Basic Concepts of Probabilty

Chapter Intended Learning Outcomes:

(i) Realize the usefulness of probability

(ii) Understand basic definitions, rules and axioms of probability

(iii) Able to calculate probability for independent and dependent events

(iv) Able to apply combinatorics in probability computation

Why Study Probability?

Probability is the chance that a given event will occur, e.g., what is the chance of getting the right answer in a multiple-choice question? What is the chance if the bus will arrive within 5 minutes?

Uncertainty surrounds us in everyday life, e.g., uncertainty about the grade of a course before result release, not sure if we can come to school before certain time.

Probability helps measuring and quantifying uncertainty, so that we can make decision under uncertainty, e.g., if you know the price of a stock will rise with a high chance of 90% very soon, you might buy it now.

Probability theory originated from gambling, where people want to get money from dice games, card games, roulette.

Galileo (1564-1642) found that for the game of rolling 3 fair dice, the chance of getting the sum of 10 is greater than that of getting 9, although both of them can be obtained in 6 ways:

10= 6+3+1 = 6+2+2 = 5+4+1 = 5+3+2 = 4+4+2 = 4+3+3
9= 6+2+1 = 5+3+1 = 5+2+2 = 4+4+1 = 4+3+2 = 3+3+3

Why sum of 10 has higher chance?

Famous scientists and mathematicians including Pascal (1623-1662), Laplace (1749-1827) and Bernoulli (1655–1705) laid the foundation for probability development as important branch of mathematics, now with wide applicability beyond gambling.

In radar, a typical task is to determine the presence or absence of an approaching aircraft.

Similarly, sonar is used to detect the presence of an enemy submarine.

Transmit pulse

Received waveform - aircraft present

Received waveform - no aircraft
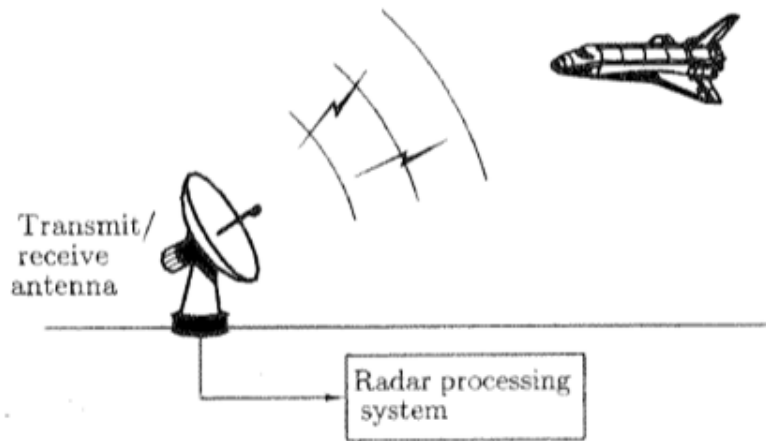
Radar antenna transmits a pulse and then receives a waveform $x(t)$, which may be reflection due to aircraft or noise due to ambient radiation.

Given $x(t)$, we need to decide whether a target is present or not.
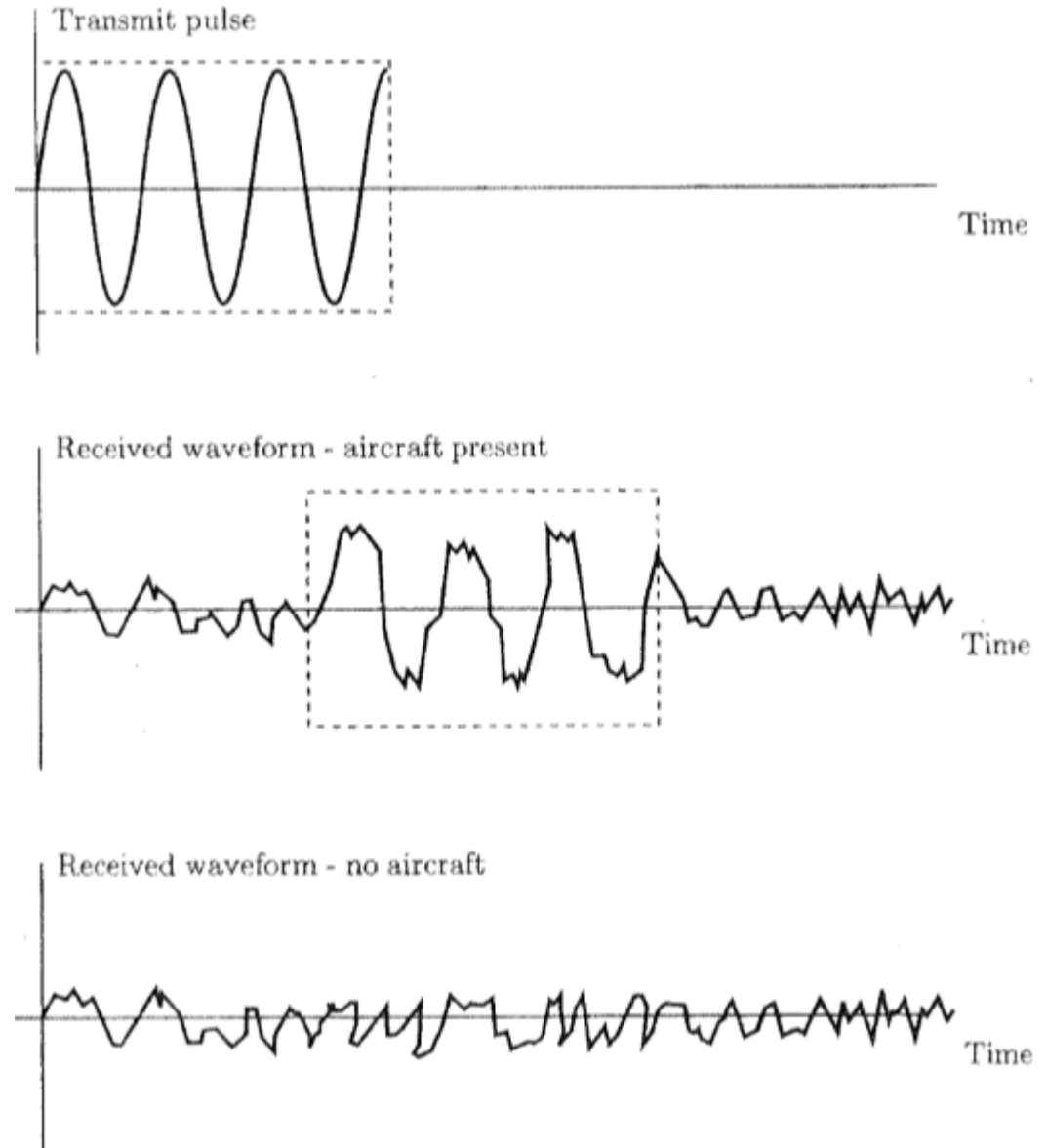
This can be generalized as the problem of signal detection in noise or binary hypothesis testing, i.e., choosing between the null (noise-only) hypothesis $\mathcal{H}_0$ and alternative (signal-present) hypothesis $\mathcal{H}_1$:
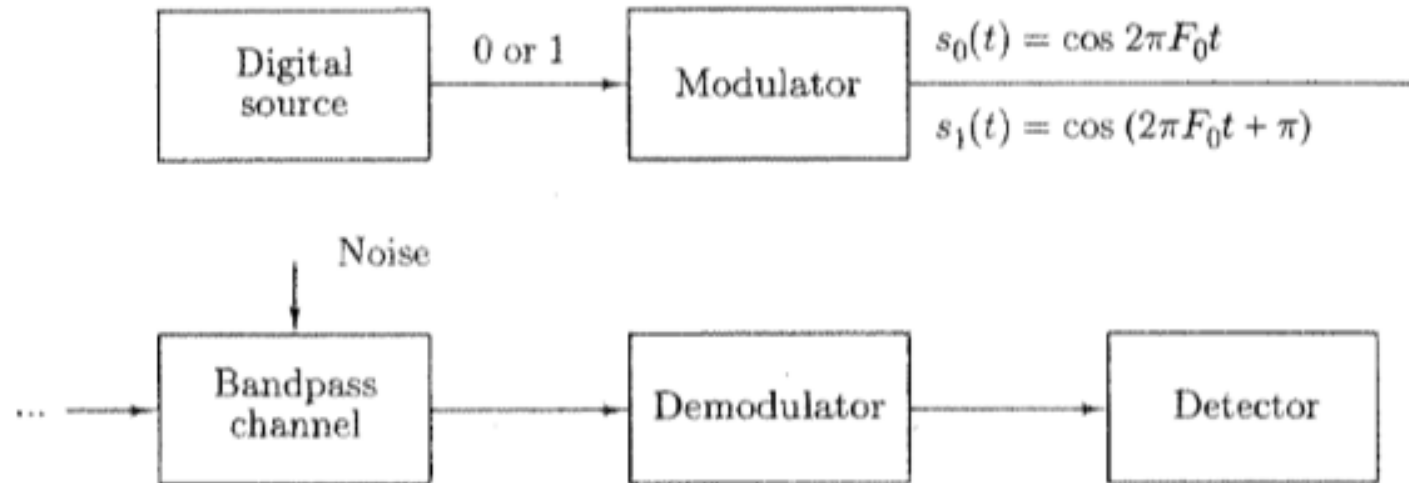
$$\mathcal{H}_0 : x(t) = n(t); \qquad \mathcal{H}_1 : x(t) = s(t) + n(t)$$

where $s(t)$ and $n(t)$ denote the signal and noise, respectively.

We base on the probabilities of $\mathcal{H}_0$ and $\mathcal{H}_1$ to make the decision, but it can be incorrect particularly in strong noise conditions.

How to make use of these two probabilities?

In digital communications, a common application is to determine the digital data sent from transmitter at receiver.

Digital source → (0 or 1) → Modulator →

$$s_0(t) = \cos 2\pi F_0 t$$
$$s_1(t) = \cos (2\pi F_0 t + \pi)$$

Noise

··· → Bandpass channel → Demodulator → Detector

Received waveform at detector input - 1 sent — Time

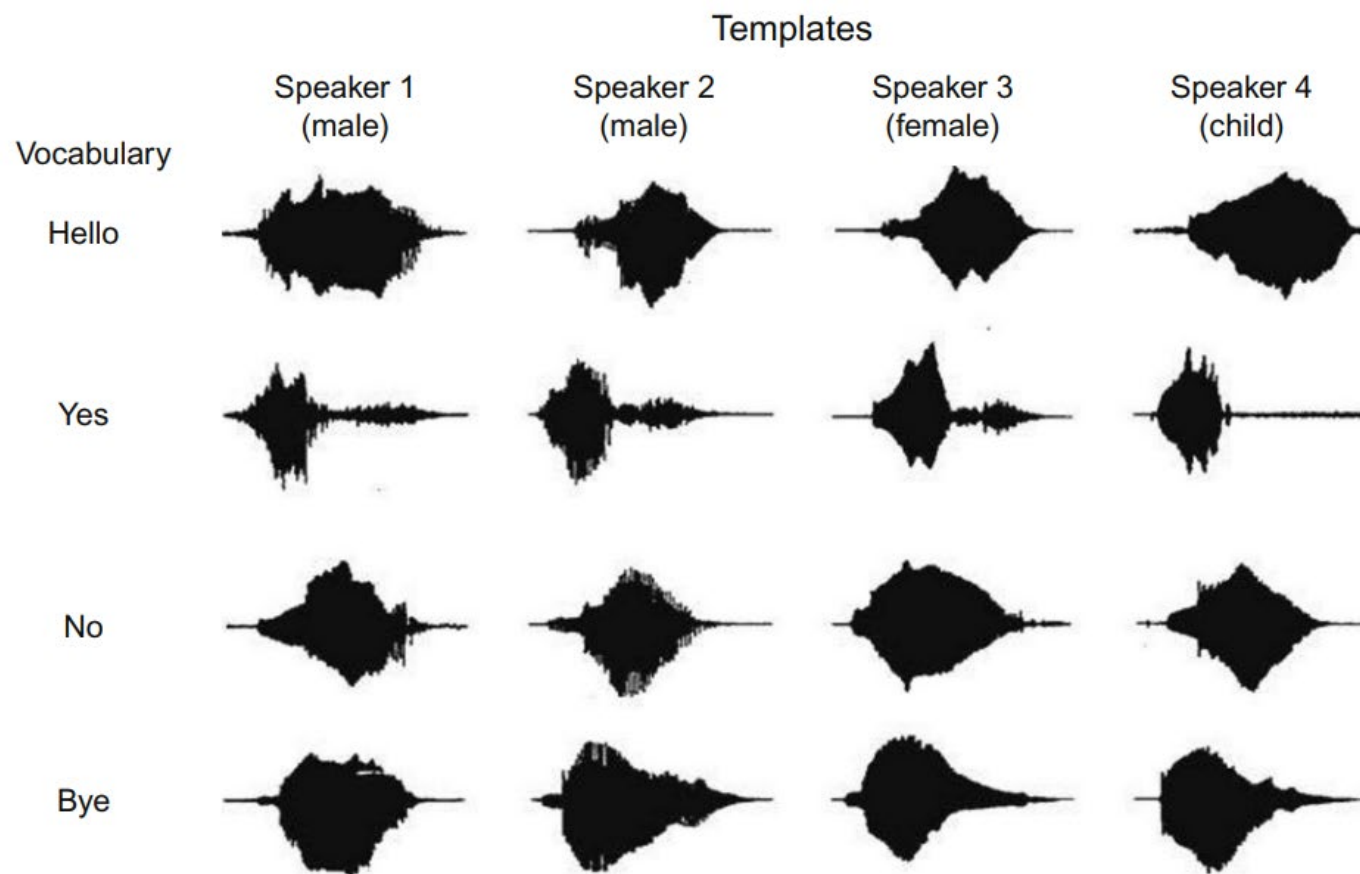Received waveform at detector input - 0 sent — Time

Data sequence with bits "0" and "1" are modulated as $s_0(t)$ and $s_1(t)$ during transmission.

At the receiver, after demodulation, probability theory can help deciding whether "0" or "1" is sent in an optimum sense.
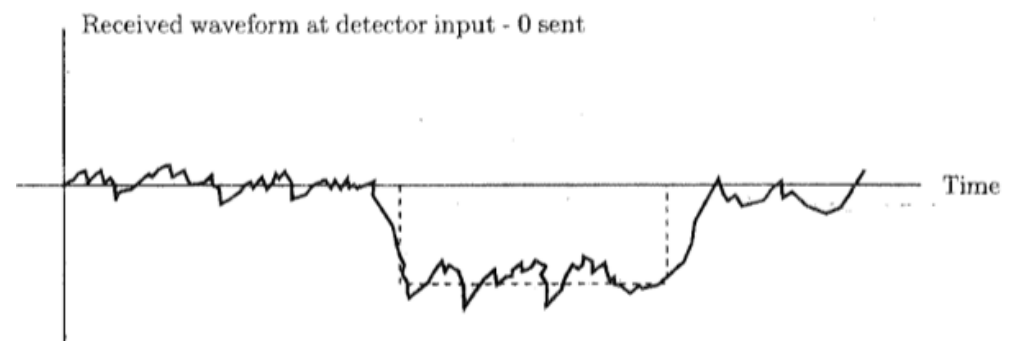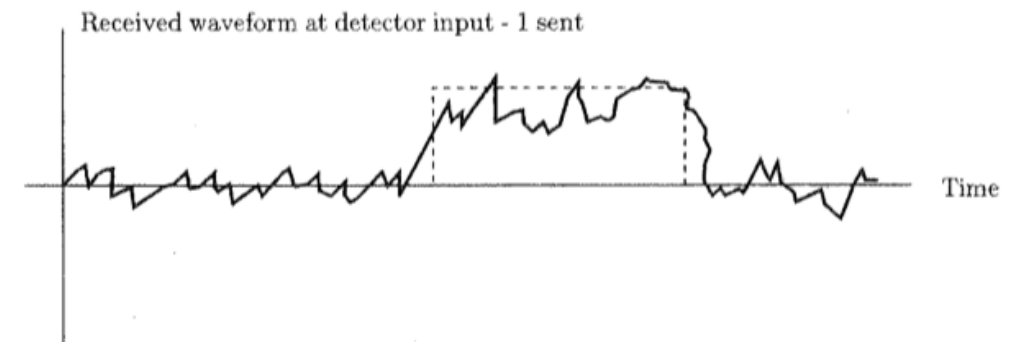
In speech processing, a simple recognition application is to decide the vocabulary from a finite set based on a voice segment.

Another task might be deciding the speaker from a group of speakers. Probability helps in both cases.

Templates

Vocabulary	Speaker 1 (male)	Speaker 2 (male)	Speaker 3 (female)	Speaker 4 (child)
Hello				
Yes				
No				
Bye				

In communication network, arrival of packets of information to a node is random. Hence each node has buffering capability to temporarily store packets when it cannot forward immediately. Probability theory can help determining how large a buffer is needed to ensure a minimal probability of buffer overflow or loss of information.
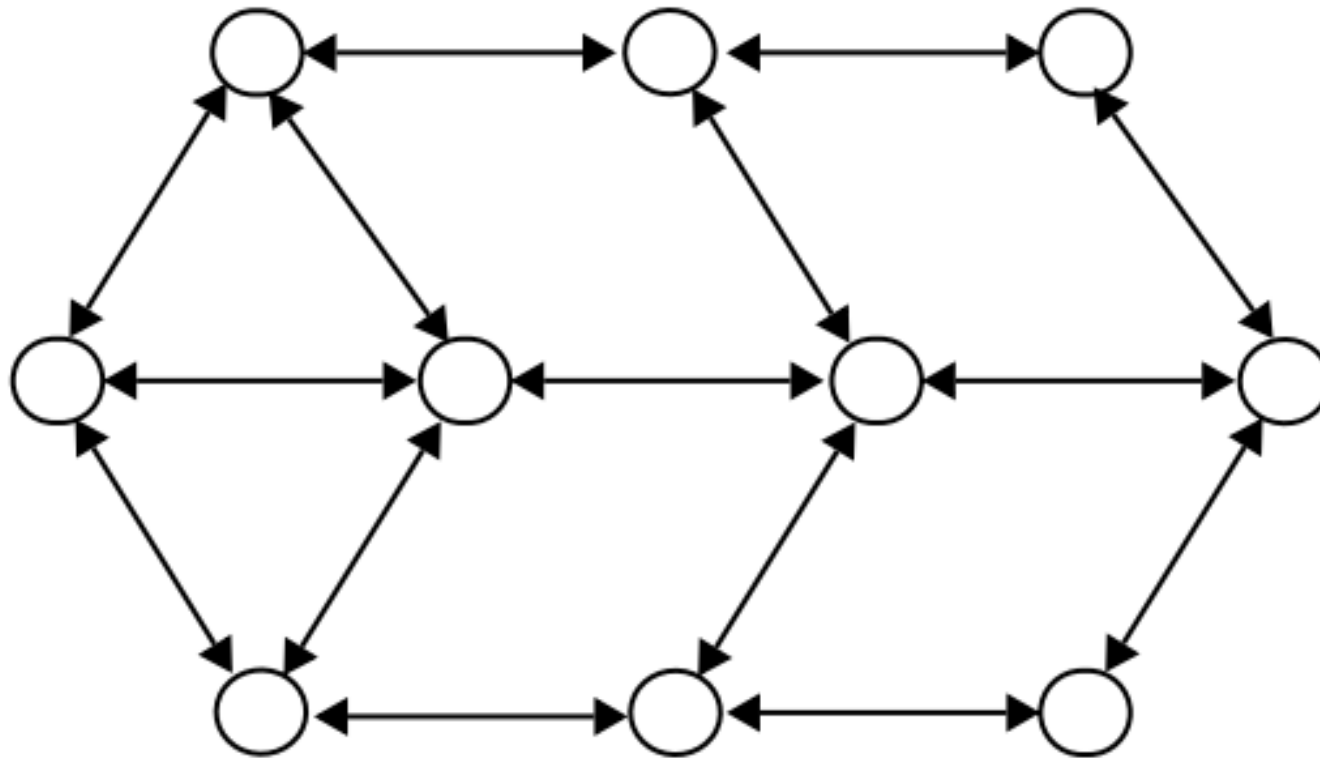
Basic Definitions in Probability

Experiment: Procedure/process/test that produces some result, e.g.,

- Roll a dice and record the number
- Draw a card from a standard deck and record its suit

Outcome: A possible result of an experiment, e.g.,

- For dice rolling, a possible result is "1"
- For card drawing, a possible result is "spade ♠"

Sample Space: Collection or set of all possible distinct outcomes of an experiment, e.g.,

- For dice rolling, the set is {1, 2, 3, 4, 5, 6}
- For card drawing, the set is {♠, ♥, ♦, ♣}

Example 1.1

Consider the following two rating scales. User needs to place a "X" for his choice. Determine their sample spaces.

1   2   3   4   5

☐   ☐   ☐   ☐   ☐

Instructions: for each pair of adjectives, place a cross at the point between them that reflects the extent to which you believe the adjectives describe the home page. You should place *only one cross* between the marks on each line.

Attractive ⊢——┴——┴——┴——┴——┴——┴——┤ Ugly

The first sample space is {1,2,3,4,5} which is discrete.

For the second, any point between "Attractive" and "Ugly" can be an outcome, which corresponds to a continuous sample space.

Example 1.2

Consider the experiment of flipping a coin until a tail occurs. Find its sample space.

Denote head and tail as H and T, respectively, possible outcomes are:

x_1: T
x_2: H,T
x_3: H,H,T
...
x_n: H,..., H,T (all H in the first $(n-1)$ experiments)

In theory, it is possible that $n \to \infty$.

We say that this sample space is countably infinite, but still referred to as discrete.

Event: Certain set of outcomes of an experiment. Hence, events are subsets of the sample space, e.g.,

- For dice rolling, event with odd number: {1, 3, 5}
- For card drawing, event with red suit {♥, ♦}

Let $P(A)$ be the probability that event A occurs, we can easily guess that if A corresponds to odd number outcome in dice rolling or red suit outcome in card drawing, $P(A) = 0.5$.

Note that a sample space of N possible outcomes yields 2^N possible events as each outcome can be included or excluded in the event.

<u>Example 1.3</u>

Consider the sample space of tossing a fair coin. There are only 2 possible outcomes, head (H) and tail (T), or {H,T}. Then there are 4 possible events:

{H,T}: the probability is 1
{H}: the probability is 0.5
{T}: the probability is 0.5
{}: the probability of no results should be 0

Set Operations

Define S as the sample space or universal set, and A is an event, we have $A \subset S$, meaning that every element (outcome) x of A is in S, i.e., $x \in A \Rightarrow x \in S$, e.g.,

- For dice rolling, A: {1, 3, 5} and B: {1, 3}, $B \subset A \subset S$
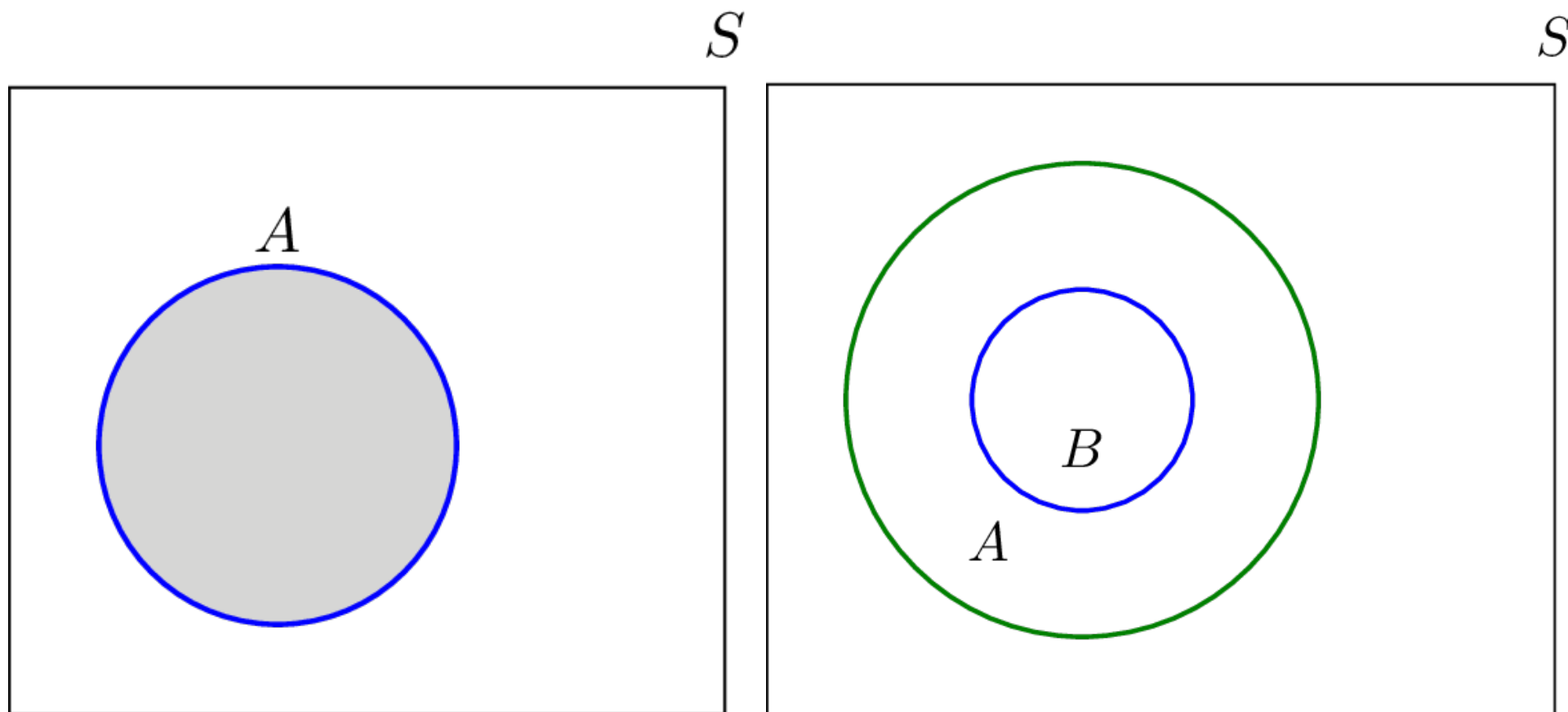
Visualization via Venn diagram: $A \subset S$ and $B \subset A$

The union of two sets (events) A and B, denoted by $A \cup B$, containing all elements (outcomes) that are in A or in B.
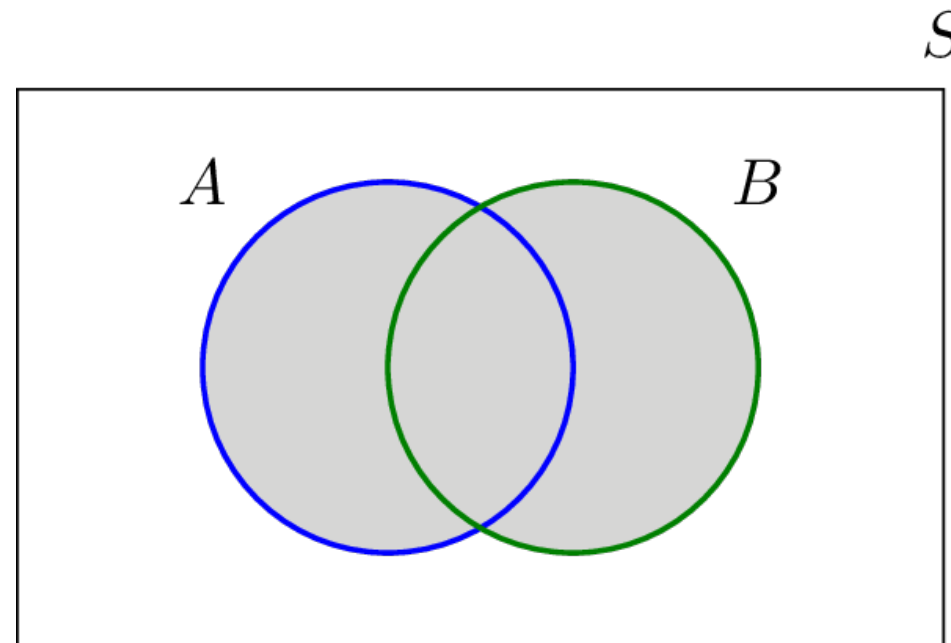
$$A \cup B = B \cup A$$

The union of $A_1, A_2, \cdots, A_N$ can be compactly written as:

$$\bigcup_{n=1}^{N} A_n = A_1 \cup A_2 \cup \cdots \cup A_N$$

The intersection of two sets A and B, denoted by $A \cap B$, containing all elements that are in both A and B.

$$A \cap B = B \cap A$$

$$A \cap B \cap C$$
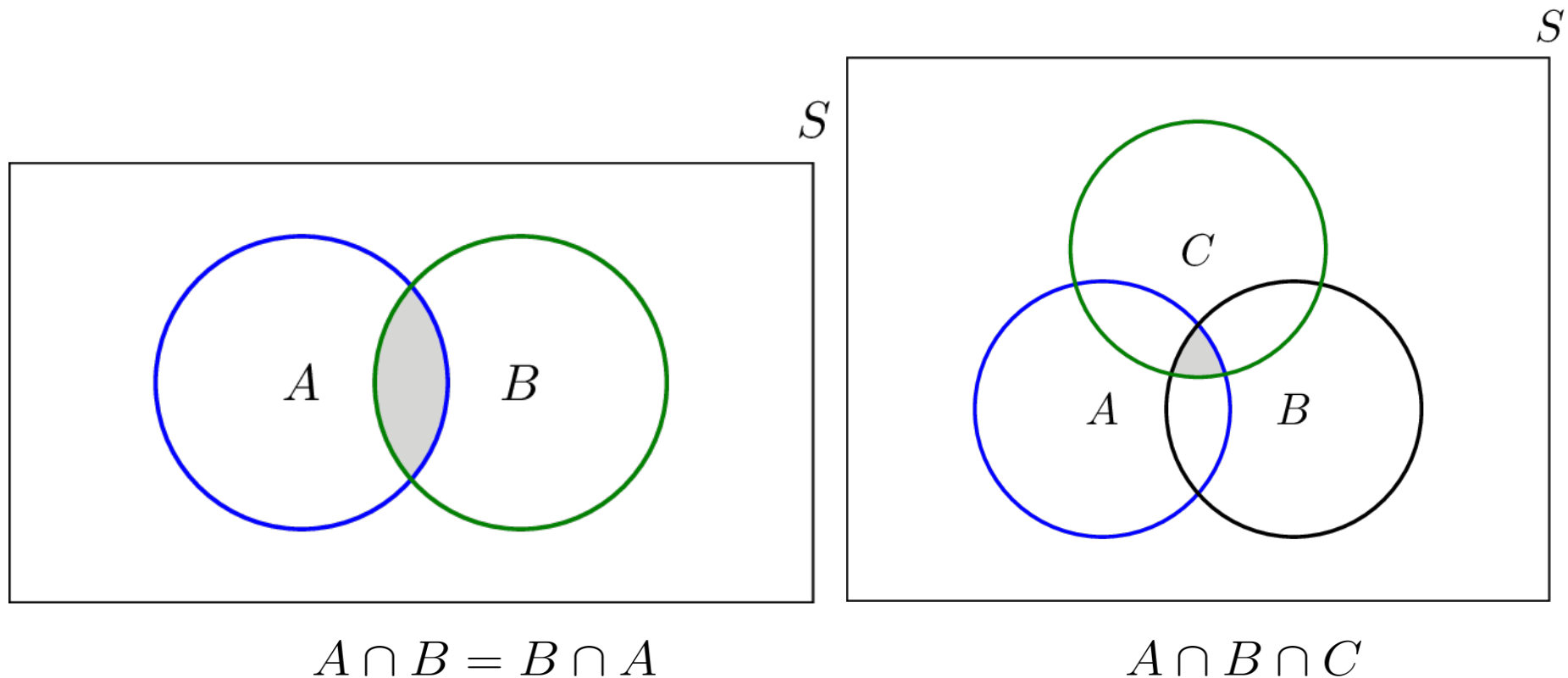
The intersection of $A_1, A_2, \cdots, A_N$ can be compactly written as:

$$\bigcap_{n=1}^{N} A_n = A_1 \cap A_2 \cap \cdots \cap A_N$$

The complement of a set A, denoted by $\overline{A}$, is the set of all elements that are in the universal set S but are not in A.

The difference of sets A and B, denoted by $A \setminus B$, consists of all elements included in A but excluded from B, corresponding to A but not B. Note that $A \setminus B = A \cap \overline{B}$.
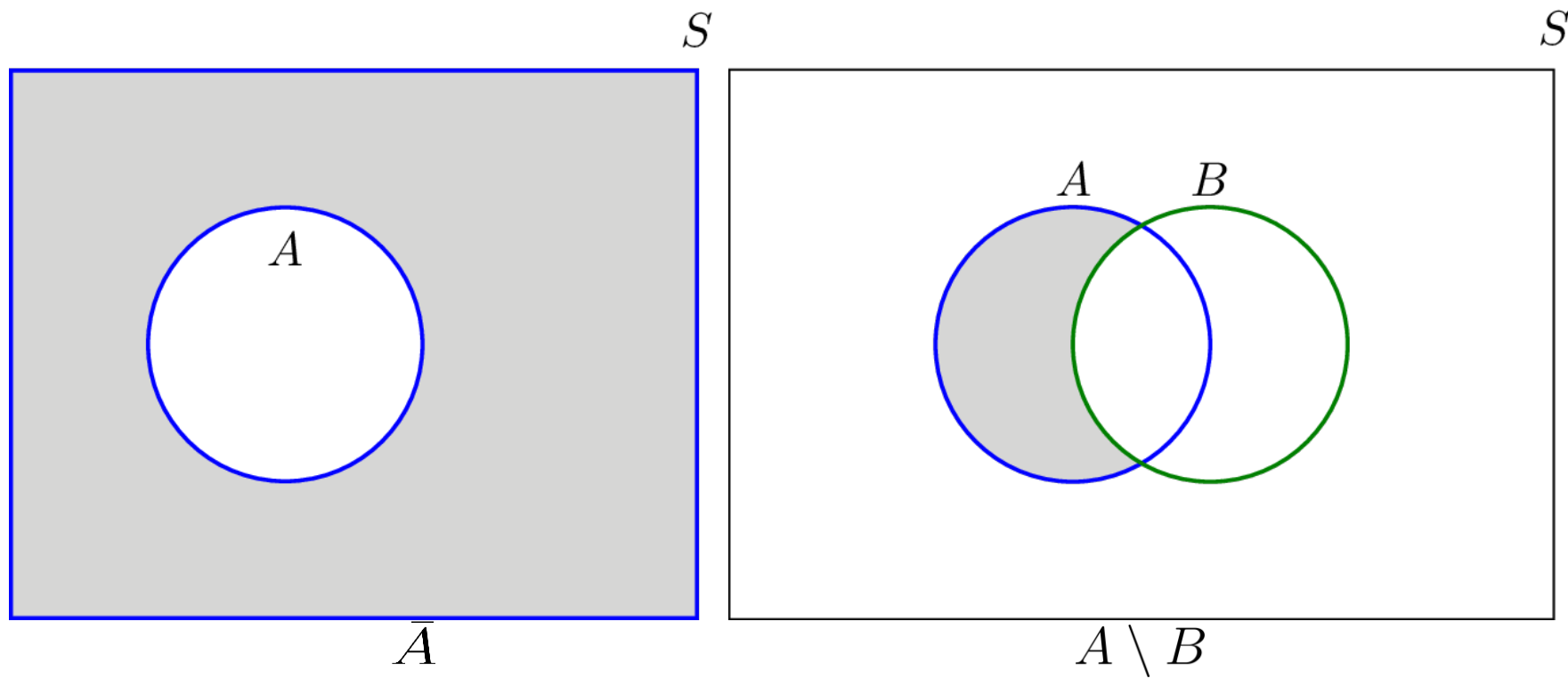
Sets A and B are disjoint if they do not have any shared elements or their intersection is empty, i.e., $A \cap B = \emptyset$. Sets A, B and C are pairwise disjoint or mutually exclusive if any two of these sets are disjoint.
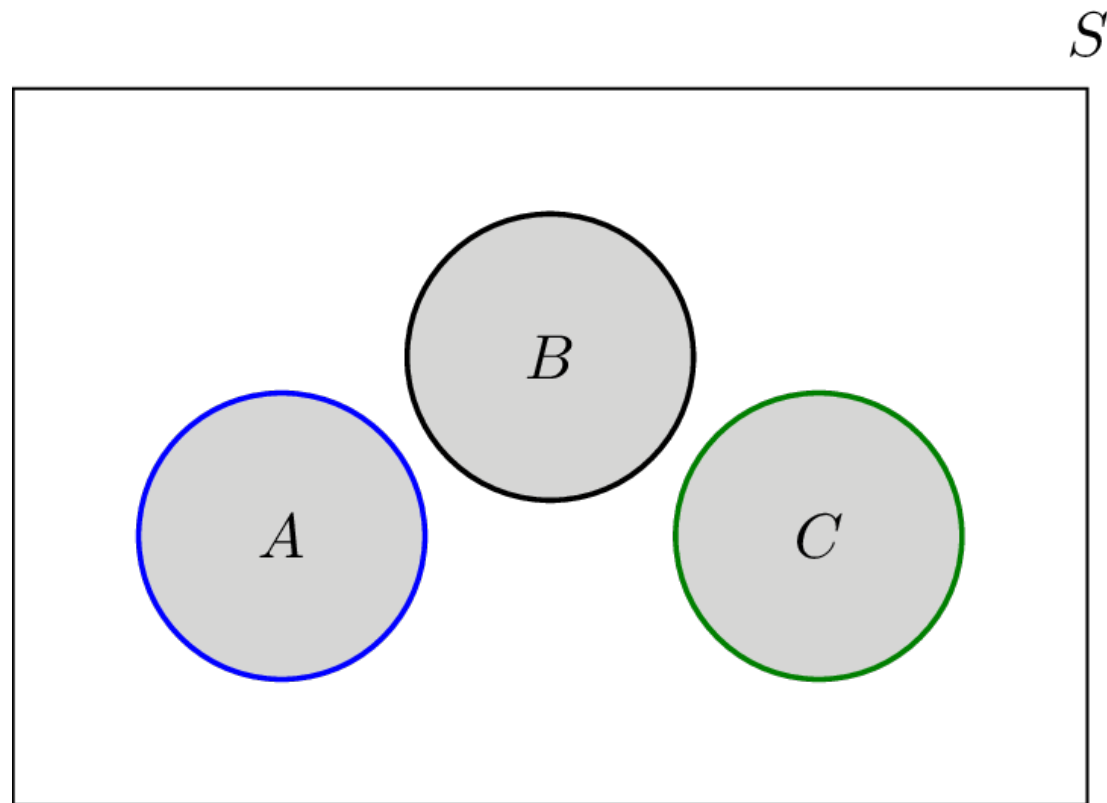
$A \cap B = \emptyset$, $B \cap C = \emptyset$ and $C \cap A = \emptyset$

Taking complement interchanges unions and intersections:

$$\overline{\bigcup_{n=1}^{N} A_n} = \overline{A_1 \cup A_2 \cup \cdots \cup A_N} = \overline{A}_1 \cap \overline{A}_2 \cap \cdots \cap \overline{A}_N = \bigcap_{n=1}^{N} \overline{A}_n \qquad (1.1)$$

$$\overline{\bigcap_{n=1}^{N} A_n} = \overline{A_1 \cap A_2 \cap \cdots \cap A_N} = \overline{A}_1 \cup \overline{A}_2 \cup \cdots \cup \overline{A}_N = \bigcup_{n=1}^{N} \overline{A}_n \qquad (1.2)$$

For any sets A, B and C, we have distributive property:

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C) \qquad (1.3)$$

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C) \qquad (1.4)$$

Example 1.4

Given the sample space S={1,2,3,4,5,6}, events A={1,2}, B ={2,4,5} and C={1,5,6}. We have:

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

$A \cap B$={2}

$\overline{A}$={3,4,5,6}

$\overline{A \cup B}$={3,6}

$\overline{A} \cap \overline{B}$={3,4,5,6}∩{1,3,6}={3,6}

$A \cap (B \cup C)$ ={1,2}∩{1,2,4,5,6}={1,2}

$(A \cap B) \cup (A \cap C)$={2}∪{1}={1,2}

Axioms of Probability

An axiom is a statement which is assumed to be true without the need for proof. There are 3 for probability:

1. Given a sample space S, each event E can be assigned to a nonnegative number $P(E) \geq 0$, as probability of E.

2. The probability of any of the outcomes happening is 1:

$$P(S) = 1$$

That is, in a given experiment, something must happen.

3. If $A_1, A_2, \cdots, A_N$ are pairwise disjoint events, then:

$$P\left(\bigcup_{n=1}^{N} A_n\right) = P(A_1) + P(A_2) + \cdots + P(A_N) = \sum_{n=1}^{N} P(A_n)$$

Based on the three axioms, we have:

$$P(\emptyset) = 0$$

Proof: Since $\emptyset$ has no elements, it follows that $\emptyset$ and S have no elements in common, thus they are disjoint. By Axiom 3:

$$P(\emptyset \cup S) = P(\emptyset) + P(S)$$

Moreover, it is clear that $\emptyset \cup S = S$, resulting in:

$$P(S) = P(\emptyset) + P(S)$$

Applying Axiom 2 that $P(S) = 1$ is finite, the conclusion immediately follows by subtraction.

For any event A, the probability of its complement is:

$$P(\overline{A}) = 1 - P(A)$$

For any event E:

$$0 \leq P(E) \leq 1$$

Proof: Noting that $E \cup \overline{E} = S$ and $E \cap \overline{E} = \emptyset$, and we apply Axioms 2 and 3 to obtain:

$$P(E \cup \overline{E}) = P(S) = 1 = P(E) + P(\overline{E})$$

Axiom 1 already tells us $P(E) \geq 0$ and $P(\overline{E}) \geq 0$. If $P(E) > 1$, if follows that $P(\overline{E}) = 1 - P(E) < 0$, which is impossible. Hence we must have $P(E) \leq 1$, and the proof is complete.

For any events A and B:

$$P(A \setminus B) = P(A \cup \overline{B}) = P(A) - P(A \cap B)$$

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

$$P(B) \leq P(A) \text{ if } B \subset A$$

To find $P(E)$, we need specific information about the experiment and apply the probability axioms.

Example 1.5
A fair dice is rolled. What is the probability of $E=\{1,5\}$?

A fair dice implies that each number has the equal chance of occurrence, namely, 1/6. A formal proof is done as follows.

As the dice is fair, all 6 possible outcomes are equally likely, i.e.,

$$P(\{1\}) = P(\{2\}) = P(\{3\}) = P(\{4\}) = P(\{5\}) = P(\{6\})$$

As all 6 events are disjoint, we apply Axioms 2 and 3 to get:

$$1 = P(S) = P(\{1\} \cup \{2\} \cdots \cup \{6\}) = P(\{1\}) + \cdots + P(\{6\}) = 6P(\{1\})$$

Using the disjoint information, $P(E) = P(\{1\}) + P(\{5\}) = 1/3$

We also refer this probability to as theoretical probability as it relies on pure mathematical reasoning without performing actual experiments.

Experimental probability is calculated based on actual results via repeating the physical experiments a sufficient number of times.

It is also known as relative frequency.

The experimental probability of a certain event is equal to the number of times this event happens divided by the total number of trials.

Instead of performing actual experiments, we can perform computer simulations for "realization" using MATLAB.

Introduction to MATLAB

MATLAB stands for "Matrix Laboratory".

It is an interactive matrix-based software for numerical and symbolic computation in scientific and engineering applications.

Its user interface is simple to use, e.g., we can use the `help` command to understand the usage and syntax of each MATLAB function.

Together with the availability of numerous toolboxes such as "5G", "Bioinformatics", "Deep Learning", "Econometrics", "Financial", "Robotics System", and "Statistics and Machine Learning", it is useful in many disciplines.

MathWorks offers MATLAB to C conversion utility.

As it is not free, it is not as popular as Python. Nevertheless, CityU has subscribed to the MathWorks Total Academic Headcount (TAH) license for MATLAB:

https://www.cityu.edu.hk/csc/deptweb/facilities/central-sw-tah.htm

Individual License for Personally-owned Computers

The TAH license allows faculty, staff, and students to install MATLAB software on their personally-owned computers and laptops. Below are the steps for getting the installer and activating the individual license:

1. Go to CityU's MATLAB Portal.
2. Click "**Sign in to get started**" in the Get MATLAB and Simulink section.
3. Sign in with your CityU EID and password.
4. Create a MathWorks account using your CityU email address.
5. Download the installer from MathWorks website.
6. Run the installer.
7. In the installer, select **Log in with a MathWorks Account** and follow the online instructions.
8. When prompted to do so, select the **Academic – Total Headcount** license labeled **Individual**.
9. Select the products to download and install.
10. After downloading and installing your products, keep the **Activate MATLAB** checkbox selected and click **Next**.
11. When asked to provide a user name, verify that the displayed user name is correct. Continue with the process until activation is complete.

After installation, you can find it at Windows Start icon or click the MATLAB icon at Desktop.

Consider generating $x[n] = A\cos(\omega n + \theta), \quad n = 0, 1, \cdots, N-1$ with $A = 1$, $\omega = 0.3$, $\theta = 1$ and $N = 21$.

MATLAB code:

```
N=21;                    % number of samples is 21
A=1;                     % tone amplitude is 1
w=0.3;                   % frequency is 0.3
p=1;                     % phase is 1
for n=1:N
   x(n)=A*cos(w*(n-1)+p); %time index should be >0
end
```

Note that x is a vector and its index should be at least 1.

An alternative approach is:

```
N=21;                        % number of samples is 21
A=1;                         % tone amplitude is 1
w=0.3;                       % frequency is 0.3
p=1;                         % phase is 1
n=0:N-1;                     % define a vector of size N
x = A.*cos(w.*n+p);  % first time index is also 1
```

Note that "." denotes element-wise operation.

Both give

```
x =
Columns 1 through 7
0.5403    0.2675   -0.0292   -0.3233   -0.5885   -0.8011   -0.9422
Columns 8 through 14
-0.9991   -0.9668   -0.8481   -0.6536   -0.4008   -0.1122    0.1865
Columns 15 through 21
0.4685    0.7087    0.8855    0.9833    0.9932    0.9144    0.7539
```

Which approach is better? Why?

To plot $x[n]$, we can either use the commands `stem(x)` and `plot(x)`.

If the time index is not specified, the default start time is $n = 1$.

Nevertheless, it is easy to include the time index vector in the `stem`/`plot` command, e.g., we can use `stem` to plot $x[n]$ with the correct time index:

```
n=0:N-1;               %n is vector of time index
stem(n,x)              %plot x versus n
```

Similarly, `plot(n,x)` can be employed to show $x[n]$.
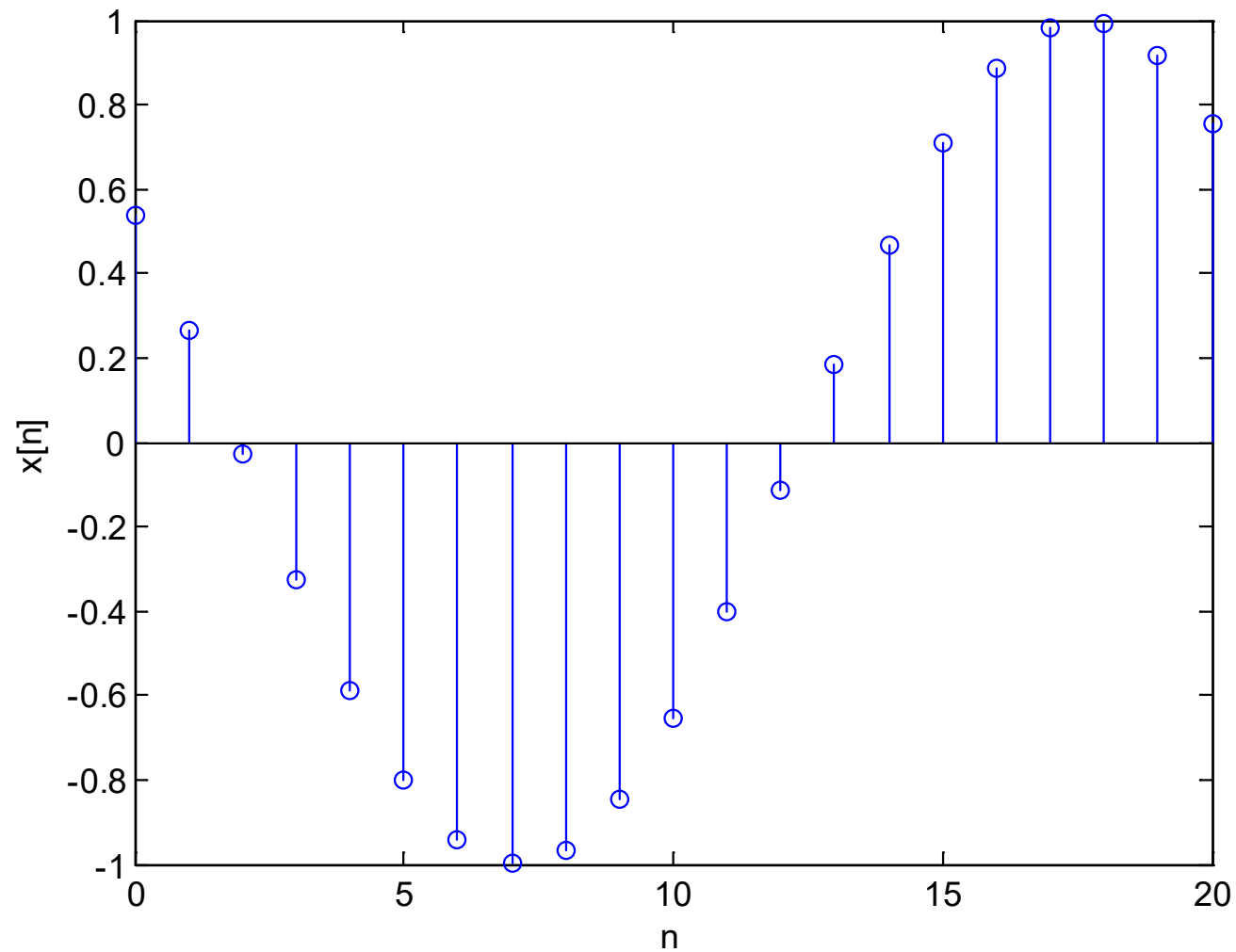
Plot of discrete-time sinusoid using `stem`

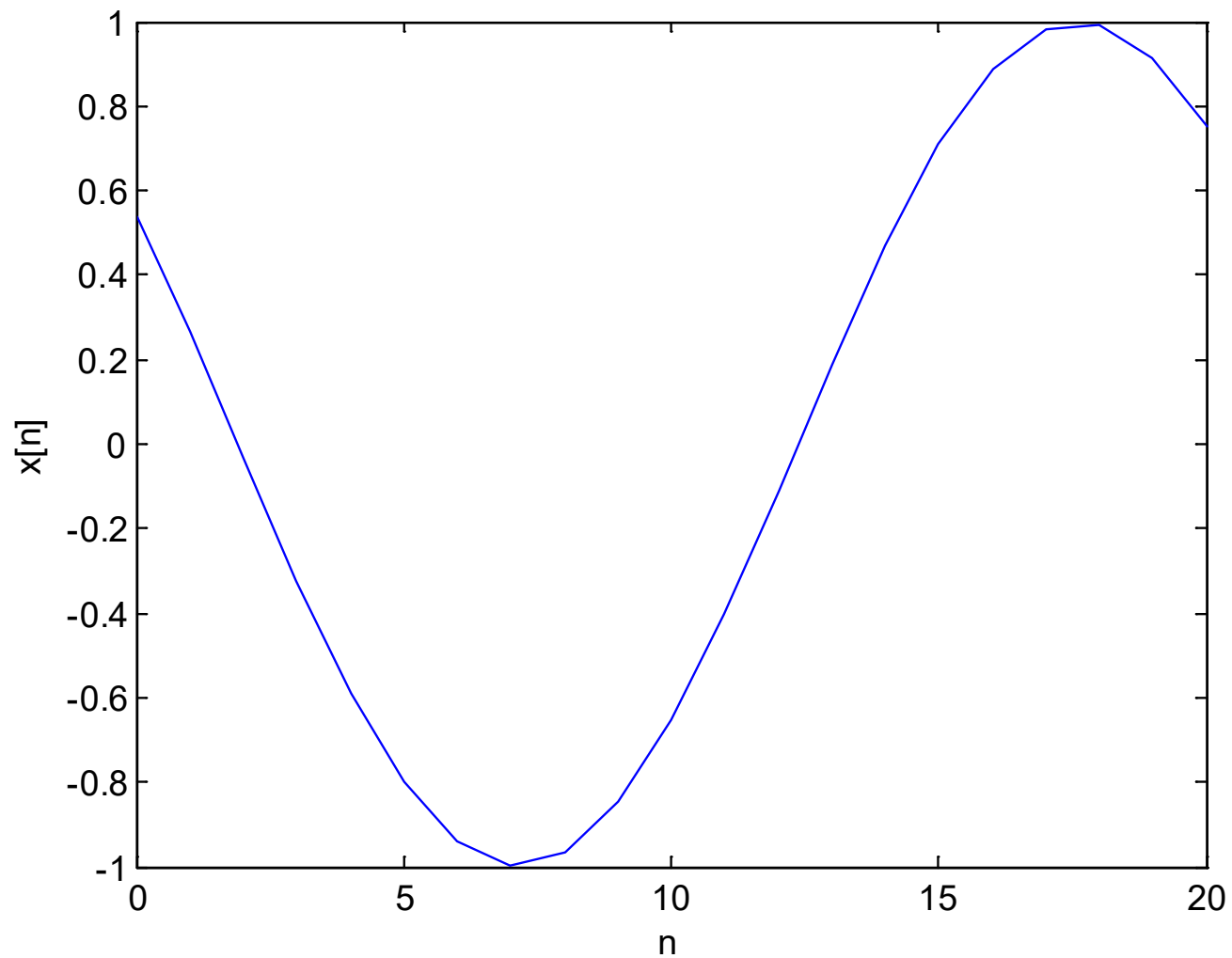Plot of discrete-time sinusoid using `plot`

Example 1.6

Consider the experiment of tossing a fair coin and record the result, which is either head (H) or tail (T). Determine the theoretical probability, and experimental probability of the event $E=\{T\}$ using MATLAB.

Apparently, the theoretical probability is $P(E) = 0.5$.

We can use MATLAB command `randi` to randomly generate 2 integers where one represents H and another represents T.

For example, we can use 0 and 1 to represent H and T, and generate a 100x1 vector to simulate 100 trials:

```
r = randi([0 1],100,1);
stem(r)
p=sum(r)/length(r) = 0.47
```
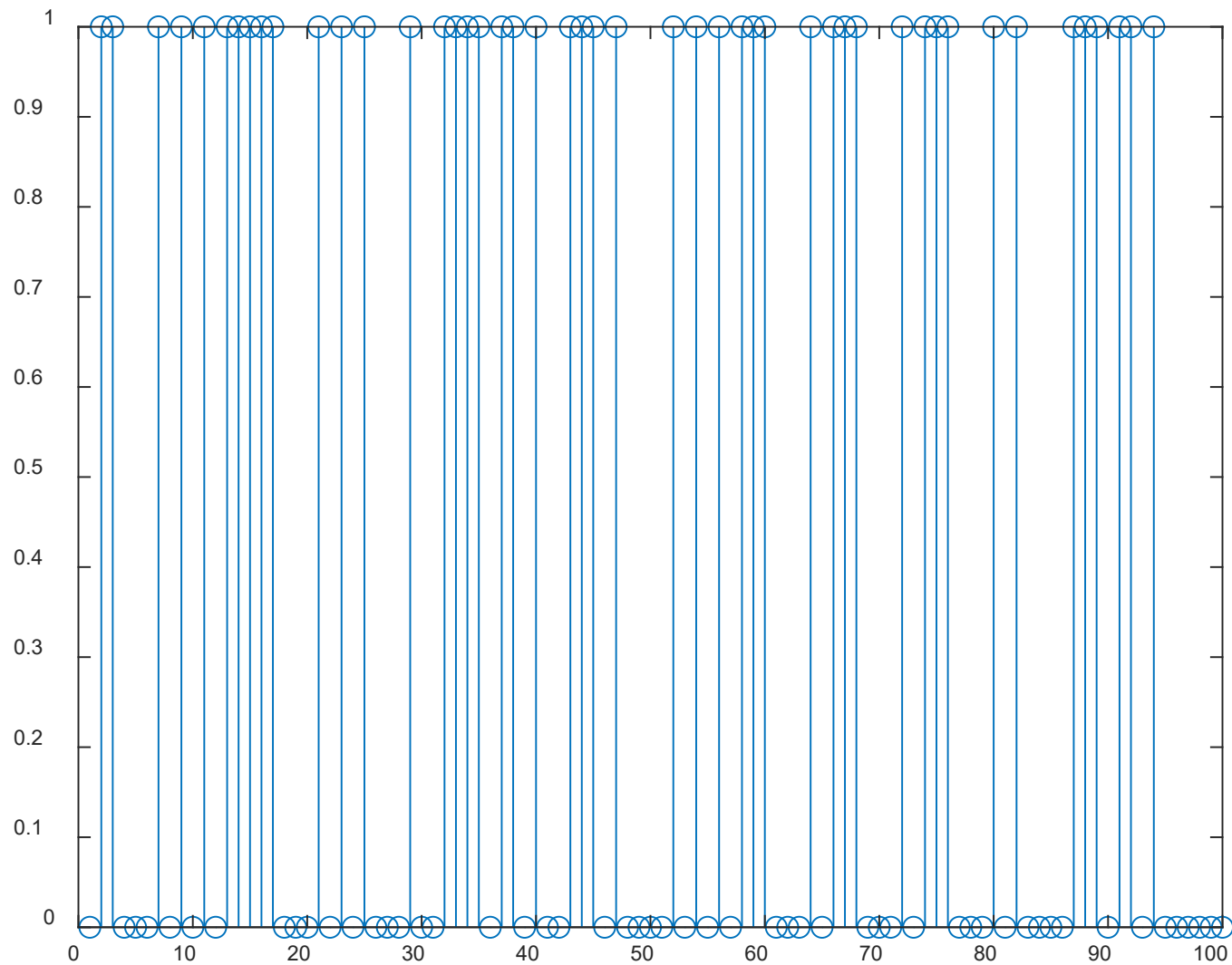
Results of 100 runs

The results produce 53 H and 47 T. Then the experimental probability is $P(E) = 0.47$.

Try 10000 runs:
```
>> r = randi([0 1],10000,1);
>> p=sum(r)/length(r)
p = 0.5033
```

Try 1000000 runs:
```
>> r = randi([0 1],1000000,1);
>> p=sum(r)/length(r)
p = 0.5004
```

The experimental probability is expected to approach the theoretical probability $P(E) \rightarrow 0.5$ when the number of trials is approaching infinity.

Example 1.7

During some construction, a network blackout occurs on Monday with probability 0.7 and on Tuesday with probability 0.5, and on both days with probability of 0.35. Find the probability of having a blackout on Monday or Tuesday.

Applying $P(A \cup B) = P(A) + P(B) - P(A \cap B)$, we easily obtain 0.7+0.5-0.35=0.85.

Alternatively, Venn diagram can be used:
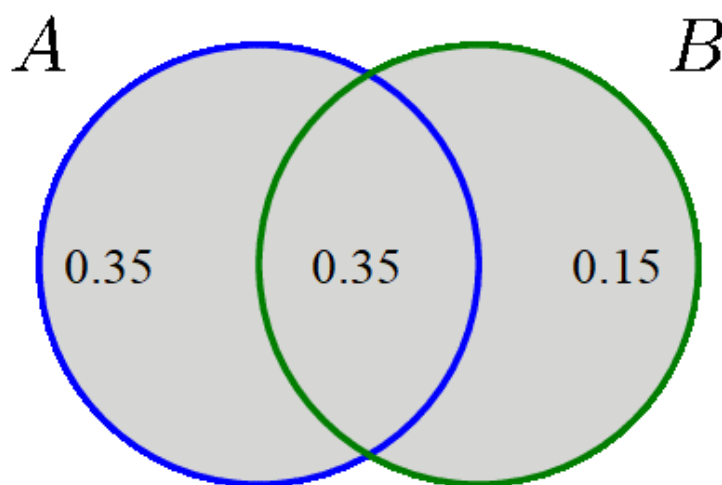
Let A and B be the probabilities that a network blackout occurs on Monday and Tuesday, respectively.

From the Venn diagram, we easily see that the overlapped region of $A \cap B$ cannot be counted twice.

We can perform:

$$P(A \setminus B) = P(A) - P(A \cap B) = 0.7 - 0.35 = 0.35$$

$$P(B \setminus A) = P(B) - P(A \cap B) = 0.5 - 0.35 = 0.15$$

$$P(A \cup B) = P(A \setminus B) + P(B \setminus A) + P(A \cap B) = 0.85$$

Note that the same solution can be obtained using:

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

Independent and Dependent Events

Two events A and B are said to be independent if the occurrence of A does not affect the probability that B occurs:

$$P(A \text{ and } B) = P(A \cap B) = P(AB) = P(A) \cdot P(B) \qquad (1.5)$$

The probability of occurrence of N independent events, $A_1, A_2, \cdots, A_N$ is:

$$P(A_1 \cap A_2 \cap \cdots \cap A_N) = P(A_1)P(A_2) \cdots P(A_N)$$

Example 1.8
Suppose a system with a hard disk having 1% probability to crash and 2 backups, each having 2% probability to crash, and they are independent. The stored information is lost only when all three devices crash. What is the probability that the information is saved or is not lost?

Let H, B_1 and B_2 be the events of hard disk, first and second backups crash, respectively. Then we have:

$$P(H) = 0.01, \quad P(B_1) = P(B_2) = 0.02$$

Applying rules for the complement and intersection of independent events, the required probability is:

$$
\begin{aligned}
P(\text{saved}) &= P(\overline{H} \cup \overline{B_1} \cup \overline{B_2}) \\
&= 1 - P(H \cap B_1 \cap B_2) = 1 - P(H)P(B_1)P(B_2) = 0.999996
\end{aligned}
$$

When the system components are connected in parallel, it is sufficient for at least one to work properly.

This probability is also known as reliability.

What is the probability without backups?

Example 1.9

Suppose that a space shuttle launch depends on 3 devices that operate independently of each other, and malfunction with probabilities 0.01, 0.02, and 0.02. If any of the devices malfunctions, the launch will be postponed. Compute the probability that the shuttle will be launched on time.

Let the malfunction events be A_1, A_2, A_3.

$$
\begin{aligned}
P(\text{on time}) &= P(\text{all devices function}) \\
&= P(\overline{A_1} \cap \overline{A_2} \cap \overline{A_3}) = P(\overline{A_1})P(\overline{A_2})P(\overline{A_3}) \\
&= (1 - P(A_1))(1 - P(A_2))(1 - P(A_3)) \\
&= (1 - 0.01)(1 - 0.02)(1 - 0.02) = 0.950796
\end{aligned}
$$

Comparing with Example 1.8, system reliability decreases when the components are connected in series.

Example 1.10

The following system is functioning properly if at least one of the links works. Compute the system reliability. Assume each component is operable with probability 0.92 independently of the other components.
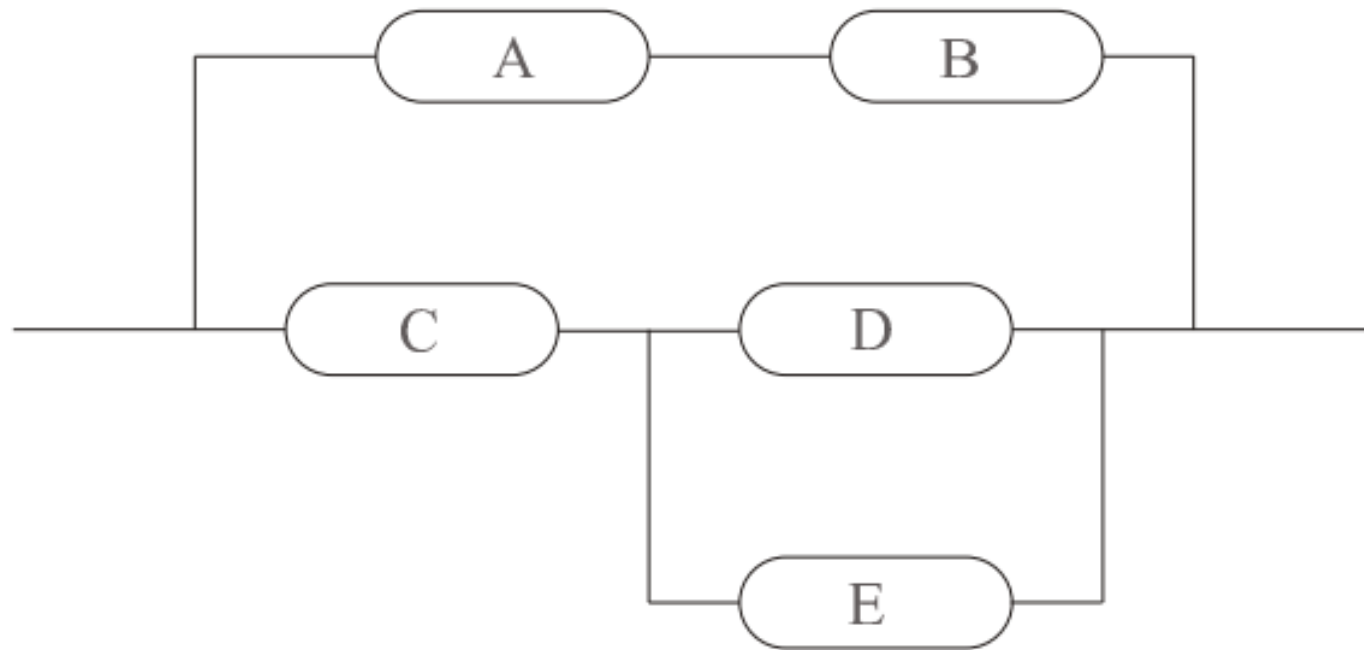
The upper link A-B works has the probability:

$$P(A \cap B) = (0.92)^2 = 0.8464$$

As in Example 1.8, we need at least one of D and E works, and the probability is:

$$P(D \cup E) = 1 - (1 - 0.92)^2 = 0.9936$$

Combining with C, the probability that this combined link works is:

$$P(C \cap (D \cup E)) = 0.92 \cdot 0.9936 = 0.9141$$

Finally, the probability that at least one link works is:

$$P(\text{work}) = 1 - (1 - 0.8464)(1 - 0.9141) = 0.9868$$

where $\text{work} = (A \cap B) \cup (C \cap (D \cup E))$.

On the other hand, if A and B are dependent, i.e., the two events affect each other, then (1.5) will be generalized as:

$$P(A \cap B) = P(A) \cdot P(B|A) = P(B) \cdot P(A|B) \qquad (1.6)$$

where $P(B|A)$ stands for the probability that B given that A occurs ($P(A|B)$ stands for the probability that A given that B occurs), which is called conditional probability because the condition of A occurs is given.

Based on (1.6), the conditional probability $P(B|A)$ is:

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \qquad (1.7)$$

Example 1.11

Consider rolling a fair dice. Let A and B be the events that the outcomes are an odd number, i.e., $A = \{1, 3, 5\}$, and at most 3, i.e., $B = \{1, 2, 3\}$. What are $P(A)$ and $P(B)$? What is $P(A \cap B)$? What is the probability of A given B, $P(A|B)$?

It is clear that:

$$P(A) = P(B) = \frac{1}{2}$$

As $A \cap B = \{1, 3\}$, we have:
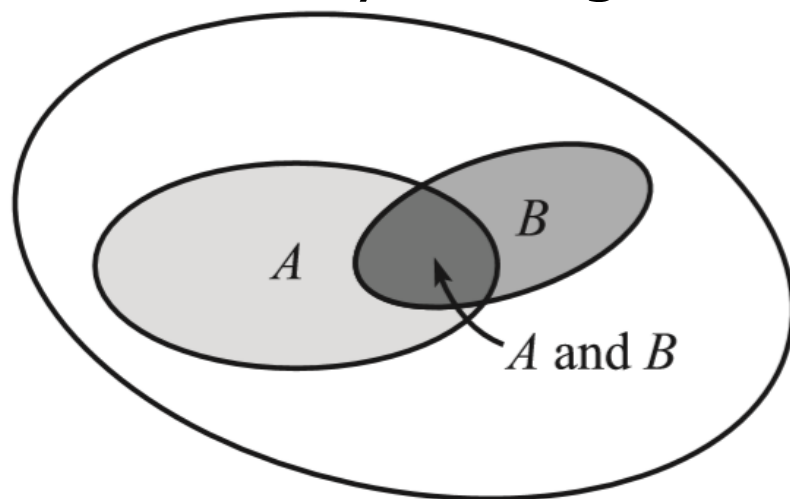
$$P(A \cap B) = \frac{2}{6} = \frac{1}{3}$$

If B occurs, the number can only be 1, 2 or 3, where 2 of them are odd. Hence:

$$P(A|B) = \frac{2}{3}$$

In a more efficient manner, we apply (1.7):

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1}{3} \div \frac{1}{2} = \frac{2}{3}$$

(1.7) can also be illustrated using the Venn diagram: $P(A \cap B)$ and $P(B)$ correspond to the darkly shaded and left shaded areas, respectively, while $P(A|B)$ is a fraction given by the darkly shaded area divided by the right shaded area.

Comparing (1.5) and (1.6), when A and B are independent, then $P(A|B) = P(A)$ and $P(B|A) = P(B)$.

Conditional probability also satisfies probability axioms:

1. For any event A, $P(A|B) \geq 0$

2. Conditional probability of B given B is 1, i.e., $P(B|B) = 1$

3. If $A_1, A_2, \cdots, A_N$ are pairwise disjoint events, then:

$$P\left(\bigcup_{n=1}^{N} A_n | B\right) = P(A_1|B) + P(A_2|B) + \cdots + P(A_N|B) = \sum_{n=1}^{N} P(A_n|B)$$

Example 1.12

90% of flights depart on time, 80% of flights arrive on time, and 75% flights depart and arrive on time.

(a) Given that a flight departed on time, what is the probability that it will arrive on time?

(b) Given that a flight arrived on time, what is the probability that it will depart on time?

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

Denote events A = {arriving on time} and D = {departing on time}, we have $P(A) = 0.8$, $P(D) = 0.9$ and $P(A \cap D) = 0.75$.

(a)
$$P(A|D) = \frac{P(A \cap D)}{P(D)} = \frac{0.75}{0.9} = 0.8333$$

(b)
$$P(D|A) = \frac{P(A \cap D)}{P(A)} = \frac{0.75}{0.8} = 0.9375$$

(c) They are not independent because $P(A|D) \neq P(A)$, $P(D|A) \neq P(D)$ and $P(A \cap D) \neq P(A) \cdot P(D)$.

Given that a flight departed on time, what is the probability that it will not arrive on time?

Applying (1.6) to 3 events as follows:

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

Also,

$$P(B \cap C) = P(B) \cdot P(C|B) \Rightarrow P(B \cap C|A) = P(B|A) \cdot P(C|A \cap B)$$

Hence we have:

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

This chain rule can be generalized as:

$$P\left(\bigcap_{n=1}^{N} A_n\right) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_2 \cap A_1) \cdots P(A_N|A_{N-1} \cap \cdots \cap A_1)$$

Example 1.13

Suppose there are 100 PCs, and 5 of which are malfunctioning. We pick 3 from them at random. What is the probability that all 3 work properly?

Let A_i be the ith chosen unit which works properly, $i = 1, 2, 3$. We start from:

$$P(A_1) = \frac{95}{100}$$

Given that the 1st chosen item was good, the 2nd item will be chosen from 94 good units and 5 defective units, thus:

$$P(A_2|A_1) = \frac{94}{99}$$

Given that the 1st and 2nd items were okay, the 3rd item will be chosen from 93 good units and 5 defective units, thus:

$$P(A_3|A_2 \cap A_1) = \frac{93}{98}$$

Hence:

$$P(A_1 \cap A_2 \cap A_3) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_2 \cap A_1) = 0.856$$

Note that we may be familiar with this product rule but are not aware that it is based on conditional probability.
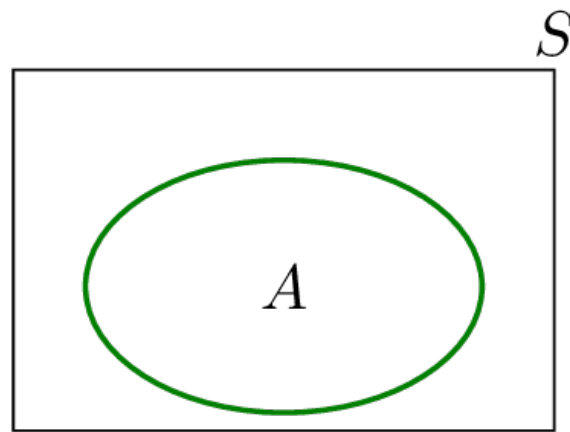
Law of Total Probability

This law relates the unconditional probability $P(A)$ with its conditional probabilities. Consider events B_n, $n = 1, \cdots, N$, with $B_1 \cup B_2 \cup \cdots \cup B_N = S$ and $B_m \cap B_n = \emptyset$, $m \neq n$. We have:

$$A = (A \cap B_1) \cup (A \cap B_2) \cup \cdots \cup (A \cap B_N)$$

where $A \cap B_n$ are also mutually exclusive. Hence:

$$P(A) = \sum_{n=1}^{N} P(A \cap B_n) = \sum_{n=1}^{N} P(B_n) \cdot P(A|B_n) \qquad (1.8)$$
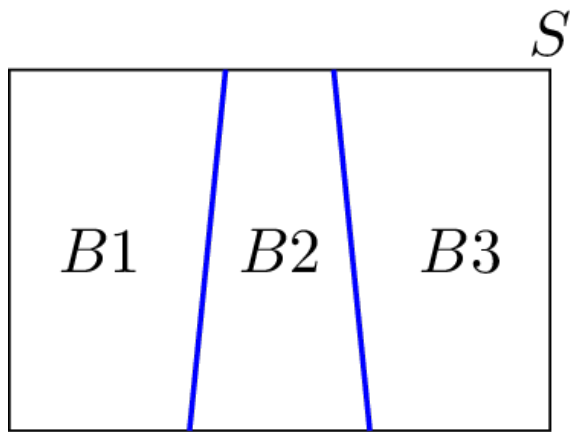
$$A \cap B_1 = A_1$$
$$\Rightarrow P(A \cap B_1) = P(A_1)$$

$$\text{Area} = \frac{A_1}{S} = \frac{B_1}{S} \cdot \frac{A_1}{B_1}$$

$$\frac{B_1}{S} : P(B_1)$$

$$\frac{A_1}{B_1} = \frac{A \cap B_1}{B_1}$$
$$: P(A_1|B_1) = P(A|B_1)$$
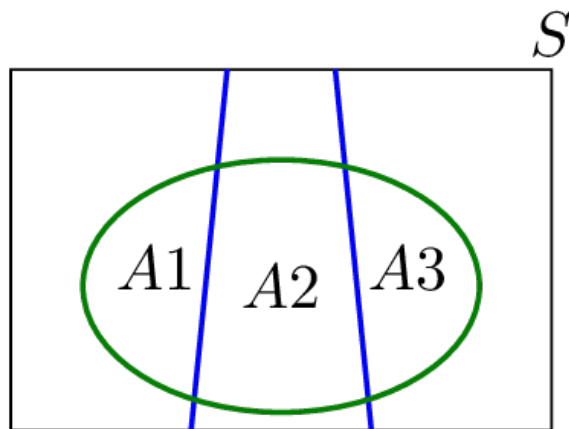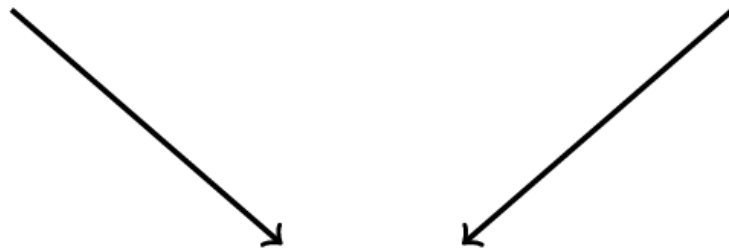
Example 1.14
A and B are two events with $P(A \cap \overline{B}) = 0.35$, $P(\overline{A} \cap B) = 0.15$, and $P(A \cup B) = 0.61$.

Draw a Venn diagram to illustrate the sample space including the events A and B with probability values. Then compute $P(A)$, $P(B)$ and $P(A|\overline{B})$.
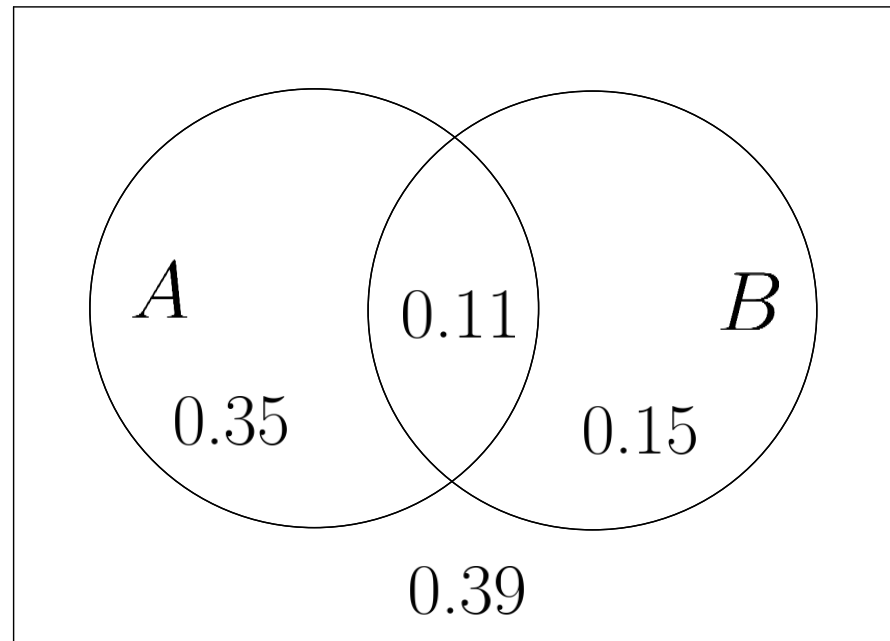
Note:

$$P(\overline{A \cup B}) = 1 - 0.61 = 0.39$$
$$P(A \cap B) = 0.61 - (0.35 + 0.15) = 0.11$$

According to the Venn diagram, we obtain:

$$P(A) = 0.35 + 0.11 = 0.46$$

$$P(B) = 0.15 + 0.11 = 0.26$$

$$P(A|\overline{B}) = \frac{P(A \cap \overline{B})}{P(\overline{B})} = \frac{0.35}{0.74} = \frac{35}{74}$$

Are events *A* and *B* independent?

Bayes' Rule

It is straightforwardly obtained from (1.6):

$$P(A) \cdot P(B|A) = P(B) \cdot P(A|B) \Rightarrow P(B|A) = \frac{P(B) \cdot P(A|B)}{P(A)} \quad (1.9)$$

Using the law of total probability, we also have:

$$P(B|A) = \frac{P(B) \cdot P(A|B)}{P(B) \cdot P(A|B) + P(\overline{B}) \cdot P(A|\overline{B})} \quad (1.10)$$

A more general form of (1.10) is:

$$P(B_k|A) = \frac{P(B_k) \cdot P(A|B_k)}{\sum_{n=1}^{N} P(B_n) \cdot P(A|B_n)} \quad (1.11)$$

Example 1.15

Suppose there are 3 bags. Bag 1 has 75 red and 25 blue balls; Bag 2 has 60 red and 40 blue balls; Bag 3 has 45 red and 55 blue balls. If one bag is chosen at random, and one ball is randomly picked from this bag, what is the probability that the ball is red?

On the other hand, if the picked ball is red, what is the probability that it is drawn from Bag 1?

Denote Bag n as B_n, and red and blue as R and B. First we need to find $P(R)$. As we choose randomly,

$$P(B_1) = P(B_2) = P(B_3) = \frac{1}{3}$$

We can easily obtain $P(R|B_1) = 75/100$ and other two. By (1.8):

$$P(R) = \sum_{n=1}^{3} P(B_n)P(R|B_n) = \frac{1}{3}\left[\frac{75}{100} + \frac{60}{100} + \frac{45}{100}\right] = 0.6$$

While for $P(B_1|R)$, we use (1.9):

$$P(B_1|R) = \frac{P(B_1) \cdot P(R|B_1)}{P(R)} = \frac{1/3(75/100)}{0.6} = \frac{5}{12}$$

Note that 5/12>1/3, aligning with $P(R|B_1) > P(R|B_2) \text{ or } P(R|B_3)$

Example 1.16
For a test on a certain disease, suppose we know:
- 2% of the overall population has the disease
- If a person has the disease, then the test has a 95% chance of correctly indicating he has it.
- If a person does not have the disease, then the test has a 10% chance of incorrectly indicating he has it.

If a patient tests positive, what is the probability that he actually has the disease?

Denote the event of having the disease as D and the event of test positive as T. Then we have:

$$P(D) = 0.02 \Rightarrow P(\overline{D}) = 0.98$$

$$P(T|D) = 0.95 \Rightarrow P(\overline{T}|D) = 0.05$$

$$P(T|\overline{D}) = 0.1 \Rightarrow P(\overline{T}|\overline{D}) = 0.9$$

To find $P(D|T)$, we apply (1.10):

$$\frac{P(D) \cdot P(T|D)}{P(D) \cdot P(T|D) + P(\overline{D}) \cdot P(T|\overline{D})} = \frac{0.02 \cdot 0.95}{0.02 \cdot 0.95 + 0.98 \cdot 0.1} = 0.1624$$

Although the patent test is positive, but the probability is not so high. Why?

True positive (TP), true negative (TN), false positive (FP) and false negative (FN) are important terms in classification:

TP: an outcome where the model correctly predicts the positive class, e.g., the patient has the disease and the test is positive.

TN: an outcome where the model correctly predicts the negative class, e.g., the patient does not have the disease and the test is negative.

FP: an outcome where the model incorrectly predicts the positive class, e.g., the patient does not have the disease but the test is positive.

FN: an outcome where the model incorrectly predicts the negative class, the patient has the disease and the test is negative.

		Predicted Class		
		Positive	**Negative**	
Actual Class	**Positive**	True Positive (TP)	False Negative (FN) **Type II Error**	**Sensitivity** $\dfrac{TP}{(TP + FN)}$
	Negative	False Positive (FP) **Type I Error**	True Negative (TN)	**Specificity** $\dfrac{TN}{(TN + FP)}$
		Precision $\dfrac{TP}{(TP + FP)}$	**Negative Predictive Value** $\dfrac{TN}{(TN + FN)}$	**Accuracy** $\dfrac{TP + TN}{(TP + TN + FP + FN)}$

Source: Data Science and Machine Learning : Confusion Matrix (manisha-sirsat.blogspot.com)

Suppose the overall population is 1000. Then:
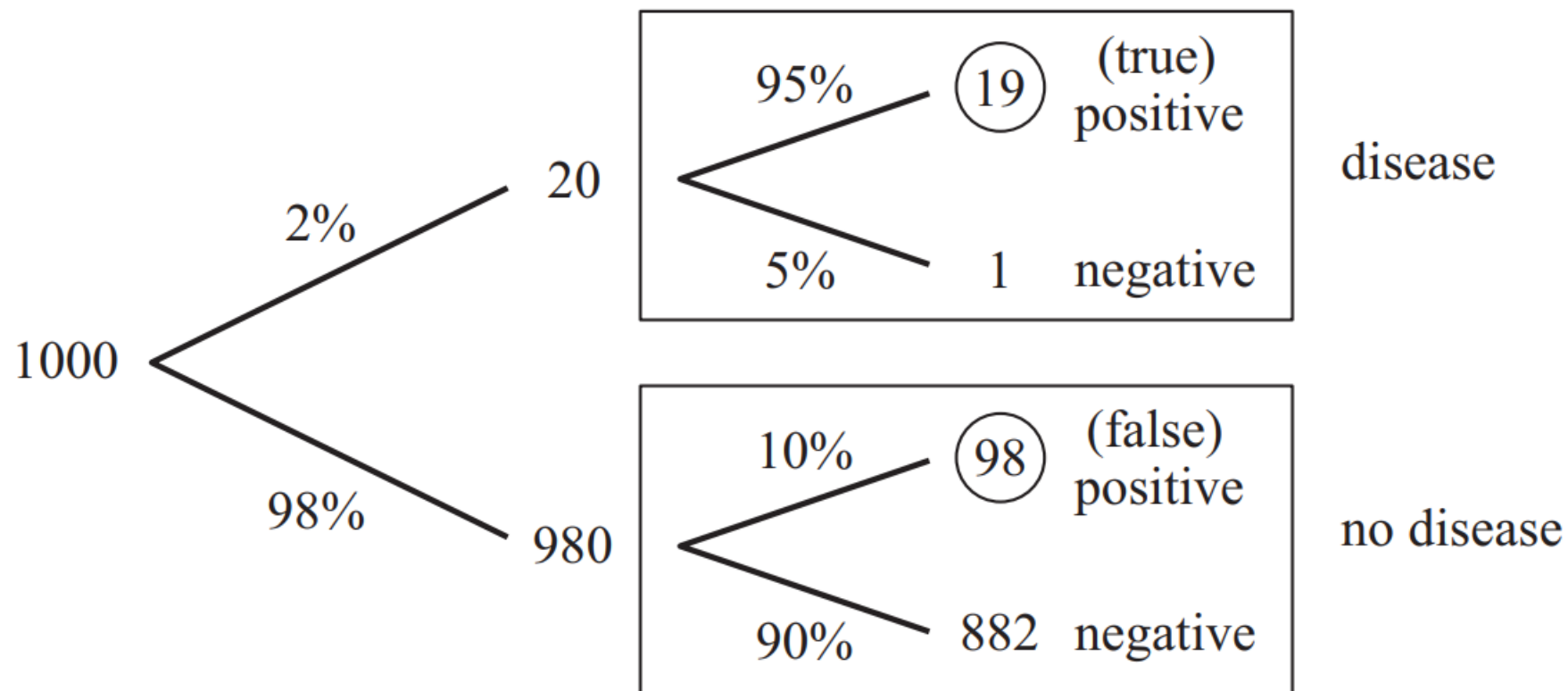
Based on this, we have

$$P(D|T) = \frac{19}{19 + 98} = \frac{19}{117} = 0.1624$$

Note that 19 corresponds to 2% x 95% and 98 corresponds to 98% x 10%, which aligns with (1.11).

In this example, the numbers of TP, TN, FP, and FN are 19, 882, 98, and 1, respectively.

Sensitivity: $= \dfrac{TP}{TP+FN} = \dfrac{19}{19+1} = 0.95$

Specificity: $= \dfrac{TP}{TP+FP} = \dfrac{882}{882+98} = 0.9$

Note that sensitivity is the probability of correctly determining that an infected person has the disease.

Specificity is the probability of correctly determining that an uninfected person does not have the disease (1-0.1).

Combinatorics

Combinatorics is the study about counting things, e.g., find combinations and permutations (different orderings).

It is useful in probability computation for equally-likely outcomes, e.g., throwing an unbiased coin or dice.

A permutation of a set of objects is a way of ordering them. For example, if there are two distinguishable objects and we label them as 1 and 2, then we can order them in 2 ways: "1 2" and "2 1" for no replacement, while 4 ways: "1 1", "2 2", "1 2" and "2 1" if replacement is allowed. For N objects:

Number of permutations without replacement is $N!$
Number of permutations with replacement is N^N

MATLAB command `perms` can illustrate the former, e.g.,

```
syms A B C v  %define symbols for symbolic computation
v = [A, B, C]; %vector with 3 distinct elements
perms(v)
=
[ C, B, A]
[ C, A, B]
[ B, C, A]
[ B, A, C]
[ A, C, B]
[ A, B, C]
```

It is clear that the number of permutations without replacement is $N! = 3! = 3 \times 2 = 6$.

How about the number of permutations of A, A, B?

The factorial command is `factorial`, e.g.,

```
factorial(3) = 6
```

Generalization to selection of k (from N) objects:

Number of permutations without replacement is:

$$P(N, k) = N(N-1)\cdots(N-k+1) = \frac{N!}{(N-k)!} = {}_N P_k \quad (1.12)$$

Number of permutations with replacement is:

$$P_r(N, k) = N^k \qquad (1.13)$$

Example 1.17
Consider choosing 2 balls from a bag containing 5 balls. Determine the numbers of permutations with and without replacement.

Number of permutations with replacement is $5^2 = 25$

Number of permutations without replacement is $5 \times 4 = 20$
Note that $k \le N$ in (1.12) but k can be larger than N in (1.13).

To compute (1.12) using MATLAB, we can use:

```
factorial(5)/factorial(3) = 20
```

or base on $P(N, k) = (N - k + 1) \cdots N$

```
prod(4:5) = 20
```

If we label 5 balls as A, B, C, D, and E, the permutations are:

AA	AB	AC	AD	AE
BA	BB	BC	BD	BE
CA	CB	CC	CD	CE
DA	DB	DC	DD	DE
EA	EB	EC	ED	EE

Example 1.18

Calculate the number of 8-character different passwords that can be created from 10 digits, 26 lower-case and 26 capital letters.

Using (1.13), the number is $P_r(62, 8) = 218, 340, 105, 584, 896$.

This means that the probability of guessing a password is $1/P_r(62, 8)$, 4.58×10^{-15}. If a spy program can try 1 million passwords per second, almost 7 years are needed to get the right one. At this speed, the spy program can test 60x60x24x7 million passwords in one week. This means that it can guess the password in 1 week is:

$$\frac{60 \times 60 \times 24 \times 7 \times 10^6}{P_r(62, 8)} \approx 0.0028$$

For faster spy program, we need to include more characters and/or change the password in a more frequent manner.

Example 1.19

In how many ways can 2 students be seated in a classroom with 10 chairs?

Students are distinguishable, and each student has a separate seat. Hence we apply (1.12):

$$P(10,2) = 10 \times 9 = 90$$

It can also be easily interpreted as follows. For the first student, there are 10 choices. Then for the second one, there are only 9 possibilities left. Hence the total number is 10x9.

When order does not matter, it refers to combination. Using the 2-object example again, without replacement, the combination number is 1 ("1 2" or "2 1"). With replacement, the number of combinations is 3 ("1 1", "2 2", "1 2"). Selecting k (from N) objects are generalized as:

Number of combinations without replacement is:

$$C(N, k) = \frac{N(N - 1) \cdots (N - k + 1)}{k(k - 1) \cdots 1} = \frac{N!}{(N - k)!k!} = \binom{N}{k} = {}_N C_k \quad (1.14)$$

Number of combinations with replacement is:

$$C_r(N, k) = \frac{(N + k - 1)!}{(N - 1)!k!} = C(N + k - 1, k) \quad (1.15)$$

(1.15) can be illustrated as follows. Consider the combinations of selecting 6 from 3 objects: A, B, and C. Two possibilities are:

A A A B C C $\rightarrow$ A A A | B | C C $\rightarrow$ * * * | * | * *
B B B B B C $\rightarrow$ | B B B B | C $\rightarrow$ | * * * * | *

Hence the number of combinations for 3 objects is identical to that of | and *, which is $C(6 + 2, 2) = C(6 + 3 - 1, 2) = 28$.

The MATLAB command is `nchoosek`, e.g,

`nchoosek(8,2) = 28`

Example 1.20

Repeat Example 1.17 when order does not matter.

Number of combinations with replacement is $C(6,2) = 15$

Number of combinations without replacement is $C(5,2) = 10$

AA	AB	AC	AD	AE
BA	BB	BC	BD	BE
CA	CB	CC	CD	CE
DA	DB	DC	DD	DE
EA	EB	EC	ED	EE

Example 1.21

How many different full-house combinations are possible in standard 5-card poker? A full-house contains 3 cards of one rank and 2 cards of another rank, e.g., AAA22. Assuming each card has the same drawing probability, determine the probability of having a full-house in drawing 5 cards?

A deck of cards consists of 13 ranks and each has 4 cards. For 3 cards of one rank, the number of combinations is:

$$C(4,3) = 4$$

Then for 2 cards of one rank, the number of combinations is:

$$C(4,2) = 6$$

We have 13 possibilities for the first rank, and then 12 for the second rank, hence full-house combination number is:

$$13 \times 12 \times 4 \times 6 = 3744$$

To select 5 from 52 without replacement, the number of combinations is:

$$C(52,5) = 2598960$$

Hence the probability of having a full-house is:

$$\frac{3744}{2598960} \approx 0.0014$$

which is about 0.1%. It can also be computed using MATLAB:

```
13*12*4*6/nchoosek(52,5) = 0.0014
```

Alternatively, to draw a particular set of 5 cards, the probability is:

$$\frac{5}{52} \times \frac{4}{51} \times \frac{3}{50} \times \frac{2}{49} \times \frac{1}{48} = \frac{1}{2598960}$$

Together with the combination number of 3744, the same result is obtained.

References:

1. D. Morin, *Probability: For the Enthusiastic Beginner*, CreateSpace Independent Publishing Platform, 2016
2. M. Baron, *Probability and Statistics for Computer Scientists*, CRC Press, 3rd Ed., 2019
3. H. Pishro-Nik, *Introduction to Probability, Statistics, and Random Processes*, Kappa Research LLC, 2014 https://www.probabilitycourse.com
4. D. G. Childers, *Probability and Random Processes: Using MATLAB with Applications to Continuous and Discrete Time Systems*, Irwin, 1997
5. M. Bollman, *Basic Gambling Mathematics: The Numbers behind the Neon*, CRC Press, 2014
6. S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*, PTR Prentice-Hall, 1998
7. S. L. Miller and D. Childers, *Probability and Random Processes: With Applications to Signal Processing and Communications*, Elsevier Science & Technology, 2012