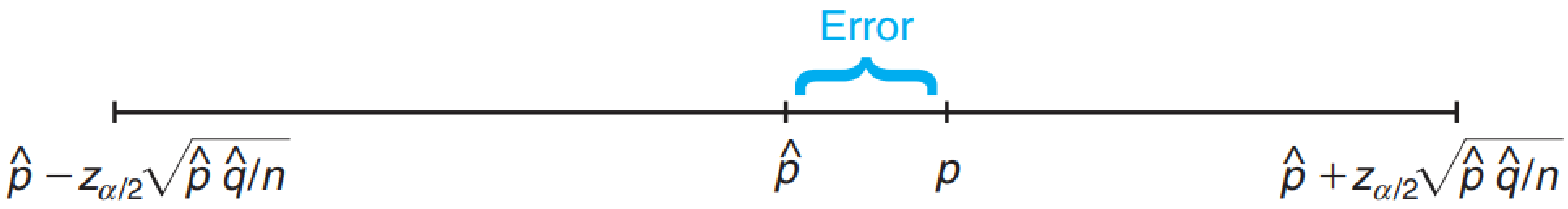
$$0.68 - 1.96\sqrt{\frac{(0.68)(0.32)}{500}} < p < 0.68 + 1.96\sqrt{\frac{(0.68)(0.32)}{500}},$$

which simplifies to $0.6391 < p < 0.7209$.

If we use method 2, we can obtain

$$\frac{0.68 + \frac{1.96^2}{(2)(500)}}{1 + \frac{1.96^2}{500}} \pm \frac{1.96}{1 + \frac{1.96^2}{500}} \sqrt{\frac{(0.68)(0.32)}{500} + \frac{1.96^2}{(4)(500^2)}} = 0.6786 \pm 0.0408,$$

which simplifies to $0.6378 < p < 0.7194$. Apparently, when n is large (500 here), both methods yield very similar results. ■



Single Sample: Estimating the Variance

Confidence Interval for σ^2

If s^2 is the variance of a random sample of size n from a normal population, a $100(1-\alpha)\%$ confidence interval for σ^2 is

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}},$$

where $\chi^2_{\alpha/2}$ and $\chi^2_{1-\alpha/2}$ are χ^2-values with $v = n-1$ degrees of freedom, leaving areas of $\alpha/2$ and $1-\alpha/2$, respectively, to the right.

An approximate $100(1-\alpha)\%$ confidence interval for σ is obtained by taking the square root of each endpoint of the interval for σ^2.

Example

The following are the weights, in decagrams, of 10 packages of grass seed distributed by a certain company: 46.4, 46.1, 45.8, 47.0, 46.1, 45.9, 45.8, 46.9, 45.2, and 46.0. Find a 95% confidence interval for the variance of the weights of all such packages of grass seed distributed by this company, assuming a normal population.

Solution:

First we find

$$s^2 = \frac{n \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}{n(n-1)}$$

$$= \frac{(10)(21,273.12) - (461.2)^2}{(10)(9)} = 0.286.$$

To obtain a 95% confidence interval, we choose $\alpha = 0.05$. Then, using Table A.5 with $v = 9$ degrees of freedom, we find $\chi^2_{0.025} = 19.023$ and $\chi^2_{0.975} = 2.700$. Therefore, the 95% confidence interval for σ^2 is

$$\frac{(9)(0.286)}{19.023} < \sigma^2 < \frac{(9)(0.286)}{2.700},$$

or simply $0.135 < \sigma^2 < 0.953$.

Estimation of a parameter.

Estimation of a population mean μ.

(a) Point estimate of μ is
$\hat{\mu} = \bar{X}$.

(b) A $(1-\alpha)$-level confidence interval for μ is

$$\left(\bar{X} \mp t_{(n-1),(\alpha/2)}\, s/\sqrt{n}\right)$$

For $n \geq 30$ the above confidence interval can be approximated as

$$\left(\bar{X} \mp z_{(\alpha/2)}\, s/\sqrt{n}\right).$$

Estimation of a population proportion p.

(a) Point estimate of p is
$\hat{p} = X/n$.

(b) A $(1-\alpha)$-level confidence interval for p is

$$\hat{p} \mp z_{(\alpha/2)} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.$$

Estimation of a population standard deviation σ.

(a) Point estimate of σ is
$\hat{\sigma} = s$.

(b) A $(1-\alpha)$-level confidence interval for σ is

$$\left(\frac{s\sqrt{n-1}}{\sqrt{\chi^{2(R)}_{(n-1),(\alpha/2)}}},\ \frac{s\sqrt{n-1}}{\sqrt{\chi^{2(R)}_{(n-1),1-(\alpha/2)}}}\right).$$

Two Samples: Estimating the Difference between Two Means

Confidence Interval for $\mu_1 - \mu_2$, σ_1^2 and σ_2^2 Known

If $\bar{x}_1$ and $\bar{x}_2$ are means of independent random samples of sizes n_1 and n_2 from populations with known variances σ_1^2 and σ_2^2, respectively, a $100(l-\alpha)\%$ confidence interval for $\mu_1 - \mu_2$ is given by

$$(\bar{x}_1 - \bar{x}_2) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}},$$

where $z_{\alpha/2}$ is the z-value leaving an area of $\alpha/2$ to the right.

standard deviation $\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}$.

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$$

Confidence Interval for $\mu_1 - \mu_2$, $\sigma_1^2 \neq \sigma_2^2$ and Both Unknown

If $\bar{x}_1$ and s_1^2 and $\bar{x}_2$ and s_2^2 are the means and variances of independent random samples of sizes n_1 and n_2, respectively, from approximately normal populations with unknown and unequal variances, an approximate $100(1-\alpha)\%$ confidence interval for $\mu_1 - \mu_2$ is given by

$$(\bar{x}_1 - \bar{x}_2) - t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}},$$

where $t_{\alpha/2}$ is the t-value with

$$v = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{[(s_1^2/n_1)^2/(n_1 - 1)] + [(s_2^2/n_2)^2/(n_2 - 1)]}$$

degrees of freedom, leaving an area of $\alpha/2$ to the right.

Chapter 10

One- and Two-Sample Tests of Hypotheses

Hypothesis testing is a very important phase of statistical inferences. It is a procedure which enables us to decide on the basis of information obtained from sample data whether to accept or reject a statement or an assumption which may or may not be true is called statistical hypothesis.

Or A statistical hypothesis is an assertion or conjecture concerning one or more populations.

We accept the hypothesis as being true, when it is supported by the sample data.

We reject the hypothesis when the sample data fail to support it

It is important to understand the terms reject or accept in hypothesis testing

The rejection means to declare it false.

The acceptance of a hypothesis is to conclude that there is not sufficient evidence to reject it.

Acceptance does not necessarily mean that the hypothesis is true

Null and Alternative Hypothesis

A null hypothesis H_0 is any hypothesis that is to be tested for possible rejection under the assumption that is true. For example, the null hypotheses is may be that the population mean in 40,$\mathrm{H}_0(\mu = 40)$

The rejection of H_0 leads to the acceptance of an alternative hypothesis, denoted H_1. We accept alternative hypotheses when the null hypothesis is rejected .

For example the population mean

$\mathrm{H}_1(\mu \neq 40)$, H_1 $(\mu > 0)$

$\mathrm{H}_0(\mu < 40)$

Test-statistic

A sample statistic that provides a basis for testing a null hypothesis (whether to accept or reject) is called test statistic. Every test-statistics has a probability or sampling distribution which gives the probability of obtaining a specified value of the test statistics when the null hypothesis is true. The sampling distribution of the most commonly used test statistic are normal t, chi-square, or F.

Acceptance and rejection region:

The region is called a critical region. The value(s) that separates the critical region from the acceptance region is called the critical value(s).

Type I and Type II Errors

Rejection of the null hypothesis when it is true is called a **type I error**.

Nonrejection of the null hypothesis when it is false is called a **type II error**.

Type I error: Rejection of H_0 when H_0 is true.
Type II error: Nonrejection of H_0 when H_1 is true.

α = probability of making a Type I error (also called the **level of significance**)
β = probability of making a Type II error

Table 10.1: Possible Situations for Testing a Statistical Hypothesis

	H_0 is true	**H_0 is false**
Do not reject H_0	Correct decision	Type II error
Reject H_0	Type I error	Correct decision

The probability of committing a type I error, also called the **level of significance**, is denoted by the Greek letter α. In our illustration, a type I error will occur when more than 8 individuals inoculated with the new vaccine surpass the 2-year period without contracting the virus and researchers conclude that the new vaccine is better when it is actually equivalent to the one in use. Hence, if X is the number of individuals who remain free of the virus for at least 2 years,

$$\alpha = P(\text{type I error}) = P\left(X > 8 \text{ when } p = \frac{1}{4}\right) = \sum_{x=9}^{20} b\left(x; 20, \frac{1}{4}\right)$$

$$= 1 - \sum_{x=0}^{8} b\left(x; 20, \frac{1}{4}\right) = 1 - 0.9591 = 0.0409.$$

We say that the null hypothesis, $p = 1/4$, is being tested at the $\alpha = 0.0409$ level of significance. Sometimes the level of significance is called the **size of the test**. A critical region of size 0.0409 is very small, and therefore it is unlikely that a type I error will be committed. Consequently, it would be most unusual for more than 8 individuals to remain immune to a virus for a 2-year period using a new vaccine that is essentially equivalent to the one now on the market.

The Probability of a Type II Error

The probability of committing a type II error, denoted by β, is impossible to compute unless we have a specific alternative hypothesis. If we test the null hypothesis that $p = 1/4$ against the alternative hypothesis that $p = 1/2$, then we are able to compute the probability of not rejecting H_0 when it is false. We simply find the probability of obtaining 8 or fewer in the group that surpass the 2-year period when $p = 1/2$. In this case,

$$\beta = P(\text{type II error}) = P\left(X \leq 8 \text{ when } p = \frac{1}{2}\right)$$

$$= \sum_{x=0}^{8} b\left(x; 20, \frac{1}{2}\right) = 0.2517.$$

This is a rather high probability, indicating a test procedure in which it is quite likely that we shall reject the new vaccine when, in fact, it is superior to what is now in use. Ideally, we like to use a test procedure for which the type I and type II error probabilities are both small.

If $p = 0.7$, this test procedure gives

$$\beta = P(\text{type II error}) = P(X \leq 8 \text{ when } p = 0.7)$$
$$= \sum_{x=0}^{8} b(x; 20, 0.7) = 0.0051.$$

With such a small probability of committing a type II error, it is extremely unlikely that the new vaccine would be rejected when it was 70% effective after a period of 2 years. As the alternative hypothesis approaches unity, the value of β diminishes to zero.

Example

The proportion of adults living in a small town who are matriculates is estimated to be $p = 0.3$. To test this hypothesis a random sample of 15 adults is selected. If the number of matriculates in our sample is anywhere from 2 to 7, we shall accept the null hypotheses that $p = 0.3$, otherwise we shall conclude that p not equal to 0.3. Evaluate α assuming $p = 0.3$. evaluate β for the alternatives $p = 0.2$ and $p = 0.4$.

Solution

Let X denotes the number of adults who are matriculates. The the test-statistic has binomial distribution with $p = 0.3$ and $n = 15$.

Acceptance region: $X = 2$ to $X = 7$. Critical region: $X < 2$ and $X > 7$

Thus, the probability of making type one error is α , so

$$\begin{aligned}
\alpha =& P(X < 2, \text{when } p = 0.3) + P(X > 7, \text{when } p = 0.3) \\
=& \sum_{x=0}^{1} b(x; 15, 0.3) + \sum_{x=8}^{15} b(x; 15, 0.3) \\
=& \sum_{x=0}^{1} b(x; 15, 0.3) + \left(1 - \sum_{x=0}^{7} b(x; 15, 0.3)\right) \\
=& 0.0353 + (1 - 0.9500) = 0.0853
\end{aligned}$$

To compute β, the probability of type II error, we need a specific alternate hypothesis test, $H_0 : p = 0.3$ and $H_1 : p = 0.2$. A type II error results when a false null hypothesis is accepted. That is a type II error occurs if any value of the distribution under $H_1 : p = 0.2$ falls in the region $X = 2$ to $X = 7$, the acceptance region of the distribution under null hypothesis $H_0 : p = 0.3$. Hence

$$\begin{aligned}\beta =& P(2 \leq X \leq 7 \text{ when } H_1 : p = 0.2)\\ =& \sum_{x=2}^{7} b(x; 15, 0.2)\\ =& \sum_{x=0}^{7} b(x; 15, 0.2) - \sum_{x=0}^{1} b(x; 15, 0.2)\\ =& 0.8858 - 0.1671 = 0.8287\end{aligned}$$

Similarly, when $H_1 : p = 0.4$, we have

$$\begin{aligned}
\beta =& P(2 \leq X \leq 7 \text{ when } H_1 : p = 0.4) \\
=& \sum_{x=2}^{7} b(x; 15, 0.4) \\
=& \sum_{x=0}^{7} b(x; 15, 0.4) - \sum_{x=0}^{1} b(x; 15, 0.4) \\
=& 0.7869 - 0.0.0052 = 0.7817
\end{aligned}$$

The **power** of a test is the probability of rejecting H_0 given that a specific alternative is true.

One- and Two-Tailed Tests

A test of any statistical hypothesis where the alternative is **one sided**, such as

$$H_0\colon\ \theta = \theta_0,$$
$$H_1\colon\ \theta > \theta_0$$

or perhaps

$$H_0\colon\ \theta = \theta_0,$$
$$H_1\colon\ \theta < \theta_0,$$

is called a **one-tailed test**. Earlier in this section, we referred to the **test statistic** for a hypothesis. Generally, the critical region for the alternative hypothesis $\theta > \theta_0$ lies in the right tail of the distribution of the test statistic, while the critical region for the alternative hypothesis $\theta < \theta_0$ lies entirely in the left tail.

A test of any statistical hypothesis where the alternative is **two sided**, such as

$$H_0\colon\ \theta = \theta_0,$$
$$H_1\colon\ \theta \neq \theta_0,$$

is called a **two-tailed test**, since the critical region is split into two parts, often having equal probabilities, in each tail of the distribution of the test statistic. The alternative hypothesis $\theta \neq \theta_0$ states that either $\theta < \theta_0$ or $\theta > \theta_0$.

Example

A manufacturer of a certain brand of rice cereal claims that the average saturated fat content does not exceed 1.5 grams per serving. State the null and alternative hypotheses to be used in testing this claim and determine where the critical region is located.

***Solution*:**

The manufacturer's claim should be rejected only if μ is greater than 1.5 milligrams and should not be rejected if μ is less than or equal to 1.5 milligrams. We test

$$H_0\colon\ \mu = 1.5,$$
$$H_1\colon\ \mu > 1.5.$$

Nonrejection of H_0 does not rule out values less than 1.5 milligrams. Since we have a one-tailed test, the greater than symbol indicates that the critical region lies entirely in the right tail of the distribution of our test statistic $\bar{X}$.

Example

A real estate agent claims that 60% of all private residences being built today are 3-bedroom homes. To test this claim, a large sample of new residences is inspected; the proportion of these homes with 3 bedrooms is recorded and used as the test statistic. State the null and alternative hypotheses to be used in this test and determine the location of the critical region.

***Solution*:**

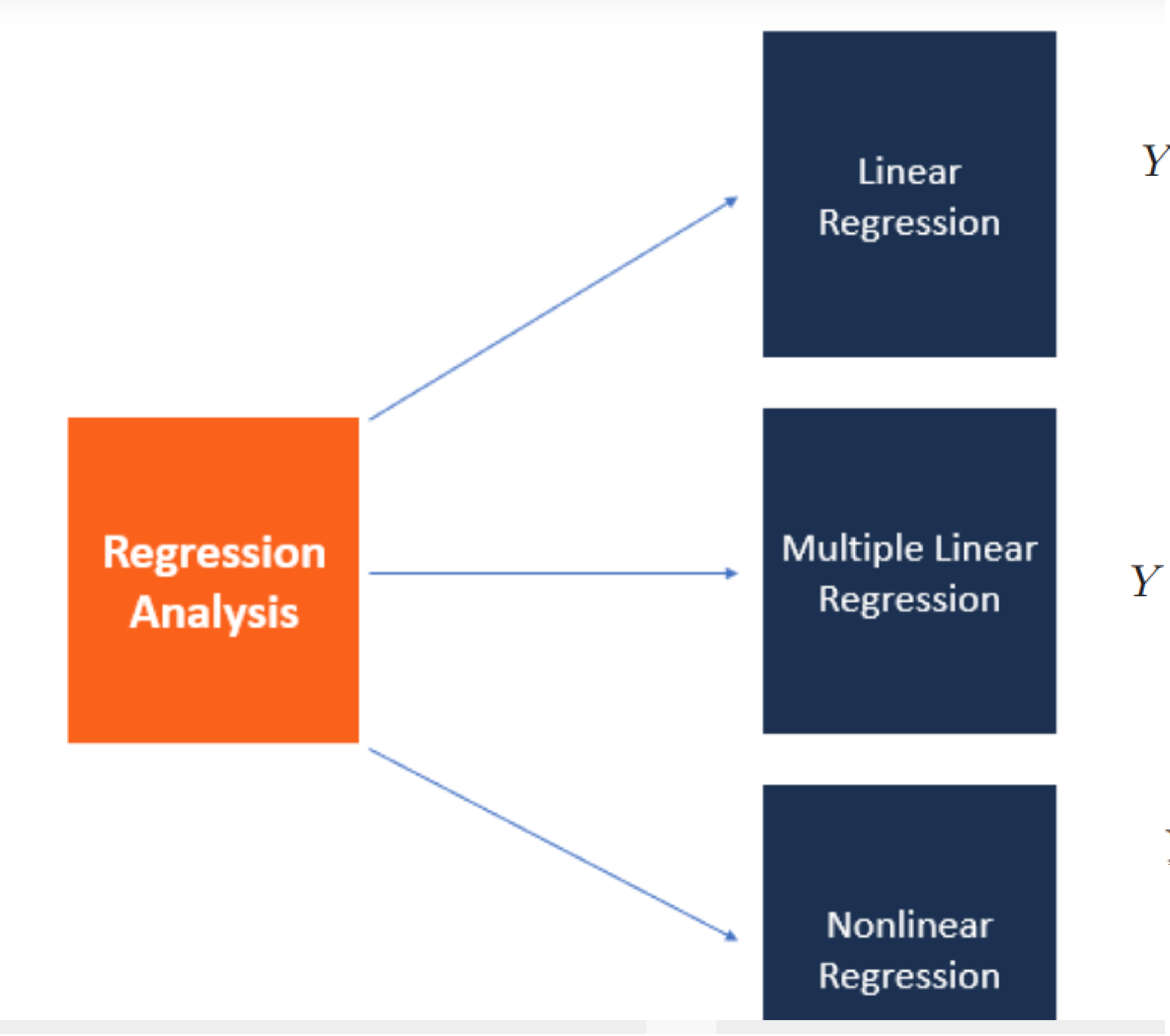
If the test statistic were substantially higher or lower than $p = 0.6$, we would reject the agent's claim. Hence, we should make the hypothesis

$$H_0\colon\ p = 0.6,$$
$$H_1\colon\ p \neq 0.6.$$

The alternative hypothesis implies a two-tailed test with the critical region divided equally in both tails of the distribution of $\widehat{P}$, our test statistic.

Chapter 11

Simple Linear Regression and Correlation



$$Y = \beta_0 + \beta_1 x,$$

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2,$$

$$Y = a\, e^{bx}$$

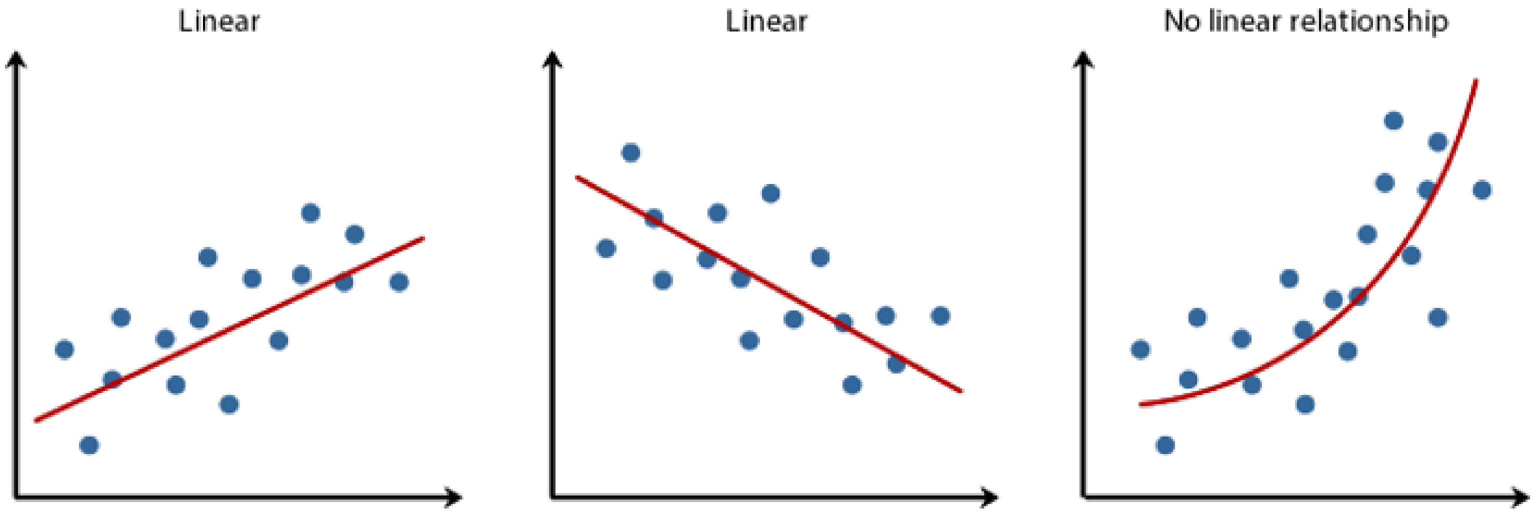
Simple Linear Regression Model

$$Y = \beta_0 + \beta_1 x + \epsilon.$$

In the above, β_0 and β_1 are unknown intercept and slope parameters, respectively, and ϵ is a random variable that is assumed to be distributed with $E(\epsilon) = 0$ and $\mathrm{Var}(\epsilon) = \sigma^2$. The quantity σ^2 is often called the error variance or residual variance.

The first step in any analysis of the relationship between the two variables is to plot the data in a **scatter plot** or **scattergram**. The predictor variable *x* is located on the horizontal axes and the response variable *y* on the vertical axis.

Creating a **scatter plot** is an important preliminary step preceding any statistical analysis of the two variables. The existence of any increasing, or decreasing, relationship becomes readily apparent.



Least squares estimates

$$\widehat{\alpha} = \bar{y} - b \cdot \bar{x} \quad \text{and} \quad \widehat{\beta} = \frac{S_{xy}}{S_{xx}}$$

Fitted (or estimated) regression line

$$\widehat{y} = \widehat{\alpha} + \widehat{\beta}\, x$$

where the $\widehat{\ }$ on y, α, and β indicates the estimated value.

The individual deviations of the observations y_i from their fitted values $\widehat{y}_i = \widehat{\alpha} + \widehat{\beta}\, x_i$ are called the **residuals**.

Residuals

$$\text{observation} - \text{fitted value} = y_i - \widehat{\alpha} - \widehat{\beta}\, x_i$$

The minimum value of the sum of squares is called the **residual sum of squares** or **error sum of squares**. Below we show that

$$\text{SSE} = \text{residual sum of squares} = \sum_{i=1}^{n} (y_i - \widehat{\alpha} - \widehat{\beta}\, x_i)^2$$

$$= S_{yy} - S_{xy}^2 / S_{xx}$$

$$S_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}$$

$$S_{yy} = \sum_{i=1}^{n} (y_i - \bar{y})^2 = \sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}$$

$$S_{xy} = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{n} x_i y_i - \frac{\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n}$$

EXAMPLE

Least squares calculations for the cooling rate data

Calculate the least squares estimates and sum of squares error for the cooling rate data.

Solution

x	y	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})^2$	$(x-\bar{x})(y-\bar{y})$	$(y-\bar{y})^2$	residual
0	25	−3	−15	9	45	225	3
1	20	−2	−20	4	40	400	−8
2	30	−1	−10	1	10	100	−4
2	40	−1	0	1	0	0	6
4	45	1	5	1	5	25	−1
4	50	1	10	1	10	100	4
5	60	2	20	4	40	400	8
6	50	3	10	9	30	100	−8
$\bar{x}=3$	$\bar{y}=40$	0	0	$S_{xx}=30$	$S_{xy}=180$	$S_{yy}=1350$	

so $\widehat{\beta} = S_{xy}/S_{xx} = 180\,/\,30 = 6$ and $\widehat{\alpha} = \bar{y} - \widehat{\beta}\,\bar{x} = 40 - 6(3) = 22$.

Since $\widehat{\beta} = 6$ and $\widehat{\alpha} = 22$, the *least squares line* is

$$\widehat{y} = \widehat{\alpha} + \widehat{\beta}\,x = 22 + 6x$$

The *residuals* are $y_i - \widehat{\alpha} - \widehat{\beta}\,x_i = y_i - 22 - 6\,x_i$, or, $25 - 22 - 6(0) = 3, -8, -4, 6, -1, 4, 8, -8$.

The sum of squares error is then

$$\text{SSE} = \sum_{i=1}^{n}(y_i - \hat{\alpha} - \hat{\beta}\,x_i)^2$$

$$= 3^2 + (-8)^2 + (-4)^2 + 6^2 + (-1)^2 + 4^2 + 8^2 + (-8)^2 = 270$$

Alternatively,

$$\text{SSE} = S_{yy} - S_{xy}^2/S_{xx} = 1350 - 180^2/30 = 270$$