Chapter 2

Probability

This chapter introduces the key concept of *probability*, its fundamental rules and properties, and discusses most basic methods of computing probabilities of various events.

intuition = sezgi
anticipate = tahmin etmek

2.1 Events and their probabilities

The concept of *probability* perfectly agrees with our intuition. In everyday life, probability of an event is understood as a *chance* that this event will happen.

Example 2.1. If a fair coin is tossed, we say that it has a 50-50 (equal) chance of turning up heads or tails. Hence, the probability of each side equals 1/2. It does not mean that a coin tossed 10 times will always produce exactly 5 heads and 5 tails. If you don't believe, try it! However, if you toss a coin 1 million times, the proportion of heads is anticipated to be very close to 1/2.

$\Diamond$

This example suggests that in a long run, probability of an event can be viewed as a *proportion* of times this event happens, or its *relative frequency*. In forecasting, it is common to speak about the probability as a *likelihood* (say, the company's profit is likely to rise during the next quarter). In gambling and lottery, probability is equivalent to *odds*. Having the winning odds of 1 to 99 (1:99) means that the probability to win is 0.01, and the probability to lose is 0.99. It also means, on a relative-frequency language, that if you play long enough, you will win about 1% of the time.

Example 2.2. If there are 5 communication channels in service, and a channel is selected at random when a telephone call is placed, then each channel has a probability $1/5 = 0.2$ of being selected. $\Diamond$

Example 2.3. Two competing software companies are after an important contract. Company A is twice as likely to win this competition as company B. Hence, the probability to win the contract equals $2/3$ for A and $1/3$ for B.

$\Diamond$

A mathematical definition of probability will be given in Section 2.2.1, after we get acquainted with a few fundamental concepts.

2.1.1 Outcomes, events, and the sample space

Probabilities arise when one considers and weighs possible results of some *experiment*. Some results are more likely than others. An experiment may be as simple as a coin toss, or as complex as starting a new business.

DEFINITION 2.1

> A collection of all elementary results, or **outcomes** of an experiment, is called a **sample space**.

DEFINITION 2.2

> Any set of outcomes is an **event**. Thus, events are subsets of the sample space.

Example 2.4. A tossed die can produce one of 6 possible outcomes: 1 dot through 6 dots. Each outcome is an event. There are other events: observing an even number of dots, an odd number of dots, a number of dots less than 3, etc. $\Diamond$

A sample space of N possible outcomes yields 2^N possible events.

PROOF: To count all possible events, we shall see how many ways an event can be constructed. The first outcome can be included into our event or excluded, so there are two possibilities. Then, every next outcome is either included or excluded, so every time the number of possibilities doubles. Overall, we have

$$\overbrace{2 \cdot 2 \cdot \ldots \cdot 2}^{N \text{ times}} = 2^N \qquad (2.1)$$

possibilities, leading to a total of 2^N possible events. $\Box$

Example 2.5. Consider a football game between the Dallas Cowboys and the New York Giants. The sample space consists of 3 outcomes,

$$\Omega = \{ \text{ Cowboys win, Giants win, they tie } \}$$

Combining these outcomes in all possible ways, we obtain the following $2^3 = 8$ events: Cowboys win, lose, tie, get at least a tie, get at most a tie, no tie, get *some* result, and get *no result*. The event "some result" is the entire sample space Ω, and by common sense, it should have probability 1. The event "no result" is empty, it does not contain any outcomes, so its probability is 0. $\Diamond$

NOTATION			
	Ω	=	sample space
	$\varnothing$	=	empty event
	$\boldsymbol{P}\{E\}$	=	probability of event E

2.1.2 Set operations

Events are *sets* of outcomes. Therefore, to learn how to compute probabilities of events, we shall discuss some *set operations*. Namely, we shall define unions, intersections, differences, and complements.

DEFINITION 2.3

A **union** of events $A, B, C, \ldots$ is an event consisting of *all* the outcomes in all these events. It occurs if *any* of $A, B, C, \ldots$ occurs, and therefore, corresponds to the word "OR": A or B or C or ... (Figure 2.1a).

Diagrams like Figure 2.1, where events are represented by circles, are called *Venn diagrams*.

DEFINITION 2.4

An **intersection** of events $A, B, C, \ldots$ is an event consisting of outcomes that are *common* in all these events. It occurs if *each* $A, B, C, \ldots$ occurs, and therefore, corresponds to the word "AND": A and B and C and ... (Figure 2.1b).

DEFINITION 2.5

A **complement** of an event A is an event that occurs every time when A does not occur. It consists of outcomes excluded from A, and therefore, corresponds to the word "NOT": not A (Figure 2.1c).

DEFINITION 2.6

A **difference** of events A and B consists of all outcomes included in A but excluded from B. It occurs when A occurs and B does not, and corresponds to "BUT NOT": A but not B (Figure 2.1d).

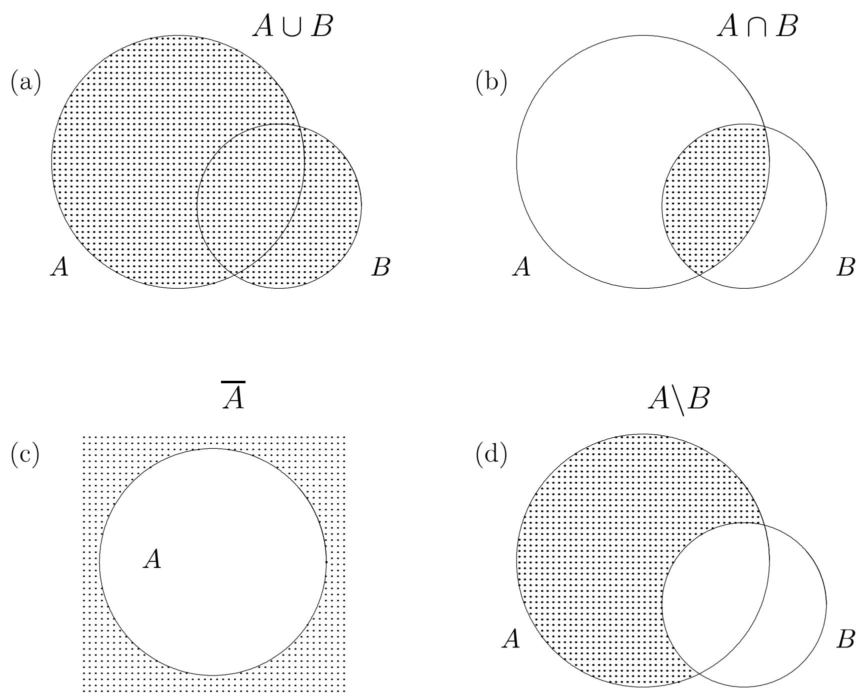

FIGURE 2.1: *Venn diagrams for (a) union, (b) intersection, (c) complement, and (d) difference of events.*

<u>NOTATION</u>

$A \cup B$	$=$	union
$A \cap B$	$=$	intersection
$\overline{A}$ or A^c	$=$	complement
$A \backslash B$	$=$	difference

DEFINITION 2.7

Events A and B are **disjoint** if their intersection is empty,

$$A \cap B = \varnothing.$$

Events $A_1, A_2, A_3, \ldots$ are **mutually exclusive** or **pairwise disjoint** if any two of these events are disjoint, i.e.,

$$A_i \cap A_j = \varnothing \quad \text{for any} \quad i \neq j\,.$$

DEFINITION 2.8

Events $A, B, C, \ldots$ are **exhaustive** if their union equals the whole sample space, i.e.,

$$A \cup B \cup C \cup \ldots = \Omega.$$

disjoint event =ayik olay
exhaustive events = birlesimleri ornek uzayin tamami ise

mutually = karsilikli olarak
exclusive = ayricalikli,ozel,haric
tutulan

Mutually exclusive events will never occur at the same time. Occurrence of any one of them eliminates the possibility for all the others to occur.

Exhaustive events cover the entire Ω, so that "there is nothing left." In other words, among any collection of exhaustive events, at least one occurs for sure.

Example 2.6. When a card is pooled from a deck at random, the four suits are at the same time disjoint and exhaustive. $\Diamond$

Example 2.7. Any event A and its complement $\overline{A}$ represent a classical example of disjoint and exhaustive events. $\Diamond$

Example 2.8. Receiving a grade of A, B, or C for some course are mutually exclusive events, but unfortunately, they are not exhaustive. $\Diamond$

As we see in the next sections, it is often easier to compute probability of an intersection than probability of a union. Taking complements converts unions into intersections, see (2.2).

$$\boxed{\overline{E_1 \cup \ldots \cup E_n} = \overline{E}_1 \cap \ldots \cap \overline{E}_n, \qquad \overline{E_1 \cap \ldots \cap E_n} = \overline{E}_1 \cup \ldots \cup \overline{E}_n} \tag{2.2}$$

PROOF OF (2.2): Since the union $E_1 \cup \ldots \cup E_n$ represents the event "at least one event occurs," its complement has the form

$$\begin{aligned} \overline{E_1 \cup \ldots \cup E_n} &= \{ \text{ none of them occurs } \} \\ &= \{ E_1 \text{ does not occur} \cap \ldots \cap E_n \text{ does not occur } \} \\ &= \overline{E}_1 \cap \ldots \cap \overline{E}_n. \end{aligned}$$

The other equality in (2.2) is left as Exercise 2.34. $\square$

Example 2.9. Graduating with a GPA of 4.0 is an *intersection* of getting an A in *each* course. Its *complement*, graduating with a GPA below 4.0, is a *union* of receiving a grade below A *at least in one* course. $\Diamond$

Rephrasing (2.2), a complement to "nothing" is "something," and "not everything" means "at least one is missing".

2.2 Rules of Probability

Now we are ready for the rigorous definition of probability. All the rules and principles of computing probabilities of events follow from this definition.

Mathematically, *probability* is introduced through several axioms.

2.2.1 Axioms of Probability

First, we choose a *sigma-algebra* $\mathfrak{M}$ of events on a sample space Ω. This is a collection of events whose probabilities we can consider in our problem.

DEFINITION 2.9

A collection $\mathfrak{M}$ of events is a **sigma-algebra** on sample space Ω if

(a) it includes the sample space,

$$\Omega \in \mathfrak{M}$$

(b) every event in $\mathfrak{M}$ is contained along with its complement; that is,

$$E \in \mathfrak{M} \quad \Rightarrow \quad \overline{E} \in \mathfrak{M}$$

(c) every finite or countable collection of events in $\mathfrak{M}$ is contained along with its union; that is,

$$E_1, E_2, \ldots \in \mathfrak{M} \quad \Rightarrow \quad E_1 \cup E_2 \cup \ldots \in \mathfrak{M}.$$

Here are a few examples of sigma-algebras.

Example 2.10 (DEGENERATE SIGMA-ALGEBRA). By conditions (a) and (b) in Definition 2.9, every sigma-algebra has to contain the sample space Ω and the empty event $\varnothing$. This minimal collection

$$\mathfrak{M} = \{\Omega, \varnothing\}$$

forms a sigma-algebra that is called *degenerate*. ◊

Example 2.11 (POWER SET). On the other extreme, what is the richest sigma-algebra on a sample space Ω? It is the collection of *all* the events,

$$\mathfrak{M} = 2^{\Omega} = \{E,\ E \subset \Omega\}.$$

As we know from (2.1), there are 2^N events on a sample space of N outcomes. This explains the notation 2^{Ω}. This sigma-algebra is called a *power set*. ◊

Example 2.12 (BOREL SIGMA-ALGEBRA). Now consider an experiment that consists of selecting a point on the real line. Then, each outcome is a point $x \in \mathbb{R}$, and the sample space is $\Omega = \mathbb{R}$. Do we want to consider a probability that the point falls in a given interval? Then define a sigma-algebra $\mathfrak{B}$ to be a collection of all the intervals, finite and infinite, open and closed, and all their finite and countable unions and intersections. This sigma-algebra is very rich, but apparently, it is much less than the power set 2^{Ω}. This is the *Borel sigma-algebra*, after the French mathematician *Émile Borel* (1871–1956). In fact, it consists of all the real sets that *have length*. ◊

Axioms of Probability are in the following definition.

DEFINITION 2.10

Assume a sample space Ω and a sigma-algebra of events $\mathfrak{M}$ on it. **Probability**

$$\boldsymbol{P} \; : \; \mathfrak{M} \to [0,1]$$

is a function of events with the domain $\mathfrak{M}$ and the range $[0,1]$ that satisfies the following two conditions,

(Unit measure) The sample space has unit probability, $\boldsymbol{P}(\Omega) = 1$.

(Sigma-additivity) For any finite or countable collection of *mutually exclusive* events $E_1, E_2, \ldots \in \mathfrak{M}$,

$$\boldsymbol{P}\left\{E_1 \cup E_2 \cup \ldots\right\} = \boldsymbol{P}(E_1) + \boldsymbol{P}(E_2) + \ldots.$$

All the rules of probability are consequences from this definition.

2.2.2 Computing probabilities of events

Armed with the fundamentals of probability theory, we are now able to compute probabilities of many interesting events.

Extreme cases

A sample space Ω consists of all possible outcomes, therefore, it occurs for sure. On the contrary, an empty event $\varnothing$ never occurs. So,

$$\boldsymbol{P}\left\{\Omega\right\} = 1 \;\text{ and }\; \boldsymbol{P}\left\{\varnothing\right\} = 0. \tag{2.3}$$

Proof: Probability of Ω is given by the definition of probability. By the same definition, $\boldsymbol{P}\left\{\Omega\right\} = \boldsymbol{P}\left\{\Omega \cup \varnothing\right\} = \boldsymbol{P}\left\{\Omega\right\} + \boldsymbol{P}\left\{\varnothing\right\}$, because Ω and $\varnothing$ are mutually exclusive. Therefore, $\boldsymbol{P}\left\{\varnothing\right\} = 0$. $\square$

Union

Consider an event that consists of some finite or countable collection of mutually exclusive outcomes,

$$E = \left\{\omega_1, \omega_2, \omega_3, \ldots\right\}.$$

Summing probabilities of these outcomes, we obtain the probability of the entire event,

$$\boxed{\boldsymbol{P}\left\{E\right\} = \sum_{\omega_k \in E} \boldsymbol{P}\left\{\omega_k\right\} = \boldsymbol{P}\left\{\omega_1\right\} + \boldsymbol{P}\left\{\omega_2\right\} + \boldsymbol{P}\left\{\omega_3\right\} \ldots}$$

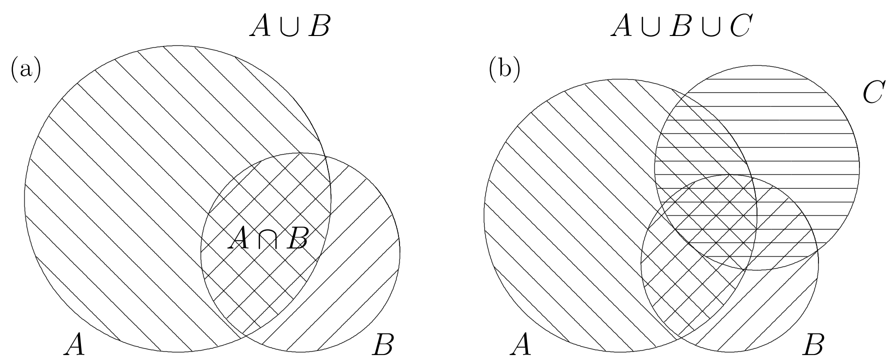

FIGURE 2.2: *(a) Union of two events. (b) Union of three events.*

Example 2.13. If a job sent to a printer appears first in line with probability 60%, and second in line with probability 30%, then with probability 90% it appears either first or second in line. ◊

It is crucial to notice that only *mutually exclusive* events (those with empty intersections) satisfy the sigma-additivity. If events intersect, their probabilities cannot be simply added. Look at the following example.

Example 2.14. During some construction, a network blackout occurs on Monday with probability 0.7 and on Tuesday with probability 0.5. Then, does it appear on Monday *or* Tuesday with probability $0.7 + 0.5 = 1.2$? Obviously not, because probability should always be between 0 and 1! Probabilities are not additive here because blackouts on Monday and Tuesday are not mutually exclusive. In other words, it is not impossible to see blackouts on both days. ◊

In Example 2.14, blind application of the rule for the union of mutually exclusive events clearly overestimated the actual probability. The Venn diagram shown in Figure 2.2a explains it. We see that in the sum $\boldsymbol{P}\{A\} + \boldsymbol{P}\{B\}$, all the common outcomes are counted *twice*. Certainly, this caused the overestimation. Each outcome should be counted only once! To correct the formula, subtract probabilities of common outcomes, which is $\boldsymbol{P}\{A \cap B\}$.

Probability of a union

$$\boxed{\begin{array}{l} \boldsymbol{P}\{A \cup B\} = \boldsymbol{P}\{A\} + \boldsymbol{P}\{B\} - \boldsymbol{P}\{A \cap B\} \\ \text{For mutually exclusive events,} \\ \boldsymbol{P}\{A \cup B\} = \boldsymbol{P}\{A\} + \boldsymbol{P}\{B\} \end{array}} \qquad (2.4)$$

Generalization of this formula is not straightforward. For 3 events,

$$\begin{aligned} \boldsymbol{P}\{A \cup B \cup C\} &= \boldsymbol{P}\{A\} + \boldsymbol{P}\{B\} + \boldsymbol{P}\{C\} - \boldsymbol{P}\{A \cap B\} - \boldsymbol{P}\{A \cap C\} \\ &\quad - \boldsymbol{P}\{B \cap C\} + \boldsymbol{P}\{A \cap B \cap C\}. \end{aligned}$$

As seen in Figure 2.2b, when we add probabilities of A, B, and C, each pairwise intersection is counted twice. Therefore, we subtract the probabilities of $\boldsymbol{P}\{A \cap B\}$, etc. Finally, consider the triple intersection $A \cap B \cap C$. Its probability is counted 3 times within each main event, then subtracted 3 times with each pairwise intersection. Thus, it is not counted at all so far! Therefore, we add its probability $\boldsymbol{P}\{A \cap B \cap C\}$ in the end.

For an arbitrary collection of events, see Exercise 2.33.

Example 2.15. In Example 2.14, suppose there is a probability 0.35 of experiencing network blackouts on both Monday and Tuesday. Then the probability of having a blackout on Monday *or* Tuesday equals

$$0.7 + 0.5 - 0.35 = 0.85.$$

$\Diamond$

Complement

Recall that events A and $\overline{A}$ are exhaustive, hence $A \cup \overline{A} = \Omega$. Also, they are disjoint, hence

$$\boldsymbol{P}\{A\} + \boldsymbol{P}\{\overline{A}\} = \boldsymbol{P}\{A \cup \overline{A}\} = \boldsymbol{P}\{\Omega\} = 1.$$

Solving this for $\boldsymbol{P}\{\overline{A}\}$, we obtain a rule that perfectly agrees with the common sense,

Complement rule $\quad \boxed{\boldsymbol{P}\{\overline{A}\} = 1 - \boldsymbol{P}\{A\}}$

Example 2.16. If a system appears protected against a new computer virus with probability 0.7, then it is exposed to it with probability $1 - 0.7 = 0.3$. $\Diamond$

Example 2.17. Suppose a computer code has no errors with probability 0.45. Then, it has at least one error with probability 0.55. $\Diamond$

Intersection of independent events

DEFINITION 2.11

> Events $E_1, \ldots, E_n$ are **independent** if they occur independently of each other, i.e., occurrence of one event does not affect the probabilities of others.

The following basic formula can serve as the criterion of independence.

Independent events $\quad \boxed{\boldsymbol{P}\{E_1 \cap \ldots \cap E_n\} = \boldsymbol{P}\{E_1\} \cdot \ldots \cdot \boldsymbol{P}\{E_n\}}$

We shall defer explanation of this formula until Section 2.4 which will also give a rule for intersections of *dependent* events.

2.2.3 Applications in reliability

Formulas of the previous Section are widely used in *reliability*, when one computes the probability for a system of several components to be functional.

Example 2.18 (RELIABILITY OF BACKUPS). There is a 1% probability for a hard drive to crash. Therefore, it has two backups, each having a 2% probability to crash, and all three components are independent of each other. The stored information is lost only in an unfortunate situation when all three devices crash. What is the probability that the information is saved?

Solution. Organize the data. Denote the events, say,

$$H = \{ \text{ hard drive crashes } \},$$

$$B_1 = \{ \text{ first backup crashes } \}, \; B_2 = \{ \text{ second backup crashes } \}.$$

It is given that H, B_1, and B_2 are independent,

$$\boldsymbol{P}\{H\} = 0.01, \text{ and } \boldsymbol{P}\{B_1\} = \boldsymbol{P}\{B_2\} = 0.02.$$

Applying rules for the complement and for the intersection of independent events,

$$\begin{aligned} \boldsymbol{P}\{ \text{ saved } \} &= 1 - \boldsymbol{P}\{ \text{ lost } \} = 1 - \boldsymbol{P}\{H \cap B_1 \cap B_2\} \\ &= 1 - \boldsymbol{P}\{H\}\,\boldsymbol{P}\{B_1\}\,\boldsymbol{P}\{B_2\} \\ &= 1 - (0.01)(0.02)(0.02) = 0.999996. \end{aligned}$$

(This is precisely the reason of having backups, isn't it? Without backups, the probability for information to be saved is only 0.99.) ◊

When the system's components are connected *in parallel*, it is sufficient for at least one component to work in order for the whole system to function. Reliability of such a system is computed as in Example 2.18. Backups can always be considered as devices connected in parallel.

At the other end, consider a system whose components are connected *in sequel*. Failure of one component inevitably causes the whole system to fail. Such a system is more "vulnerable." In order to function with a high probability, it needs each component to be reliable, as in the next example.

Example 2.19. Suppose that a shuttle's launch depends on three key devices that operate independently of each other and malfunction with probabilities 0.01, 0.02, and 0.02, respectively. If any of the key devices malfunctions, the launch will be postponed. Compute the probability for the shuttle to be launched on time, according to its schedule.

<u>Solution</u>. In this case,

$$\begin{aligned}
\boldsymbol{P}\{\text{ on time }\} &= \boldsymbol{P}\{\text{ all devices function }\} \\
&= \boldsymbol{P}\left\{\overline{H} \cap \overline{B}_1 \cap \overline{B}_2\right\} \\
&= \boldsymbol{P}\left\{\overline{H}\right\} \boldsymbol{P}\left\{\overline{B}_1\right\} \boldsymbol{P}\left\{\overline{B}_2\right\} \quad (\textit{independence}) \\
&= (1-0.01)(1-0.02)(1-0.02) \quad (\textit{complement rule}) \\
&= 0.9508.
\end{aligned}$$

Notice how with the same probabilities of individual components as in Example 2.18, the system's reliability decreased because the components were connected sequentially. ◊

Many modern systems consist of a great number of devices connected in sequel and in parallel.

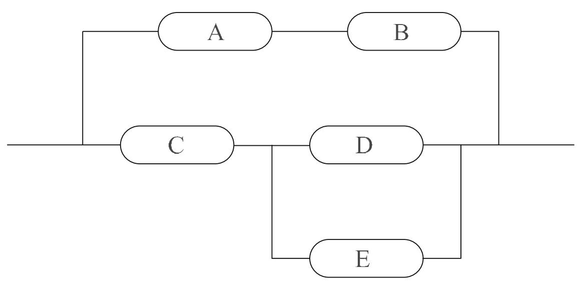

FIGURE 2.3: *Calculate reliability of this system (Example 2.20).*

Example 2.20 (Techniques for solving reliability problems). Calculate reliability of the system in Figure 2.3 if each component is operable with probability 0.92 independently of the other components.

<u>Solution</u>. This problem can be simplified and solved "step by step."

1. The upper link A-B works if both A and B work, which has probability

$$\boldsymbol{P}\{A \cap B\} = (0.92)^2 = 0.8464.$$

   We can represent this link as one component F that operates with probability 0.8464.

2. By the same token, components D and E, connected in parallel, can be replaced by component G, operable with probability

$$\boldsymbol{P}\{D \cup E\} = 1 - (1-0.92)^2 = 0.9936,$$

   as shown in Figure 2.4a.

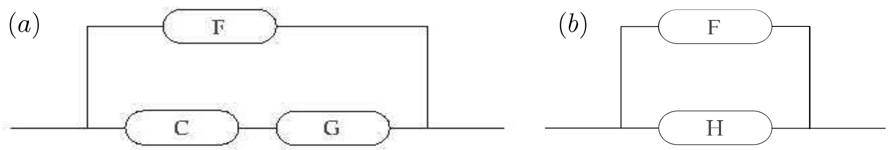

FIGURE 2.4: *Step by step solution of a system reliability problem.*

3. Components C and G, connected sequentially, can be replaced by component H, operable with probability $\boldsymbol{P}\{C \cap G\} = 0.92 \cdot 0.9936 = 0.9141$, as shown in Figure 2.4b.

4. Last step. The system operates with probability

$$\boldsymbol{P}\{F \cup H\} = 1 - (1 - 0.8464)(1 - 0.9141) = \underline{0.9868},$$

which is the final answer.

In fact, the event "the system is operable" can be represented as $(A \cap B) \cup \{C \cap (D \cup E)\}$, whose probability we found step by step. $\Diamond$

2.3 Combinatorics

2.3.1 Equally likely outcomes

A simple situation for computing probabilities is the case of *equally likely outcomes*. That is, when the sample space Ω consists of n possible outcomes, $\omega_1, \ldots, \omega_n$, each having the same probability. Since

$$\sum_1^n \boldsymbol{P}\{\omega_k\} = \boldsymbol{P}\{\Omega\} = 1,$$

we have in this case $\boldsymbol{P}\{\omega_k\} = 1/n$ for all k. Further, a probability of any event E consisting of t outcomes, equals

$$\boldsymbol{P}\{E\} = \sum_{\omega_k \in E} \left(\frac{1}{n}\right) = t\left(\frac{1}{n}\right) = \frac{\text{number of outcomes in } E}{\text{number of outcomes in } \Omega}.$$

The outcomes forming event E are often called "favorable." Thus we have a formula

Equally likely outcomes

$$\boxed{\boldsymbol{P}\{E\} = \frac{\text{number of favorable outcomes}}{\text{total number of outcomes}} = \frac{\mathcal{N}_F}{\mathcal{N}_T}} \tag{2.5}$$

where index "F" means "favorable" and "T" means "total."

Example 2.21. Tossing a die results in 6 equally likely possible outcomes, identified by the number of dots from 1 to 6. Applying (2.5), we obtain,

$$\boldsymbol{P}\{1\} = 1/6, \quad \boldsymbol{P}\{\text{ odd number of dots }\} = 3/6, \quad \boldsymbol{P}\{\text{ less than 5 }\} = 4/6.$$

$\Diamond$

The solution and even the answer to such problems may depend on our choice of outcomes and a sample space. Outcomes should be defined in such a way that they appear equally likely, otherwise formula (2.5) does not apply.

Example 2.22. A card is drawn from a bridge 52-card deck at random. Compute the probability that the selected card is a spade.

First solution. The sample space consists of 52 equally likely outcomes—cards. Among them, there are 13 favorable outcomes—spades. Hence, $\boldsymbol{P}\{\text{spade}\} = 13/52 = 1/4$.

Second solution. The sample space consists of 4 equally likely outcomes—suits: clubs, diamonds, hearts, and spades. Among them, one outcome is favorable—spades. Hence, $\boldsymbol{P}\{\text{spade}\} = 1/4$. $\Diamond$

These two solutions relied on different sample spaces. However, in both cases, the defined outcomes were equally likely, therefore (2.5) was applicable, and we obtained the same result.

However, the situation may be different.

Example 2.23. A young family plans to have two children. What is the probability of two girls?

Solution 1 (wrong). There are 3 possible families with 2 children: two girls, two boys, and one of each gender. Therefore, the probability of two girls is 1/3.

Solution 2 (right). Each child is (supposedly) equally likely to be a boy or a girl. Genders of the two children are (supposedly) independent. Therefore,

$$\boldsymbol{P}\{\text{two girls}\} = \left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = 1/4.$$

$\Diamond$

The second solution implies that the sample space consists of four, not three, equally likely outcomes: two boys, two girls, a boy and a girl, a girl and a boy. Each outcome in this sample has probability 1/4. Notice that the last two outcomes are counted separately, with the meaning, say, "the first child is a boy, the second one is a girl" and "the first child is a girl, the second one is a boy."

It is all right to define a sample space as in Solution 1. However, one must know in this case that the defined outcomes are *not equally likely*. Indeed, from Solution 2, we see that having one child of each gender is the most likely outcome, with probability of $1/4 + 1/4 = 1/2$. It was a mistake to apply (2.5) for such a sample space in Solution 1.

Example 2.24 (Paradox). There is a simple but controversial "situation" about a family with two children. Even some graduate students often fail to resolve it.

A family has two children. You met one of them, Leo, and he is a boy. What is the probability that the other child is also a boy?

On one hand, why would the other child's gender be affected by Leo? Leo should have a brother or a sister with probabilities 1/2 and 1/2.

On the other hand, see Example 2.23. The sample space consists of 4 equally likely outcomes, $\{GG, BB, BG, GB\}$. You have already met one boy, thus the first outcome is automatically eliminated: $\{BB, BG, GB\}$. Among the remaining three outcomes, Jimmy has a brother in one case and a sister in two cases. Thus, isn't the probability of a boy equal 1/3?

Where is the catch? Apparently, the sample space Ω has not been clearly defined in this example. The experiment is more complex than in Example 2.23 because we are now concerned not only about the gender of children but also about meeting one of them. What is the mechanism, what are the probabilities for you to meet one or the other child? And once you met Leo, do the outcomes $\{BB, BG, GB\}$ remain equally likely?

A complete solution to this paradox is broken into steps in Exercise 2.30. $\Diamond$

In reality, business-related, sports-related, and political events are typically not equally likely. One outcome is usually more likely that another. For example, one team is always stronger than the other. Equally likely outcomes are usually associated with conditions of "a fair game" and "selected at random." In fair gambling, all cards, all dots on a die, all numbers in a roulette are equally likely. Also, when a survey is conducted, or a sample is selected "at random," the outcomes are "as close as possible" to being equally likely. This means all the subjects have the same chance to be selected into a sample (otherwise, it is not a fair sample, and it can produce "biased" results).

2.3.2 Permutations and combinations

Formula (2.5) is simple, as long as its numerator and denominator can be easily evaluated. This is rarely the case; often the sample space consists of a multitude of outcomes. *Combinatorics* provides special techniques for the computation of $\mathcal{N}_T$ and $\mathcal{N}_F$, the total number and the number of favorable outcomes.

We shall consider a generic situation when objects are selected *at random* from a set of n. This general model has a number of useful applications.

The objects may be selected with replacement or without replacement. They may also be *distinguishable* or *indistinguishable*.

DEFINITION 2.12

Sampling **with replacement** means that every sampled item is replaced into the initial set, so that any of the objects can be selected with probability $1/n$ at any time. In particular, the same object may be sampled more than once.

DEFINITION 2.13

Sampling **without replacement** means that every sampled item is removed from further sampling, so the set of possibilities reduces by 1 after each selection.

DEFINITION 2.14

Objects are **distinguishable** if sampling of exactly the same objects *in a different order* yields a different outcome, that is, a different element of the sample space. For **indistinguishable** objects, the order is not important, it only matters which objects are sampled and which ones aren't. Indistinguishable objects arranged in a different order do not generate a new outcome.

Example 2.25 (COMPUTER-GENERATED PASSWORDS). When random passwords are generated, the order of characters is important because a different order yields a different password. Characters are distinguishable in this case. Further, if a password has to consist of different characters, they are sampled from the alphabet without replacement. $\Diamond$

Example 2.26 (POLLS). When a sample of people is selected to conduct a poll, the same participants produce the same responses regardless of their order. They can be considered indistinguishable. $\Diamond$

Permutations with replacement

Possible selections of k *distinguishable* objects from a set of n are called *permutations*. When we sample with replacement, each time there are n possible selections, and the total number of permutations is

Permutations with replacement

$$\boxed{P_r(n,k) = \overbrace{n \cdot n \cdot \ldots \cdot n}^{k \text{ terms}} = n^k}$$

Example 2.27 (BREAKING PASSWORDS). From an alphabet consisting of 10 digits, 26 lower-case and 26 capital letters, one can create $P_r(62, 8) = 218{,}340{,}105{,}584{,}896$ (over 218 trillion) different 8-character passwords. At a speed of 1 million passwords per second, it will take a spy program almost 7 years to try all of them. Thus, on the average, it will guess your password in about 3.5 years.

At this speed, the spy program can test 604,800,000,000 passwords within 1 week. The probability that it guesses your password in 1 week is

$$\frac{\mathcal{N}_F}{\mathcal{N}_T} = \frac{\text{number of favorable outcomes}}{\text{total number of outcomes}} = \frac{604{,}800{,}000{,}000}{218{,}340{,}105{,}584{,}896} = 0.00277.$$

However, if capital letters are not used, the number of possible passwords is reduced to $P_r(36, 8) = 2{,}821{,}109{,}907{,}456$. On the average, it takes the spy only 16 days to guess such a password! The probability that it will happen in 1 week is 0.214. A wise recommendation to include all three types of characters in our passwords and to change them once a year is perfectly clear to us now... $\Diamond$

Permutations without replacement

During sampling without replacement, the number of possible selections reduces by 1 each time an object is sampled. Therefore, the number of permutations is

Permutations without replacement

$$\boxed{P(n,k) = \overbrace{n(n-1)(n-2)\cdot\ldots\cdot(n-k+1)}^{k \text{ terms}} = \frac{n!}{(n-k)!}}$$

where $n! = 1 \cdot 2 \cdot \ldots n$ (*n-factorial*) denotes the product of all integers from 1 to n.

The number of permutations without replacement also equals the number of possible allocations of k distinguishable objects among n available slots.

Example 2.28. In how many ways can 10 students be seated in a classroom with 15 chairs?

Solution. Students are distinguishable, and each student has a separate seat. Thus, the number of possible allocations is the number of permutations without replacement, $P(15, 10) = 15 \cdot 14 \cdot \ldots \cdot 6 = 1.09 \cdot 10^{10}$. Notice that if students enter the classroom one by one, the first student has 15 choices of seats, then one seat is occupied, and the second student has only 14 choices, etc., and the last student takes one of 6 chairs available at that time. $\Diamond$

Combinations without replacement

Possible selections of k *indistinguishable* objects from a set of n are called *combinations*. The number of combinations without replacement is also called "n choose k" and is denoted by $C(n, k)$ or $\binom{n}{k}$.

The only difference from $P(n, k)$ is disregarding the order. Now the same objects sampled in a different order produce the same outcome. Thus, $P(k,k) = k!$ different permutations (rearrangements) of the same objects yield only 1 combination. The total number of combinations is then

Combinations without replacement

$$\boxed{C(n,k) = \binom{n}{k} = \frac{P(n,k)}{P(k,k)} = \frac{n!}{k!(n-k)!}} \tag{2.6}$$

Example 2.29. An antivirus software reports that 3 folders out of 10 are infected. How many possibilities are there?

Solution. Folders A, B, C and folders C, B, A represent the same outcome, thus, the order is not important. A software clearly detected 3 different folders, thus it is sampling without replacement. The number of possibilities is

$$\binom{10}{3} = \frac{10!}{3!\,7!} = \frac{10 \cdot 9 \cdot \ldots \cdot 1}{(3 \cdot 2 \cdot 1)(7 \cdot \ldots \cdot 1)} = \frac{10 \cdot 9 \cdot 8}{3 \cdot 2 \cdot 1} = 120.$$

$\Diamond$

Computational shortcuts

Instead of computing $C(n,k)$ directly by the formula, we can simplify the fraction. At least, the numerator and denominator can both be divided by either $k!$ or $(n-k)!$ (choose the larger of these for greater reduction). As a result,

$$C(n,k) = \binom{n}{k} = \frac{n \cdot (n-1) \cdot \ldots \cdot (n-k+1)}{k \cdot (k-1) \cdot \ldots \cdot 1},$$

the top and the bottom of this fraction being products of k terms. It is also handy to notice that

$$\begin{aligned} C(n,k) &= C(n,n-k) \quad \text{for any } k \text{ and } n \\ C(n,0) &= 1 \\ C(n,1) &= n \end{aligned}$$

Example 2.30. There are 20 computers in a store. Among them, 15 are brand new and 5 are refurbished. Six computers are purchased for a student lab. From the first look, they are indistinguishable, so the six computers are selected at random. Compute the probability that among the chosen computers, two are refurbished.

Solution. Compute the total number and the number of favorable outcomes. The *total* number of ways in which 6 computers are selected from 20 is

$$\mathcal{N}_T = \binom{20}{6} = \frac{20 \cdot 19 \cdot 18 \cdot 17 \cdot 16 \cdot 15}{6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}.$$

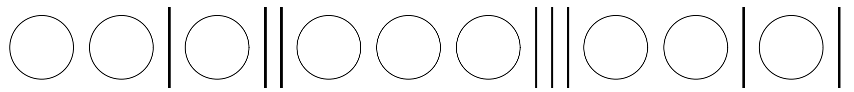

FIGURE 2.5: *Counting combinations with replacement. Vertical bars separate different classes of items.*

We applied the mentioned computational shortcut. Next, for the number of *favorable* outcomes, 2 refurbished computers are selected from a total of 5, and the remaining 4 new ones are selected from a total of 15. There are

$$\mathcal{N}_F = \binom{5}{2}\binom{15}{4} = \left(\frac{5\cdot 4}{2\cdot 1}\right)\left(\frac{15\cdot 14\cdot 13\cdot 12}{4\cdot 3\cdot 2\cdot 1}\right)$$

favorable outcomes. With further reduction of fractions, the probability equals

$$\boldsymbol{P}\{\text{ two refurbished computers }\} = \frac{\mathcal{N}_F}{\mathcal{N}_T} = \frac{7\cdot 13\cdot 5}{19\cdot 17\cdot 4} = 0.3522.$$

$\Diamond$

Combinations with replacement

For combinations with replacement, the order is not important, and each object may be sampled more than once. Then each outcome consists of counts, how many times each of n objects appears in the sample. In Figure 2.5, we draw a circle for each time object #1 is sampled, then draw a separating bar, then a circle for each time object #2 is sampled, etc. Two bars next to each other mean that the corresponding object has never been sampled.

The resulting picture has to have k circles for a sample of size k and $(n-1)$ bars separating n objects. Each picture with these conditions represents an outcome. How many outcomes are there? It is the number of allocations of k circles and $(n-1)$ bars among $(k+n-1)$ slots available for them. Hence,

Combinations with replacement

$$\boxed{C_r(n,k) = \binom{k+n-1}{k} = \frac{(k+n-1)!}{k!(n-1)!}}$$

NOTATION

$$\begin{aligned} P_r(n,k) &= \text{number of permutations with replacement} \\ P(n,k) &= \text{number of permutations without replacement} \\ C_r(n,k) &= \text{number of combinations with replacement} \\ \left.\begin{matrix} C(n,k) \\ \binom{n}{k} \end{matrix}\right\} &= \text{number of combinations without replacement} \end{aligned}$$

2.4 Conditional probability and independence

Conditional probability

Suppose you are meeting someone at an airport. The flight is likely to arrive on time; the probability of that is 0.8. Suddenly it is announced that the flight departed one hour behind the schedule. Now it has the probability of only 0.05 to arrive on time. New information affected the probability of meeting this flight on time. The new probability is called *conditional probability*, where the new information, that the flight departed late, is a *condition*.

DEFINITION 2.15

> **Conditional probability** of event A given event B is the probability that A occurs when B is *known to occur*.

NOTATION $\boldsymbol{P}\{A \mid B\}$ = conditional probability of A given B

How does one compute the conditional probability? First, consider the case of equally likely outcomes. In view of the new information, occurrence of the condition B, only the outcomes contained in B still have a non-zero chance to occur. Counting only such outcomes, the *unconditional probability* of A,

$$\boldsymbol{P}\{A\} = \frac{\text{number of outcomes in } A}{\text{number of outcomes in } \Omega},$$

is now replaced by the *conditional probability* of A given B,

$$\boldsymbol{P}\{A \mid B\} = \frac{\text{number of outcomes in } A \cap B}{\text{number of outcomes in } B} = \frac{\boldsymbol{P}\{A \cap B\}}{\boldsymbol{P}\{B\}}.$$

This appears to be the general formula.

Conditional probability

$$\boxed{\boldsymbol{P}\{A \mid B\} = \frac{\boldsymbol{P}\{A \cap B\}}{\boldsymbol{P}\{B\}}} \qquad (2.7)$$

Rewriting (2.7) in a different way, we obtain the general formula for the probability of intersection.

Intersection, general case

$$\boxed{\boldsymbol{P}\{A \cap B\} = \boldsymbol{P}\{B\}\,\boldsymbol{P}\{A \mid B\}} \qquad (2.8)$$

Independence

Now we can give an intuitively very clear definition of *independence*.

DEFINITION 2.16

> Events A and B are **independent** if occurrence of B does not affect the probability of A, i.e.,
> $$\boldsymbol{P}\{A \mid B\} = \boldsymbol{P}\{A\}.$$

According to this definition, *conditional* probability equals *unconditional* probability in case of independent events. Substituting this into (2.8) yields

$$\boldsymbol{P}\{A \cap B\} = \boldsymbol{P}\{A\} \boldsymbol{P}\{B\}.$$

This is our old formula for independent events.

Example 2.31. Ninety percent of flights depart on time. Eighty percent of flights arrive on time. Seventy-five percent of flights depart on time and arrive on time.

(a) You are meeting a flight that departed on time. What is the probability that it will arrive on time?

(b) You have met a flight, and it arrived on time. What is the probability that it departed on time?

(c) Are the events, departing on time and arriving on time, independent?

<u>Solution</u>. Denote the events,

$$\begin{aligned} A &= \{\text{arriving on time}\}, \\ D &= \{\text{departing on time}\}. \end{aligned}$$

We have:

$$\boldsymbol{P}\{A\} = 0.8, \quad \boldsymbol{P}\{D\} = 0.9, \quad \boldsymbol{P}\{A \cap D\} = 0.75.$$

(a) $\boldsymbol{P}\{A \mid D\} = \dfrac{\boldsymbol{P}\{A \cap D\}}{\boldsymbol{P}\{D\}} = \dfrac{0.75}{0.9} = \underline{0.8333}.$

(b) $\boldsymbol{P}\{D \mid A\} = \dfrac{\boldsymbol{P}\{A \cap D\}}{\boldsymbol{P}\{A\}} = \dfrac{0.75}{0.8} = \underline{0.9375}.$

(c) Events are not independent because

$$\boldsymbol{P}\{A \mid D\} \neq \boldsymbol{P}\{A\}, \quad \boldsymbol{P}\{D \mid A\} \neq \boldsymbol{P}\{D\}, \quad \boldsymbol{P}\{A \cap D\} \neq \boldsymbol{P}\{A\} \boldsymbol{P}\{D\}.$$

Actually, any one of these inequalities is sufficient to prove that A and D are dependent. Further, we see that $\boldsymbol{P}\{A \mid D\} > \boldsymbol{P}\{A\}$ and $\boldsymbol{P}\{D \mid A\} > \boldsymbol{P}\{D\}$. In other words, departing on time increases the probability of arriving on time, and vice versa. This perfectly agrees with our intuition. $\Diamond$

Bayes Rule

The last example shows that two conditional probabilities, $\boldsymbol{P}\{A \mid B\}$ and $\boldsymbol{P}\{B \mid A\}$, are not the same, in general. Consider another example.

Example 2.32 (RELIABILITY OF A TEST). There exists a test for a certain viral infection (including a virus attack on a computer network). It is 95% reliable for infected patients and 99% reliable for the healthy ones. That is, if a patient has the virus (event V), the test shows that (event S) with probability $\boldsymbol{P}\{S \mid V\} = 0.95$, and if the patient does not have the virus, the test shows that with probability $\boldsymbol{P}\left\{\overline{S} \mid \overline{V}\right\} = 0.99$.

Consider a patient whose test result is positive (i.e., the test shows that the patient has the virus). Knowing that sometimes the test is wrong, naturally, the patient is eager to know the probability that he or she indeed has the virus. However, this conditional probability, $\boldsymbol{P}\{V \mid S\}$, is not stated among the given characteristics of this test. ♢

This example is applicable to any testing procedure including software and hardware tests, pregnancy tests, paternity tests, alcohol tests, academic exams, etc. The problem is to connect the given $\boldsymbol{P}\{S \mid V\}$ and the quantity in question, $\boldsymbol{P}\{V \mid S\}$. This was done in the eighteenth century by English minister *Thomas Bayes* (1702–1761) in the following way.

Notice that $A \cap B = B \cap A$. Therefore, using (2.8), $\boldsymbol{P}\{B\}\boldsymbol{P}\{A \mid B\} = \boldsymbol{P}\{A\}\boldsymbol{P}\{B \mid A\}$.

Solve for $\boldsymbol{P}\{B \mid A\}$ to obtain

Bayes Rule

$$\boxed{\boldsymbol{P}\{B \mid A\} = \frac{\boldsymbol{P}\{A \mid B\}\boldsymbol{P}\{B\}}{\boldsymbol{P}\{A\}}} \tag{2.9}$$

Example 2.33 (SITUATION ON A MIDTERM EXAM). On a midterm exam, students X, Y, and Z forgot to sign their papers. Professor knows that they can write a good exam with probabilities 0.8, 0.7, and 0.5, respectively. After the grading, he notices that two unsigned exams are good and one is bad. Given this information, and assuming that students worked independently of each other, what is the probability that the bad exam belongs to student Z?

Solution. Denote good and bad exams by G and B. Also, let GGB denote two good and one bad exams, XG denote the event "student X wrote a good exam," etc. We need to find $\boldsymbol{P}\{ZB \mid GGB\}$ given that $\boldsymbol{P}\{G \mid X\} = 0.8$, $\boldsymbol{P}\{G \mid Y\} = 0.7$, and $\boldsymbol{P}\{G \mid Z\} = 0.5$.

By the *Bayes Rule*,

$$\boldsymbol{P}\{ZB \mid GGB\} = \frac{\boldsymbol{P}\{GGB \mid ZB\}\boldsymbol{P}\{ZB\}}{\boldsymbol{P}\{GGB\}}.$$

Given ZB, event GGB occurs only when both X and Y write good exams. Thus, $\boldsymbol{P}\{GGB \mid ZB\} = (0.8)(0.7)$.

Event GGB consists of three outcomes depending on the student who wrote the bad exam.

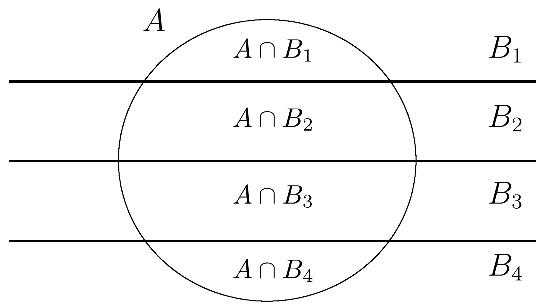

FIGURE 2.6: *Partition of the sample space Ω and the event A.*

Adding their probabilities, we get

$$\begin{aligned} &\boldsymbol{P}\{GGB\} \\ &= \boldsymbol{P}\{XG \cap YG \cap ZB\} + \boldsymbol{P}\{XG \cap YB \cap ZG\} + \boldsymbol{P}\{XB \cap YG \cap ZG\} \\ &= (0.8)(0.7)(0.5) + (0.8)(0.3)(0.5) + (0.2)(0.7)(0.5) = 0.47. \end{aligned}$$

Then

$$\boldsymbol{P}\{ZB \mid GGB\} = \frac{(0.8)(0.7)(0.5)}{0.47} = \underline{0.5957}.$$

$\Diamond$

In the Bayes Rule (2.9), the denominator is often computed by the Law of Total Probability.

Law of Total Probability

This law relates the unconditional probability of an event A with its conditional probabilities. It is used every time when it is easier to compute conditional probabilities of A given additional information.

Consider some partition of the sample space Ω with mutually exclusive and exhaustive events $B_1, \ldots, B_k$. It means that

$$B_i \cap B_j = \varnothing \text{ for any } i \neq j \text{ and } B_1 \cup \ldots \cup B_k = \Omega.$$

These events also partition the event A,

$$A = (A \cap B_1) \cup \ldots \cup (A \cap B_k),$$

and this is also a union of mutually exclusive events (Figure 2.6). Hence,

$$\boldsymbol{P}\{A\} = \sum_{j=1}^{k} \boldsymbol{P}\{A \cap B_j\},$$

and we arrive to the following rule.

Law of Total Probability

$$\boxed{\begin{array}{l} \boldsymbol{P}\{A\} = \displaystyle\sum_{j=1}^{k} \boldsymbol{P}\{A \mid B_j\}\, \boldsymbol{P}\{B_j\} \\ \text{In case of two events } (k=2), \\ \boldsymbol{P}\{A\} = \boldsymbol{P}\{A \mid B\}\, \boldsymbol{P}\{B\} + \boldsymbol{P}\{A \mid \overline{B}\}\, \boldsymbol{P}\{\overline{B}\} \end{array}} \tag{2.10}$$

Together with the Bayes Rule, it makes the following popular formula

Bayes Rule for two events

$$\boxed{\boldsymbol{P}\{B \mid A\} = \frac{\boldsymbol{P}\{A \mid B\}\, \boldsymbol{P}\{B\}}{\boldsymbol{P}\{A \mid B\}\, \boldsymbol{P}\{B\} + \boldsymbol{P}\{A \mid \overline{B}\}\, \boldsymbol{P}\{\overline{B}\}}}$$

Example 2.34 (Reliability of a test, continued). Continue Example 2.32. Suppose that 4% of all the patients are infected with the virus, $\boldsymbol{P}\{V\} = 0.04$. Recall that $\boldsymbol{P}\{S \mid V\} = 0.95$ and $\boldsymbol{P}\{\overline{S} \mid \overline{V}\} = 0.99$. If the test shows positive results, the (conditional) probability that a patient has the virus equals

$$\begin{aligned} \boldsymbol{P}\{V \mid S\} &= \frac{\boldsymbol{P}\{S \mid V\}\, \boldsymbol{P}\{V\}}{\boldsymbol{P}\{S \mid V\}\, \boldsymbol{P}\{V\} + \boldsymbol{P}\{S \mid \overline{V}\}\, \boldsymbol{P}\{\overline{V}\}} \\ &= \frac{(0.95)(0.04)}{(0.95)(0.04) + (1-0.99)(1-0.04)} = \underline{0.7983}. \end{aligned}$$

◊

Example 2.35 (Diagnostics of computer codes). A new computer program consists of two modules. The first module contains an error with probability 0.2. The second module is more complex; it has a probability of 0.4 to contain an error, independently of the first module. An error in the first module alone causes the program to crash with probability 0.5. For the second module, this probability is 0.8. If there are errors in both modules, the program crashes with probability 0.9. Suppose the program crashed. What is the probability of errors in both modules?

Solution. Denote the events,

$$A = \{\text{errors in module I}\}, \quad B = \{\text{errors in module II}\}, \quad C = \{\text{crash}\}.$$

Further,

$$\begin{aligned} \{\text{errors in module I alone}\} &= A \backslash B = A \backslash (A \cap B) \\ \{\text{errors in module II alone}\} &= B \backslash A = B \backslash (A \cap B). \end{aligned}$$

It is given that $\boldsymbol{P}\{A\} = 0.2$, $\boldsymbol{P}\{B\} = 0.4$, $\boldsymbol{P}\{A \cap B\} = (0.2)(0.4) = 0.08$, by independence,

$\boldsymbol{P}\{C \mid A\backslash B\} = 0.5$, $\boldsymbol{P}\{C \mid B\backslash A\} = 0.8$, and $\boldsymbol{P}\{C \mid A \cap B\} = 0.9$.

We need to compute $\boldsymbol{P}\{A \cap B \mid C\}$. Since A is a union of disjoint events $A\backslash B$ and $A \cap B$, we compute

$$\boldsymbol{P}\{A\backslash B\} = \boldsymbol{P}\{A\} - \boldsymbol{P}\{A \cap B\} = 0.2 - 0.08 = 0.12.$$

Similarly,

$$\boldsymbol{P}\{B\backslash A\} = 0.4 - 0.08 = 0.32.$$

Events $(A\backslash B)$, $(B\backslash A)$, $A \cap B$, and $\overline{(A \cup B)}$ form a partition of Ω, because they are mutually exclusive and exhaustive. The last of them is the event of no errors in the entire program. Given this event, the probability of a crash is 0. Notice that A, B, and $(A \cap B)$ are neither mutually exclusive nor exhaustive, so they cannot be used for the Bayes Rule. Now organize the data.

Location of errors	Probability of a crash
$\boldsymbol{P}\{A\backslash B\} = 0.12$	$\boldsymbol{P}\{C \mid A\backslash B\} = 0.5$
$\boldsymbol{P}\{B\backslash A\} = 0.32$	$\boldsymbol{P}\{C \mid B\backslash A\} = 0.8$
$\boldsymbol{P}\{A \cap B\} = 0.08$	$\boldsymbol{P}\{C \mid A \cap B\} = 0.9$
$\boldsymbol{P}\left\{\overline{A \cup B}\right\} = 0.48$	$\boldsymbol{P}\left\{C \mid \overline{A \cup B}\right\} = 0$

Combining the Bayes Rule and the Law of Total Probability,

$$\boldsymbol{P}\{A \cap B \mid C\} = \frac{\boldsymbol{P}\{C \mid A \cap B\}\,\boldsymbol{P}\{A \cap B\}}{\boldsymbol{P}\{C\}},$$

where

$$\begin{aligned}\boldsymbol{P}\{C\} &= \boldsymbol{P}\{C \mid A\backslash B\}\,\boldsymbol{P}\{A\backslash B\} + \boldsymbol{P}\{C \mid B\backslash A\}\,\boldsymbol{P}\{B\backslash A\} \\ &\quad + \boldsymbol{P}\{C \mid A \cap B\}\,\boldsymbol{P}\{A \cap B\} + \boldsymbol{P}\left\{C \mid \overline{A \cup B}\right\}\boldsymbol{P}\left\{\overline{A \cup B}\right\}.\end{aligned}$$

Then

$$\boldsymbol{P}\{A \cap B \mid C\} = \frac{(0.9)(0.08)}{(0.5)(0.12) + (0.8)(0.32) + (0.9)(0.08) + 0} = \underline{0.1856}.$$

$\Diamond$

Summary and conclusions

Probability of any event is a number between 0 and 1. The empty event has probability 0, and the sample space has probability 1. There are rules for computing probabilities of unions, intersections, and complements. For a union of disjoint events, probabilities are added. For an intersection of independent events, probabilities are multiplied. Combining these rules, one evaluates reliability of a system given reliabilities of its components.

In the case of equally likely outcomes, probability is a ratio of the number of favorable outcomes to the total number of outcomes. Combinatorics provides tools for computing these numbers in frequent situations involving permutations and combinations, with or without replacement.

Given occurrence of event B, one can compute conditional probability of event A. Unconditional probability of A can be computed from its conditional probabilities by the Law of Total Probability. The Bayes Rule, often used in testing and diagnostics, relates conditional probabilities of A given B and of B given A.

Exercises

2.1. Out of six computer chips, two are defective. If two chips are randomly chosen for testing (without replacement), compute the probability that both of them are defective. List all the outcomes in the sample space.

2.2. Suppose that after 10 years of service, 40% of computers have problems with motherboards (MB), 30% have problems with hard drives (HD), and 15% have problems with both MB and HD. What is the probability that a 10-year old computer still has fully functioning MB and HD?

2.3. A new computer virus can enter the system through e-mail or through the internet. There is a 30% chance of receiving this virus through e-mail. There is a 40% chance of receiving it through the internet. Also, the virus enters the system simultaneously through e-mail and the internet with probability 0.15. What is the probability that the virus does not enter the system at all?

2.4. Among employees of a certain firm, 70% know C/C++, 60% know Fortran, and 50% know both languages. What portion of programmers

(a) does not know Fortran?

(b) does not know Fortran and does not know C/C++?

(c) knows C/C++ but not Fortran?

(d) knows Fortran but not C/C++?

(e) If someone knows Fortran, what is the probability that he/she knows C/C++ too?

(f) If someone knows C/C++, what is the probability that he/she knows Fortran too?

2.5. A computer program is tested by 3 *independent* tests. When there is an error, these tests will discover it with probabilities 0.2, 0.3, and 0.5, respectively. Suppose that the program contains an error. What is the probability that it will be found by at least one test?

2.6. Under good weather conditions, 80% of flights arrive on time. During bad weather, only 30% of flights arrive on time. Tomorrow, the chance of good weather is 60%. What is the probability that your flight will arrive on time?

2.7. A system may become infected by some spyware through the internet or e-mail. Seventy percent of the time the spyware arrives via the internet, thirty percent of the time via e-mail. If it enters via the internet, the system detects it immediately with probability 0.6. If via e-mail, it is detected with probability 0.8. What percentage of times is this spyware detected?

2.8. A shuttle's launch depends on three key devices that may fail independently of each other with probabilities 0.01, 0.02, and 0.02, respectively. If any of the key devices fails, the launch will be postponed. Compute the probability for the shuttle to be launched on time, according to its schedule.

2.9. Successful implementation of a new system is based on three independent modules. Module 1 works properly with probability 0.96. For modules 2 and 3, these probabilities equal 0.95 and 0.90. Compute the probability that at least one of these three modules fails to work properly.

2.10. Three computer viruses arrived as an e-mail attachment. Virus A damages the system with probability 0.4. Independently of it, virus B damages the system with probability 0.5. Independently of A and B, virus C damages the system with probability 0.2. What is the probability that the system gets damaged?

2.11. A computer program is tested by 5 independent tests. If there is an error, these tests will discover it with probabilities 0.1, 0.2, 0.3, 0.4, and 0.5, respectively. Suppose that the program contains an error. What is the probability that it will be found

(a) by at least one test?

(b) by at least two tests?

(c) by all five tests?

2.12. A building is examined by policemen with four dogs that are trained to detect the scent of explosives. If there are explosives in a certain building, and each dog detects them with probability 0.6, independently of other dogs, what is the probability that the explosives will be detected by at least one dog?

2.13. An important module is tested by three independent teams of inspectors. Each team detects a problem in a defective module with probability 0.8. What is the probability that at least one team of inspectors detects a problem in a defective module?

2.14. A spyware is trying to break into a system by guessing its password. It does not give up until it tries 1 million different passwords. What is the probability that it will guess the password and break in if by rules, the password must consist of

(a) 6 different lower-case letters

(b) 6 different letters, some may be upper-case, and it is case-sensitive

(c) any 6 letters, upper- or lower-case, and it is case-sensitive

(d) any 6 characters including letters and digits

2.15. A computer program consists of two blocks written independently by two different programmers. The first block has an error with probability 0.2. The second block has an error with probability 0.3. If the program returns an error, what is the probability that there is an error in both blocks?

2.16. A computer maker receives parts from three suppliers, S1, S2, and S3. Fifty percent come from S1, twenty percent from S2, and thirty percent from S3. Among all the parts supplied by S1, 5% are defective. For S2 and S3, the portion of defective parts is 3% and 6%, respectively.

(a) What portion of all the parts is defective?

(b) A customer complains that a certain part in her recently purchased computer is defective. What is the probability that it was supplied by S1?

2.17. A computer assembling company receives 24% of parts from supplier X, 36% of parts from supplier Y, and the remaining 40% of parts from supplier Z. Five percent of parts supplied by X, ten percent of parts supplied by Y, and six percent of parts supplied by Z are defective. If an assembled computer has a defective part in it, what is the probability that this part was received from supplier Z?

2.18. A problem on a multiple-choice quiz is answered correctly with probability 0.9 if a student is prepared. An unprepared student guesses between 4 possible answers, so the probability of choosing the right answer is 1/4. Seventy-five percent of students prepare for the quiz. If Mr. X gives a correct answer to this problem, what is the chance that he did not prepare for the quiz?

2.19. At a plant, 20% of all the produced parts are subject to a special electronic inspection. It is known that any produced part which was inspected electronically has no defects with probability 0.95. For a part that was not inspected electronically this probability is only 0.7. A customer receives a part and find defects in it. What is the probability that this part went through an electronic inspection?

2.20. All athletes at the Olympic games are tested for performance-enhancing steroid drug use. The imperfect test gives positive results (indicating drug use) for 90% of all steroid-users but also (and incorrectly) for 2% of those who do not use steroids. Suppose that 5% of all registered athletes use steroids. If an athlete is tested negative, what is the probability that he/she uses steroids?

2.21. In the system in Figure 2.7, each component fails with probability 0.3 independently of other components. Compute the system's reliability.

2.22. Three highways connect city A with city B. Two highways connect city B with city C. During a rush hour, each highway is blocked by a traffic accident with probability 0.2, independently of other highways.

(a) Compute the probability that there is at least one open route from A to C.

(b) How will a new highway, also blocked with probability 0.2 independently of other highways, change the probability in (a) if it is built