CHAPTER 22

Data Analysis. Probability Theory

We first show how to handle data numerically or graphically (in terms of figures), in order to see what properties they may have and what kind of information we can extract from them. If data are influenced by chance effect (e.g., weather data, properties of steel, stock prices, etc.), they may suggest and motivate concepts and rules of probability theory because this is the theoretical counterpart of the observable reality whenever "chance" is at work. This theory gives us mathematical models of such chance processes (briefly called "experiments"; Sec. 22.2). In any such experiment we observe a "random variable" X, a function whose values in the experiment occur "by chance" (Sec. 22.5), which is characterized by a probability distribution (Secs. 22.5–22.8). Or we observe more than one random variable, for example, height and weight of persons, hardness and tensile strength of copper. This is discussed in Sec. 22.9, which will also give the basis for the mathematical justification of statistical methods in Chap. 23.

Prerequisite for this chapter: Calculus.
References: Appendix 1, Part G.
Answers to problems: Appendix 2.

22.1 Data: Representation, Average, Spread

Data can be represented numerically or graphically in various ways. For instance, your daily newspaper may contain tables of stock prices and money exchange rates, curves or bar charts illustrating economical or political developments, or pie charts showing how your tax dollar is spent. And there are numerous other representations of data for special purposes.

In this section we discuss the use of standard representations of data in statistics. (For these, software packages, such as DATA DESK and MINITAB, are available, and MAPLE or MATHEMATICA may also be helpful.) We explain corresponding concepts and methods in terms of typical examples, beginning with

(1) 89 84 87 81 89 86 91 90 78 89 87 99 83 89.

These are $n = 14$ measurements of the tensile strength of sheet steel in kg/mm^2, recorded in the order obtained and rounded to integer values. To see what is going on, we **sort** these data, that is, we order them by size,

(2) 78 81 83 84 86 87 87 89 89 89 89 90 91 99.

Sorting is a standard process on the computer; see Ref. [E13], Chap. 8, listed in Appendix 1.

Graphical Representation of Data

We shall now discuss standard graphical representations used in statistics for obtaining information on properties of data.

Stem-and-Leaf Plot

This is one of the simplest but most useful representations of data. For (1) it is shown in Fig. 475. The numbers in (1) range from 78 to 99 [see (2)]. We divide these numbers into 5 groups, 75–79, 80–84, 85–89, 90–94, 95–99. The integers in the tens position of the groups are 7, 8, 8, 9, 9. These form the *stem* in Fig. 475. The first *leaf* is 8 (representing 78). The second leaf is 134 (representing 81, 83, 84), and so on.

Leaf unit = 1.0

1	7	8
4	8	134
11	8	6779999
13	9	01
14	9	9

Fig. 475. Stem-and-leaf plot of the data in (1) and (2)

The number of times a value occurs is called its **absolute frequency.** Thus 78 has absolute frequency 1, the value 89 has absolute frequency 4, etc. The column to the extreme left in Fig. 475 shows the **cumulative absolute frequencies,** that is, the sum of the absolute frequencies of the values up to the line of the leaf. Thus, the number 4 in the second line on the left shows that (1) has 4 values up to and including 84. The number 11 in the next line shows that there are 11 values not exceeding 89, etc. Dividing the cumulative absolute frequencies by n (= 14 in Fig. 475) gives the **cumulative relative frequencies.**

Histogram

For large sets of data, histograms are better in displaying the distribution of data than stem-and-leaf plots. The principle is explained in Fig. 476. (An application to a larger

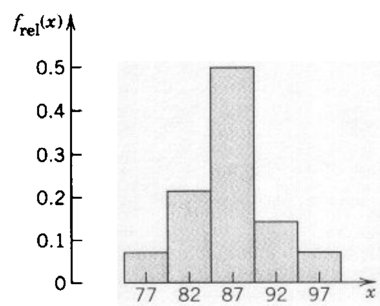


Fig. 476. Histogram of the data in (1) and (2) (grouped as in Fig. 475)

data set is shown in Sec. 23.7). The bases of the rectangles in Fig. 476 are the x-intervals (known as **class intervals**) 74.5–79.5, 79.5–84.5, 84.5–89.5, 89.5–94.5, 94.5–99.5, whose midpoints (known as **class marks**) are $x = 77, 82, 87, 92, 97$, respectively. The height of a rectangle with class mark x is the **relative class frequency** $f_{\text{rel}}(x)$, defined as the number of data values in that class interval, divided by n (= 14 in our case). Hence the areas of the rectangles are proportional to these relative frequencies, so that histograms give a good impression of the distribution of data.

Center and Spread of Data: Median, Quartiles

As a center of the location of data values we can simply take the **median,** the data value that falls in the middle when the values are ordered. In (2) we have 14 values. The seventh of them is 87, the eighth is 89, and we split the difference, obtaining the median 88. (In general, we would get a fraction.)

The spread (variability) of the data values can be measured by the **range** $R = x_{\max} - x_{\min}$, the largest minus the smallest data values, $R = 99 - 78 = 21$ in (2).

Better information gives the **interquartile range** $\text{IQR} = q_U - q_L$. Here the **upper quartile** q_U is the middle value among the data values *above* the median. The **lower quartile** q_L is the middle value among the data values *below* the median. Thus in (2) we have $q_U = 89$ (the fourth value from the end), $q_L = 84$ (the fourth value from the beginning), and $\text{IQR} = 89 - 84 = 5$. The median is also called the **middle quartile** and is denoted by q_M. The rule of "splitting the difference" (just applied to the middle quartile) is equally well used for the other quartiles if necessary.

Boxplot

The **boxplot** of (1) in Fig. 477 is obtained from the five numbers $x_{\min}, q_L, q_M, q_U, x_{\max}$ just determined. The box extends from q_L to q_U. Hence it has the height IQR. The position of the median in the box shows that the data distribution is not symmetric. The two lines extend from the box to $x_{\min}$ below and to $x_{\max}$ above. Hence they mark the range R.

Boxplots are particularly suitable for making comparisons. For example, Fig. 477 shows boxplots of the data sets (1) and

$$(3) \qquad 91 \quad 89 \quad 93 \quad 91 \quad 87 \quad 94 \quad 92 \quad 85 \quad 91 \quad 90 \quad 96 \quad 93 \quad 89$$

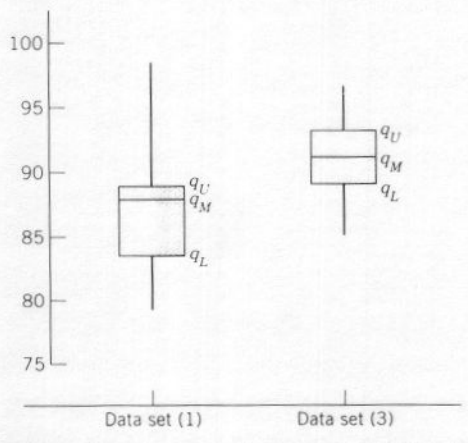


Fig. 477. Boxplots of data sets (1) and (3)

(consisting of $n = 13$ values). Ordering gives

(4) 85 87 89 89 90 91 91 91 92 93 93 94 96

(tensile strength, as before). From the plot we immediately see that the box of (3) is shorter than the box of (1) (indicating the higher quality of the sheets!) and that q_M is located in the middle of the box (showing the more symmetric form of the distribution). Finally, $x_{\max}$ is closer to q_U for (3) than for (1), a fact that we shall discuss later.

For plotting the box of (3) we took from (4) $x_{\min} = 85$, $q_L = 89$, $q_M = 91$, $q_U = 93$, $x_{\max} = 96$.

Outliers

An **outlier** is a value that appears to be uniquely different from the rest of the data set. It might indicate that something went wrong with the data collection process. In connection with quartiles an outlier is conventionally defined as a value more than a distance of 1.5 IQR from either end of the box.

For the data in (1) we have IQR = 5, $q_L = 84$, $q_U = 89$. Hence outliers are smaller than $84 - 7.5$ or larger than $89 + 7.5$, so that 99 is an outlier [see (2)]. The data (3) have no outliers, as you can readily verify.

Mean. Standard Deviation. Variance

Medians and quartiles are easily obtained by ordering and counting, practically without calculation. But they do not give full information on data: you can move data values around to some extent without changing the median. Similarly for the quartiles.

The average size of the data values can be measured in a more refined way by the **mean**

$$\bar{x} = \frac{1}{n}\sum_{j=1}^{n} x_j = \frac{1}{n}(x_1 + x_2 + \cdots + x_n). \tag{5}$$

This is the arithmetic mean of the data values, obtained by taking their sum and dividing by the data *size* n. Thus in (1),

$$\bar{x} = \tfrac{1}{14}(89 + 84 + \cdots + 89) = \tfrac{611}{7} \approx 87.3.$$

Every data value contributes, and changing one of them will change the mean.

Similarly, the spread (variability) of the data values can be measured in a more refined way by the **standard deviation** s or by its square, the **variance**

$$s^2 = \frac{1}{n-1}\sum_{j=1}^{n}(x_j - \bar{x})^2 = \frac{1}{n-1}\left[(x_1 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2\right]. \tag{6}$$

Thus, to obtain the variance of the data, take the difference $x_j - \bar{x}$ of each data value from the mean, square it, take the sum of these n squares, and divide it by $n - 1$ (not n, as we motivate in Sec. 23.2). To get the standard deviation s, take the square root of s^2.

For example, using $\bar{x} = 611/7$, we get for the data (1) the variance

$$s^2 = \tfrac{1}{13}\left[\left(89 - \tfrac{611}{7}\right)^2 + \left(84 - \tfrac{611}{7}\right)^2 + \cdots + \left(89 - \tfrac{611}{7}\right)^2\right] = \tfrac{176}{7} \approx 25.14.$$

Hence the standard deviation is $s = \sqrt{176/7} \approx 5.014$. Note that the standard deviation has the same dimension (kg/mm^2) as the data values. On the other hand, the variance is more advantageous than the standard deviation in developing statistical methods.

Caution! Your CAS (MAPLE, for instance) may use $1/n$ instead of $1/(n - 1)$ in (6), but the latter is better when n is small (see Sec. 23.2).

PROBLEM SET 22.1

Representation of Data

Represent the following data by a stem-and-leaf plot, a histogram, and a boxplot.

1. 12 11 9 5 12 6 7 9 11 11

2. 17 18 17 16 17 16 18 16

3. 46 48 44 23 31 20 34 27 41 36 46 28 28 39 29

4. −0.51 0.12 −0.47 0.95 0.25 −0.18 −0.54

5. 50.6 50.9 49.1 51.3 50.5 49.7 51.5 49.8 51.1 48.9 50.3 49.2 51.2 50.4 52.8

6. 13.1 11.0 13.4 11.5 10.2 18.2 12.4 12.8 15.7 10.9

7. Release time [sec] of a relay

1.3 1.2 1.4 1.5 1.3 1.3 1.4 1.1 1.5 1.4

1.6 1.3 1.5 1.1 1.4 1.2 1.3 1.5 1.4 1.4

8. Carbon content [%] of coal

86 87 86 81 77 85 87 86 85 87

82 84 83 79 82 73 86 84 83 83

9. Miles per gallon of gasoline required by six cars of the same make

15.0 15.5 14.5 15.0 15.5 15.0

10. Weight of filled bags [grams] in an automatic filling process

203 199 198 201 200 201 201

Average and Spread

Find the mean and compare it with the median. Find the standard deviation and compare it with the interquartile range.

11. The data in Prob. 1

12. The data in Prob. 2

13. The data in Prob. 5

14. 5 22 7 23 6. Why is $|\bar{x} - q_M|$ so large?

15. Construct the simplest possible data with $\bar{x} = 100$ but $q_M = 0$.

16. **(Mean)** Prove that $\bar{x}$ must always lie between the smallest and the largest data values.

17. **(Outlier, reduced data)** Calculate s for the data 4, 1, 3, 10, 2. Then reduce the data by deleting the outlier and calculate s. Comment.

18. **WRITING PROJECT. Average and Spread.** Compare Q_M, IQR and $\bar{x}$, s, illustrating the advantages and disadvantages with examples and plots of your own.

22.2 Experiments, Outcomes, Events

We now turn to **probability theory.** This theory has the purpose of providing mathematical models of situations affected or even governed by "chance effects," for instance, in weather forecasting, life insurance, quality of technical products (computers, batteries, steel sheets, etc.), traffic problems, and, of course, games of chance with cards or dice. And the accuracy of these models can be tested by suitable observations or experiments—this is the purpose of **statistics** in Chap. 23.

We begin by defining some standard terms. An **experiment** is a process of measurement or observation, in a laboratory, in a factory, on the street, in nature, or wherever; so "experiment" is used in a rather general sense. Our interest is in experiments that involve **randomness,** chance effects, so that we cannot predict a result exactly. A **trial** is a single performance of an experiment. Its result is called an **outcome** or a **sample point.** n trials then give a **sample** of **size** n consisting of n sample points. The **sample space** S of an experiment is the set of all possible outcomes. Examples are

EXAMPLES 1–6

(1) Inspecting a lightbulb. $S = \{\text{Defective, Nondefective}\}$.
(2) Rolling a die. $S = \{1, 2, 3, 4, 5, 6\}$.
(3) Measuring tensile strength of wire. S the numbers in some interval.
(4) Measuring copper content of brass. S: 50% to 90%, say.
(5) Counting daily traffic accidents in New York. S the integers in some interval.
(6) Asking for opinion about a new car model. $S = \{\text{Like, Dislike, Undecided}\}$.

The subsets of S are called **events** and the outcomes **simple events.**

EXAMPLE 7 Events

In (2), events are $A = \{1, 3, 5\}$ (*"Odd number"*), $B = \{2, 4, 6\}$ (*"Even number"*), $C = \{5, 6\}$, etc. Simple events are $\{1\}, \{2\}, \cdots, \{6\}$. ◀

If in a trial an outcome a happens and $a \in A$ (*a is an element of A*), we say that A happens. For instance, if a die turns up a 3, the event A: *Odd number* happens. Similarly, if C happens (meaning 5 or 6 turns up), then $D = \{4, 5, 6\}$ happens. Also note that S happens in each trial, meaning that *some* event of S always happens. All this is quite natural.

Unions, Intersections, Complements of Events

In connection with basic probability laws we shall need the following concepts and facts about events (subsets) $A, B, C, \cdots$ of a given sample space S.

The **union** $A \cup B$ of A and B consists of all points in A or B or both.

The **intersection** $A \cap B$ of A and B consists of all points that are in both A and B.

If A and B have no points in common, we write

$$A \cap B = \varnothing$$

where $\varnothing$ is the *empty set* (set with no elements) and we call A and B **mutually exclusive** (or **disjoint**) because the occurrence of A *excludes* that of B (and conversely)—if your die turns up an odd number, it cannot turn up an even number in the same trial. Similarly, a coin cannot turn up *Head* and *Tail* at the same time.

The **complement**[1] A^c of A consists of all the points of S ***not*** in A. Thus,

$$A \cap A^c = \varnothing, \qquad A \cup A^c = S.$$

In Example 7 we have $A^c = B$, hence $A \cup A^c = \{1, 2, 3, 4, 5, 6\} = S$.

Unions and intersections of more events are defined similarly. The **union**

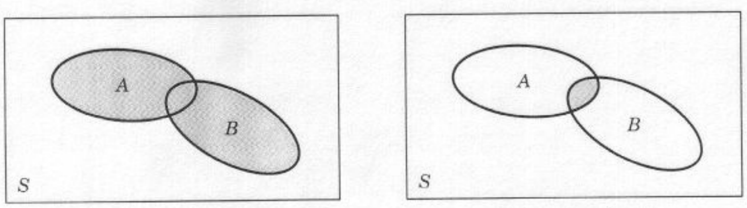
$$\bigcup_{j=1}^{m} A_j = A_1 \cup A_2 \cup \cdots \cup A_m$$

of events $A_1, \cdots, A_m$ consists of all points that are in at least one A_j. Similarly for the union $A_1 \cup A_2 \cup \cdots$ of infinitely many subsets $A_1, A_2, \cdots$ of an *infinite* sample space S (that is, S consists of infinitely many points). The **intersection**

$$\bigcap_{j=1}^{m} A_j = A_1 \cap A_2 \cap \cdots \cap A_m$$

of $A_1, \cdots, A_m$ consists of the points of S that are in each of these events. Similarly for the intersection $A_1 \cap A_2 \cap \cdots$ of infinitely many subsets of S.

Working with events can be illustrated and facilitated by **Venn diagrams**[2] for showing unions, intersections, and complements, as in Figs. 478 and 479, which are typical examples that give the idea.



Union $A \cup B$ Intersection $A \cap B$

Fig. 478. Venn diagrams showing two events A and B in a sample space S and their union $A \cup B$ (colored) and intersection $A \cap B$ (colored)

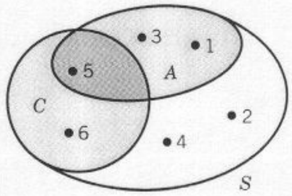


Fig. 479. Venn diagram for the experiment of rolling a die, showing S, $A = \{1, 3, 5\}$, $C = \{5, 6\}$, $A \cup C = \{1, 3, 5, 6\}$, $A \cap C = \{5\}$

[1] Or $\overline{A}$, but we shall not use this because in set theory it is used to denote the closure of A.

[2] JOHN VENN (1834—1923), English mathematician.

EXAMPLE 8 Unions and intersections of 3 events

In rolling a die, consider the events

A: *Number greater than* 3, B: *Number less than* 6, C: *Even number.*

Then $A \cap B = \{4, 5\}$, $B \cap C = \{2, 4\}$, $C \cap A = \{4, 6\}$, $A \cap B \cap C = \{4\}$. Can you sketch a Venn diagram of this? Furthermore, $A \cup B = S$, hence $A \cup B \cup C = S$ (why?), etc. ◀

PROBLEM SET 22.2

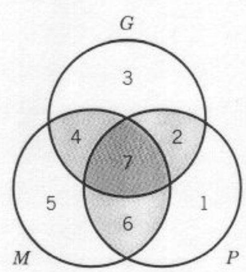
Sample Spaces, Events

Graph a sample space for the following experiments.

1. Drawing three screws from a lot of right-handed and left-handed screws
2. Rolling two dice
3. Tossing two coins
4. Rolling a die until the first 6 appears
5. Drawing bolts from a lot of 10, containing 1 defective D, until D is drawn, assuming **sampling without replacement,** that is, bolts drawn are not returned to the lot.
6. In Prob. 1 let A, B, C, D mean 1 right-handed, 1 left-handed, 2 right-handed, 2 left-handed, respectively, among the 3 screws drawn. Are A and B mutually exclusive? C and D?
7. In rolling two dice, are A: *Sum divisible by* 3, B: *Sum divisible by* 5 mutually exclusive? Answer the same question for rolling three dice.
8. In Prob. 2 circle and mark the events A: *Faces are equal,* B: *Sum of faces less than* 5, $A \cup B$, $A \cap B$, A^c, B^c.
9. List all eight subsets of $S = \{a, b, c\}$.
10. In Prob. 4 list the outcomes that make up the event E: *First "Six" in rolling at most* 5 *times.* Describe E^c.

Venn Diagrams

11. In connection with a trip to Europe by some students, consider the events P that they see Paris, G that they have a good time, and M that they run out of money, and describe in words the events $1, \cdots, 7$ in the diagram.



Problem 11

12. In a lot of 20 gaskets, 7 have no defect, 3 have a T-defect (too thin), 6 have an L-defect (too large), and 4 have both defects. Show this in a Venn diagram, also giving the number in each set.
13. (**De Morgan's laws**) Using Venn diagrams, graph and check *De Morgan's laws*

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

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

14. Using Venn diagrams, graph and check the rules

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

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

15. Show that, by the definition of complement, for any subset A of a sample space S,

$$(A^c)^c = A, \quad S^c = \varnothing, \quad \varnothing^c = S, \quad A \cup A^c = S, \quad A \cap A^c = \varnothing.$$

22.3 Probability

The "probability" of an event A in an experiment is supposed to measure how frequently A is about to occur if we make many trials. If we flip a coin, then heads H and tails T will appear *about* equally often—we say that H and T are **"equally likely."** Table 22.1 confirms this. Similarly, for a regularly shaped die of homogeneous material (**"fair die"**) each of the six outcomes $1, \cdots, 6$ will be equally likely. These are examples of experiments in which the sample space S consists of finitely many outcomes (points) that for reasons of some symmetry can be regarded as equally likely. This suggests the following definition.

Definition 1. Probability

If the sample space S of an experiment consists of finitely many outcomes (points) that are equally likely, then the probability $P(A)$ of an event A is

(1) $$P(A) = \frac{\text{Number of points in } A}{\text{Number of points in } S}.$$

Thus, in particular,

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

as follows directly from (1).

EXAMPLE 1 Fair die

In rolling a fair die, what is the probability $P(A)$ of A of obtaining at least a 5? The probability of B: *"Even number"*?

Solution. The six outcomes are equally likely, so that each has probability 1/6. Thus $P(A) = 2/6 = 1/3$ because $A = \{5, 6\}$ has 2 points, and $P(B) = 3/6 = 1/2$. ◀

Table 22.1
Coin Tossing

Experiments by	Number of Throws	Number of Heads	Relative Frequency of Heads
BUFFON	4,040	2,048	0.5069
K. PEARSON	12,000	6,019	0.5016
K. PEARSON	24,000	12,012	0.5005

Definition 1 takes care of many games as well as some practical applications, as we shall see, but certainly not of all experiments, simply because in many problems we do not have finitely many equally likely outcomes. To arrive at a more general definition of probability, we regard ***probability as the counterpart of relative frequency.*** Recall from Sec. 22.1 that the **absolute frequency** $f(A)$ of an event A in n trials is the number of times A occurs, and the **relative frequency** of A in these trials is $f(A)/n$; thus

$$f_{\text{rel}}(A) = \frac{f(A)}{n} = \frac{\text{Number of times } A \text{ occurs}}{\text{Number of trials}} \tag{3}$$

Now if A did not occur, then $f(A) = 0$. If A always occurred, then $f(A) = n$. These are the extreme cases. Division by n gives

$$0 \leqq f_{\text{rel}}(A) \leqq 1. \tag{4*}$$

In particular, for $A = S$ we have $f(S) = n$ because S always occurs (meaning that some event always occurs; if necessary, see Sec. 22.2, after Example 7). Division by n gives

$$f_{\text{rel}}(S) = 1.$$

Finally, if A and B are mutually exclusive, they cannot occur together. Hence the absolute frequency of their union $A \cup B$ must equal the sum of the absolute frequencies of A and B. Division by n gives the same relation for the relative frequencies,

$$f_{\text{rel}}(A \cup B) = f_{\text{rel}}(A) + f_{\text{rel}}(B) \qquad (A \cap B = \varnothing).$$

We are now ready to extend the definition of probability to experiments in which equally likely outcomes are not available. Of course, the extended definition should include Definition 1. Since probabilities are supposed to be the theoretical counterpart of relative frequencies, we choose the properties in (4*), (5*), (6*) as axioms. (Historically, such a choice is the result of a long process of gaining experience on what might be best and most practical.)

Definition 2. Probability

Given a sample space S, with each event A of S (subset of S) there is associated a number $P(A)$, called the **probability** of A, such that the following **axioms of probability** are satisfied.

1. For every A in S,

$$0 \leqq P(A) \leqq 1. \tag{4}$$

2. The entire sample space S has the probability

$$P(S) = 1.$$

3. For mutually exclusive events A and B ($A \cap B = \varnothing$; see Sec. 22.2),

$$P(A \cup B) = P(A) + P(B) \qquad (A \cap B = \varnothing). \tag{6}$$

If S is infinite (has infinitely many points), Axiom 3 has to be replaced by[3]

3′. For mutually exclusive events $A_1, A_2, \cdots$,

$$P(A_1 \cup A_2 \cup \cdots) = P(A_1) + P(A_2) + \cdots. \tag{6'}$$

Basic Theorems for Probability

We shall see that the axioms of probability will enable us to build up probability theory and its application to statistics. We begin with three basic theorems. The first of them is useful if we can get the probability of the complement A^c more easily than $P(A)$ itself.

THEOREM 1 (Complementation rule)

For an event A and its complement A^c in a sample space S,

$$P(A^c) = 1 - P(A). \tag{7}$$

PROOF. By the definition of complement (Sec. 22.2) we have $S = A \cup A^c$ and $A \cap A^c = \varnothing$. Hence by Axioms 2 and 3,

$$1 = P(S) = P(A) + P(A^c), \qquad \text{thus} \qquad P(A^c) = 1 - P(A). \quad ◀$$

EXAMPLE 2 Coin tossing

Five coins are tossed simultaneously. Find the probability of the event A: *At least one head turns up.* Assume that the coins are fair.

Solution. Since each coin can turn up heads or tails, the sample space consists of $2^5 = 32$ outcomes. Since the coins are fair, we may assign the same probability (1/32) to each outcome. Then the event A^c (*No heads turn up*) consists of only 1 outcome. Hence $P(A^c) = 1/32$, and the answer is $P(A) = 1 - P(A^c) = 31/32$. ◀

The next theorem is a simple extension of Axiom 3, which you can readily prove by induction:

THEOREM 2 (Addition rule for mutually exclusive events)

For mutually exclusive events $A_1, \cdots, A_m$ in a sample space S,

$$P(A_1 \cup A_2 \cup \cdots A_m) = P(A_1) + P(A_2) + \cdots + P(A_m). \tag{8}$$

EXAMPLE 3 Mutually exclusive events

If the probability that on any workday a garage will get 10–20, 21–30, 31–40, over 40 cars to service is 0.20, 0.35, 0.25, 0.12, respectively, what is the probability that on a given workday the garage gets at least 21 cars to service?

Solution. Since these are mutually exclusive events, Theorem 2 gives the answer $0.35 + 0.25 + 0.12 = 0.72$. Check this by the complementation rule. ◀

[3]In the infinite case, for a *theoretical* restriction of the subsets of S, of no *practical* consequence to us, see "σ-algebra," for example, in Ref. [8] listed in Appendix 1.

In many cases, events will not be mutually exclusive. Then we have

THEOREM 3 **(Addition rule for arbitrary events)**

For events A and B in a sample space,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B). \tag{9}$$

PROOF. C, D, E in Fig. 480 make up $A \cup B$ and are mutually exclusive (disjoint). Hence by Theorem 2,

$$P(A \cup B) = P(C) + P(D) + P(E).$$

This gives (9) because on the right $P(C) + P(D) = P(A)$ by Axiom 3 and disjointness; and $P(E) = P(B) - P(D) = P(B) - P(A \cap B)$, also by Axiom 3 and disjointness. ◀

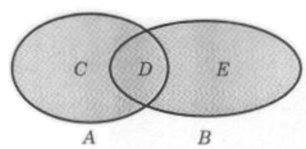


Fig. 480. Proof of Theorem 3

Note that for mutually exclusive events A and B we have $A \cap B = \varnothing$ by definition and, by comparing (9) and (6),

$$\boxed{P(\varnothing) = 0.} \tag{10}$$

(Can you also prove this by (5) and (7)?)

EXAMPLE 4 Union of arbitrary events

In tossing a fair die, what is the probability of getting an odd number or a number less than 4?

Solution. Let A be the event "*Odd number*" and B the event "*Number less than* 4." Then Theorem 3 gives the answer

$$P(A \cup B) = \tfrac{3}{6} + \tfrac{3}{6} - \tfrac{2}{6} = \tfrac{2}{3}$$

because $A \cap B$ = "*Odd number less than* 4" = $\{1, 3\}$. ◀

Conditional Probability. Independent Events

Often it is required to find the probability of an event B under the condition that an event A occurs. This probability is called the **conditional probability** ***of B given A*** and is denoted by $P(B|A)$. In this case A serves as a new (reduced) sample space, and that probability is the fraction of $P(A)$ which corresponds to $A \cap B$. Thus

$$\boxed{P(B|A) = \frac{P(A \cap B)}{P(A)}} \qquad [P(A) \neq 0]. \tag{11}$$

Similarly, the *conditional probability of A given B* is

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \qquad [P(B) \neq 0]. \tag{12}$$

Solving (11) and (12) for $P(A \cap B)$, we obtain

THEOREM 4 (Multiplication rule)

If A and B are events in a sample space S and $P(A) \neq 0$, $P(B) \neq 0$, *then*

$$P(A \cap B) = P(A)P(B|A) = P(B)P(A|B). \tag{13}$$

EXAMPLE 5 Multiplication rule

In producing screws, let A mean "screw too slim" and B "screw too short." Let $P(A) = 0.1$ and let the conditional probability that a slim screw is also too short be $P(B|A) = 0.2$. What is the probability that a screw that we pick randomly from the lot produced will be both too slim and too short?

Solution. $P(A \cap B) = P(A)P(B|A) = 0.1 \cdot 0.2 = 0.02 = 2\%$, by Theorem 4. ◀

Independent events. If events A and B are such that

$$P(A \cap B) = P(A)P(B), \tag{14}$$

they are called **independent events.** Assuming $P(A) \neq 0$, $P(B) \neq 0$, we see from (11)–(13) that in this case

$$P(A|B) = P(A), \qquad P(B|A) = P(B).$$

This means that the probability of A does not depend on the occurrence or nonoccurrence of B, and conversely. This justifies the term "independent."

Independence of m events. Similarly, m events $A_1, \cdots, A_m$ are called **independent** if

$$P(A_1 \cap \cdots \cap A_m) = P(A_1) \cdots P(A_m) \tag{15a}$$

as well as for every k different events $A_{j_1}, A_{j_2}, \cdots, A_{j_k}$

$$P(A_{j_1} \cap A_{j_2} \cap \cdots \cap A_{j_k}) = P(A_{j_1})P(A_{j_2}) \cdots P(A_{j_k}) \tag{15b}$$

where $k = 2, 3, \cdots, m - 1$.

Accordingly, three events A, B, C are independent if

$$\begin{aligned} P(A \cap B) &= P(A)P(B), \\ P(B \cap C) &= P(B)P(C), \\ P(C \cap A) &= P(C)P(A), \\ P(A \cap B \cap C) &= P(A)P(B)P(C). \end{aligned} \tag{16}$$

Sampling. Our next example has to do with randomly drawing objects, *one at a time,* from a given set of objects. This is called **sampling from a population,** and there are two ways of sampling, as follows.

1. In **sampling with replacement,** the object that was drawn at random is placed back to the given set and the set is mixed thoroughly. Then we draw the next object at random.

2. In **sampling without replacement** the object that was drawn is put aside.

EXAMPLE 6 Sampling with and without replacement

A box contains 10 screws, three of which are defective. Two screws are drawn at random. Find the probability that none of the two screws is defective.

Solution. We consider the events

A: First drawn screw nondefective.

B: Second drawn screw nondefective.

Clearly, $P(A) = \frac{7}{10}$ because 7 of the 10 screws are nondefective and we sample at random, so that each screw has the same probability ($\frac{1}{10}$) of being picked. If we sample with replacement, the situation before the second drawing is the same as at the beginning, and $P(B) = \frac{7}{10}$. The events are independent, and the answer is

$$P(A \cap B) = P(A)P(B) = 0.7 \cdot 0.7 = 0.49 = 49\%.$$

If we sample without replacement, then $P(A) = \frac{7}{10}$, as before. If A has occurred, then there are 9 screws left in the box, 3 of which are defective. Thus $P(B|A) = \frac{6}{9} = \frac{2}{3}$, and Theorem 4 yields the answer

$$P(A \cap B) = \tfrac{7}{10} \cdot \tfrac{2}{3} \approx 47\%.$$

Is it intuitively clear that this value must be smaller than the preceding one? ◀

PROBLEM SET 22.3

1. In rolling two fair dice, what is the probability of obtaining a sum greater than 3 but not exceeding 6?

2. In Prob. 1, what is the probability of obtaining a sum not exceeding 10?

3. If a box contains 10 left-handed and 20 right-handed screws, what is the probability of obtaining at least one right-handed screw in drawing 2 screws with replacement?

4. Will the probability in Prob. 3 increase or decrease if we draw without replacement. First guess, then calculate.

5. Three screws are drawn at random from a lot of 100 screws, 10 of which are defective. Find the probability of the event that all 3 screws drawn are nondefective, assuming that we draw (a) with replacement, (b) without replacement.

6. What is the probability of obtaining at least one Six in rolling three fair dice?

7. Under what conditions will it make *practically* no difference whether we sample with or without replacement?

8. Two boxes contain ten chips each, numbered from 1 to 10, and one chip is drawn from each box. Find the probability of the event E that the sum of the numbers on the drawn chips is greater than 4.

9. If a certain kind of tire has a life exceeding 30 000 miles with probability 0.90, what is the probability that a set of these tires on a car will last longer than 30 000 miles?

10. A batch of 200 iron rods consists of 50 oversized rods, 50 undersized rods, and 100 rods of the desired length. If two rods are drawn at random without replacement, what is the probability