Core course for *App.AI*, *Bioinfo.*, *Data Sci.&Eng.*, *Dec.Analytics*, *Q.Fin.*, *Risk Mgmt.* and *Stat.* Majors:

STAT2601A Probability and Statistics I (2023-2024 First Semester)

Chapter 3: Probability

3.1 The Mathematical Theory of Probability

In 1933, a Russian mathematician named Andreï Nikolayevich Kolmogorov published the axiomatic structure of probability theory. A review of the essential concepts of set algebra is useful in understanding Kolmogorov's theory, since every uncertain situation involves a collection, a set, of possible outcomes.

3.1.1 Language of Sets

Set a precisely specified collection of objects.
Element an object in the set.

We may specify a set by listing its elements or by describing them.

Example 3.1.

$A = \{1, 2, 3, 4, 5, 6\} = \{\text{the six faces of a die}\}$

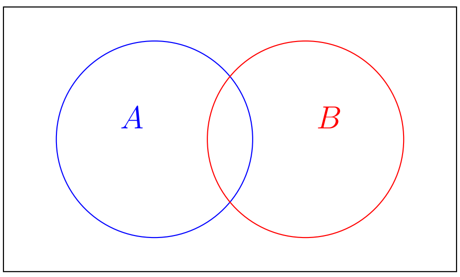
$B = \{(H, H), (H, T), (T, H), (T, T)\} = \{\text{possible outcomes of flipping two coins with regard to order}\}$

$C = \{\text{all the students taking STAT2601}\}$

★

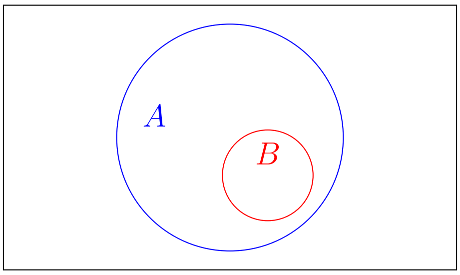
3.1.1.1 Venn Diagram

Often relationships among sets may be better understood by picturing them using Venn diagrams. The relationship between two sets A and B can be drawn as



The two circles overlap, indicating that some, but not all, of the elements of A is also elements of B.

The relationship would be drawn as



That is, all elements of B are elements of A. But some of elements of A are not elements of B.

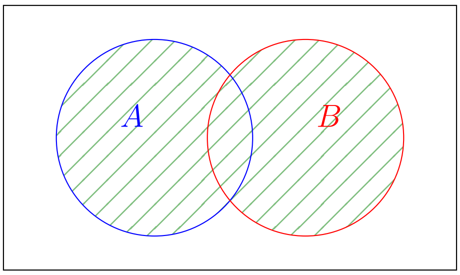
3.1.1.2 Set Relations and Operations

Belong

If x is an element of A, then we write $x \in A$. For example, $3 \in \{1, 2, 3, 4, 5, 6\}$.

Union

The set of all object which are either elements of A or elements of B is called the *union* of A and B, written as $A \cup B$. This area is shaded in the following Venn diagram.

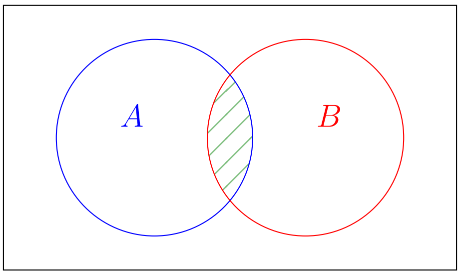


In mathematical notations, $x \in A \cup B \iff x \in A$ or $x \in B$.

For example, $x \in \{2, 3, 5, 7\} \iff x \in \{2, 5\}$ or $x \in \{3, 7\}$.

Intersection

The set of all objects which are both elements of A and elements of B is called the *intersection* of A and B, written as $A \cap B$.

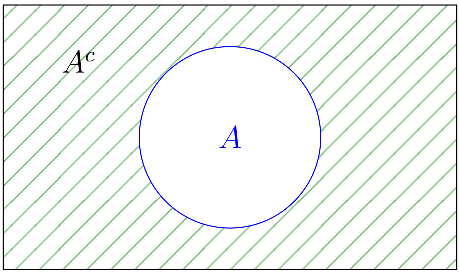


In mathematical notations, $x \in A \cap B \iff x \in A$ and $x \in B$.

For example, $x \in \{3, 5\} \iff x \in \{2, 3, 5\}$ and $x \in \{3, 5, 7\}$.

Complement

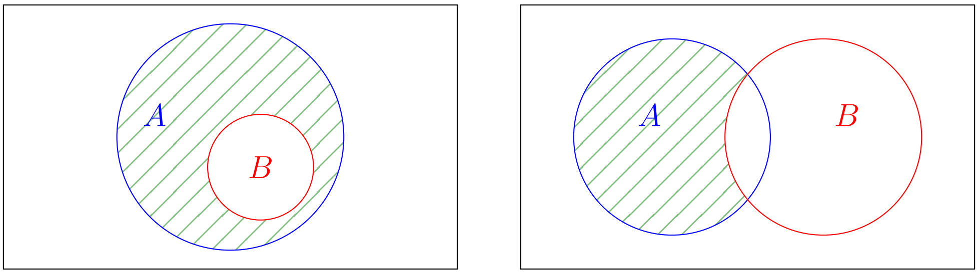
The set of all objects under consideration which are not elements of A is called the *complement* of A, written as A^c, or $\overline{A}$, or A'.



In mathematical notations, $x \in A^c \iff x \notin A$.

Difference

The set of elements in A that are not in B is called the *set difference* and is denoted as $A \backslash B$ or $A - B$, i.e., $A \backslash B = A - B = A \cap B^c$.

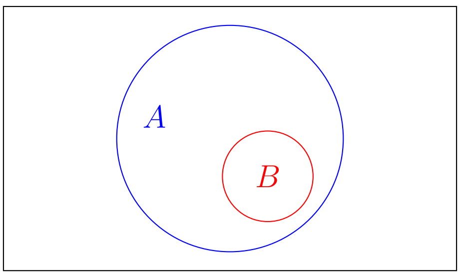


In mathematical notations, $x \in A \backslash B \iff x \in A$ and $x \notin B$.

For example, $x \in \{2, 4\} \iff x \in \{2, 4, 6\}$ and $x \notin \{3, 6, 9\}$.

Subset and *Superset*

If all the elements of a set B are elements of another set A, then B is said to be a *subset* of A, written as $B \subset A$. In the meanwhile, A is a *superset* of B, written as $A \supset B$.



In symbols, $B \subset A$ (or $A \supset B$) is equivalent to $x \in B \implies x \in A$. For example, $\{2, 4\} \subset \{2, 4, 6\}$.

If in addition, $A \subset B$, then we say that A is the same as B, i.e., $A = B$. This is the equality of sets.

Null Set

The set containing no element is called a *null set* or an *empty set*, written as $\varnothing$.

3.1.1.3 Properties

Suppose A, B, C, and E_i for all i are sets.

1. $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
2. $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
3. $A \cup (E_1 \cap E_2 \cap \cdots) = (A \cup E_1) \cap (A \cup E_2) \cap \cdots$
4. $A \cap (E_1 \cup E_2 \cup \cdots) = (A \cap E_1) \cup (A \cap E_2) \cup \cdots$
5. $A \cup B = A \cup (B \cap \overline{A}) = B \cup (A \cap \overline{B})$
6. De Morgan's Law
   (a) Two sets: $\overline{A \cup B} = \overline{A} \cap \overline{B}, \qquad \overline{A \cap B} = \overline{A} \cup \overline{B}$
   (b) Generalized form:
   $$\left(\bigcup_{i=1}^{n} E_i \right)^c = \bigcap_{i=1}^{n} E_i^c, \qquad \left(\bigcap_{i=1}^{n} E_i \right)^c = \bigcup_{i=1}^{n} E_i^c$$

3.1.2 Language of Probability

Sample space — The set containing all possible outcomes, usually denoted as Ω, e.g., $\Omega = \{\text{outcomes of rolling a six-faced die}\} = \{1, 2, 3, 4, 5, 6\}$.

Event — A subset of Ω, e.g., $E = \{\text{even number}\} = \{2, 4, 6\}$.

Empty event — event $E = \varnothing$, e.g., $E = \{\text{larger than 6 on a face of a die}\}$.

Non-empty event — event $E \neq \varnothing$, e.g., $E = \{\text{smaller than 3}\} = \{1, 2\}$.

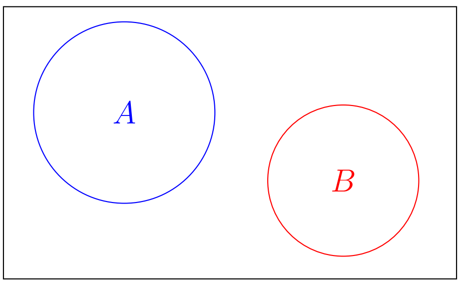
Elementary event and *Compound event*

A non-empty event that cannot be decomposed into a number of non-empty events is called an *elementary event*; otherwise it is called a *compound event*. For example, $A = \{1\}$, $B = \{2\}$ are elementary events, $C = \{1, 5\}$ is a compound event.

3.1.2.1 Event Terminologies

Disjoint

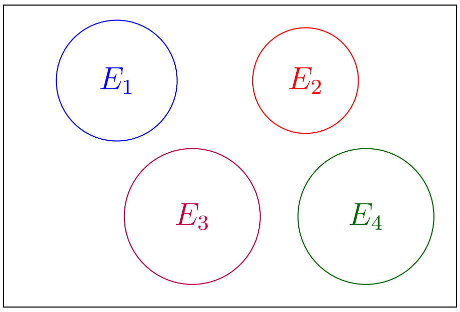
Two events A and B are said to be *disjoint* if $A \cap B = \varnothing$, i.e., they have no common elements.



e.g., $A = \{1, 2\}$, $B = \{3, 5\}$, $C = \{2, 4\}$; then A and B are disjoint; A and C are not disjoint as $A \cap C = \{2\}$.

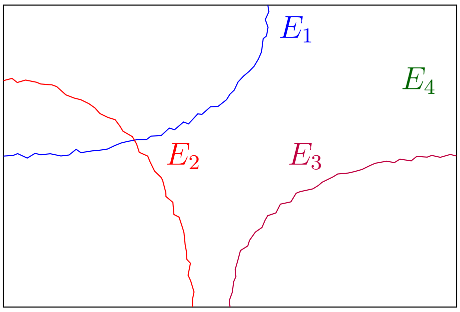
Mutually exclusive

A group of events $E_1, E_2, \ldots, E_n$ are said to be *mutually exclusive* if any two events are disjoint, i.e., $E_i \cap E_j = \varnothing$ for all $i \neq j$.



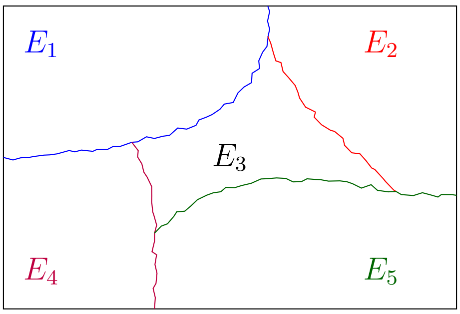
Exhaustive

A group of events $E_1, E_2, \ldots, E_n$ are said to be *exhaustive* if the union of them is the sample space, i.e., $E_1 \cup E_2 \cup \cdots \cup E_n = \Omega$.



Partition

A group of events $E_1, E_2, \ldots, E_n$ is called a *partition* if the events are mutually exclusive and exhaustive.



Complement

The *complement* of event E is the collection of the outcomes not contained in E, that is, $E^c = \Omega \backslash E$.

Occurrence

An event E is said to *occur* if the outcome belongs to E. For example, if E represents the event that an even number is on the face of a die, i.e., $E = \{2, 4, 6\}$, then E occurs if we throw a 2 or 4 or 6.

3.1.2.2 Meaning of Event Operations

1. $A \cup B$ occurs if either A or B occurs. $A \cap B$ occurs if both A and B occur.
2. A^c occurs if A does not occur.
3. If two events are disjoint ($A \cap B = \varnothing$), then it is impossible that both of them will occur at the same time.
4. If a group of events are exhaustive ($A_1 \cup A_2 \cup \cdots \cup A_n = \Omega$), then at least one of the events $A_1, A_2, \ldots, A_n$ will occur.
5. If a group of events form a partition (i.e., mutually exclusive and exhaustive), then exactly one and only one of the events $A_1, A_2, \ldots, A_n$ will occur.

3.1.3 Kolmogorov's Axioms of Probability

Definition 3.1.
A probability on the sample space Ω is an assignment of a value, $\Pr(E)$ say, to each event E such that

1. $\Pr(E) \geq 0$ for any event E.
2. $\Pr(\Omega) = 1$.
3. For any sequence of mutually exclusive events $E_1, E_2, \ldots$ (i.e., events for which $E_i \cap E_j = \varnothing$ when $i \neq j$),
$$\Pr\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} \Pr(E_i).$$
(According to this axiom, $\Pr(\cdot)$ is called *countably additive*.)

◀

Example 3.2.
$\Omega = \{1, 2, 3, 4, 5, 6\}$, $A = \{1, 3\}$, $B = \{2, 4\}$
$$\begin{aligned} \Pr(A) &= \frac{2}{6} = \frac{1}{3} > 0, \\ \Pr(\Omega) &= \frac{6}{6} = 1. \end{aligned}$$
Obviously, A and B are disjoint.
$$\begin{aligned} \Pr(A \cup B) &= \Pr(\{1, 2, 3, 4\}) = \frac{4}{6} = \frac{2}{3}, \\ \Pr(A) + \Pr(B) &= \frac{1}{3} + \frac{1}{3} = \frac{2}{3} = \Pr(A \cup B). \end{aligned}$$

★

3.1.3.1 Properties

Suppose A, B and E_i for all i's are events, i.e., subsets of a sample space Ω.

1. $\Pr(\varnothing) = 0$.
2. $\Pr(A^c) = 1 - \Pr(A)$.
3. If $A \subset B$, then $\Pr(A) \leq \Pr(B)$.
4. $\Pr(A) \leq 1$ for any event A.
5. $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$. (Special case of 6. when $n = 2$.)
6. (Inclusion-exclusion principle)

$$\begin{aligned}
\Pr(E_1 \cup E_2 \cup \cdots \cup E_n) &= \sum_{j=1}^{n} (-1)^{j+1} \sum_{i_1 < i_2 < \cdots < i_j} \Pr(E_{i_1} \cap E_{i_2} \cap \cdots \cap E_{i_j}) \\
&= \sum_{i=1}^{n} \Pr(E_i) - \sum_{i_1 < i_2} \Pr(E_{i_1} \cap E_{i_2}) + \cdots \\
&\quad + (-1)^{j+1} \sum_{i_1 < i_2 < \cdots < i_j} \Pr(E_{i_1} \cap E_{i_2} \cap \cdots \cap E_{i_j}) \\
&\quad + \cdots \\
&\quad + (-1)^{n+1} \sum_{i_1 < i_2 < \cdots < i_n} \Pr(E_{i_1} \cap E_{i_2} \cap \cdots \cap E_{i_n}).
\end{aligned}$$

7. (Boole's Inequality)

$$\Pr\left(\bigcup_{i=1}^{\infty} E_i\right) \leq \sum_{i=1}^{\infty} \Pr(E_i).$$

8. (Continuity)

Increasing sequence: $E_1 \subset E_2 \subset \cdots \subset E_n \subset E_{n+1} \subset \cdots$, define $\lim_{n \to \infty} E_n = \bigcup_{i=1}^{\infty} E_i$.

Decreasing sequence: $E_1 \supset E_2 \supset \cdots \supset E_n \supset E_{n+1} \supset \cdots$, define $\lim_{n \to \infty} E_n = \bigcap_{i=1}^{\infty} E_i$.

If $\{E_n, n \geq 1\}$ is an increasing (or decreasing) sequence of events, then

$$\Pr\left(\lim_{n \to \infty} E_n\right) = \lim_{n \to \infty} \Pr(E_n).$$

Proof. Suppose $\{E_n,\ n \geq 1\}$ is increasing, let

$$F_1 = E_1, \qquad F_2 = E_2 \cap E_1^c, \qquad F_3 = E_3 \cap E_2^c, \qquad \ldots \qquad F_n = E_n \cap E_{n-1}^c, \qquad \ldots$$

so that $\{F_n,\ n \geq 1\}$ are mutually exclusive and $\bigcup_{i=1}^{n} F_i = E_n$ and $\bigcup_{i=1}^{\infty} F_i = \bigcup_{i=1}^{\infty} E_i$.

According to the definition of the limit of E_n,

$$\begin{aligned}
\Pr\left(\lim_{n\to\infty} E_n\right) &= \Pr\left(\bigcup_{i=1}^{\infty} E_i\right) \\
&= \Pr\left(\bigcup_{i=1}^{\infty} F_i\right) \\
&= \sum_{i=1}^{\infty} \Pr(F_i) \qquad \text{(the 3rd axiom)} \\
&= \lim_{n\to\infty} \sum_{i=1}^{n} \Pr(F_i) \\
&= \lim_{n\to\infty} \Pr\left(\bigcup_{i=1}^{n} F_i\right) \qquad \text{(the 3rd axiom)} \\
&= \lim_{n\to\infty} \Pr(E_n).
\end{aligned}$$

The proof for decreasing sequence of events is in a similar manner. □

Example 3.3. (Inclusion-exclusion principle)

Suppose A_1, A_2, A_3 and A_4 are four events.

$$\begin{aligned}
&\Pr(A_1 \cup A_2 \cup A_3 \cup A_4) \\
&= \Pr(A_1) + \Pr(A_2) + \Pr(A_3) + \Pr(A_4) \\
&\quad -\{\Pr(A_1 \cap A_2) + \Pr(A_1 \cap A_3) + \Pr(A_1 \cap A_4) + \Pr(A_2 \cap A_3) + \Pr(A_2 \cap A_4) + \Pr(A_3 \cap A_4)\} \\
&\quad +\{\Pr(A_1 \cap A_2 \cap A_3) + \Pr(A_1 \cap A_2 \cap A_4) + \Pr(A_1 \cap A_3 \cap A_4) + \Pr(A_2 \cap A_3 \cap A_4)\} \\
&\quad - \Pr(A_1 \cap A_2 \cap A_3 \cap A_4).
\end{aligned}$$

★

Example 3.4.

Suppose that A, B, C are three events for which $\Pr(A) = \Pr(B) = \Pr(C) = 0.99$. Then

$$\Pr(\overline{A} \cup \overline{B} \cup \overline{C}) \leq \Pr(\overline{A}) + \Pr(\overline{B}) + \Pr(\overline{C}) = 0.01 + 0.01 + 0.01 = 0.03,$$

$$\Pr(A \cap B \cap C) = 1 - \Pr(\overline{A} \cup \overline{B} \cup \overline{C}) \geq 1 - 0.03 = 0.97.$$

★

Example 3.5.

A secretary types four letters to four people and addresses the four envelopes. If he inserts the letters at random, one in each envelope, what is the probability that none of the letters will go into the correct envelopes?

Solution:

Let E_i be the event that the i-th letter goes into the correct envelope. Then the desired probability is $1 - \Pr\left(\bigcup_{i=1}^{4} E_i\right)$.

Consider
$$\Pr(E_1) = \Pr(E_2) = \Pr(E_3) = \Pr(E_4) = \frac{3!}{4!} = \frac{1}{4},$$
$$\Pr(E_1 \cap E_2) = \Pr(E_1 \cap E_3) = \cdots = \Pr(E_3 \cap E_4) = \frac{2!}{4!} = \frac{1}{12},$$
$$\Pr(E_1 \cap E_2 \cap E_3) = \Pr(E_1 \cap E_2 \cap E_4) = \cdots = \Pr(E_2 \cap E_3 \cap E_4) = \frac{1}{4!} = \frac{1}{24},$$
$$\Pr(E_1 \cap E_2 \cap E_3 \cap E_4) = \frac{1}{4!} = \frac{1}{24}.$$

By the inclusion-exclusion principle,

$$\begin{aligned}
\Pr\left(\bigcup_{i=1}^{4} E_i\right) &= \sum_{i=1}^{4} \frac{1}{4} - \sum_{i_1<i_2} \frac{1}{12} + \sum_{i_1<i_2<i_3} \frac{1}{24} - \frac{1}{24} \\
&= 1 - \binom{4}{2}\frac{2!}{4!} + \binom{4}{3}\frac{1!}{4!} - \frac{1}{4!} \\
&= 1 - \frac{1}{2!} + \frac{1}{3!} - \frac{1}{4!} \\
&= \frac{5}{8}.
\end{aligned}$$

Hence the probability that none of the letters matched is $\frac{3}{8}$.

Remark:

In general, if n letters are inserted at random to n envelopes, the probability that none of them will go into correct envelope is equal to

$$\frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \cdots + (-1)^n \frac{1}{n!},$$

which for large n is approximately equal to $e^{-1} \approx 0.3678794412$.

★

Thought Question:

Continue with the above example, if n letters are inserted at random to n envelopes, what is the probability that *exactly* k letters go to the correct envelopes?

3.2 Conditional Probability

When part of the outcome is fixed, i.e., we have some information in advance, then the calculation of probability of its occurrence should be restricted to a smaller sample space based on the new information.

Example 3.6.

Consider 10 objects with the following classifications regarding shape and color.

	Red	Blue	Total
Ball	5	1	6
Cube	1	3	4
Total	6	4	10

Suppose we draw an object at random from these 10 objects. We can evaluate the probabilities of many events that we are interested:

$$\Pr(\text{ball}) = \frac{6}{10}, \quad \Pr(\text{blue}) = \frac{4}{10}, \quad \Pr(\text{blue ball}) = \frac{1}{10}, \quad \Pr(\text{red cube}) = \frac{1}{10}, \quad \ldots$$

Suppose when the object was drawn we found that it was blue by a glimpse. Then this information will alter our uncertainty about its shape. Denote the probability that it is a ball given that it is blue by $\Pr(\text{ball}|\text{blue})$. Then,

$$\Pr(\text{ball}|\text{blue}) = \frac{\text{no. of blue balls}}{\text{no. of blue objects}} = \frac{1}{4}.$$

Similarly,

$$\Pr(\text{ball}|\text{red}) = \frac{5}{6}, \qquad \Pr(\text{blue}|\text{ball}) = \frac{1}{6}, \qquad \ldots$$

These are called *conditional probabilities*.

It can be easily observed that

$$\begin{aligned}\Pr(\text{ball}|\text{blue}) &= \frac{\text{no. of blue balls/total no. of objects}}{\text{no. of blue objects/total no. of objects}} \\ &= \frac{\Pr(\text{ball} \cap \text{blue})}{\Pr(\text{blue})}.\end{aligned}$$

★

Hence, we have the following definition.

Definition 3.2.
For any two events A and B, the conditional probability of A given the occurrence of B is written as $\Pr(A|B)$ and is defined as

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)},$$

provided that $\Pr(B) > 0$.

◀

Example 3.7. (Boy or Girl paradox)

A mother has two kids. You ask, "is anyone of them a boy?" The mother says "Yes". What is the probability that they are both boys?

Solution:

Take $\Omega = \{BB, BG, GB, GG\}$ as the sample space of equally likely outcomes.

Let $A = \{\text{at least one of them is a boy}\} = \{BB, BG, GB\}$,

$B = \{\text{both kids are boys}\} = \{BB\}$.

$$\Pr(B|A) = \frac{\Pr(B \cap A)}{\Pr(A)} = \frac{1/4}{3/4} = \frac{1}{3}.$$

★

Example 3.8. (Boy or Girl paradox)

A mother has two kids. You ask, "is your elder kid a boy?" The mother says "Yes". What is the probability that they are both boys?

Solution:

Take $\Omega = \{BB, BG, GB, GG\}$ as the sample space of equally likely outcomes.

Let $A = \{\text{the elder kid is a boy}\} = \{BB, BG\}$,

$B = \{\text{both kids are boys}\} = \{BB\}$.

$$\Pr(B|A) = \frac{\Pr(B \cap A)}{\Pr(A)} = \frac{1/4}{2/4} = \frac{1}{2}.$$

★

As can be seen, one should pay close attention to what information is actually given in order to correctly solve the problem involving conditional probability calculations.

Example 3.9. (Simpson's paradox)

There are 2 treatments for a disease, A and B. Applying A to some patients and B to others results in the following observations:

	Men (M)		Women (W)	
	Recovered (R)	Dead (D)	Recovered (R)	Dead (D)
Treatment A	20	80	40	20
Treatment B	50	160	15	5

For men, the probabilities of recovery given the two treatments are

$$\begin{aligned}\Pr(R|M\cap A) &= \frac{\Pr(R\cap M\cap A)}{\Pr(M\cap A)} = \frac{20}{100} = \frac{1}{5},\\ \Pr(R|M\cap B) &= \frac{\Pr(R\cap M\cap B)}{\Pr(M\cap B)} = \frac{50}{210} = \frac{5}{21} > \Pr(R|M\cap A).\end{aligned}$$

For women, the probabilities of recovery given the two treatments are

$$\begin{aligned}\Pr(R|W\cap A) &= \frac{\Pr(R\cap W\cap A)}{\Pr(W\cap A)} = \frac{40}{60} = \frac{2}{3},\\ \Pr(R|W\cap B) &= \frac{\Pr(R\cap W\cap B)}{\Pr(W\cap B)} = \frac{15}{20} = \frac{3}{4} > \Pr(R|W\cap A).\end{aligned}$$

Therefore for both men and women, B is a better treatment than A.

However, if we combine the two tables for the two genders, i.e., for all patients, we have

$$\begin{aligned}\Pr(R|A) &= \frac{\Pr(R\cap A)}{\Pr(A)} = \frac{20+40}{100+60} = \frac{3}{8},\\ \Pr(R|B) &= \frac{\Pr(R\cap B)}{\Pr(B)} = \frac{50+15}{210+20} = \frac{13}{46} < \frac{3}{8} = \Pr(R|A),\end{aligned}$$

which implies that A is better!

This example illustrates what has come to be known as the *Simpson's paradox*, a reversal of the direction of a comparison or an association when data from several groups are combined to form a single group.

★

Theorem 3.1. (Multiplication Theorem)

For any two events A and B with $\Pr(B) > 0$,

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

For any three events A, B, C with $\Pr(B\cap C) > 0$,

$$\Pr(A\cap B\cap C) = \Pr(C)\Pr(B|C)\Pr(A|B\cap C).$$

▲

Example 3.10.

What is the probability of drawing three aces in a row from a poker deck?

Solution:

Denote A_i, $i = 1, 2, 3$, as the event that the i-th drawn card is an ace. Then,

$$\begin{aligned}\Pr(A_1 \cap A_2 \cap A_3) &= \Pr(A_1) \times \Pr(A_2|A_1) \times \Pr(A_3|A_1 \cap A_2) \\ &= \frac{4}{52} \times \frac{3}{51} \times \frac{2}{50} = \frac{1}{5525}.\end{aligned}$$

★

Definition 3.3.
Two events A and B are called independent if and only if

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

If $\Pr(A) > 0$, then A and B are independent if and only if

$$\Pr(B|A) = \Pr(B).$$

◀

Example 3.11.

Opinion poll on building an incinerator in Hong Kong:

	Support (S)	Opposed (S^c)
HKU students (H)	0.459	0.441
Non-HKU students (H^c)	0.051	0.049

If a person is drawn at random, then

$$\begin{aligned}\Pr(H) &= 0.459 + 0.441 = 0.9, \\ \Pr(S) &= 0.459 + 0.051 = 0.51, \\ \Pr(H \cap S) &= 0.459, \\ \Pr(H)\Pr(S) &= 0.9 \times 0.51 = 0.459 = \Pr(H \cap S).\end{aligned}$$

Hence being a HKU student or not and opinion are independent in this poll.

★

Definition 3.4.
The events $A_1, A_2, \ldots, A_k$ are (mutually) independent if and only if the probability of the intersection of any combination of them is equal to the product of the probabilities of the corresponding single events.

◀

For example, A_1, A_2, A_3 are independent if and only if

$$\begin{aligned}
\Pr(A_1 \cap A_2) &= \Pr(A_1)\Pr(A_2), \\
\Pr(A_1 \cap A_3) &= \Pr(A_1)\Pr(A_3), \\
\Pr(A_2 \cap A_3) &= \Pr(A_2)\Pr(A_3), \text{ and} \\
\Pr(A_1 \cap A_2 \cap A_3) &= \Pr(A_1)\Pr(A_2)\Pr(A_3).
\end{aligned}$$

Example 3.12.

Two fair coins are tossed. Denote A as the event that the first coin lands on head, B as the event that the second coin lands on tail, C as the event that both coin land on the same face.

Are these events independent?

Solution:

$\Omega = \{HH, HT, TH, TT\}$, assuming equally likely outcomes.

$A = \{HH, HT\}$, $B = \{HT, TT\}$, $C = \{HH, TT\}$.

Obviously, $\Pr(A) = \Pr(B) = \Pr(C) = \frac{1}{2}$.

$\Pr(A \cap B) = \Pr(\{HT\}) = \frac{1}{4} = \Pr(A)\Pr(B) \quad \Longrightarrow A$ and B are independent.

$\Pr(A \cap C) = \Pr(\{HH\}) = \frac{1}{4} = \Pr(A)\Pr(C) \quad \Longrightarrow A$ and C are independent.

$\Pr(B \cap C) = \Pr(\{TT\}) = \frac{1}{4} = \Pr(B)\Pr(C) \quad \Longrightarrow B$ and C are independent.

However, $\Pr(A \cap B \cap C) = \Pr(\varnothing) = 0 \neq \Pr(A)\Pr(B)\Pr(C)$.

Thus, the events A, B, and C are not mutually independent, they are just *pairwisely independent*.

★

3.2.1 Law of Total Probability and Bayes' Theorem

Theorem 3.2. (Law of Total Probability)
If $0 < \Pr(B) < 1$, then

$$\Pr(A) = \Pr(A|B)\Pr(B) + \Pr(A|B^c)\Pr(B^c)$$

for any event A.

▲

Proof.

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

Since $(A \cap B) \cap (A \cap B^c) = \varnothing$, by the third axiom of probability,

$\Pr(A) = \Pr(A \cap B) + \Pr(A \cap B^c) = \Pr(A|B)\Pr(B) + \Pr(A|B^c)\Pr(B^c)$. □

Example 3.13.

When coded messages are sent, there may be errors in the transmission. In particular, Morse code used "dots" and "dashes", which are known to occur in the proportion of 3:4. Suppose there is interference on the transmission line, and with probability 1/8 a dot is mistakenly received as a dash, and vice versa. If a single signal is sent to us, what is the probability that we will receive a dot?

Solution:

Denote B as the event that the original signal sent is a dot and A as the event that we receive a dot. Then we have

$$\Pr(B) = \frac{3}{7}, \qquad \Pr(A^c|B) = \Pr(A|B^c) = \frac{1}{8}.$$

Using the law of total probability,

$$\begin{aligned}\Pr(A) &= \Pr(A|B)\Pr(B) + \Pr(A|B^c)\Pr(B^c) \\ &= \left(1 - \frac{1}{8}\right) \times \frac{3}{7} + \frac{1}{8} \times \left(1 - \frac{3}{7}\right) \\ &= \frac{25}{56}.\end{aligned}$$

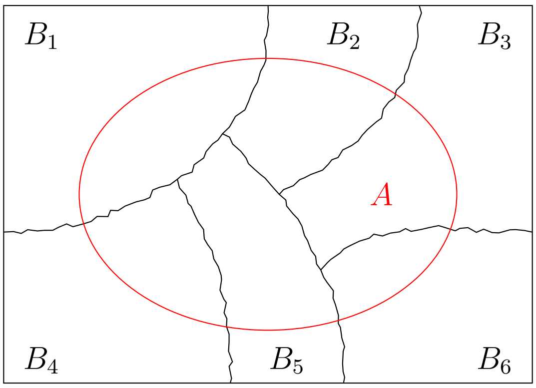
★

Theorem 3.3. (Law of Total Probability – General Version)
If $B_1, B_2, \ldots, B_k$ are mutually exclusive and exhaustive events (i.e., a partition of the sample space), then for any event A,

$$\Pr(A) = \sum_{j=1}^{k} \Pr(A|B_j)\Pr(B_j).$$

▲

Proof. From the Venn diagram,



we observed that $A = (A \cap B_1) \cup (A \cap B_2) \cup \cdots \cup (A \cap B_k)$, and the events $A \cap B_1$, $A \cap B_2$, …, $A \cap B_k$ are mutually exclusive. Hence,

$$\begin{aligned}\Pr(A) &= \Pr(A \cap B_1) + \Pr(A \cap B_2) + \cdots + \Pr(A \cap B_k) \\ &= \Pr(A|B_1)\Pr(B_1) + \Pr(A|B_2)\Pr(B_2) + \cdots + \Pr(A|B_k)\Pr(B_k).\end{aligned}$$

□

Theorem 3.4. (Bayes' Theorem, or Bayes' Rule, or Bayes' Law)
For any two events A and B with $\Pr(A) > 0$ and $\Pr(B) > 0$,

$$\Pr(B|A) = \Pr(A|B)\frac{\Pr(B)}{\Pr(A)}.$$

▲

Remarks

1. This theorem related the conditional probability $\Pr(A|B)$ to its "inverse" counterpart, $\Pr(B|A)$. As can be seen, $\Pr(A|B)$ and $\Pr(B|A)$ are totally different. They are the same if and only if $\Pr(A) = \Pr(B)$. Hence if we observed that a large proportion of lung cancer patients are smokers, we should not jump to the conclusion that smokers will have large chance to have lung cancer.

2. $\Pr(B)$ is called the *prior probability* of B. It is "prior" in the sense that it does not take into account any information about A.

3. $\Pr(B|A)$, the conditional probability of B given A, is also called the *posterior probability* in contrast to the prior probability $\Pr(B)$. It is "posterior" in the sense that some additional information (occurrence of A) has been taken into account.

4. Hence the Bayes' theorem described the way in which one's belief about B are updated by having observed A.

Example 3.14.
According to the Morse code example, if we received a dot, what is the probability that the actual symbol sent was really a dot?

Solution:

Using Bayes' theorem, $\Pr(B|A) = \dfrac{\Pr(A|B)\Pr(B)}{\Pr(A)} = \dfrac{\left(1-\frac{1}{8}\right)\times\frac{3}{7}}{\frac{25}{56}} = \dfrac{21}{25} = 0.84.$

★

Together with the law of total probability, the Bayes' theorem is commonly expressed in the following form:

Theorem 3.5. (Bayes' Theorem)
If $B_1, B_2, \ldots, B_k$ are mutually exclusive and exhaustive events (i.e., a partition of the sample space), and A is any event with $\Pr(A) > 0$, then for any event B_j,

$$\Pr(B_j|A) = \frac{\Pr(A|B_j)\Pr(B_j)}{\Pr(A)} = \frac{\Pr(B_j)\Pr(A|B_j)}{\displaystyle\sum_{i=1}^{k}\Pr(B_i)\Pr(A|B_i)}.$$

▲

This form enables us to compute the posterior probabilities $\Pr(B_j|A)$ from the prior probabilities $\Pr(B_i)$ using $\Pr(A|B_i)$, $i = 1, \ldots, k$, which may be easier to obtain.

Example 3.15. (Diagnosis test)

Suppose 0.3% individuals out of a certain population are carrying a particular virus. A powerful diagnostic test should be as accurate as possible, i.e., with small error rate. There can be two different types of diagnostic errors:

False positive: a positive result is obtained on a non-virus-carrier

False negative: a negative result is obtained on a virus carrier

Suppose the rate of false positive and false negative are 2% and 1%, respectively. These error rates can be expressed in terms of probabilities.

Let $+$ be the event that the test shows a positive result as carrier of a particular virus,

V be the event that a person is infected by this virus.

Also denote $-$ as the complement of $+$ (i.e., $\overline{+} = -$).

The given information can be written as

$\Pr(V) = 0.003$ (0.3% of virus carriers in the population) $\Longrightarrow \Pr(\overline{V}) = 0.997$.

$\Pr(+|\overline{V}) = 0.02$ (2% of false positive) $\Longrightarrow \Pr(-|\overline{V}) = 0.98$.

$\Pr(-|V) = 0.01$ (1% of false negative) $\Longrightarrow \Pr(+|V) = 0.99$.

A common question in diagnostic test is: if the test shows a positive result on me, how likely is that I am really a virus carrier?

The prior probability of being a virus carrier before the test is $\Pr(V) = 0.003$. To answer the question, we need to compute the following posterior probability:

$$\Pr(\text{virus carrier}|\text{positive result}) = \Pr(V|+).$$

Using Bayes' theorem,

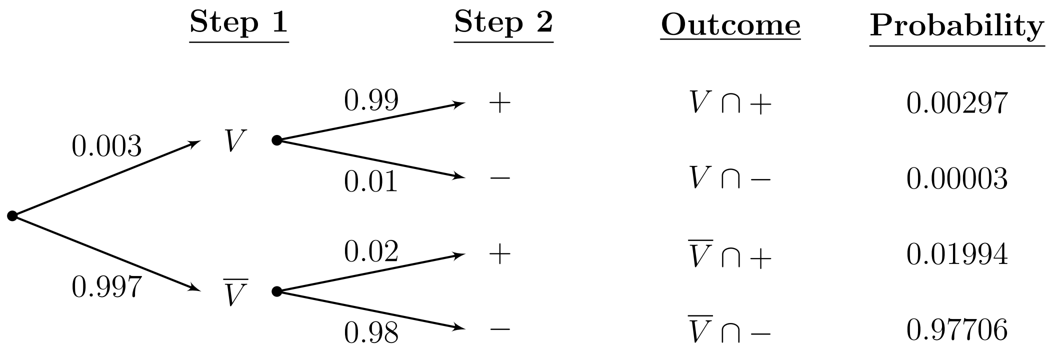
$$\begin{aligned}
\Pr(V|+) &= \frac{\Pr(V)\Pr(+|V)}{\Pr(V)\Pr(+|V) + \Pr(\overline{V})\Pr(+|\overline{V})} \\
&= \frac{0.003 \times 0.99}{0.003 \times 0.99 + 0.997 \times 0.02} \\
&= 0.1296.
\end{aligned}$$

That is, even the test shows positive on an individual, he/she will have only about 13% chance to be infected by the virus.

★

3.2.2 Tree Diagram

Definition 3.5.
A *tree diagram* is a useful graphical display that shows the outcomes of a set of events. Each node represents a particular outcome and the probabilities are written on the branches of the tree. The following diagram shows the situations according to **Example 3.14.**:



The probabilities shown on the edges are conditional on the previous node. The probability of climbing from the base to any of the branches is obtained by multiplying the probabilities.

◀

For example, there are two branches corresponding to the outcome of a negative test result. Adding up the probabilities corresponding to these two paths gives

$$\Pr(-\text{ve result}) = 0.003 \times 0.01 + 0.997 \times 0.98 = 0.97709.$$

To evaluate the conditional probability of a non-virus-carrier given a negative test result, we can simply calculate the weight of the path which passed through $\overline{V}$ out of all the paths that end at $-$ve result, i.e.,

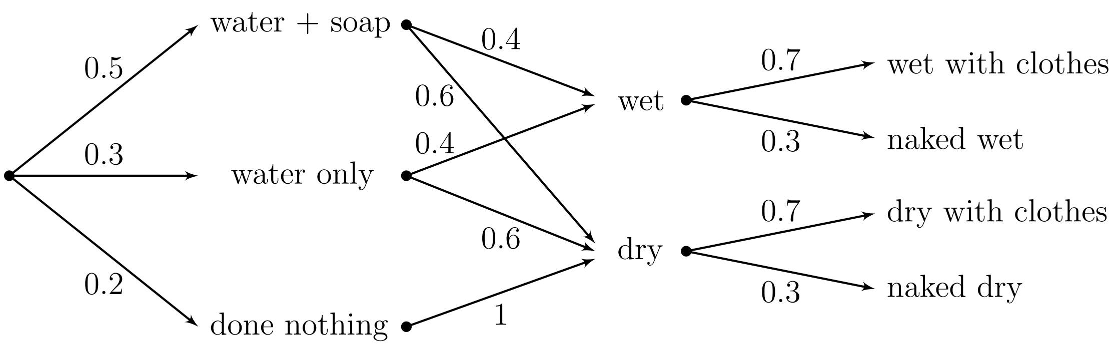
$$\Pr(\overline{V}|-) = \frac{0.997 \times 0.98}{0.003 \times 0.01 + 0.997 \times 0.98} = 0.99997.$$

Hence a negative test result indicates the individual as a non-virus-carrier almost certainly.

Example 3.16.

A mathematician goes inside a bathroom to take a bath. After taking off his clothes, he has a probability 0.5 of using soap and water to clean his body, and a probability 0.2 of doing nothing so that his body remains dry. If his body is wet, he will dry himself using a towel with probability 0.6 but will forget to dry himself with probability 0.4. Before he leaves the bathroom, he will put on his clothes with probability 0.7 but will forget to put them on with probability 0.3.

A tree diagram would be very helpful for the probability calculations in such problem.



$\Pr(\text{dry with clothes on}) = (0.5 \times 0.6 + 0.3 \times 0.6 + 0.2 \times 1) \times 0.7 = 0.476.$

$\Pr(\text{naked wet}) = (0.5 \times 0.4 + 0.3 \times 0.4) \times 0.3 = 0.096.$

$\Pr(\text{water + soap}|\text{dry with clothes on}) = \dfrac{0.5 \times 0.6 \times 0.7}{0.476} = 0.441.$

$\Pr(\text{water + soap}|\text{naked wet}) = \dfrac{0.5 \times 0.4 \times 0.3}{0.096} = 0.625.$

$\Pr(\text{done nothing}|\text{naked dry}) = \dfrac{0.2 \times 1 \times 0.3}{(0.5 \times 0.6 + 0.3 \times 0.6 + 0.2 \times 1) \times 0.3} = 0.294.$

3.2.3 Other Examples of Conditional Probabilities

Example 3.17. (Monty Hall problem)

Suppose you are on a game show, and you are given the choice of three boxes. In one box is a key to a new BMW while empty in others. You pick a box, say box A. Then the host, Monty Hall, who knows what are inside the boxes, opens another box, say box B, which is empty. He then says to you, "Do you want to abandon your box and pick box C?"

Is it to your advantage to switch your choice?

Solution:

Let A, B, C be the event that the keys are in box A, B, C respectively. Let M be the event that Monty Hall opens box B. Then consider the three different cases:

If the key is in box A, then both B and C are empty and Monty can randomly open any one of them at his own choice, therefore $\Pr(M|A) = \frac{1}{2}$.

If the key is in box B, then Monty surely won't open box B, therefore $\Pr(M|B) = 0$.

If the key is in box C, then Monty can only open box B because surely he won't open A or C, therefore $\Pr(M|C) = 1$.

Using Bayes' rule,

$$\Pr(A|M) = \frac{\Pr(M|A)\Pr(A)}{\Pr(M|A)\Pr(A) + \Pr(M|B)\Pr(B) + \Pr(M|C)\Pr(C)} = \frac{\left(\frac{1}{2}\right)\left(\frac{1}{3}\right)}{\left(\frac{1}{2}\right)\left(\frac{1}{3}\right) + (0)\left(\frac{1}{3}\right) + (1)\left(\frac{1}{3}\right)} = \frac{1}{3}.$$

Obviously, $\Pr(C|M) = 1 - \Pr(A|M) = 1 - \frac{1}{3} = \frac{2}{3}$.

Hence given the information that Monty Hall opened box B, the probability that the key is in box C is $2/3$ which doubles the probability that the key is in box A. Therefore it is a reasonable decision to trade box A for box C.

Remark:

Many people would think that since we know that box B is empty after Monty had opened box B, the chance that the key is in either boxes will become half and it would make no difference to switch or not. However, this kind of reasoning actually misinterpreted the information we obtained from Monty's action and lead to the following incorrect calculation:

$$\Pr(A|B^c) = \frac{\Pr(A \cap B^c)}{\Pr(B^c)} = \frac{\Pr(A)}{1 - \Pr(B)} = \frac{\frac{1}{3}}{1 - \frac{1}{3}} = \frac{1}{2}.$$

However, M is not the same as B^c. What we should calculate is $\Pr(A|M)$ but not $\Pr(A|B^c)$.

Note that the above calculation was based on the assumption that $\Pr(M|A) = \frac{1}{2}$, i.e., the host have no preference on choosing the box to open if he have two choices. In general, if we assume that $\Pr(M|A) = p$ (small p means he would prefer opening box C than box B), then the general result would become

$$\Pr(A|M) = \frac{p}{1+p}, \qquad \Pr(C|M) = \frac{1}{1+p}.$$

Since $\frac{1}{1+p} > \frac{1}{2}$ for all $0 \le p < 1$, trading box A for box C is always a reasonable decision as long as $p \neq 1$.

Extension:

If there are more boxes, the benefit of trading the initially selected box for the remaining box is clearer.

Using similar notations for five boxes A, B, C, D, E, events A, B, C, D, E refer to the cases that the key is in the corresponding boxes. Suppose again the player chooses box A, and Monty Hall opens boxes B, C and D (denote by M this event). The analysis can be carried out similarly as follows.

$$\Pr(M|A) = \frac{1}{4}, \qquad \Pr(M|B) = \Pr(M|C) = \Pr(M|D) = 0, \qquad \Pr(M|E) = 1.$$

With $\Pr(A) = \Pr(B) = \Pr(C) = \Pr(D) = \Pr(E) = \frac{1}{5}$, the Bayes' rule gives

$$\Pr(A|M) = \frac{\Pr(M|A)\Pr(A)}{\Pr(M|A)\Pr(A) + \cdots + \Pr(M|E)\Pr(E)} = \frac{\left(\frac{1}{4}\right)\left(\frac{1}{5}\right)}{\left(\frac{1}{4}\right)\left(\frac{1}{5}\right) + 0 + (1)\left(\frac{1}{5}\right)} = \frac{1}{5}.$$

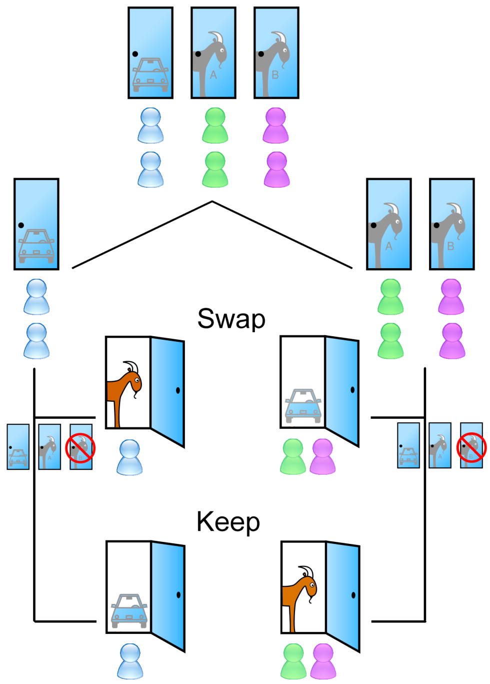
Thus, $\Pr(E|M) = 1 - \Pr(A|M) = \frac{4}{5}$.

The probability that the key is NOT in the initially selected box is very large. This effect is signified when the number of boxes increases.

Conclusion:

No matter which box is chosen, Monty Hall is able to open all boxes remaining your initial choice and one other box. That means, the probability that the key is in your initially chosen box is the probability that you have selected the box with the key initially among all the boxes, which is rather low. On the other hand, the probability that the key is in the other remained box is the probability that you have not selected the box with the key initially among all the boxes, which is rather high, since there are more boxes not containing the key.

Therefore, it is beneficial to change the choice of box.



Example 3.18.

Consider the following scenario. Suppose there is a bag originally containing a ball which can either be black or white, with equal chances. We put a white ball in the bag and then randomly draw one ball. If the randomly picked ball is white, what is the chance that the remaining ball is also white?

Note:

A common misinterpretation on the given information may lead to the answer of 0.5 as there are only two possible color combinations based on the observed information: two white balls, one white ball one black ball. However if we examine the conditional probability carefully, we can see that given our observation, these two possibilities are not equally likely.

Solution:

Let WW be the event that the original ball is white and BW be the event that the original ball is black. Let A be the observed event that the randomly drawn ball is white. Then we want to compute the conditional probability $\Pr(WW|A)$.

By using the Bayes' theorem,

$$\begin{aligned}\Pr(WW|A) &= \frac{\Pr(WW)\Pr(A|WW)}{\Pr(WW)\Pr(A|WW)+\Pr(BW)\Pr(A|BW)} \\ &= \frac{0.5\times 1}{0.5\times 1+0.5\times 0.5} \\ &= \frac{2}{3}.\end{aligned}$$

Hence it is more likely for the remaining ball to be a white ball than a black ball.

From this simple example, we see that conditional probabilities can be quite slippery and require careful interpretation.

3.3 Recurrence Relation

Consider a sequence $\{p_n\}$, if there is a linear relationship between consecutive terms, such as $p_{n+1} = ap_n + b$ or $ap_{n+2} + bp_{n+1} + cp_n = 0$, where a, b, c are independent of n, then the sequence satisfies a *recurrence relation* or a *difference equation.*

There are situations where the probability of the n-th event depends on the result of the previous event(s). In such cases, the calculation is made easier by deriving some recurrence relation linking probabilities of related events. The following examples help to illustrate the techniques of construction and manipulation of the recurrence relations.

Example 3.19.

The probability of a team winning a match is 0.8 and of losing it is 0.2 if the previous match was won. If the previous match was lost the corresponding probabilities are 0.4 and 0.6. There is an even chance that the team wins in the first match.

(a) Find the probabilities that the team wins the second, third, and fourth match.

(b) Find the probability that the team wins the n-th match.

(c) In the long run, how often will the team win?

Solution:

Let p_n denote the probability that the team wins the n-th match. It is given that $p_1 = 0.5$.

(a) The probability that the team wins the second match is

$$p_2 = 0.5(0.8) + 0.5(0.4) = 0.6.$$

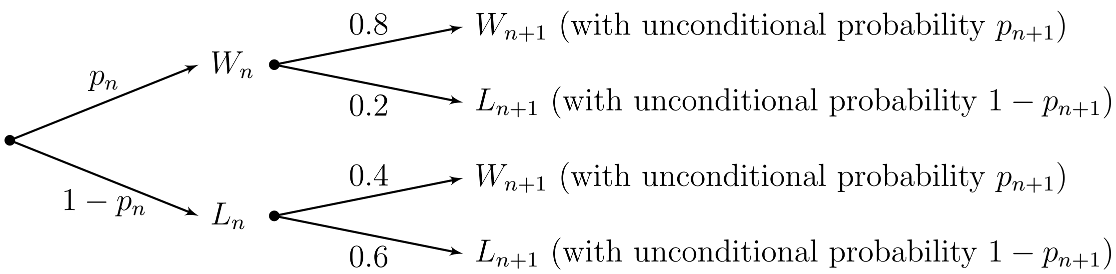
The probability that the team wins the third match is

$$p_3 = 0.5(0.8)(0.8) + 0.5(0.2)(0.4) + 0.5(0.4)(0.8) + 0.5(0.6)(0.4) = 0.64.$$

The probability that the team wins the fourth match is

$$\begin{aligned} p_4 &= 0.5(0.8)(0.8)(0.8) + 0.5(0.8)(0.2)(0.4) + 0.5(0.2)(0.4)(0.8) + 0.5(0.2)(0.6)(0.4) \\ &\quad +0.5(0.4)(0.8)(0.8) + 0.5(0.4)(0.2)(0.4) + 0.5(0.6)(0.4)(0.8) + 0.5(0.6)(0.6)(0.4) \\ &= 0.656. \end{aligned}$$

These probabilities can also be obtained by using a recurrence relation. The following tree diagram shows the relationship between p_n and $1 - p_n$. Let W_n and L_n respectively denote the events that the team wins and loses in the n-th match. (L_n is the complement of W_n.)



From the tree diagram,

$$p_{n+1} = p_n \times 0.8 + (1 - p_n) \times 0.4,$$

or

$$p_{n+1} = 0.4 + 0.4p_n \qquad (\star).$$

Putting $n = 1, 2, 3$ in the recurrence relation $(\star)$ gives

$$\begin{aligned} p_2 &= 0.4 + 0.4p_1 = 0.4 + 0.4(0.5) = 0.6, \\ p_3 &= 0.4 + 0.4p_2 = 0.4 + 0.4(0.6) = 0.64, \\ p_4 &= 0.4 + 0.4p_3 = 0.4 + 0.4(0.64) = 0.656. \end{aligned}$$

(b) The probability that the team wins the n-th match is

$$\begin{aligned}
p_n &= 0.4 + 0.4p_{n-1} \\
&= 0.4 + 0.4\left(0.4 + 0.4p_{n-2}\right) \\
&= 0.4 + 0.4^2 + 0.4^2 p_{n-2} \\
&= 0.4 + 0.4^2 + 0.4^2\left(0.4 + 0.4p_{n-3}\right) \\
&= 0.4 + 0.4^2 + 0.4^3 + 0.4^3 p_{n-3} \\
&= \cdots \\
&= 0.4 + 0.4^2 + 0.4^3 + \cdots + 0.4^{n-2} + 0.4^{n-1} + 0.4^{n-1}p_1 \\
&= \frac{0.4(1-0.4^{n-1})}{1-0.4} + 0.4^{n-1}p_1 \\
&= \frac{2}{3}(1-0.4^{n-1}) + \frac{1}{2}(0.4)^{n-1} \\
&= \frac{2}{3} - \frac{5}{12}(0.4)^n,
\end{aligned}$$

which is the solution to ($\star$).

(c) The required long-run proportion of winning is

$$\begin{aligned}
\lim_{n\to\infty}\frac{p_1+\cdots+p_n}{n} &= \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}p_i = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\left[\frac{2}{3}-\frac{5}{12}(0.4)^i\right] \\
&= \frac{2}{3}-\frac{5}{12}\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}0.4^i = \frac{2}{3}-\frac{5}{12}\lim_{n\to\infty}\frac{0.4(1-0.4^n)}{0.6n} = \frac{2}{3}-\frac{5}{12}(0) = \frac{2}{3}.
\end{aligned}$$

Hence, in the long run, the team wins two-thirds of the time.

Alternatively, we can simply calculate $\lim_{n\to\infty} p_n$ (called the limiting probability) which always gives the same result:

$$\lim_{n\to\infty} p_n = \lim_{n\to\infty}\left[\frac{2}{3}-\frac{5}{12}(0.4)^n\right] = \frac{2}{3}-\frac{5}{12}(0) = \frac{2}{3}.$$

Remarks:

1. A mathematical result of the Cesàro mean states that for a convergent sequence $\{a_n\}$,
$$\lim_{n\to\infty}\frac{a_1+\cdots+a_n}{n} = \lim_{n\to\infty} a_n.$$
2. The limiting probability can also be obtained by taking limits in ($\star$). Let $p = \lim_{n\to\infty} p_n$, then
$$p = 0.4 + 0.4p \qquad \Longrightarrow 0.6p = 0.4 \qquad \Longrightarrow p = \frac{0.4}{0.6} = \frac{2}{3}.$$
3. For recurrence relation $p_{n+1} = a + rp_n$, with $p_1 = A$, the solution is
$$p_n = \frac{a(1-r^{n-1})}{1-r} + Ar^{n-1},$$
and this can easily be proved by mathematical induction.

★

Example 3.20.

Let p_n be the probability that in n tosses of a fair coin no run of two consecutive heads appears.

(a) Show that, for $n \geq 2$, $p_n = \frac{1}{2}p_{n-1} + \frac{1}{4}p_{n-2}$.

(b) Find the values of p_5, p_6 and p_7.

Solution:

(a) Any such run must begin either with $T\ldots\ldots$ or $HT\ldots\ldots$. Thus,

$$\begin{aligned} & \Pr(\text{no } HH \text{ in } n \text{ tosses}) \\ = \; & \Pr(T)\Pr(\text{no } HH \text{ in } n-1 \text{ tosses}) + \Pr(HT)\Pr(\text{no } HH \text{ in } n-2 \text{ tosses}). \end{aligned}$$

In other words,

$$\begin{aligned} p_n &= \frac{1}{2} \times p_{n-1} + \frac{1}{2} \times \frac{1}{2} \times p_{n-2} \\ &= \frac{1}{2}p_{n-1} + \frac{1}{4}p_{n-2}. \end{aligned}$$

(b) Note that

$$\begin{aligned} p_1 &= 1, \\ p_2 &= 1 - \frac{1}{4} = \frac{3}{4}, \\ p_3 &= \frac{1}{2}p_2 + \frac{1}{4}p_1 = \frac{1}{2}\left(\frac{3}{4}\right) + \frac{1}{4}(1) = \frac{5}{8}, \\ p_4 &= \frac{1}{2}p_3 + \frac{1}{4}p_2 = \frac{1}{2}\left(\frac{5}{8}\right) + \frac{1}{4}\left(\frac{3}{4}\right) = \frac{1}{2}, \\ p_5 &= \frac{1}{2}p_4 + \frac{1}{4}p_3 = \frac{1}{2}\left(\frac{1}{2}\right) + \frac{1}{4}\left(\frac{5}{8}\right) = \frac{13}{32}, \\ p_6 &= \frac{1}{2}p_5 + \frac{1}{4}p_4 = \frac{1}{2}\left(\frac{13}{32}\right) + \frac{1}{4}\left(\frac{1}{2}\right) = \frac{21}{64}, \\ p_7 &= \frac{1}{2}p_6 + \frac{1}{4}p_5 = \frac{1}{2}\left(\frac{21}{64}\right) + \frac{1}{4}\left(\frac{13}{32}\right) = \frac{17}{64}. \end{aligned}$$

Remark:

$$p_n = \left(\frac{5+3\sqrt{5}}{10}\right)\left(\frac{1+\sqrt{5}}{4}\right)^n + \left(\frac{5-3\sqrt{5}}{10}\right)\left(\frac{1-\sqrt{5}}{4}\right)^n.$$

★

3.3.1 First Step Analysis

The *first step analysis* is a powerful technique for solving probability problems with hierarchical structures. When the conditional situation after given the first stage (or some particular intermediate stage) repeats or resembles the original situation of the problem, we can use the law of total probability to set up some recursion equations. By solving the recursion equation, the desired probabilities can be easily calculated.

Example 3.21.

Two players, Paul and Quinn, take turns to flip a fair coin. Paul plays first. The first player who flips a head wins. What is the probability that Paul wins?

Solution:

Consider the possible scenarios after the first flip by Paul.

Case 1: The first flip by Paul is head.

Paul wins. Therefore $\Pr(\text{Paul wins}|\text{Head}) = 1$.

Case 2: The first flip by Paul is tail.

It becomes Quinn's turn. If Quinn's flip is head, Paul will lose. If Quinn's flip is tail, it will become Paul's turn again and the game will be reset to the original situation.

Therefore, $\Pr(\text{Paul wins}|\text{Tail}) = \frac{1}{2}\Pr(\text{Paul wins})$.

Combining these two scenarios using law of total probability, we have

$$\begin{aligned}
\Pr(\text{Paul wins}) &= \Pr(\text{Head})\Pr(\text{Paul wins}|\text{Head}) + \Pr(\text{Tail})\Pr(\text{Paul wins}|\text{Tail}) \\
&= \frac{1}{2}\times 1 + \frac{1}{2}\times\frac{1}{2}\Pr(\text{Paul wins}) \\
&= \frac{1}{2} + \frac{1}{4}\Pr(\text{Paul wins}).
\end{aligned}$$

Solving this equation gives $\Pr(\text{Paul wins}) = \frac{2}{3}$.

★

Example 3.22.

A single cell can either die, with probability 0.25, or split into two cells, with probability 0.75. Each cell in the new generation dies or splits into two cells independently with the same probabilities as the initial cell. What is the probability that eventually the whole population of these cells will become extinct?

Solution:

Denote p_n the probability that the n-th generation of a cell does not exist (i.e., all the offspring of the cell had died before the n-th generation). Consider the possible outcomes on the initial cell.

Case 1: The initial cell dies.

The population extincts and there will be no n-th generation of this cell for any $n > 1$.

Case 2: The initial cell splits into two cells.

Since the n-th generation of the initial cell are the $(n-1)$-th generation of the new cells, there will be no n-th generation of the initial cell if and only if both the two new cells has no $(n-1)$-th generation offspring.

Combining these two scenarios using law of total probability, we have

$$\begin{aligned} p_n &= \frac{1}{4} \times 1 + \frac{3}{4} \times p_{n-1} \times p_{n-1} \\ &= \frac{1}{4}\left(1 + 3p_{n-1}^2\right). \end{aligned}$$

Solving this recursion equation is a difficult mathematical problem. Fortunately, we just need to determine the limit, say $P = \lim_{n\to\infty} p_n$, and it must satisfy

$$P = \frac{1}{4}\left(1 + 3P^2\right),$$

which gives two roots, $P = \frac{1}{3}$ or $P = 1$. To determine which root is the correct limit, we may rewrite the recursion equation as

$$\begin{aligned} p_n - \frac{1}{3} &= \frac{3}{4}\left(p_{n-1}^2 - \frac{1}{9}\right) \\ &= \frac{3}{4}\left(p_{n-1} + \frac{1}{3}\right)\left(p_{n-1} - \frac{1}{3}\right), \end{aligned}$$

which reveals that if $p_{n-1} < \frac{1}{3}$ then $p_n < \frac{1}{3}$. Since $p_2 = \frac{1}{4} < \frac{1}{3}$, we have $p_n < \frac{1}{3}$ for all $n > 1$. As a result,

$$P = \lim_{n\to\infty} p_n = \frac{1}{3}.$$

That is, the population will become extinct with probability $\frac{1}{3}$.

★

Example 3.23.

Suppose n fair dice are rolled. What is the probability that the sum of their faces is divisible by 7?

Solution:

In **Example 2.25.** in **Chapter 2**, we have calculated that the probability is $\frac{31}{216}$ for $n = 4$. Here we can make use of the first step approach to determine the probability for general n. Denote p_n the probability that the sum of n rolled dice is divisible by 7. Consider the following two scenarios, given the outcome of the first $n-1$ dice.

Case 1: The sum of the first $n-1$ dice is divisible by 7.

In this case, the sum of the n dice must not be divisible by 7 as the outcome of the last die can be 1 to 6 only.

Case 2: The sum of the first $n-1$ dice is not divisible by 7.

In this case, the remainder after the sum of the first $n-1$ dice is divided by 7 must be from 1 to 6. No matter what the remainder is, there will be one and only one outcome of the last dice that would result in an overall sum divisible by 7, and the conditional probability is equal to $\frac{1}{6}$.

Combining these two scenarios using the law of total probability, we have

$$\begin{aligned} p_n &= p_{n-1} \times 0 + (1 - p_{n-1}) \times \frac{1}{6} \\ &= \frac{1}{6} - \frac{1}{6} p_{n-1}. \end{aligned}$$

Solving this recursion equation, with $p_1 = 0$, gives

$$p_n = \frac{1}{7}\left[1 - \left(-\frac{1}{6}\right)^{n-1}\right], \qquad n = 1, 2, 3, \ldots$$

In particular, when $n = 4$,

$$p_4 = \frac{1}{7}\left[1 - \left(-\frac{1}{6}\right)^{3}\right] = \frac{31}{216},$$

which is consistent with the result obtained in **Example 2.25.**.

★

∼ End of Chapter 3 ∼