2

BASIC CONCEPTS IN PROBABILITY

We see that the theory of probability is at bottom only common sense reduced to calculation; it makes us appreciate with exactitude what reasonable minds feel by a sort of instinct, often without being able to account for it. . . . It is remarkable that this science, which originated in the consideration of games of chance, should become the most important object of human knowledge.

Laplace

This chapter covers fundamental topics on probabilities of events.

Main Topics

- Basics of Set Theory
- Fundamental Concepts in Probability
- Conditional Probability
- Independent Events
- Total Probability Theorem and Bayes' Rule
- Combined Experiments and Bernoulli Trials

The materials covered in this chapter are essential for the study of the remaining chapters. Emphasis should be put on the understanding of concepts and how they can be applied.

2.1 Basics of Set Theory

2.1.1 Basic Definitions

- ***set*** = a collection of objects, denoted by an upper case Latin letter
  Example: $A = \{1, 4\}, \quad D = \{d_1, d_2, d_3\}$.
- ***element*** = an object in a set, denoted by a lower case Latin letter
  We say "a is an element of A," "a is in A," or "a belongs to A," denoted as $a \in A$.
- ***empty set*** = ***null set*** = a set with no elements, denoted by $\emptyset$
- ***space*** = the set with all the elements for the problem *under consideration* (sometimes called ***universal set***), denoted by S

Convention:

$$\text{Upper case Latin letter} = \text{set}$$
$$\text{Lower case Latin letter} = \text{element}$$

If *every* element of set A is also an element of set B, then A is said to be a ***subset*** of B, denoted as $A \subset B$ or $B \supset A$. Set A is said to be ***equal*** to set B if $A \subset B$ *and* $A \supset B$, denoted as $A = B$. In this case, A and B have exactly the same elements. Two sets are said to be ***disjoint*** if they do not have any element in common.

Example 2.1: Consider the set of all positive integers smaller than 7:

$$\begin{aligned} \text{rule method}: \quad & A = \{x : 0 < x < 7, x \text{ an integer}\} \\ \text{tabular method}: \quad & A = \{1, 2, 3, 4, 5, 6\} \\ & \quad = \text{the space (universal set) of a 6-face die} \end{aligned}$$

Tabular form is not universally applicable.

Example 2.2: Consider the set of all positive numbers smaller than 6:

$$\text{rule method}: \quad B = \{x : 0 < x < 6, x \text{ a real number}\}$$

There is no tabular form for this set because it is *uncountable*.

Example 2.3: Consider the set of all positive integers:

$$C = \{x : x > 0, x \text{ an integer}\} = \{1, 2, 3, \ldots\}$$

Example 2.4: The set of human genders G = {female, male}

2.1.2 Basic Set Operations

Definitions:

- The set of all elements of A *or* B is called the ***union*** (or ***sum***) of A and B, denoted as $A \cup B$ or $A + B$. Union of *disjoint* sets A and B may be denoted as $A \uplus B$.
  Convention: "A or B" = "either A or B or both."
- The set of all elements *common* to A *and* B is called the ***intersection*** (or ***product***) of A and B, denoted as $A \cap B$ or AB.
- The set of all elements of A that are not in B is called the ***difference*** of A and B, denoted as $A - B$.
- The set of all elements in the space S but not in A is called the ***complement*** of A, denoted as $\overline{A}$. It is equal to $S - A$.

A simple and instructive way of illustrating the relationships among sets is the so-called ***Venn diagram***, as illustrated below.

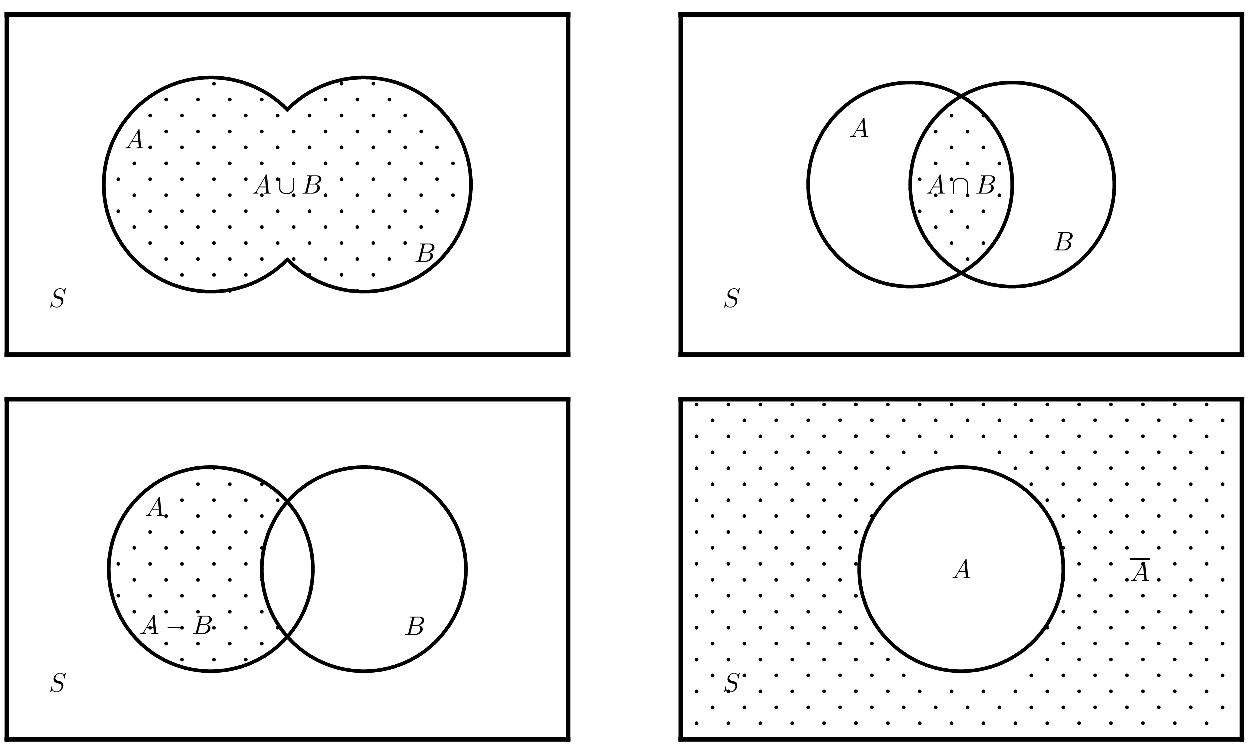


Figure 2.1: Basic set operations.

Example 2.5: Set Operations

For A, B, and C considered in Examples 2.1, 2.2, and 2.3:

$A \subset C$

$A \cup B = \{x : 0 < x \leq 6, x \text{ a real number}\}$

$A \cup C = C$

$B \cup C = \{x : x \text{ a positive integer or a real number satisfying } 0 < x < 6\}$

This set has a mixed type.

$A \cap B = \{1, 2, 3, 4, 5\}$

$A \cap C = A$

$B \cap C = \{1, 2, 3, 4, 5\}$

$A - B = \{6\}$

$A - C = \emptyset$

$B - A = \{x : 0 < x < 6, x \text{ a noninteger real number}\}$

$B - C = \{x : 0 < x < 6, x \text{ a noninteger real number}\}$

$C - A = \{x : x \geq 7, x \text{ an integer}\} = \{7, 8, 9, \ldots\}$

$C - B = \{x : x \geq 6, x \text{ an integer}\} = \{6, 7, 8, \ldots\}$

Space S depends on what we are considering. If we are considering only positive real numbers, then $S = \{x : x > 0, x \text{ real}\}$. Thus,

$$\overline{A} = \{x : x \text{ a positive real number other than } 1, 2, 3, 4, 5, 6\}$$
$$\overline{B} = \{x : x \geq 6, x \text{ a real number}\}$$
$$\overline{C} = \{x : x \text{ a noninteger positive real number}\}$$

If, however, we are considering all real numbers, then $S = \{x : x \text{ real}\}$. Thus

$$\overline{A} = \{x : x \text{ a real number other than } 1, 2, 3, 4, 5, 6\}$$
$$\overline{B} = \{x : x \leq 0 \text{ or } x \geq 6, x \text{ a real number}\}$$
$$\overline{C} = \{x : x \leq 0 \text{ or } x \text{ a noninteger positive real number}\}$$

2.1.3 Basic Algebra of Sets

	Algebra of sets	*Algebra of numbers*
	Union $\cup$	sum "+"
	Intersection $\cap$	product "·"
1	$A \cup B = B \cup A$	$a + b = b + a$
2	$A \cap B = B \cap A$	$a \cdot b = b \cdot a$
3	$A \cup (B \cup C) = A \cup B \cup C$	$a + (b + c) = a + b + c$
4	$A \cap (B \cap C) = A \cap B \cap C$	$a \cdot (b \cdot c) = a \cdot b \cdot c$
5	$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$	$a \cdot (b + c) = a \cdot b + a \cdot c$
6	$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$	see below

Since $A \cap A = A$, $(A \cap B) \subset A$, $(A \cap C) \subset A$, and

$$(a + b)(a + c) = a \cdot a + a \cdot b + a \cdot c + b \cdot c$$

Line 6 in the table above follows from

$$(A \cup B) \cap (A \cup C) = \underbrace{(A \cap A) \cup (A \cap B) \cup (A \cap C)}_{=A} \cup (B \cap C) = A \cup (B \cap C)$$

This illustrates that set algebra has its own rules.

De Morgan's laws:

$$\overline{A \cup B} = \overline{A} \cap \overline{B} \tag{2.1}$$

$$\overline{A \cap B} = \overline{A} \cup \overline{B} \tag{2.2}$$

Similarly,

$$\overline{A \cup B \cup C} = \overline{A \cup D} = \overline{A} \cap \overline{D} = \overline{A} \cap \overline{B \cup C} = \overline{A} \cap (\overline{B} \cap \overline{C}) = \overline{A} \cap \overline{B} \cap \overline{C}$$

$$\overline{A_1 \cup \cdots \cup A_n} = \overline{A_1} \cap \cdots \cap \overline{A_n}$$

$$\overline{A_1 \cap \cdots \cap A_n} = \overline{A_1} \cup \cdots \cup \overline{A_n}$$

$$\overline{(A_1 \cup A_2) \cap (A_3 \cup A_4)} = (\overline{A_1} \cap \overline{A_2}) \cup (\overline{A_3} \cap \overline{A_4})$$

Rules: (1) interchange $\cup$ and $\cap$; (2) interchange $(*)$ and $\overline{(*)}$. However, care should be taken when dealing with multiple nests, as demonstrated below.

Example 2.6:

$$\overline{(\overline{\underbrace{\overline{A} \cap B}_{D}}) \cup \overline{C}} = \overline{\overline{D} \cup \overline{C}} = D \cap C = (\overline{A} \cap B) \cap C = \overline{A \cup \overline{B}} \cap C \tag{2.3}$$

2.2 Fundamental Concepts in Probability

2.2.1 Definitions

- ***random experiment*** = experiment (action) whose result is uncertain (cannot be predicted with certainty) before it is performed
- ***trial*** = single performance of the random experiment
- ***outcome*** = result of a trial
- ***sample space*** S = the set of all possible outcomes of a random experiment
- ***event*** = subset of the sample space S (to which a probability can be assigned)
  = a collection of possible outcomes
- ***sure event*** = sample space S (an event for sure to occur)
- ***impossible event*** = empty set $\emptyset$ (an event impossible to occur)

We say *an event has* ***occurred*** *if and only if the outcome observed belongs to the set of the event,* as explained below.

Example 2.7: Die-Rolling Events

Rolling a die is a random experiment. An outcome can be any number from 1 to 6. Sample space = $\{1,2,3,4,5,6\}$. Some possible events are

$$
\begin{aligned}
A &= \{\text{an even number shows up}\} = \{2,4,6\} \qquad \text{(3 outcomes)}\\
B &= \{\text{a number greater than 5 shows up}\}\\
&= \{6\} \qquad \text{(single outcome)}\\
C &= \{\text{2 shows up}\} = \{2\} \qquad \text{(single outcome)}\\
D &= \{\text{a number greater than 6 shows up}\} = \emptyset \qquad \text{(no outcome)}\\
E &= \{\text{2 and 4 show up}\} = \emptyset \qquad \text{(no outcome)}\\
F &= \{\text{2 or 4 shows up}\} = \{2,4\} \qquad \text{(2 outcomes)}\\
G &= \{\text{a number from 1 to 6 shows up}\}\\
&= \{1,2,3,4,5,6\} = S \qquad \text{(all outcomes)}
\end{aligned}
$$

Thus, if “2” showed up, then we say that events A, C, F, and G have all *occurred*.

2.2.2 Probability of an Event

Traditional definitions of the probability of an event A:

$$\textbf{\textit{Classical:}}\ P\{A\} = \frac{\text{\# of possible outcomes for event } A}{\text{\# of possible outcomes for space } S}$$

$$\textbf{\textit{Relative frequency:}}\ P\{A\} = \lim_{N\to\infty} \frac{\text{\# of occurrences of event } A}{N \text{ (total \# of trials)}}$$

$$\textbf{\textit{Geometric:}}\ P\{A\} = \frac{\text{geometric measure of set } A}{\text{geometric measure of space } S}$$

These definitions are very natural but limited:

- The classical definition is virtually applicable only to events with finitely (or countably) many outcomes that are equally probable.
- The geometric definition is an extension for events with uncountably many outcomes that are uniformly probable.
- The relative-frequency definition is more general than the other two definitions but is still limited. It is difficult to be applied to problems in which outcomes are not equally probable.

Example 2.8: Classical Probability: Die Rolling

Consider Example 2.7. The probabilities of events are

$$\begin{aligned}
P\{A\} &= P\{\text{an even number shows up}\} = \frac{1+1+1}{6} = \frac{1}{2} \\
P\{E\} &= P\{\text{2 and 4 show up}\} = 0/6 = 0 \\
&\qquad \Longrightarrow \textit{Impossible event has zero probability} \\
P\{F\} &= P\{\text{2 or 4 shows up}\} = 2/6 = 1/3 \\
P\{G\} &= P\{\text{a number from 1 to 6 shows up}\} = 6/6 = 1 \\
&\qquad \Longrightarrow \textit{Sure event has unity probability}
\end{aligned}$$

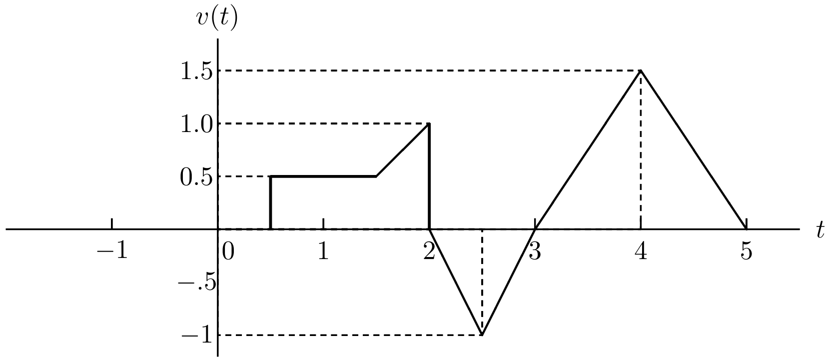
Note, however, as illustrated in the next example,

- An event of zero probability is not necessarily an impossible event.
- An event of unity probability is not necessarily a sure event.

These counter-intuitive results are possible only when the sample space has infinitely many elements.

Example 2.9: Geometric Probability: Waveform Sampling

The following voltage waveform is to be sampled at a random time τ over the period $-1 \leq t \leq 5$.



(a) Determine the probability that the sampled value $v(\tau) < -0.25$:

$$A = \{\text{sampled value } v(\tau) < -0.25\}$$
$$P\{A\} = \frac{\text{time in which sampled value } v(\tau) < -0.25}{\text{total time}} = \frac{3/4}{6} = \frac{1}{8}$$

(b) Determine the probability that the sampled value $v(\tau) \geq 1.0$:

$$B = \{\text{sampled value } v(\tau) \geq 1.0\}, \quad P\{B\} = \frac{(1/3)(5-3)}{6} = \frac{1}{9}$$

(c) Determine the probability that the sampled value $v(\tau) = 0.5$:

$$C = \{\text{sampled value } v(\tau) = 0.5\}, \quad P\{C\} = (1.5 - 0.5)/6 = 1/6$$

(d) Determine the probability that the sampled value $v(\tau) = 1.2$:

$$D = \{\text{sampled value } v(\tau) = 1.2\}, \quad P\{D\} = 0/6 = 0$$

(e) Determine the probability that the sampled value satisfies $-1 \leq v(\tau) < 1.5$ but not equal to -0.5:

$$E = \{-1 \leq v(\tau) < 1.5, v(\tau) \neq -0.5\}, \quad P\{E\} = 6/6 = 1$$

Note:

- D is not an impossible event but $P\{D\} = 0$.
- E is not a sure event but $P\{E\} = 1$.

2.2.3 Axioms of Probability Theory

A set of events A_1, A_2, $\ldots$, A_n is said to be ***mutually exclusive*** or ***disjoint*** if

$$A_i \cap A_j = \emptyset \qquad \forall i \neq j$$

That is, at most one event can occur (if one occurs, any other cannot occur).

In the contemporary theory of probability, the following properties have been identified as fundamental (necessary and sufficient) for probability as a measure, which are taken as axioms:

- *Axiom* 1 (nonnegativity): Probability of any event A is bounded by 0 and 1:

$$\boxed{0 \leq P\{A\} \leq 1} \tag{2.4}$$

- *Axiom* 2 (unity): Any sure event (the sample space) has unity probability:

$$\boxed{P\{S\} = 1} \tag{2.5}$$

- *Axiom* 3 (finite additivity): If A_1, A_2, $\ldots$, A_n are disjoint events, then

$$P\{A_1 \uplus A_2 \uplus \cdots \uplus A_n\} \triangleq P\Big\{\biguplus_{i=1}^{n} A_i\Big\} = \sum_{i=1}^{n} P\{A_i\} \tag{2.6}$$

- *Axiom* 3' (countable additivity): If A_1, A_2, $\ldots$ are disjoint events, then

$$\boxed{P\Big\{\biguplus_{i=1}^{\infty} A_i\Big\} = \sum_{i=1}^{\infty} P\{A_i\}} \tag{2.7}$$

All other probability laws can be derived from these axioms. Keep in mind that (2.6) and (2.7) are valid only for mutually exclusive events.

These axioms imply that probability can be interpreted as *mass* associated with various events. They are clearly reasonable from relative-frequency perspective:

$$\begin{aligned} 0 \leq P\{A\} &= \frac{N_A}{N} \leq 1 \\ P\{S\} &= \frac{N}{N} = 1 \\ P\Big\{\biguplus_{i=1}^{n} A_i\Big\} &= \frac{N_{A_1} + \cdots + N_{A_n}}{N} = \frac{N_{A_1}}{N} + \cdots + \frac{N_{A_n}}{N} = \sum_{i=1}^{n} P\{A_i\} \end{aligned}$$

2.2.4 Probability of the Union of Two Events

The union of events A and B in space S is the set of all outcomes of A or B (or both). In other words, if any outcome of either A or B occurs, then we say the ***union of events*** A and B, denoted by $A \cup B$, occurs.

By Axiom 3, if $A \cap B = \emptyset$ (A and B cannot both occur), then

$$P\{\text{either } A \text{ or } B \text{ occurs}\} = P\{A \uplus B\} = P\{A\} + P\{B\}$$

What if $A \cap B \neq \emptyset$ (A and B may both occur)? Note that

$$A \cup B = A \uplus (\overline{A} \cap B)$$

which can be shown easily by Venn diagram. Clearly, $A \cap (\overline{A} \cap B) = \emptyset$. Hence,

$$P\{A \cup B\} = P\{A \uplus (\overline{A} \cap B)\} \overset{\text{Axiom 3}}{=} P\{A\} + P\{\overline{A} \cap B\} \tag{2.8}$$

Similarly, $B = (A \cap B) \uplus (\overline{A} \cap B)$ and $(A \cap B) \cap (\overline{A} \cap B) = \emptyset$, and thus

$$P\{B\} = P\{(A \cap B) \uplus (\overline{A} \cap B)\} \overset{?}{=} P\{A \cap B\} + P\{\overline{A} \cap B\} \tag{2.9}$$

Thus, (2.8) $-$ (2.9) yields the ***probability of union of two events*** or ***addition rule of probability***:

$$\boxed{P\{A \cup B\} = P\{A\} + P\{B\} - P\{A \cap B\}} \tag{2.10}$$

This is intuitively correct:

$$\begin{aligned} P\{\text{either } A \text{ or } B \text{ or both occur}\} = {} & P\{A \text{ occurs}\} + P\{B \text{ occurs}\} \\ & - \underbrace{P\{A \text{ and } B \text{ both occur}\}}_{\text{double counted}} \end{aligned}$$

since $P\{A \text{ and } B \text{ both occur}\}$ is double counted in $P\{A \text{ occurs}\} + P\{B \text{ occurs}\}$.

2.2.5 Probability of the Complement of an Event

If A is an event in space S, then $S = A \uplus \overline{A}$, and

$$1 \overset{\text{Axiom 2}}{=} P\{S\} = P\{A \uplus \overline{A}\} \overset{\text{Axiom 3}}{=} P\{A\} + P\{\overline{A}\}$$

Thus, the ***probability of the complement of an event*** is

$$\boxed{P\{\overline{A}\} = 1 - P\{A\}} \tag{2.11}$$

2.2.6 Joint Probability

The ***joint events*** A and B is the intersection of sets A and B, which is the set of outcomes common to both A and B. As such, the ***joint probability*** of events A and B is the probability that they both occur. This probability is denoted by

$$P\{A \cap B\} \triangleq P\{A, B\} \triangleq P\{AB\} \tag{2.12}$$

which is given by, from (2.10),

$$\boxed{P\{A \cap B\} = P\{A\} + P\{B\} - P\{A \cup B\}} \tag{2.13}$$

This can be interpreted as follows:

$$\begin{aligned} P\{A \text{ and } B \text{ both occur}\} = P\{&A \text{ occurs regardless } B \text{ occurs or not}\} \\ &+ P\{B \text{ occurs regardless } A \text{ occurs or not}\} \\ &- P\{A \text{ or } B \text{ or both occur}\} \end{aligned}$$

If $A \cap B = \emptyset$, then

$$P\{A \cap B\} = P\{A\} + P\{B\} - P\{A \uplus B\} \overset{\text{Axiom 3}}{=} 0$$

or

$$P\{A \cap B\} = P\{\emptyset\} = P\{\overline{S}\} \overset{(2.11)}{=} 1 - P\{S\} = 0$$

which makes sense: "*mutually exclusive events have zero joint probability*" and "*impossible events have zero probability*."

Example 2.10: Axiomatic Probabilities of Die-Rolling Events

Rolling a die is a random experiment whose outcomes can be any face from 1 to 6. Let F_i be the event that face i shows up. Assume the die is a fair one — each outcome is equiprobable. Let us use the probability axioms to (determine how to) assign a probability to every outcome.

$$1 \overset{\text{Axiom 2}}{=} P\{S\} = P\Big\{\biguplus_{i=1}^{6} F_i\Big\} \overset{\text{Axiom 3}}{=} \sum_{i=1}^{6} P\{F_i\} \overset{\text{fair die}}{=} 6P\{F_i\}$$

which yields $P\{F_i\} = 1/6, i = 1, \ldots, 6$ and hence:

$$P\{\text{an even numbered face shows up}\} \overset{\text{Axiom 3}}{=} P\{F_2\} + P\{F_4\} + P\{F_6\} = 1/2$$

2.3 Conditional Probability

The probability of an event A *under the condition* that event B has occurred is called the ***conditional probability*** of A given B (or probability of A conditioned on B), defined as

$$\boxed{P\{A|B\} \triangleq \frac{P\{A \cap B\}}{P\{B\}}} \qquad \text{if } P\{B\} \neq 0 \tag{2.14}$$

The relative-frequency interpretation of this is

$$\begin{aligned} P\{A|B\} &= \frac{\text{\# of outcomes for event } (A \cap B)}{\text{\# of outcomes for event } B} = \frac{N_{A\cap B}}{N_B} \\ &= \frac{N_{A\cap B}/N}{N_B/N} = \frac{P\{A \cap B\}}{P\{B\}} \end{aligned}$$

Since $A \cap S = A$ and $P\{S\} = 1$, we have $P\{A\} = \frac{P\{A\cap S\}}{P\{S\}} = P\{A|S\}$. A comparison of this with (2.14) indicates that *conditional probability of A given B is simply the probability of A assuming the entire sample space is B*, which is reasonable since event B has occurred (and the occurrence of any outcome outside B is impossible), as illustrated below.

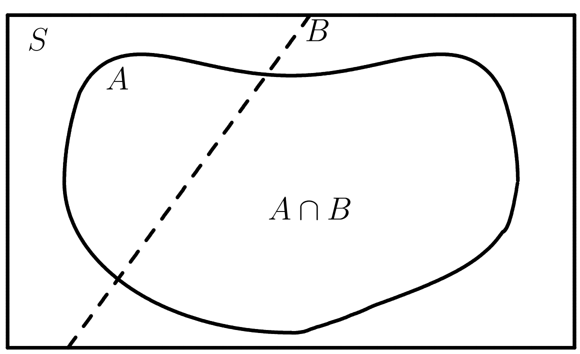


Figure 2.3: Illustration of conditional probability.

(2.14) implies that the following ***product rule*** or ***multiplication rule*** holds

$$\boxed{P\{AB\} \triangleq P\{A \cap B\} = P\{A|B\}P\{B\}} \qquad \text{if } P\{B\} \neq 0 \tag{2.15}$$

Example 2.12: Double Luck with Lottery

The probability that a person will win the lottery twice can be calculated as follows. Let $A_i = \{\text{win the } i\text{th time}\}$. Then

$$P\{\text{win twice}\} = P\{A_1, A_2\} = P\{A_2|A_1\}P\{A_1\} \stackrel{?}{=} P\{A_2\}P\{A_1\}$$

where $P\{A_2|A_1\}$ is the probability of winning the second time after winning the first time, which is equal to $P\{A_2\}$ in reality. When this is the case, we say that the two wins are statistically independent, to be studied later.

Example 2.13: A pair of resistor and capacitor was chosen at random from a box of nine pairs of resistors and capacitors with the following resistance and capacitance, respectively.

Pair	1	2	3	4	5	6	7	8	9
R ($k\Omega$)	1	1	1	2.5	2.5	2.5	5	5	5
C (μF)	20	40	50	20	40	50	20	40	50
$\tau = RC$ (ms)	20	40	50	50	100	125	100	200	250

(a) What is the probability that the time constant $\tau < 50$ or $\tau > 125$?
Let $T = \{50 \le \tau \le 125\}$. Then

$$P\{\tau < 50 \text{ or } \tau > 125\} = P\{\overline{T}\} = 1 - P\{T\} \stackrel{\text{table}}{=} 1 - \frac{5}{9} = \frac{4}{9}$$

(b) Determine $P\{\overline{T}|C \neq 20\}$:

$$P\{\overline{T}|C \neq 20\} = \frac{P\{\overline{T} \cap (C \neq 20)\}}{P\{C \neq 20\}} \stackrel{\text{table}}{=} \frac{3/9}{6/9} = \frac{1}{2}$$

(c) Determine $P\{C \neq 20|\overline{T}\}$:

$$P\{C \neq 20|\overline{T}\} = \frac{P\{\overline{T} \cap (C \neq 20)\}}{P\{\overline{T}\}} \stackrel{\text{(a),(b)}}{=} \frac{3/9}{4/9} = \frac{3}{4}$$

(d) Determine $P\{C \neq 20, R \neq 5|\overline{T}\}$:

$$P\{C \neq 20, R \neq 5|\overline{T}\} = \frac{P\{\overline{T} \cap (C \neq 20) \cap (R \neq 5)\}}{P\{\overline{T}\}} = \frac{1/9}{4/9} = \frac{1}{4}$$

Example 2.14: Life Expectancy

Let t be the age of a person when he/she dies. The probability that he/she dies at an age not older than t_0 is given by

$$P\{t \leq t_0\} = \int_0^{t_0} \alpha(t)dt$$

where $\alpha(t)$ is a function determined from mortality records by

$$\alpha(t) = \begin{cases} 3 \times 10^{-9} t^2 (100 - t)^2 & 0 \leq t \leq 100 \text{ years} \\ 0 & \text{otherwise} \end{cases}$$

(This model is approximate — it implies a zero probability to have a life expectancy longer than 100 years.)

(a) The probability that a person will die between the ages of 60 and 70 is

$$P\{60 \leq t \leq 70\} = \frac{\text{\# of people who die between 60 and 70}}{\text{total population}}$$
$$= \int_0^{70} \alpha(t)dt - \int_0^{60} \alpha(t)dt = \int_{60}^{70} \alpha(t)dt = 0.154$$

(b) The probability that a person will die between the ages of 60 and 70 assuming that his/her current age is 60 is

$$P\{60 \leq t \leq 70 | t \geq 60\} = \frac{\text{\# of people who die between 60 and 70}}{\text{total population of age older than 60}}$$
$$= \frac{P\{60 \leq t \leq 70\}}{P\{t \geq 60\}} = \frac{\int_{60}^{70} \alpha(t)dt}{\int_{60}^{100} \alpha(t)dt} = 0.486$$

(c) The probability that a person will die between the ages of 20 and 50 assuming that his/her current age is 60 is

$$P\{20 \leq t \leq 50 | t \geq 60\} = \frac{P\{(20 \leq t \leq 50) \cap (t \geq 60)\}}{P\{t \geq 60\}} = \frac{P\{\emptyset\}}{P\{t \geq 60\}} = 0$$

The results of this example are useful for e.g., determination of the premiums of a life insurance policy.

Example 2.15: Winning Strategy for a TV Game

A TV game that was popular in some European countries is as follow. A car is behind one of three doors. A player chooses one door. The player wins the car if it turns out to be behind the selected door. After the player chooses, the TV host will open another door which does not have the car because the host knows where the car is. After the host opens that door, the player is allowed to switch to choose the third door or stick to the original choice.

Question: Is it better to switch to the third door?

Assume for notational simplicity that the player has chosen door B and the host has opened door A. Then, since by now we know that door A does not have the car, switching to choose door C will win if and only if door B does not have the car, denoted by $\overline{B}$; that is,

$$P\{\text{Winning by switching}\} = P\{C|\overline{A},\overline{B}\}P\{\overline{B}\} = P\{\overline{B}\} = 1 - P\{B\} = 2/3$$

Clearly, switching and not switching are mutually exclusive because they cannot both win. Thus, we have

$$P\{\text{Winning by not switching}\} \leq 1 - P\{\text{Winning by switching}\} = 1/3$$

In fact, the probability of winning by choosing door B in the first place is $1/3$. By not switching, the probability of winning stays unchanged because the new information that door A does not have the car is not utilized since it would be the same if instead door C was opened by the host. Thus,

$$P\{\text{Winning by not switching}\} = P\{B\} = 1/3$$

Consequently, the chance of winning is doubled by switching! This answer would be hard to come by without a good understanding of probability concepts.

The above analysis can be extended to the general n-door problem as follows. Assume for simplicity that the original choice of the player was door B, door A was opened by the host, and door C is chosen if the player changes his or her choice. Then, switching will win if and only if door B does not have the car and door C turns out to have the car among the $n-2$ remaining doors; that is,

$$P\{\text{Winning by switching}\} = P\{C|\overline{A},\overline{B}\}P\{\overline{B}\} = \frac{1 - P\{B\}}{n-2} = \frac{1}{n-2}\frac{n-1}{n}$$

which is greater than $P\{\text{Winning by not switching}\} = P\{C\} = \frac{1}{n}$.

2.4 Independent Events

Two events A and B are said to be ***independent*** if the probability of occurrence of one event is not affected by the occurrence of the other event, that is,

$$P\{A|B\} = P\{A\} \qquad \text{and} \qquad P\{B|A\} = P\{B\} \tag{2.17}$$

where $P\{A\}$ and $P\{B\}$ are assumed nonzero. An equivalent but more compact form of (2.17) is

$$P\{A \cap B\} = P\{A\}P\{B\} \tag{2.18}$$

Thus, formally two events A and B are said to be statistically (or probabilistically) ***independent*** if (2.18) holds. The equivalence of (2.17) (with $P\{A\} \neq 0 \neq P\{B\}$) and (2.18) can be seen as follows:

$$\text{(2.17)} \implies P\{A \cap B\} \overset{(2.14)}{=} P\{A|B\}P\{B\} \overset{(2.17)}{=} P\{A\}P\{B\} \implies (2.18)$$

$$\text{(2.18)} \implies \left\{ \begin{matrix} P\{A\}P\{B\} \overset{(2.18)}{=} P\{A \cap B\} \overset{(2.14)}{=} P\{A|B\}P\{B\} \\ P\{A\}P\{B\} \overset{(2.18)}{=} P\{B \cap A\} \overset{(2.14)}{=} P\{B|A\}P\{A\} \end{matrix} \right\} \implies (2.17)$$

Events are said to be statistically ***dependent*** if they are not independent. Independence simplifies the calculation of joint probability greatly:

$$\text{joint probability} \overset{\text{if independent}}{=} \text{product of probabilities}$$

Independence of n Events

For n events $A_1, A_2, \ldots, A_n$, if

$$\begin{aligned} P\{A_i \cap A_j\} &= P\{A_i\}P\{A_j\} \qquad \forall i \neq j \\ P\{A_i \cap A_j \cap A_k\} &= P\{A_i\}P\{A_j\}P\{A_k\} \qquad \forall i \neq j \neq k \\ &\vdots \\ P\{A_1 \cap A_2 \cap \cdots \cap A_n\} &= P\{A_1\}P\{A_2\} \cdots P\{A_n\} \end{aligned}$$

then events $A_1, A_2, \ldots, A_n$ are said to be statistically ***independent***. Otherwise, they are dependent.

Example 2.16: Independent vs. Disjoint Events

Can two events be both independent and ***disjoint*** (i.e., ***mutually exclusive***)? Note that

$$P\{A\}P\{B\} \overset{\text{independent}}{=} P\{A \cap B\} \overset{\text{disjoint}}{=} 0$$

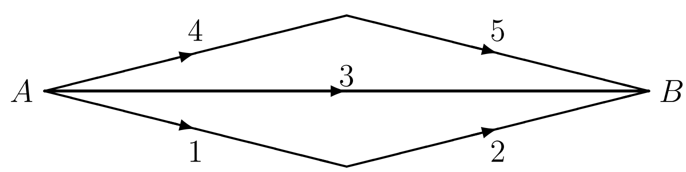
indicates that if two events are both independent and disjoint, then at least one of them has zero probability — *nonzero-probability events cannot be both independent and disjoint*. Intuitively, if two events are disjoint, the occurrence of one precludes the other and thus they cannot be independent. Note the difference:

$$\text{joint probability} \overset{\text{if independent}}{=} \text{product of probabilities} \tag{2.19}$$

$$\text{union probability} \overset{\text{if disjoint}}{=} \text{sum of probabilities} \tag{2.20}$$

Example 2.17: Reliability of Communication Channel

Consider the following communication network. Assume the links are independent and the probability that a link is operational is 0.95.



Since independence of links implies that paths are independent, the probability of being able to transmit from A to B can be calculated as follows:

$$\begin{aligned}
P\{\text{path 1-2 OK}\} &\overset{?}{=} P\{\text{link 1 OK}\}P\{\text{link 2 OK}\} = 0.95 \times 0.95 = 0.9025 \\
P\{\text{path 1-2 fails}\} &= 1 - P\{\text{path 1-2 OK}\} = 1 - 0.9025 = 0.0975 \\
P\{\text{path 3 fails}\} &= 1 - P\{\text{link 3 OK}\} = 1 - 0.95 = 0.05 \\
P\{\text{all paths fail}\} &\overset{?}{=} P\{\text{path 1-2 fails}\}P\{\text{path 4-5 fails}\}P\{\text{path 3 fails}\} \\
&= 0.0975 \times 0.0975 \times 0.05 = 0.000475
\end{aligned}$$

Finally,

$$\begin{aligned}
P\{\text{able to transmit from } A \text{ to } B\} &= 1 - P\{\text{all paths fail}\} \\
&= 1 - 0.000475 = 0.999525 \quad \text{(very high)}
\end{aligned}$$

Example 2.18: Security of Nuclear Power Plant

A nuclear power plant will shut down if systems A and B or A and C fail simultaneously. $A, B,$ and C are independent systems and their probabilities of failure are 0.01, 0.03, and 0.02, respectively.

(a) What is the probability that the plant will stay on line?
Let Fail = 0 and OK = 1. Then

Case	A	B	C	*Plant* P	*Probabilities*
0	0	0	0	0	$(0.01)(0.03)(0.02)$
1	0	0	1	0	$(0.01)(0.03)(1-0.02)$
2	0	1	0	0	$(0.01)(1-0.03)(0.02)$
3	0	1	1	1	$(0.01)(1-0.03)(1-0.02)$
4	1	0	0	1	$(1-0.01)(0.03)(0.02)$
5	1	0	1	1	$(1-0.01)(0.03)(1-0.02)$
6	1	1	0	1	$(1-0.01)(1-0.03)(0.02)$
7	1	1	1	1	$(1-0.01)(1-0.03)(1-0.02)$

$$\begin{aligned}
P\{\text{plant shut down}\} &\stackrel{?}{=} (0.01)(0.03)(0.02) + (0.01)(0.03)(1-0.02) \\
&\quad +(0.01)(1-0.03)(0.02) \\
&= 0.000494 \\
P\{\text{plant on line}\} &= 1 - P\{P = \text{“0”}\} = 1 - 0.000494 \\
\text{or} &= P\{\text{Cases 3 through 7}\} \\
&= 0.999506 \quad \text{(very high)}
\end{aligned}$$

(b) What is the probability that the plant stays on line given that A failed?

$$\begin{aligned}
P\{\text{plant on line}|A \text{ failed}\} &= P\{P = \text{“1”}|A = \text{“0”}\} \\
&= \frac{P\{(P = \text{“1”}) \cap (A = \text{“0”})\}}{P\{A = \text{“0”}\}} \\
&= \frac{(0.01)(1-0.03)(1-0.02)}{0.01} \\
&= 0.9506 \quad \text{(still not low)}
\end{aligned}$$

2.5 Total Probability Theorem and Bayes' Rule

2.5.1 Total Probability Theorem

A set of events $A_1, A_2, \cdots, A_n$ is said to be

- ***mutually exclusive*** (or ***disjoint***) if $A_i \cap A_j = \emptyset, \forall i \neq j$, meaning that *at most one* event can occur (if one occurs then any other cannot occur).
- (collectively) ***exhaustive*** if $A_1 \cup A_2 \cup \cdots \cup A_n = S$, meaning that *at least one* of the events will occur.
- a ***partition*** of sample space S if *one and only one* of the events will occur. Symbolically,

$$\text{partition} = \text{mutually exclusive} + \text{exhaustive}$$

Clearly, the probabilities of the member events of a partition $A_1, A_2, \cdots, A_n$ sum up to unity:

$$P\{A_1\} + \cdots + P\{A_n\} = P\{A_1 \uplus A_2 \uplus \cdots \uplus A_n\} = P\{S\} = 1 \quad (2.21)$$

Consider an event B in S and a partition $A_1, A_2, \cdots, A_n$ of S. Clearly,

$$B = B \cap S = B \cap (A_1 \uplus \cdots \uplus A_n) = (B \cap A_1) \uplus \cdots \uplus (B \cap A_n) \quad (2.22)$$

Since $(B \cap A_1), \ldots, (B \cap A_n)$ are mutually exclusive, we have

$$\begin{aligned} P\{B\} &= P\{(B \cap A_1) \uplus (B \cap A_2) \uplus \cdots \uplus (B \cap A_n)\} \\ &\overset{\text{Axiom 3}}{=} P\{B \cap A_1\} + P\{B \cap A_2\} + \cdots + P\{B \cap A_n\} \end{aligned}$$

But, for $P\{A_i\} \neq 0$,

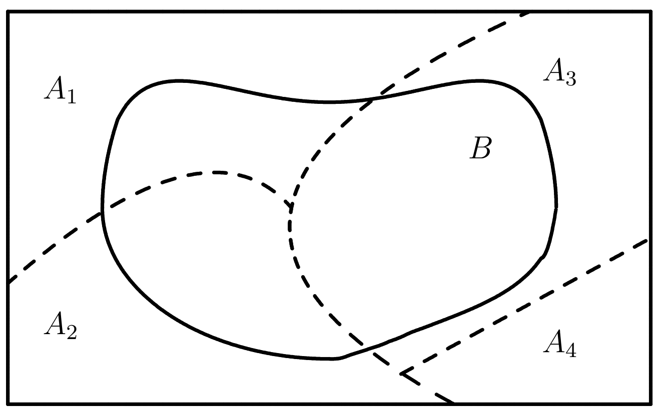
$$P\{B \cap A_i\} \overset{(2.15)}{=} P\{B|A_i\}P\{A_i\}$$

Hence, we have the following result, known as ***total probability theorem***:

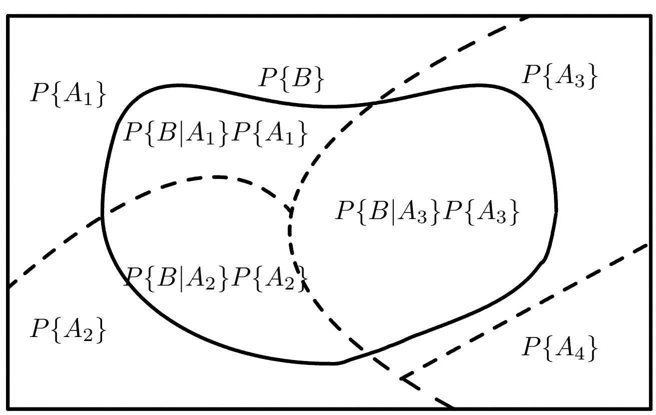
$$\boxed{P\{B\} = P\{B|A_1\}P\{A_1\} + P\{B|A_2\}P\{A_2\} + \cdots + P\{B|A_n\}P\{A_n\}} \quad (2.23)$$

This theorem is valid for any event B and any partition $A_1, A_2, \cdots, A_n$ of the sample space S. It facilitates greatly the calculation of $P\{B\}$ in many situations because both $P\{B|A_i\}$ and $P\{A_i\}$ may be much easier to calculate than a direct calculation of $P\{B\}$.

Total probability theorem is often useful for the calculation of the unconditional probability of an event $P\{B\}$ knowing various conditional probabilities of the events $P\{B|A_i\}$ and the probabilities of the conditioning events $P\{A_i\}$. Intuitively, it provides a way to find an "effect" from its "causes": It calculates the probability of an "effect" (event B) from the probabilities of all its possible "causes" (events A_i's) and the relationships between these possible "causes" and "effect" ($P\{B|A_i\}$).



(a) sample space partitioning and event B



(b) probability decomposition

Figure 2.4: Illustration of total probability theorem.

2.5.2 Bayes' Rule

Since

$$P\{A|B\}P\{B\} \overset{(2.15)}{=} P\{A \cap B\} = P\{B \cap A\} \overset{(2.15)}{=} P\{B|A\}P\{A\}$$

we have the ***Bayes' rule***, ***Bayes' formula***, or ***Bayes' theorem***:

$$\boxed{P\{A|B\} = \frac{P\{B|A\}P\{A\}}{P\{B\}}} \tag{2.24}$$

In particular, for any partition $A_1, A_2, \cdots, A_n$ of the sample space S, we have

$$\begin{aligned} P\{A_i|B\} &= \frac{P\{B|A_i\}P\{A_i\}}{P\{B\}} \\ &\overset{(2.23)}{=} \frac{P\{B|A_i\}P\{A_i\}}{P\{B|A_1\}P\{A_1\} + \cdots + P\{B|A_n\}P\{A_n\}} \end{aligned} \tag{2.25}$$

How Bayes' rule should be interpreted divides statistics into two schools: Bayesian and non-Bayesian. The Bayesian school interprets the various probabilities involved in the Bayes' rule as follows:

$P\{A_i\}$ = ***a priori probability*** of event A_i
= probability of event A_i without knowing event B has occurred
$P\{A_i|B\}$ = ***a posteriori probability*** of event A_i
= probability of event A_i knowing event B has occurred

In this sense, Bayes' rule provides a way of calculating the a posteriori probability by combining the a priori probability with the evidence from the current experiment in which the occurrence of event B has been observed.

Bayes' rule is usually used to find the conditional probability $P\{A|B\}$ of event A given another event B knowing the reversely conditional probability $P\{B|A\}$. Intuitively, it is often used to find a "cause" from the "effect": given an "effect" (event B), it calculates the probability of a particular possible "cause" (event A_i) among all its possible "causes" (events $A_1, \ldots, A_n$).

Total probability theorem and Bayes' rule are two of the most important and useful probability laws.

Example 2.19: Probability of Correct Communication

A binary (with element 0 or 1) digital communication channel has the following error probabilities

$$P\{R_1|S_0\} = 0.1$$
$$P\{R_0|S_1\} = 0.05$$

where

$$S_0 = \{\text{“0” sent}\} \qquad R_0 = \{\text{“0” received}\}$$
$$S_1 = \{\text{“1” sent}\} \qquad R_1 = \{\text{“1” received}\}$$

Since only “0” or “1” can be received, we have

$$P\{R_0|S_0\} = P\{\overline{R_1}|S_0\} = 1 - P\{R_1|S_0\} = 0.9$$
$$P\{R_1|S_1\} = P\{\overline{R_0}|S_1\} = 1 - P\{R_0|S_1\} = 0.95$$

Suppose that it is discovered that “0” is received with probability 0.8 (i.e., $P\{R_0\} = 0.8$).

(a) Determine the probability that “1” is sent: Let $x = P\{S_1\}$. Since $S_0 \uplus S_1 = S$ (sample space), $P\{S_0\} = 1 - P\{S_1\} = 1 - x$. Then by total probability theorem,

$$0.8 = P\{R_0\} \overset{(2.23)}{=} P\{R_0|S_1\}P\{S_1\} + P\{R_0|S_0\}P\{S_0\} = 0.05x + 0.9(1-x)$$

Solving the above equation yields

$$P\{\text{“1” sent}\} = P\{S_1\} = x = \frac{0.1}{0.85} = 0.1176$$

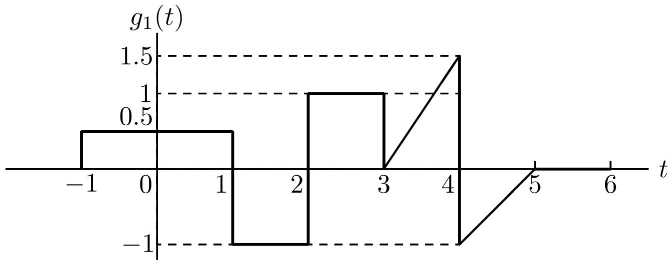
(b) Determine the probability that “1” was sent given that “1” is received:

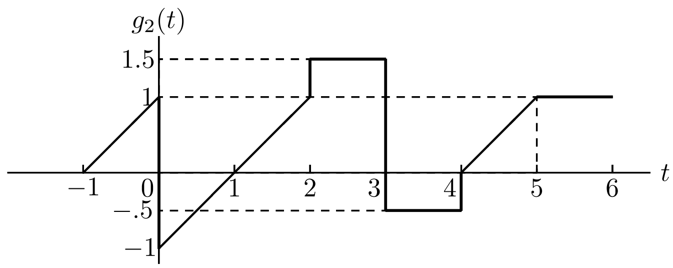
$$P\{S_1|R_1\} \overset{(2.24)}{=} \frac{P\{R_1|S_1\}P\{S_1\}}{P\{R_1\}} = \frac{(0.95)(0.1176)}{0.2} = 0.5588$$

(c) Determine the probability that “0” was sent given that “0” is received:

$$P\{S_0|R_0\} = \frac{P\{R_0|S_0\}P\{S_0\}}{P\{R_0\}} = \frac{(0.9)(1 - 1176)}{0.8} = 0.9926$$

Note that it is much more reliable to transmit “0” for this channel.





Example 2.20: Random Selection and Sampling of Waveform

Consider the waveforms and the following random experiment:

S1. First, select one waveform at random.
S2. Then, sample the selected waveform at a random time τ, $-1 \leq \tau < 6$.

If the sampled value $g(\tau) \geq 0.5$, what is the probability that it was sampled from $g_1(t)$? Let

$$A_i = \{\text{waveform } g_i(t) \text{ is sampled}\}, \qquad i = 1, 2 \Longrightarrow \text{ a partition}$$
$$B = \{\text{sampled value } g(\tau) \geq 0.5\}$$

By total probability theorem,

$$\begin{aligned} P\{B\} &= P\{B|A_1\}P\{A_1\} + P\{B|A_2\}P\{A_2\} \\ &= \frac{2+1+2/3}{7} \times \frac{1}{2} + \frac{1/2+1/2+1+1/2+1}{7} \times \frac{1}{2} = 0.5119 \end{aligned}$$

Thus, by Bayes' rule,

$$\begin{aligned} P\{g_1(t) \text{ sampled}|g(\tau) \geq 0.5\} = P\{A_1|B\} &= \frac{P\{B|A_1\}P\{A_1\}}{P\{B\}} \\ &= \frac{(11/21)(1/2)}{0.5119} = 0.5116 \end{aligned}$$

Example 2.21: Random Selection of Capacitor

Given the table below, consider the following random experiment: (1) first select a box; and (2) then choose a capacitor from the box.

	Box #			
Capacitance (μF)	1	2	3	*Total*
0.1	35	25	40	100
0.5	75	95	70	240
1.0	60	10	65	135
Total	170	130	175	475

Assume that the box selection and the capacitor selection are both with equal probability. If a $0.1\mu F$ capacitor is selected, what is the probability that it came from box 3? Let

$$A_i = \{\text{capacitors in box } i\} \quad \Longrightarrow \quad P\{A_i\} = \frac{1}{3}, \quad i = 1, 2, 3$$
$$B = \{0.1\mu F \text{ chosen}\}$$

Then, from the table,

$$P\{B|A_1\} = \frac{35}{170}, \qquad P\{B|A_2\} = \frac{25}{130}, \qquad P\{B|A_3\} = \frac{40}{175}$$

A_1, A_2, A_3 form a partition:

- A capacitor cannot be in both A_i and A_j — mutually exclusive
- A capacitor must be in one of A_i — exhaustive.

Thus, by total probability theorem and Bayes' rule,

$$\begin{aligned} P\{B\} &= P\{B|A_1\}P\{A_1\} + P\{B|A_2\}P\{A_2\} + P\{B|A_3\}P\{A_3\} \\ &= \frac{35}{170}\frac{1}{3} + \frac{25}{130}\frac{1}{3} + \frac{40}{175}\frac{1}{3} = 0.2089 \neq \frac{100}{475} \quad \text{(why?)} \end{aligned}$$

$$P\{\text{box } 3|0.1\mu F\} = P\{A_3|B\} = \frac{P\{B|A_3\}P\{A_3\}}{P\{B\}} = \frac{(40/175)(1/3)}{0.2089} = 0.3647$$

Note that $P\{\text{box } 3|0.1\mu F\} \neq 40/100$. What if we assume (unrealistically for this problem) the following?

$$P\{A_1\} = \frac{170}{475}, \qquad P\{A_2\} = \frac{130}{475}, \qquad P\{A_3\} = \frac{175}{475}$$

See Example 2.31.

2.6 Combined Experiments and Bernoulli Trials

2.6.1 Combined Experiments

Consider the following random experiment:

S1. First roll a die.
S2. Then toss a coin (independent from Step 1).

The sample space of these two individual experiments are, respectively,

$$S_1 = \{1, 2, 3, 4, 5, 6\}$$
$$S_2 = \{H, T\}$$

This is a ***combined experiment*** with sample space

$$\begin{aligned} S &= S_1 \times S_2 \qquad\qquad \text{(Cartesian product of } S_1 \text{ and } S_2\text{)} \\ &= \{(1,H),(1,T),(2,H),(2,T),(3,H),(3,T), \\ &\quad (4,H),(4,T),(5,H),(5,T),(6,H),(6,T)\} \end{aligned}$$

What is the probability that an even number and head H will show up?

$$A = \{(2,H),(4,H),(6,H)\}$$
$$P\{A\} = \frac{3}{12}$$

Alternatively, this can be obtained as follows:

$$\begin{aligned} A_1 &= \{\text{an even number shows up in die rolling}\} \\ A_2 &= \{\text{head shows up in coin tossing}\} \\ P\{A\} &= P\{A_1 \times A_2\} \\ &\stackrel{?}{=} P\{A_1\}P\{A_2\} = \frac{3}{6} \times \frac{1}{2} \\ &= \frac{3}{12} \end{aligned}$$

The combined experiment of more than two experiments can be handled similarly.

2.6.2 Bernoulli Trials

Bernoulli trials are a special random combined experiment for which

- Only two outcomes (A and $\overline{A}$) are possible on any single trial.
- Repeated trials are independent (from trial to trial).

A typical problem is: What is the probability that A occurs exactly k times out of N (independent) trials? Let

$$\begin{aligned} p &= P\{A\} \text{ (on a single trial)} \\ q &= P\{\overline{A}\} = 1 - p \text{ (on a single trial)} \\ B &= \{A \text{ occurs exactly } k \text{ times out of } N \text{ trials}\} \end{aligned}$$

Consider events

$$B_1 = \{\underbrace{A, A, \ldots, A}_{k \text{ times}}, \underbrace{\overline{A}, \overline{A}, \ldots, \overline{A}}_{(N-k) \text{ times}}\} = \{A \text{ occurs on and only on each of first } k \text{ trials}\}$$

$$B_2 = \{\underbrace{A, A, \ldots, A}_{k-1 \text{ times}}, \overline{A}, A, \underbrace{\overline{A}, \overline{A}, \ldots, \overline{A}}_{(N-k-1) \text{ times}}\}$$

$$\vdots$$

$$B_M = \{\underbrace{\overline{A}, \overline{A}, \ldots, \overline{A}}_{(N-k) \text{ times}}, \underbrace{A, A, \ldots, A}_{k \text{ times}}\}$$

where M is the number of distinct orders for choosing k out of N trials:

$$M = \binom{N}{k} = \frac{N!}{k!(N-k)!} = \textbf{\textit{binomial coefficient}} \text{ of } k \text{ out of } N \qquad (2.26)$$

All these events have identical probability, given by, for $i = 1, \ldots, M$,

$$P\{B_i\} = P\{A\} \cdots P\{A\} P\{\overline{A}\} \cdots P\{\overline{A}\} = p^k (1-p)^{N-k} = p^k q^{N-k}$$

Since such sequences consist of mutually exclusive events, we have finally

$$\begin{aligned} P\{B\} &= P\{A \text{ occurs exactly } k \text{ times in } \textit{any order} \text{ out of } N \text{ trials}\} \\ &= P\{B_1 \uplus B_2 \uplus \cdots \uplus B_M\} = \sum_{i=1}^{M} P\{B_i\} = M P\{B_i\} = \binom{N}{k} p^k q^{N-k} \end{aligned}$$

or

$$\boxed{P\{A \text{ occurs exactly } k \text{ times in } N \text{ trials}\} = \tbinom{N}{k} [P\{A\}]^k [P\{\overline{A}\}]^{N-k}} \qquad (2.27)$$

Example 2.23: Repeated Die Rolling

A die will be rolled 5 times.

(a) What is the probability that "3" will show up exactly twice?
First identify for (2.27) that $k = 2$, $N = 5$, and

$$p = P\{A\} = \frac{1}{6} \qquad q = P\{\overline{A}\} = \frac{5}{6}$$

Then

$$\begin{aligned} P\{\text{“3” shows up exactly twice in 5 trials}\} &= \binom{5}{2} p^2 q^{5-2} = \frac{5!}{2!3!}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^3 \\ &= 0.16075 \end{aligned}$$

(b) What is the probability that "4" will show up at least twice?

$$\begin{aligned} P\{\text{“4” shows up at least twice}\} &= 1 - P\{\text{“4” does not show up}\} \\ &\quad - P\{\text{“4” shows up once}\} \\ &= 1 - \binom{5}{0} p^0 q^{5-0} - \binom{5}{1} p^1 q^{5-1} \\ &= 1 - \frac{5!}{0!5!}\left(\frac{1}{6}\right)^0\left(\frac{5}{6}\right)^5 - \frac{5!}{1!4!}\left(\frac{1}{6}\right)\left(\frac{5}{6}\right)^4 \\ &= 0.1962 \end{aligned}$$

Alternatively,

$$\begin{aligned} P\{\text{“4” shows up at least twice}\} &= P\{\text{“4” shows up twice}\} \\ &\quad + P\{\text{“4” shows up 3 times}\} \\ &\quad + P\{\text{“4” shows up 4 times}\} \\ &\quad + P\{\text{“4” shows up 5 times}\} \\ &= \binom{5}{2} p^2 q^3 + \binom{5}{3} p^3 q^2 + \binom{5}{4} p^4 q + \binom{5}{5} p^5 \\ &= 0.1962 \end{aligned}$$

(c) What is the probability that "4" shows up at least 5 times?

$$\begin{aligned} P\{\text{“4” shows up at least 5 times}\} &= P\{\text{“4” shows up 5 times}\} = \binom{5}{5} p^5 \\ &= 0.0001286 \end{aligned}$$

Example 2.24: Probability of Typos

A typist makes an error while typing a letter 0.3% of the time. There are two types of errors. Type A and B errors occur 80% and 20% of the time, respectively, whenever an error occurs.

(a) What is the probability of no error in 10 letters? Let

$$E_i^N = \{i \text{ errors in } N \text{ letters}\} \quad \Longrightarrow \quad P\{E_1^1\} = 0.003$$
$$A_i^N = \{i \text{ type A errors in } N \text{ letters}\} \quad \Longrightarrow \quad P\{A_1^1|E_1^1\} = 0.80$$

Then, from (2.27),

$$P\{E_0^{10}\} = P\{\text{no error in 10 letters}\} = \binom{10}{0}(P\{E_1^1\})^0(1-P\{E_1^1\})^{10}$$
$$= \binom{10}{0}(0.003)^0(1-0.003)^{10} = 0.9704$$

(b) What is the probability of no type A error in 10 letters? By (2.23),

$$P\{A_1^1\} = P\{A_1^1|E_1^1\}P\{E_1^1\} + \underbrace{P\{A_1^1|\overline{E_1^1}\}}_{=0 \text{ (why?)}} P\{\overline{E_1^1}\} = (0.8)(0.003) + 0 = 0.0024$$

Then,

$$P\{A_0^{10}\} = P\{\text{no type A error in 10 letters}\} = \binom{10}{0}(P\{A_1^1\})^0(1-P\{A_1^1\})^{10}$$
$$= \binom{10}{0}(0.0024)^0(1-0.0024)^{10} = 0.9763 > P\{E_0^{10}\} \quad \text{(why?)}$$

(c) Given that exactly one error has occurred in 10 letters, what is the probability that it is a type A error? By Bayes' rule,

$$P\{A_1^{10}|E_1^{10}\} = \frac{P\{E_1^{10}|A_1^{10}\}P\{A_1^{10}\}}{P\{E_1^{10}\}} = \frac{P\{A_1^{10}\}}{P\{E_1^{10}\}} \qquad \text{(since } P\{E_1^{10}|A_1^{10}\} = 1\text{)}$$
$$= \frac{\binom{10}{1}(P\{A_1^1\})^1(1-P\{A_1^1\})^9}{\binom{10}{1}(P\{E_1^1\})^1(1-P\{E_1^1\})^9} = \frac{(0.0024)(1-0.0024)^9}{(0.003)(1-0.003)^9} = 0.8043$$

Note that

$$P\{A_1^{10}|E_1^{10}\} = 0.8043 \neq 0.8 = P\{A_1^1|E_1^1\} = P\{\text{a type A error given an error}\}$$

You are invited to provide an explanation (see problems 2.38 and 2.39).