Probability and Statistics

Fall semester, 2019

Chapter 1　Introduction to Probability

1.1 Random Experiments

- **Deterministic phenomenon**

  For example, the sun rises every morning and sets every evening.

- **Random phenomenon**

  Examples include tossing a coin, rolling a die, knowing one’s score before taking the exam, tomorrow’s weather, etc.

Here are some typical examples：

Example 1.1.1. Tossing a coin and observing whether it lands heads or tails.

Example 1.1.2. Tossing a coin twice and observing the sequence of heads and tails.

Example 1.1.3. Tossing a die and observing the number of spots.

Example 1.1.4. Measuring the lifetime of a light bulb.

These trials have the following characteristics:

(1) The experiment can be repeated indefinitely under essentially the same conditions.

(2) The possible outcomes may not be unique and can be identified ahead of time.

(3) The outcome is uncertain before the experiment.

A random experiment is real or hypothetical process satisfying the above three conditions.

1.2 Sample Space

Definition 1.2.1.

- The collection of all possible outcomes of an experiment is called the **sample space** of the experiment, denoted by Ω or S.
- The elements in sample space are called **sample points** or **basic events**

Example 1.2.1. Rolling a Die. The sample space is $S = \{1,2,3,4,5,6\}$.

Example 1.2.2.
Tossing a coin twice and observing the sequence of heads(H) and tails(T)
The sample space is $S = \{HH, HT, TH, TT\}$.

Definition 1.2.2.

- An **event** is a well-defined subset of sample space of the experiment, denoted by A, B, C, etc.
- Sample space S is guaranteed to happen and called **the certain event**.
- The empty set $\varnothing$ that cannot happen is called **the impossible event**.

Example 1.2.1. Rolling a Die.

(1) One event A is that an even number is obtained: $A = \{2,4,6\}$.

(2) The event B is that a number greater than 2 is obtained: $B = \{3,4,5,6\}$.

Example 1.2.2.

Tossing a coin twice and observing the sequence of heads(H) and tails(T)

One event A is that the first toss results in a head: $A = \{HH,HT\}$.

1.3 Relations and Operations between Events

1. Containment

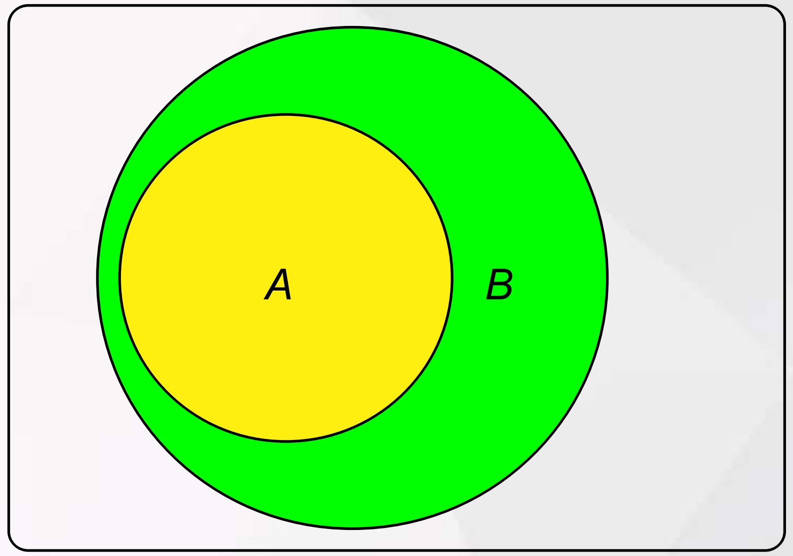
Definition 1.3.1.

For events, $A \subset B$ means that if A occurs then so does B.

Theorem 1.3.1.

Let A, B, and C be events.

- Then $A \subset S$. If $A \subset B$ and $B \subset A$, then $A = B$.
- If $A \subset B$ and $B \subset C$, then $A \subset C$.



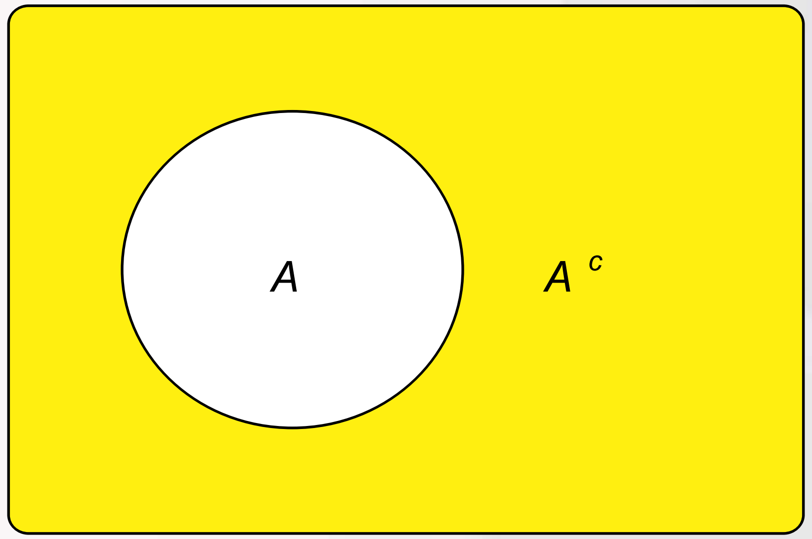
Example 1.3.1.

Rolling a Die. $A = \{2,4,6\}$, $C = \{2,3,4,5,6\}$, then $A \subset C$.

2. Complement

Definition 1.3.2.

The **complement** of a set A is defined to be the set that contains all elements of the sample space S that do not belong to A, denoted by A^c or $\overline{A}$.



Theorem 1.3.2.

Let A be an event, then $(A^c)^c = A$, $\varnothing^c = S$, $S^c = \varnothing$

Example 1.3.2.

Rolling a Die. Suppose A is the event that an even number is rolled, then $A = \{2,4,6\}$, and $A^c = \{1,3,5\}$.

3. Union

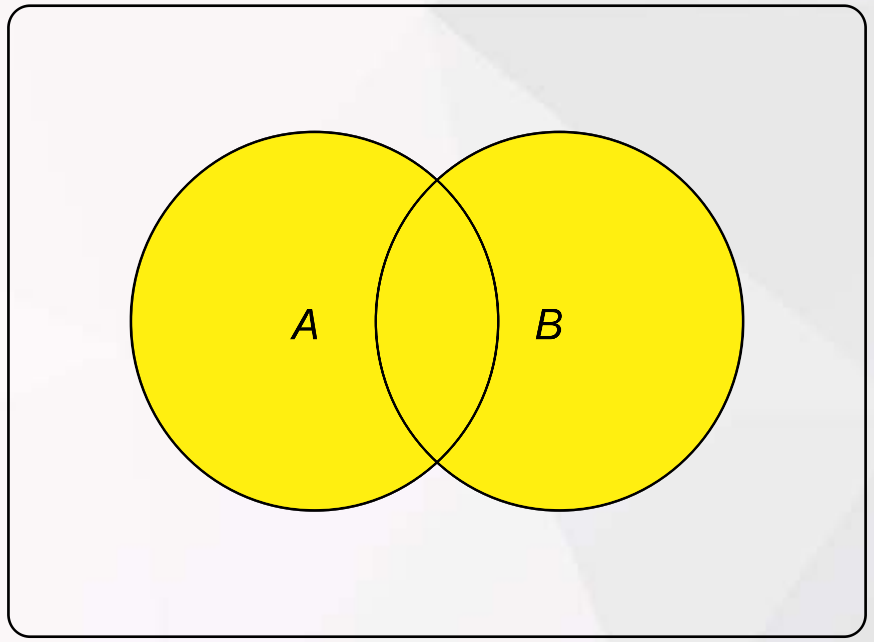
Definition 1.3.3.

The notation for the **union** of A and B is $A \cup B$. For events, $A \cup B$ means that A occurs, or B occurs, or both occur.

Remark

for all sets A and B,

- $A \cup B = B \cup A$, $A \cup A = A$, $A \cup A^c = S$, $A \cup \varnothing = A$, $A \cup S = S$.
- if $A \subset B$, then $A \cup B = B$.



The concept of union can be extended to more than two sets.

4. The union of n sets or an infinite sequence of sets

Definition 1.3.4.

- **The union of n sets $A_1,\cdots,A_n$** is defined to be the set that contains all outcomes that belong to at least one of these n sets.

$$A_1 \cup A_2 \cup \cdots \cup A_n, \text{ or } \bigcup_{i=1}^{n} A_i$$

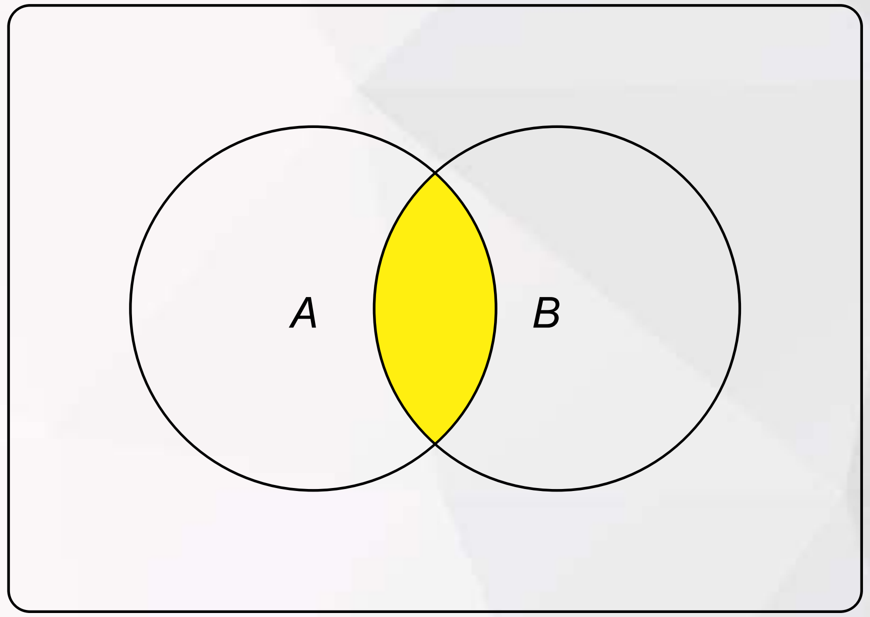
- **The union of an infinite sequence of sets $A_1, A_2, \cdots$** is the set that contains all outcomes that belong to at least one of the events in the sequence.

$$\bigcup_{i=1}^{\infty} A_i$$

5. Intersection

Definition 1.3.5.

The intersection of A and B is defined to be the set that contains all outcomes that belong both to A and to B, denoted by **$A \cap B$ or AB.**



Remark

for all sets A and B,

- $A \cap B = B \cap A, A \cap A = A, A \cap A^c = \varnothing$, $A \cap \varnothing = \varnothing$, $A \cap S = A$.
- if $A \subset B$, then $A \cap B = A$.

The concept of intersection can also be extended to more than two sets.

6. The intersection of n sets or an infinite sequence of sets

Definition 1.3.4.

- **The intersection of n sets $A_1, \cdots, A_n$** is defined to be the set that contains the elements that are common to all these n sets.

$$A_1 \cap A_2 \cap \cdots \cap A_n, \text{ or } \bigcap_{i=1}^{n} A_i$$

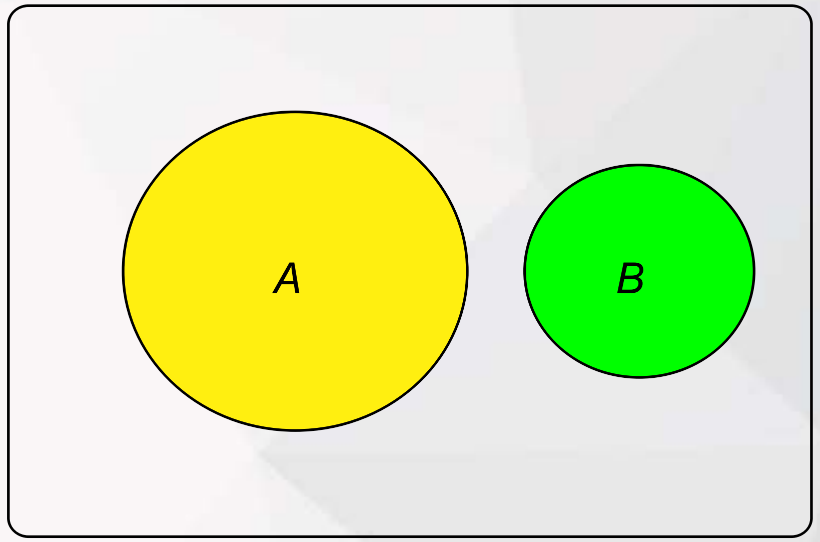
- **The intersection of an infinite sequence of sets $A_1, A_2, \cdots$** is the set that contains all outcomes that that are common to all the events in the sequence.

$$\bigcap_{i=1}^{\infty} A_i$$

7. Mutually exclusive

Definition 1.3.7.

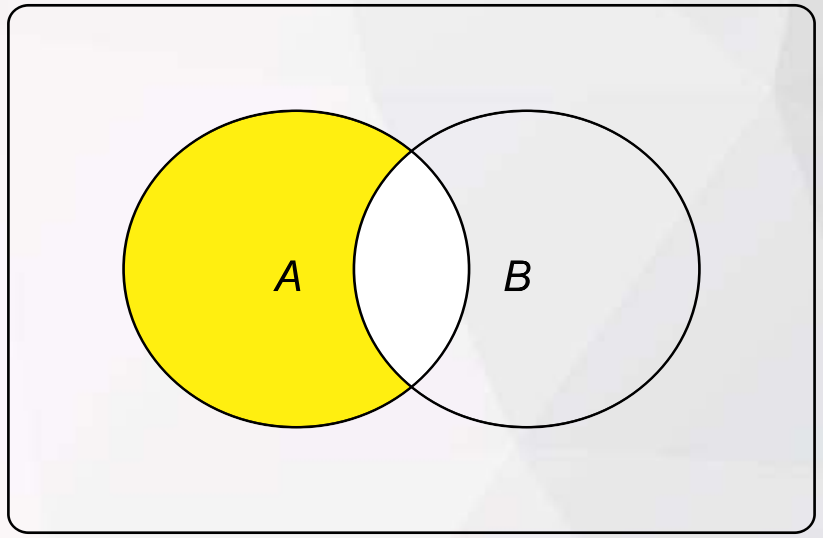
- It is said that two sets A and B are **disjoint, or mutually exclusive**, if **$A \cap B = AB = \varnothing$.**
- For n sets $A_1, \cdots, A_n$, if any two of them are disjoint, then they are called **pairwise disjoint events**.



8. Difference

Definition 1.3.8.

The **difference** of two sets A and B, which we write as **$A - B$**, is the set formed by all the elements that belong to A, but do not belong to B.



Remark

(1)$A - B = AB^c = A - AB$. (2)$A - B$ is not same as $B - A$.

Theorem 1.3.3.

For events $A, B, C, A_1, A_2, \cdots$, the operations satisfy the following properties:

- **Associative Property**

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

- **Distributive Properties**

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

- **De Morgan's Laws**

$$(A \cup B)^c = A^c \cap B^c, (A \cap B)^c = A^c \cup B^c$$
$$(\cap A_i)^c = \cup A_i^c, \quad (\cup A_i)^c = \cap A_i^c$$

Example 1.3.3.

Suppose that a coin is tossed three times. Then the sample space S contains the following eight possible outcomes: (H-head and T-tail) $S = \{HHH, THH, HTH, HHT, HTT, THT, TTH, TTT\}$

Let A be the event that at least one head is obtained in the three tosses;
let B be the event that a head is obtained on the second toss;
let C be the event that a tail is obtained on the third toss;
and let D be the event that no heads are obtained.

Find: $A^c, B \cap D, A \cup C, B \cap C, (B \cup C)^c, A \cap (B \cup C)$.

Solution:

$A = \{HHH, THH, HTH, HHT, HTT, THT, TTH\}$
$B = \{HHH, THH, HHT, THT\}$
$C = \{HHT, HTT, THT, TTT\}$
$D = \{TTT\}$

$A^c = D = \{TTT\}$,
$B \cap D = \varnothing$,
$A \cup C = S$,
$B \cap C = \{HHT, THT\}$,
$(B \cup C)^c = \{HTH, TTH\}$,
$A \cap (B \cup C) = \{HHH, THH, HHT, HTT, THT\}$.

Example 1.3.4.

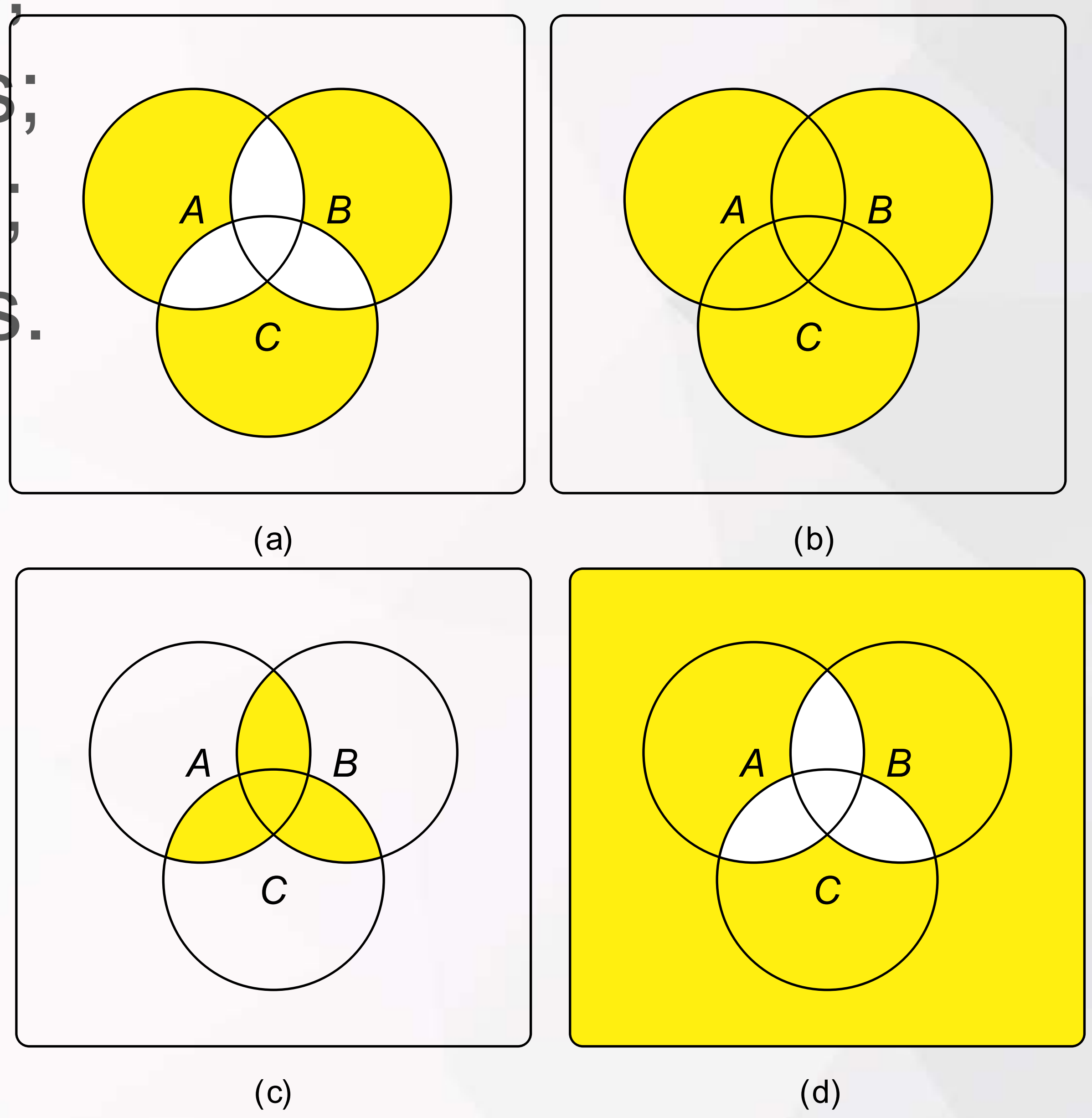
Express each of the following events in terms of the events A, B and C

(a) exactly one of the events A, B, C occurs;

(b) at least one of the events A, B, C occurs;

(c) at least two of the events A, B, C occurs;

(d) at most one of the events A, B, C occurs.



Solution:

(a) $AB^cC^c \cup A^cBC^c \cup A^cB^cC$.

(b) $A \cup B \cup C$.

(c) $AB \cup AC \cup BC$.

(d) $AB^cC^c \cup A^cBC^c \cup A^cB^cC \cup A^cB^cC^c$

or $A^cB^c \cup A^cC^c \cup B^cC^c$.

1.4 The Definition of Probability

Definition 1.4.1.

In a given experiment, for any **event A** in the **sample space S**, a real number P(A) is assigned. **P(A) indicates the probability that A will occur** if the set function P(•) **satisfies the following three specific axioms:**

1. Nonnegativity : For every event A, P(A) ≥ 0,
2. Normality : P(S) = 1,
3. Countable additivity : For every infinite sequence of disjoint events $A_1, A_2, \cdots$,

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

Proposition 1.4.1. $P(\varnothing) = 0$.

Proof.

$$P(\emptyset) = P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i) = \sum_{i=1}^{\infty} P(\emptyset)$$

Since $P(\varnothing) \geq 0$, then $P(\varnothing) = 0$.

Proposition 1.4.2.

For every finite sequence of n disjoint events $A_1, \cdots, A_n$, $P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$

Proof.

Consider the infinite sequence of events $A_1, A_2, \cdots$,

- $A_1, \cdots, A_n$ are the n given disjoint events;
- $A_i = \varnothing$ for i = n+1,n+2,….

$$\begin{aligned} P\left(\bigcup_{i=1}^{n} A_i\right) &= P\left(\bigcup_{i=1}^{\infty} A_i\right) \\ &= \sum_{i=1}^{\infty} P(A_i) \\ &= \sum_{i=1}^{n} P(A_i) + 0 \\ &= \sum_{i=1}^{n} P(A_i) \end{aligned}$$

Proposition 1.4.3. For every event A, $P(A^c) = 1 - P(A)$.

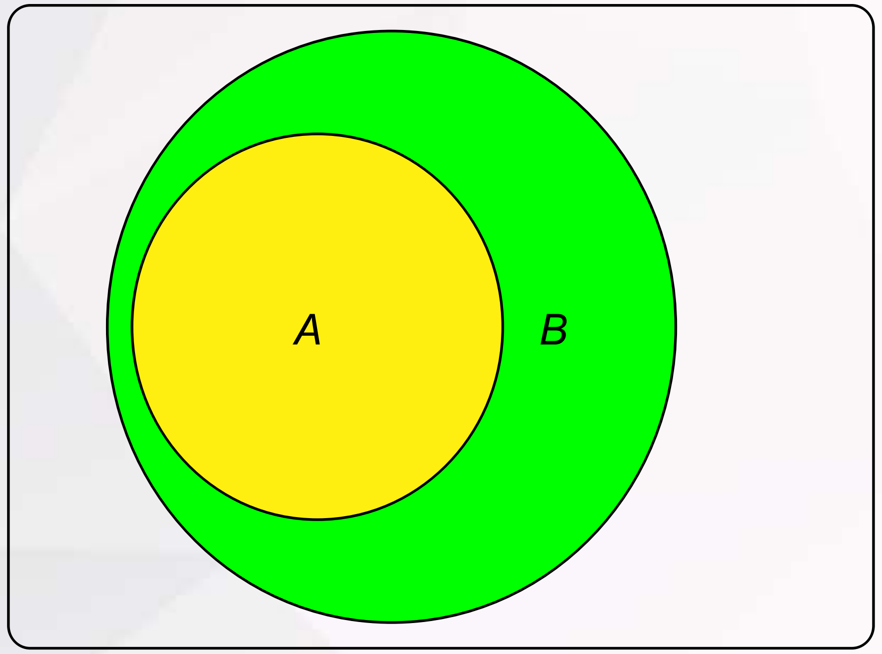
Proof.

Since A and A^c are disjoint events and $A \cup A^c = S$, $1 = P(S) = P(A \cup A^c) = P(A) + P(A^c)$

$$P(A^c) = 1 - P(A)$$

Proposition 1.4.4. If $A \subset B$, then $P(B - A) = P(B) - P(A)$, $P(A) \leq P(B)$.

Proof.



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

A and $B - A$ are disjoint, therefore,
$P(B) = P(A) + P(B - A) \geq P(A)$.

Proposition 1.4.5. For every event A, $0 \le P(A) \le 1$.

Proof.

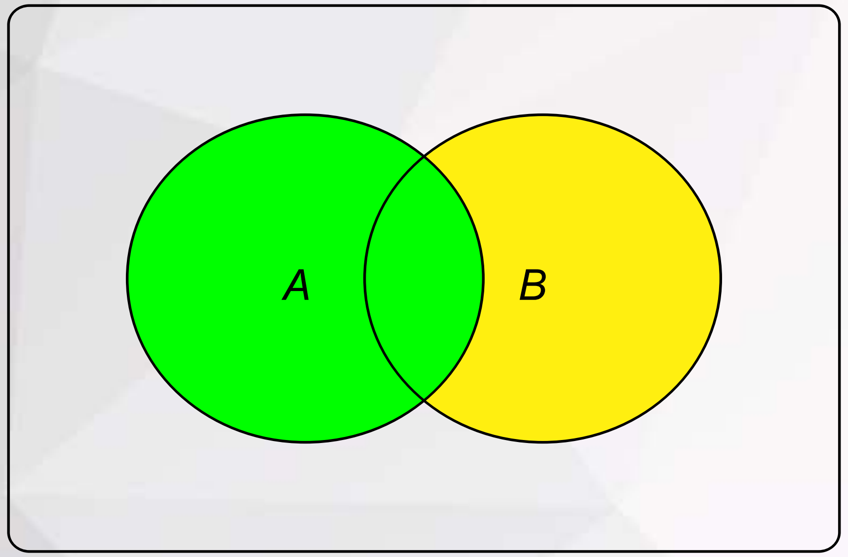
$A \subset S$,

by Proposition 1.4.4 and Axiom 2, $P(A) \le P(S) = 1$.

Proposition 1.4.6.

For every two events A and B, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Proof.



$A \cup B = A \cup BA^c$, where A and BA^c are disjoint,

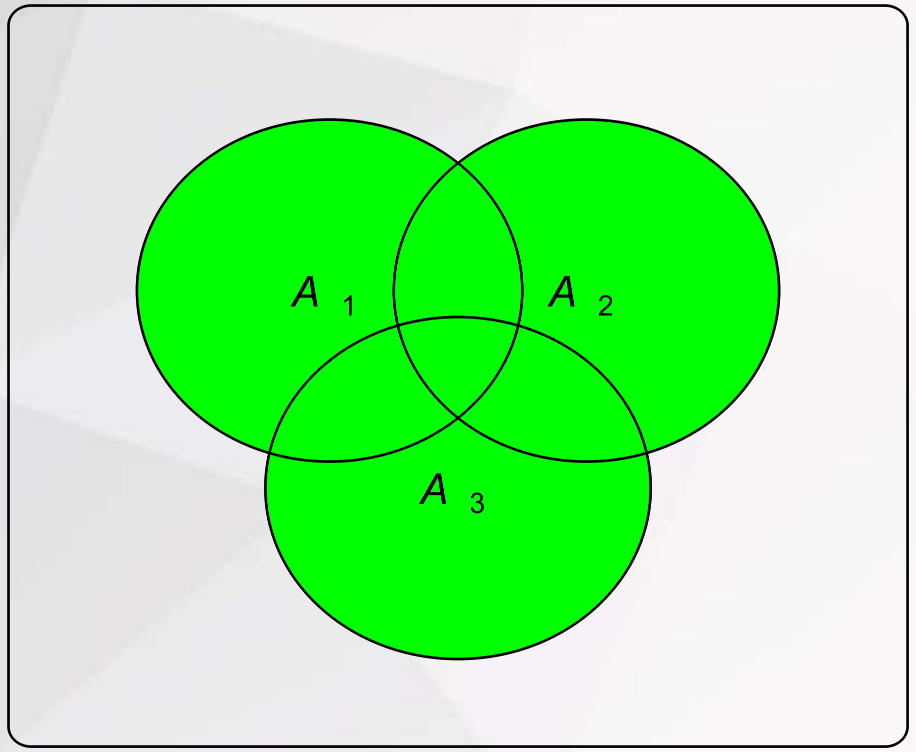
Hence, by Proposition 1.4.2 and 1.4.4,

$P(A \cup B) = P(A) + P(BA^c) = P(A) + P(B) - P(AB)$.

Proposition 1.4.7. For every three events A_1, A_2, and A_3,

$$P(A_1 \cup A_2 \cup A_3) = P(A_1) + P(A_2) + P(A_3) - [P(A_1A_2) + P(A_2A_3) + P(A_1A_3)] + P(A_1A_2A_3).$$

Proof.



$$P(A_1 \cup A_2 \cup A_3)$$
$$= P(A_1) + P(A_2 \cup A_3) - P[A_1 \cap (A_2 \cup A_3)]$$
$$= P(A_1) + P(A_2 \cup A_3) - P[A_1A_2 \cup A_1A_3]$$
$$= P(A_1) + P(A_2) + P(A_3) - P(A_2A_3) - [P(A_1A_2) + P(A_1A_3) - P(A_1A_2A_3)]$$
$$= P(A_1) + P(A_2) + P(A_3) - [P(A_1A_2) + P(A_2A_3) + P(A_1A_3)] + P(A_1A_2A_3).$$

Proposition 1.4.8. For every n events $A_1, \cdots, A_n$,

$$P(\cup_{i=1}^{n} A_i)$$

$$=\sum_{i=1}^{n} P(A_i) - \sum_{1\le i<j\le n} P(A_iA_j) + \sum_{1\le i<j<k\le n} P\left(A_iA_jA_k\right) - \ldots$$

$$+(-1)^{n+1}P(A_1A_2 \ldots A_n).$$

Proof.

The conclusion can be proved by mathematical induction.

1.5 Equally Likely Outcomes Model

We consider the **equally likely outcomes model** (**classical model**). The experiments have the two features:

(1) The sample space S contains only a finite number of points $s_1, \cdots, s_n$.

(2)The probability assigned to each of the outcomes $s_1, \cdots, s_n$ is $\frac{1}{n}$

The sample space is called **simple sample space**.

Proposition 1.5.1.

If an event A contains exactly m outcomes, then

$$P(A) = \frac{\text{number of outcomes in } A}{\text{number of outcomes in } S} = \frac{m}{n}$$

Example 1.5.1.

Suppose that **three fair coins** are tossed **simultaneously**. We shall determine **the probability of obtaining exactly two heads.** (H-head and T-tail)

Solution:

The sample space is $S = \{HHH, THH, HTH, HHT, HTT, THT, TTH, TTT\}$.

The probabilities of the basic events are $P(\{HHH\}) = \cdots = P(\{TTT\}) = \frac{1}{8}$

A=obtaining exactly two heads={HHT,HTH,THH}, $P(A) = \frac{3}{8}$

Example 1.5.2.

Two fair Dice are rolled. The sample space S comprises the following 36 outcomes:

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

(1) What is the probability that **the sum of the two numbers is 6**?

(2) What is the probability that **the sum of the two numbers that appear is larger than 2**?

Solution:

(1) $A = \{(1,5),(2,4),(3,3),(4,2),(5,1)\}$ $\quad P(A) = \frac{5}{36}$

(2) $B^c = \{(1,1)\}$. $\quad P(B) = 1 - P(B^c) = 1 - \frac{1}{36} = \frac{35}{36}$

Proposition 1.5.2. (Multiplication Rule)

Suppose that an experiment has k steps ($k \geq 2$), that the i-th step can have n_i possible outcomes ($i = 1, \cdots, k$), then the total number of elements in S will be equal to the product $n_1 \times n_2 \times \cdots \times n_k$.

Example 1.5.3.

Suppose that we toss six coins.

Since there are two possible outcomes for each of the six coins, the number of sample points in S is $2^6 = 64$.

The event that observing exactly one head contains 6 sample points. Therefore,

$$P(\text{exactly one head}) = \frac{6}{64} = \frac{3}{32}.$$

Proposition 1.5.3. (Permutations)

Selecting k of the elements one at a time without replacement from a set of n elements. Each such outcome is called a permutation, the number of distinct such permutations is

$$P_n^k = n \times (n-1) \times \cdots \times (n-k+1)$$

When $k = n$,

$$P_n^n = n! = n \times (n-1) \times \cdots \times 2 \times 1$$

$0! = 1.$

Example 1.5.4.

Suppose that a club consists of 25 members and that a president and a secretary are to be chosen from the membership. Find: the total possible number of ways.

Solution:

$$P_{25}^2 = 25 \times 24 = 600$$

Example 1.5.5.

Suppose that six different books are to be arranged on a shelf. The number of possible permutations of the books is

Solution:

$$6! = 720.$$

Proposition 1.5.4. (Combinations)

Consider a set with n elements. Each subset of size k chosen from this set is called a combination, the number of distinct such combinations is

$$C_n^k = \frac{P_n^k}{k!} = \frac{n!}{k!(n-k)!}$$

Example 1.5.6.

Consider the set {a,b,c,d} containing the four different letters. We want to count the number of distinct subsets of size two.

Solution:

{a,b},{a,c},{a,d},{b,c},{b,d},{c,d}.

$$C_4^2 = \frac{4!}{2! \times 2!} = 6$$

This is different from counting permutaions because {a,b} and {b,a} are the same subset.

Example 1.5.7.

(1) A committee composed of 8 people is to be selected from a group of 20 people.

$$C_{20}^8 = \frac{20!}{8! \times 12!} = 125970$$

(2) The eight people in the committee each get a different job to perform on the committee.

$$P_{20}^8 = C_{20}^8 \times 8! = 5,078,110,400$$

Remark

The number C_n^k is also denoted by

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = C_n^k$$

Be named as → Binomial coefficient

For all numbers x and y and each positive integer n,

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$$

Proposition 1.5.5.

The number of different arrangements of n objects consisting of k similar objects of one type and n − k similar objects of a second type is

$$C_n^k = \binom{n}{k}$$

Example 1.5.8.

Suppose that a fair coin is to be tossed 10 times, and it is desired to determine the probability of obtaining exactly three heads.

Solution:

The total possible number of different sequences of 10 heads and tails is 2^{10}, and it may be assumed that each of these sequences is equally probable.

The number of these sequences that contain exactly three heads will be equal to the number of different arrangements that can be formed with three heads and seven tails $\binom{10}{3}$

$$P = \frac{\binom{10}{3}}{2^{10}} = 0.1172.$$

Example 1.5.9.

A box contains 4 white balls and 2 black balls. Two balls are drawn with replacement.

(1) What is the probability of getting all white balls?

(2) If two balls are drawn without replacement, what is the probability of getting all white balls?

Solution:

(1) For the first time, there are 6 balls to be selected. Since it is sampling with replacement, there are also 6 balls to be selected for the second time. By multiplication rule, there are 6×6 ways in sample space. Each of these possible ways is equally probable. For the event that getting all white balls, there are 4×4 ways. Therefore,

$$P = \frac{4 \times 4}{6 \times 6} = \frac{4}{9}$$

Example 1.5.9.

(2) If two balls are drawn without replacement, what is the probability of getting all white balls?

Solution:

(2) For the first time, there are 6 balls to be selected. Since it is sampling without replacement, there are only 5 balls to be selected for the second time. By multiplication rule, there are 6×5 ways in sample space. Each of these possible ways is equally probable. For the event that getting all white balls, there are 4×3 ways. Therefore,

$$P = \frac{4 \times 3}{6 \times 5} = \frac{2}{5}$$

Example 1.5.9.

(2) If two balls are drawn without replacement, what is the probability of getting all white balls?

Solution:

Another method:

Suppose we don't distinguish between balls of the same color and don't care about the order in which the balls are selected, then there are C_6^2 ways to select 2 balls from 6 balls. And there are C_4^2 ways to select 2 white balls from 6 white balls.

$$P = \frac{C_4^2}{C_6^2} = \frac{2}{5}$$

Example 1.5.10.

Suppose that there are 25 people in a room, and each person's birthday is uniformly distributed over the 365 days (ignore leap year) of the year and independent of the other peoples' birthday. What is the probability that no two people share a birthday.

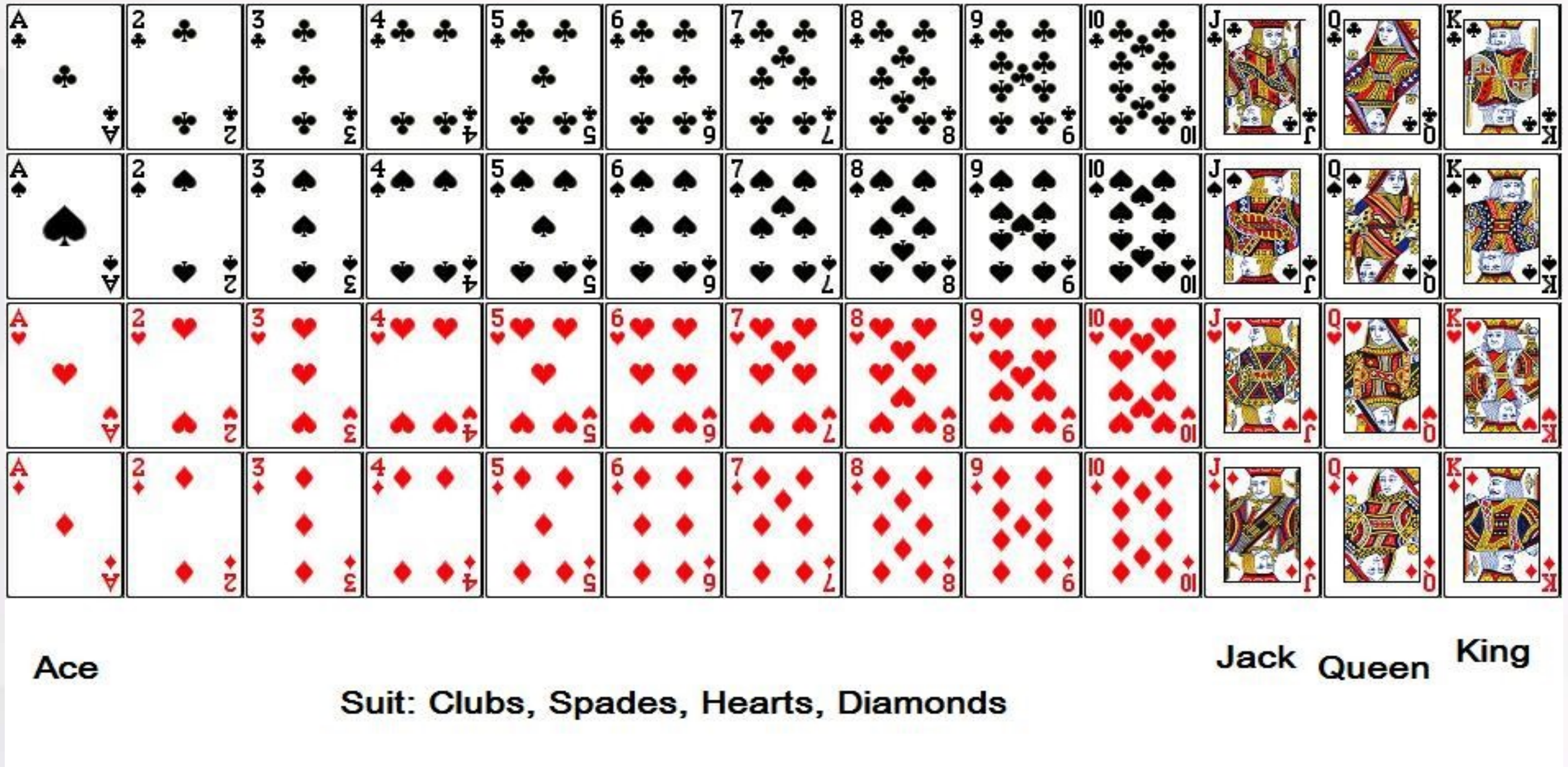
Solution:

By multiplication rule, there are 365^{25} ways for 25 people's birthdays. Since no two people share a birthday, the first person has a possibility of 365 days as being his birthday. The second person has 364 different days that could be his and not shared with the first person. The next person has only 363 open days. Therefore,

$$P = \frac{365 \times 364 \times \cdots \times (365 - 25 + 1)}{365^{25}}$$

Example 1.5.11.

Suppose that a deck of 52 cards containing four aces is shuffled thoroughly and the cards are then distributed among four players so that each player receives 13 cards. We shall determine the probability that each player will receive one ace.



Solution:

Method 1:

Using unordered samples, the number of possible combinations of the four positions occupied by the four aces is $\binom{52}{4}$

If each player is to receive one ace, then there are 13 possible positions for the ace that the first player is to receive, 13 other possible positions for the ace that the second player is to receive, and so on. Therefore, exactly 13^4 *combinations will lead to the desired result. Hence,*

$$P = \frac{13^4}{\binom{52}{4}} = 0.1055$$

Solution:

Method 2:

Using ordered samples, we distinguish outcomes with the same cards in different orders, there are 13^4 ways to choose the four positions for the four aces, 4! ways to arrange the four aces in these four positions, there are 48! ways to arrange the remaining 48 cards in the 48 remaining positions.

$$P = \frac{4! \times 13^4 \times 48!}{52!} = 0.1055$$

Solution:

Method 3:

Consider the four players, for the first player, there are $\binom{52}{13}$ ways to receive the 13 cards, for the second player, there are $\binom{39}{13}$ ways to receive the 13 cards from the remaining 39 cards, and so on, the sample space contains $\binom{52}{13} \times \binom{39}{13} \times \binom{26}{13} \times \binom{13}{13}$ elements.

If each player is to receive one ace, 4! ways to distribute the four aces to the four players. There are $\binom{48}{12}$ ways to choose 12 cards and distribute them to the first player to ensure that he has 13 cards. There are $\binom{36}{12}$ ways to choose 12 cards and distribute them to the second player, and so on. Therefore,

$$P = \frac{4! \times \binom{48}{12} \times \binom{36}{12} \times \binom{24}{12} \times \binom{12}{12}}{\binom{52}{13} \times \binom{39}{13} \times \binom{26}{13} \times \binom{13}{13}} = 0.1055$$

Example 1.5.12.

Suppose that a person types n letters, types the corresponding addresses on n envelopes, and then places the n letters in the n envelopes in a random manner. Determine the probability that at least one letter will be placed in the correct envelope.

Solution:

Let A_i be the event that letter i is placed in the correct envelope

(i = 1,⋯ ,n), we shall determine the value $P_n = P(\bigcup_{i=1}^{n} A_i)$

$$\sum_{i=1}^{n} P(A_i) = n \times \frac{1}{n} = 1$$

$$\sum_{1\le i<j\le n} P(A_iA_j) = \binom{n}{2}\frac{1}{n(n-1)} = \frac{1}{2!}$$

$$\sum_{1\le i<j<k\le n} P(A_iA_jA_k) = \binom{n}{3}\frac{1}{n(n-1)(n-2)} = \frac{1}{3!}$$

$$\cdots = \cdots$$

Solution:

Therefore, by Proposition 1.4.8,

$$
\begin{aligned}
P_n &= \sum_{i=1}^{n} P(A_i) - \sum_{1\le i<j\le n} P(A_iA_j) \\
&\quad + \sum_{1\le i<j<k\le n} P(A_iA_jA_k) - \cdots + (-1)^{n+1} P(A_1A_2\cdots A_n) \\
&= 1 - \frac{1}{2!} + \frac{1}{3!} - \frac{1}{4!} + \cdots + (-1)^{n+1}\frac{1}{n!}.
\end{aligned}
$$

Example 1.5.13.

Suppose that 7 guests check their hats when they arrive at a restaurant, and that these hats are returned to them in a random order when they leave. Determine the probability that

(1) no guest will receive the proper hat.

(2) exactly 2 guests will receive their own hats.

Solution:

(1) Let A_i be the event that guest i receives his own hat
(i = 1, ⋯ ,n), we shall determine the value

$$\begin{aligned}
P(n) &= P(\bar{A}_1 \cap \bar{A}_2 \cdots \cap \bar{A}_n) \\
&= P(\overline{\cup_{i=1}^{n} A_i}) \\
&= 1 - P(\cup_{i=1}^{n} A_i) \\
&= 1 - [1 - \frac{1}{2!} + \frac{1}{3!} - \frac{1}{4!} + \cdots + (-1)^{n+1}\frac{1}{n!}] \\
&= \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} + \cdots + (-1)^{n}\frac{1}{n!}
\end{aligned}$$

Solution:

(1) *Then, n = 7, the probability that no guest will receive the proper hat is*

$$P(7) = \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \frac{1}{5!} + \frac{1}{6!} - \frac{1}{7!} = \frac{103}{280} = 0.3679$$

(2) *Let X be the number of guests who receive their own hats.*

$$\begin{aligned} P(X=2) &= C_7^2 \times \frac{1}{7} \times \frac{1}{6} \times P(5) \\ &= \frac{1}{2}\left[\frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \frac{1}{5!}\right] \\ &= \frac{11}{60} = 0.1833 \end{aligned}$$

Solution:

(2) *In fact, we can compute all the probabilities* $P(X=k),\ k=0,1,\cdots,7.$

$$
\begin{aligned}
P(X=0) &= \frac{1}{2!}-\frac{1}{3!}+\frac{1}{4!}-\frac{1}{5!}+\frac{1}{6!}-\frac{1}{7!}=\frac{103}{280}\\
P(X=1) &= C_7^2\times\frac{1}{7}\times[\frac{1}{2!}-\frac{1}{3!}+\frac{1}{4!}-\frac{1}{5!}+\frac{1}{6!}]=\frac{53}{144}\\
P(X=2) &= C_7^2\times\frac{1}{7}\times\frac{1}{6}\times[\frac{1}{2!}-\frac{1}{3!}+\frac{1}{4!}-\frac{1}{5!}]=\frac{11}{60}\\
P(X=3) &= C_7^3\times\frac{1}{7}\times\frac{1}{6}\times\frac{1}{5}\times[\frac{1}{2!}-\frac{1}{3!}+\frac{1}{4!}]=\frac{1}{16}\\
P(X=4) &= C_7^4\times\frac{1}{7}\times\frac{1}{6}\times\frac{1}{5}\times\frac{1}{4}\times[\frac{1}{2!}-\frac{1}{3!}]=\frac{1}{72}\\
P(X=5) &= C_7^5\times\frac{1}{7}\times\frac{1}{6}\times\frac{1}{5}\times\frac{1}{4}\times\frac{1}{3}\times[\frac{1}{2!}]=\frac{1}{240}\\
P(X=6) &= 0\\
P(X=7) &= \frac{1}{7!}=\frac{1}{5040}
\end{aligned}
$$

Proposition 1.5.6.

Suppose the sample space of an experiment is a given domain $\Omega \in R^n$. The outcomes are points in Ω and are equally likely to occur. Let A be any subregion of Ω. Then the probability that basic events fall into A is

$$P(A) = \frac{S_A}{S_\Omega}$$

where S_A, S_Ω are the geometric measures of A and Ω, respectively.

Example 1.5.14.

Two friends agree to meet up in a station. They arrive to the station uniformly between 12:00pm and 13:00pm. Suppose that after arriving randomly, each waits 30 minutes for the other person before leaving. What is the chance the two will meet?

Solution:

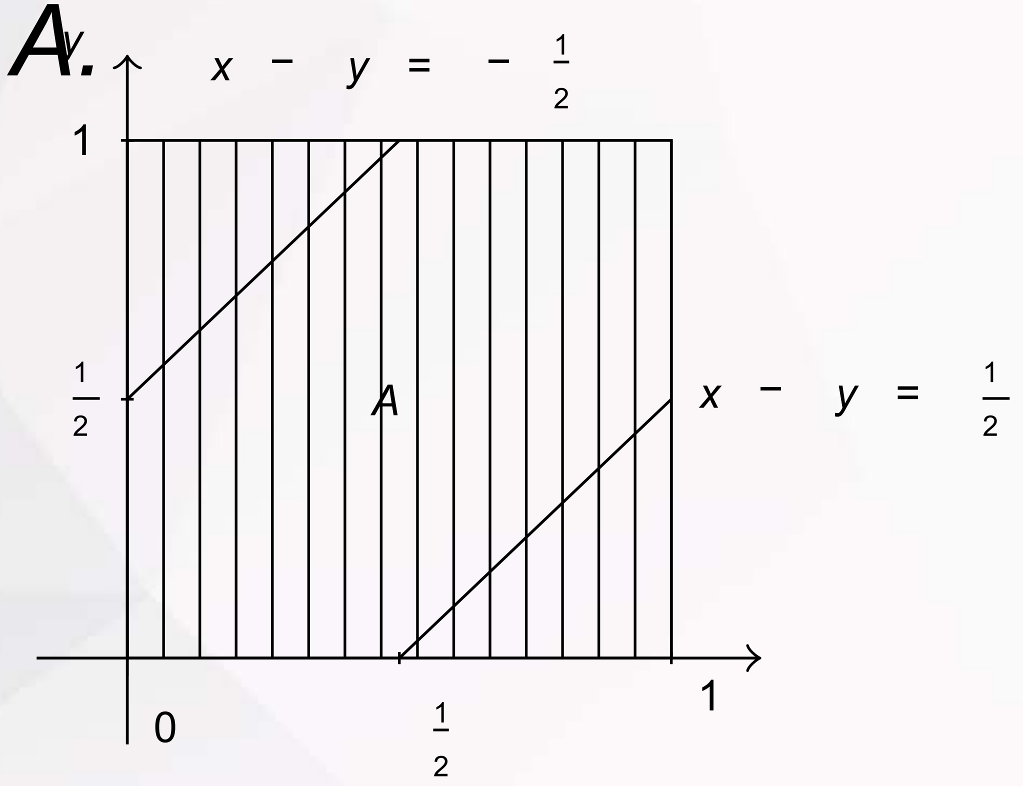
Let x denote the time one person arrives, and y the other, we can use the notation (x,y) to denote the time in hours, after 12:00pm, that each person arrives, where $0 \leq x \leq 1, 0 \leq y \leq 1$. The sample space Ω is a square with side length 1. Any coordinate in the square represents a time that the pair arrives.

Solution:

What is the set of coordinates for which the two will meet? We want the two coordinates to be within 30 minutes or $\frac{1}{2}$ hours of each other.

That is, we want all coordinates such that $|x-y| \le \frac{1}{2}$.

We draw the lines $y = x + \frac{1}{2}$ and $y = x - \frac{1}{2}$ and shade the area in between these two lines to denote the event A.



$$P(A) = \frac{S_A}{S_\Omega} = \frac{3/4}{1} = \frac{3}{4}$$

Remark

What is the chance the two will arrive at the same time? The set is a line $x = y$ which has no area. The probability that the two will arrive at the same time is zero. But clearly they can arrive at the same time, for example, both arrive at 12:30pm, etc. It implies that zero probability does not mean impossible.

1.6 Conditional Probability

Example 1.6.1.

In a group of 100 people, the numbers who do or do not smoke and do or do not have cancer are reported as shown in the table.

	Smoke	Not Smoke	Total
Cancer	20	5	25
No Cancer	30	45	75
Total	50	50	100

If a person is selected at random and it is found that he smokes, what is the probability that he has a cancer?

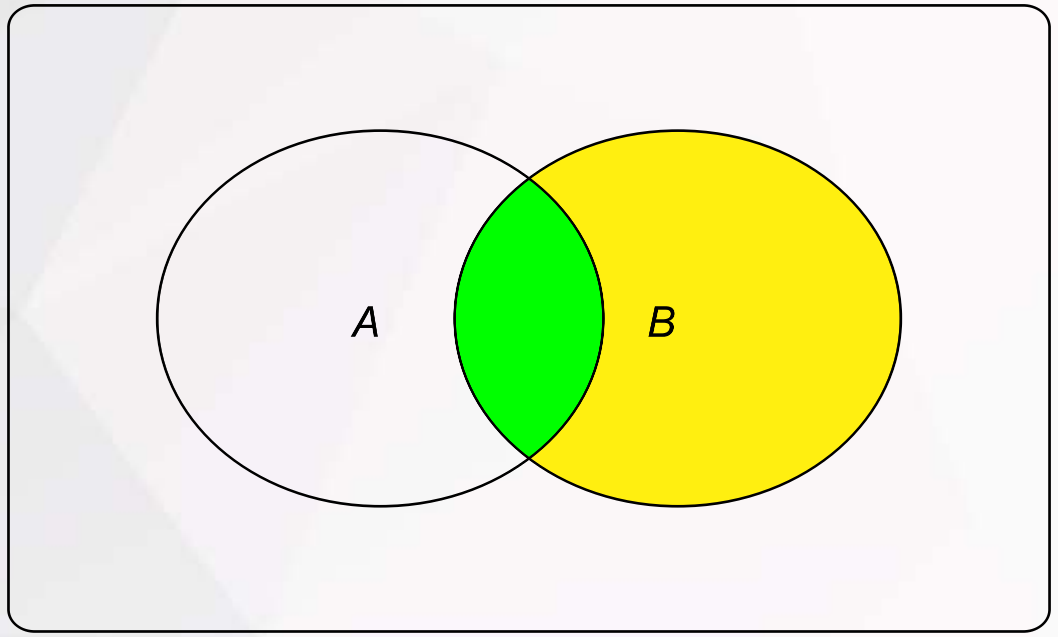
Solution:

The sample space contains 100 people. Since additional information is given, the sample space is reduced. There are only 50 elements left. The probability is $\frac{20}{50} = 0.4$

Definition 1.6.1.

Suppose that we learn that an event B has occurred and that we wish to compute the probability of another event A taking into account that we know that B has occurred.

The new probability of A is called **the conditional probability** of the event A given that the event B has occurred and is denoted as **P(A|B)**. If P(B) > 0, we compute this probability as



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

Remark

Unconitional probability P(A) can be regarded as the conditional probability of the event A given that the certain event or sample space Ω has occurred, since

$$P(A|\Omega) = \frac{P(A\Omega)}{P(\Omega)} = P(A).$$

Example 1.6.2.

You know only that your colleague has three children. You ask him if he is lucky enough to have at least one daughter and he says yes. What is the probability that all three children are girls?

Solution:

Let A be the event that all three children are girls and let B be the event that at least one daughter. We shall determine P(A|B).

$$A \subset B, A \cap B = A, P(AB) = P(A) = \frac{1}{8}$$

$$P(B) = 1 - P(B^c) = 1 - \frac{1}{8} = \frac{7}{8}$$

$$P(A|B) = \frac{P(AB)}{P(B)} = \frac{1/8}{7/8} = \frac{1}{7}$$

Example 1.6.3.

Suppose that two dice were rolled and it was observed that the sum T of the two numbers was odd. We shall determine the probability that T was less than 8.

Solution:

If we let A be the event that T < 8 and let B be the event that T is odd, then A ∩ B is the event that T is 3, 5, or 7.

$$\begin{aligned} P(A\cap B) &= P(T=3 \text{ or } 5 \text{ or } 7) \\ &= P(T=3)+P(T=5)+P(T=7) \\ &= \frac{2}{36}+\frac{4}{36}+\frac{6}{36} \\ &= \frac{1}{3} \end{aligned}$$

$$\begin{aligned} P(B) &= P(T=3 \text{ or } 5 \ \cdots \text{ or } 11) \\ &= P(T=3)+P(T=5)\cdots+P(T=11) \\ &= \frac{2}{36}+\frac{4}{36}+\cdots+\frac{2}{36} \\ &= \frac{1}{2} \end{aligned}$$

$$P(A|B)=\frac{P(AB)}{P(B)}=\frac{1/3}{1/2}=\frac{2}{3}$$

Example 1.6.4.

Suppose that two dice are to be rolled repeatedly and the sum T of the two numbers is to be observed for each roll. We shall determine the probability p that the value T = 7 will be observed before the value T = 8 is observed.

Solution: (Method 1):

We could assume that the sample space S contains all sequences of outcomes that terminate as soon as either the sum T = 7 or the sum T = 8 is obtained. Then we could find the sum of the probabilities of all the sequences that terminate when the value T = 7 is obtained. For each roll, we have

$$P(T=7)=\frac{1}{6}, P(T=8)=\frac{5}{36}, P(T\neq 7 \text{ and } T\neq 8)=1-\frac{11}{36}=\frac{25}{36}.$$

Solution:

Let p_i be the probability that the first 7 occurs on the i-th position and no 7 or 8 occurs before the i-th position.

$$p_1 = \frac{1}{6}, p_2 = \frac{25}{36} \times \frac{1}{6}, p_3 = (\frac{25}{36})^2 \times \frac{1}{6}, \cdots$$

Therefore,

$$p = p_1 + p_2 + \cdots = \frac{1}{6} \times \frac{1}{1 - \frac{25}{36}} = \frac{6}{11}$$

Solution: (Method 2):

If we repeat the experiment until either the sum T = 7 or the sum T = 8 is obtained, the effect is to restrict the outcome of the experiment to one of these two values. When we get either 7 or 8, we get the answer to the question 'Is 7 observed before 8?' Once we get either a 7 or an 8, we know the result of the run. Any rolls we make subsequently will not affect the outcome, so we might ignore them.

Hence, the problem can be restated as follows: Given that the outcome of the experiment is either T = 7 or T = 8, determine the probability p that the outcome is actually T = 7.

$$p = P(T = 7 | T = 7 \text{ or } T = 8) = \frac{P(T = 7)}{P(T = 7 \text{ or } T = 8)} = \frac{1/6}{5/36 + 1/6} = \frac{6}{11}$$

Solution: **(Method 3)**

Solution: Let

- A be the event that $T = 7$ is observed before the value $T = 8$ is observed.
- B_i be the event that T is not 7 nor 8 for the $1, \ldots, (i-1)$th roll, but $T = 7$ or 8 for the ith roll.

It is obvious that $B_1, B_2, \ldots$ are disjoint. Let $B = \bigcup_{i=1}^{\infty} B_i$, then $A \subset B$ and

$$P(B) = \sum_{i=1}^{\infty} P(B_i) = \sum_{i=1}^{\infty} (1 - \frac{11}{36})^{i-1} \frac{11}{36} = 1.$$

$$P(A) = P(AB) = \sum_{i=1}^{\infty} P(AB_i) = \sum_{i=1}^{\infty} (1 - \frac{11}{36})^{i-1} \frac{1}{6} = \frac{6}{11},$$

which is the desired probability.

Proposition 1.6.1.

Similar to probability, conditional probability satisfies the three axioms: if P(B) > 0,

1. Nonnegativity: For every event A, P(A|B) ≥ 0,
2. Normality: P(S|B) = 1,
3. Countable additivity: For every infinite sequence of disjoint events $A_1, A_2, \cdots$,

$$P\left(\bigcup_{i=1}^{\infty} A_i | B\right) = \sum_{i=1}^{\infty} P(A_i|B)$$

In addition, conditional probability also holds the main properties of probability such as $P(A|B) + P(A^c|B) = 1.$

Theorem 1.6.1.
(The Multiplication Rule for Conditional Probabilities)

- Let A and B be events. If P(B) > 0, then $P(A \cap B) = P(B) \times P(A|B)$
- If P(A) > 0, then $P(A \cap B) = P(A) \times P(B|A)$

Example 1.6.5.

Consider a bag of marbles, containing 10 red, 20 blue, and 15 green marbles. Suppose that two marbles are drawn without replacement. What is the probability that both marbles drawn are red?

Solution:

Let A be the event that the first marble is red, and let B be the event that the second marble is red.

$$P(A \cap B) = P(A) \times P(B|A) = \frac{10}{45} \times \frac{9}{44} = \frac{1}{22}$$

Theorem 1.6.2.

Suppose that $A_1, A_2, \cdots, A_n$ are events such that $P(A_1 \cap A_2 \cdots \cap A_{n-1}) > 0$. Then

$$P(A_1 \cap A_2 \cdots \cap A_n) = P(A_1)P(A_2|A_1)\cdots P(A_n|A_1A_2\cdots A_{n-1}).$$

Example 1.6.6.

Three cards are dealt successively at random and without replacement from a standard deck of 52 playing cards. What is the probability of receiving, in order, a king, a queen, and a jack?

Solution:

Let A_i $(i = 1,2,3)$ denote the event that receiving, in order, a king, a queen, and a jack. Then

$$\begin{aligned} P(A_1 \cap A_2 \cap A_3) &= P(A_1)P(A_2|A_1)P(A_3|A_1A_2) \\ &= \frac{4}{52} \times \frac{4}{51} \times \frac{4}{50} \\ &= \frac{64}{132600}. \end{aligned}$$

1.7 Total Probability and Bayes' Theorem

- **Total probability**

  to obtain the desired result by calculating the probability of these simple events

- **Bayesian theorem**

  to discuss the possibility of the occurrence of each simple event under the condition that the result is given

Definition 1.7.1.

Let S denote the sample space of some experiment, and consider n events $B_1, \cdots, B_n$ in S such that

(1) $B_1, \cdots, B_k$ are disjoint; (2) $\cup_{i=1}^{n} B_i = S$

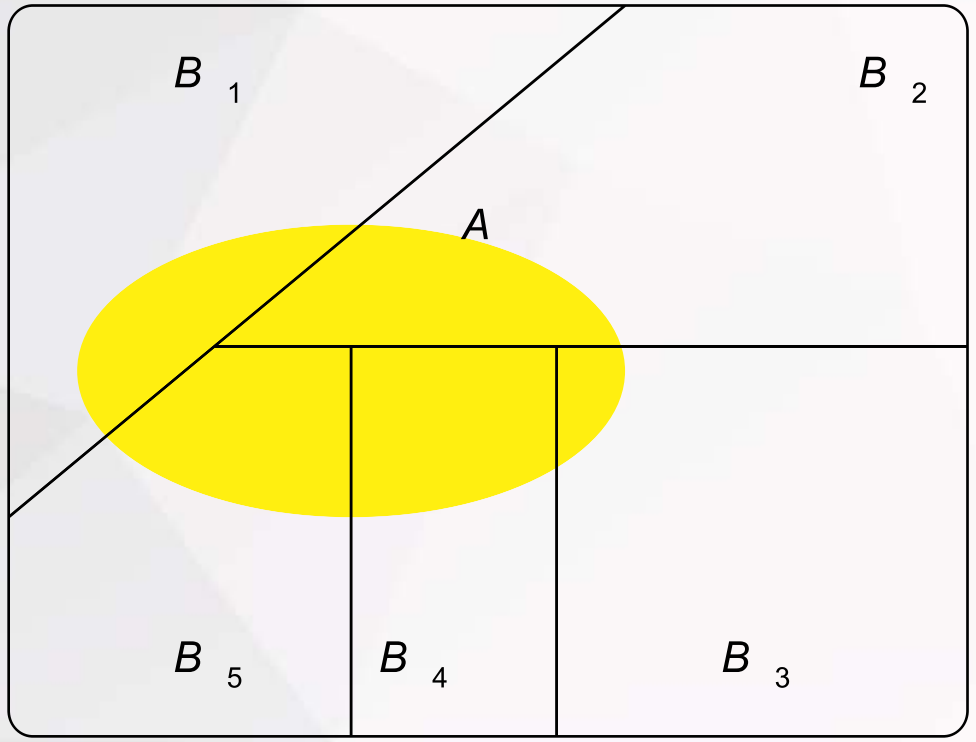
Then these events are called a partition of S.

Theorem 1.7.1. (Total Probability)

Suppose that the events $B_1, \cdots, B_n$ form a partition of the space S and $P(B_i) > 0$ for $i = 1, \cdots, n$. Then, for every event A in S,

$$P(A) = \sum_{i=1}^{n} P(B_i)P(A|B_i)$$

Proof.



$$\begin{aligned} P(A) &= P(A \cap S) \\ &= P[A \cap (\cup_{i=1}^{n} B_i)] \\ &= P[\cup_{i=1}^{n} (A \cap B_i)] \\ &= \sum_{i=1}^{n} P(A \cap B_i) \\ &= \sum_{i=1}^{n} P(B_i)P(A|B_i) \end{aligned}$$

Theorem 1.7.2. **(Bayes' Theorem)**

Suppose that the events $B_1, \cdots, B_n$ form a partition of the space S and P(B_i) > 0 for $i = 1, \cdots, n$. and let A be an event such that P(A) > 0, Then,

$$P(B_i|A) = \frac{P(B_i)P(A|B_i)}{\sum_{i=1}^{n} P(B_i)P(A|B_i)}, \quad i = 1, \cdots, n.$$

Proof.

$$\begin{aligned} P(B_i|A) &= \frac{P(AB_i)}{P(A)} \\ &= \frac{P(B_i)P(A|B_i)}{P(A)} \\ &= \frac{P(B_i)P(A|B_i)}{\sum_{i=1}^{n} P(B_i)P(A|B_i)}, \quad i = 1, \cdots, n. \end{aligned}$$

Remark

In the case where B and B^c form a partition of S, the simplified versions of the Total Probability and Bayes' Theorem are as shown below:

$$P(A) = P(B)P(A|B) + P(B^c)P(A|B^c),$$

$$P(B|A) = \frac{P(B)P(A|B)}{P(B)P(A|B) + P(B^c)P(A|B^c)}$$

Example 1.7.1.

A factory production line is manufacturing bolts using three machines, A,B and C. Of the total output, machine A is responsible for 25%, machine B for 35% and machine C for the rest. It is known from previous experience with the machines that 5% of the output from machine A is defective, 4% from machine B and 2% from machine C. A bolt is chosen at random from the production line.

(1) What is the probability that it is defective?

(2) If we learn that the bolt selected is defective, what is the probability that it came from machine A?

Solution:

Let D = {the bolt is defective}, A = {the bolt is from machine A}, B = {the bolt is from machine B}, C = {the bolt is from machine C}. We know that

$$P(A) = 0.25, P(B) = 0.35, P(C) = 0.4, P(D|A) = 0.05,$$
$$P(D|B) = 0.04, P(D|C) = 0.02.$$

(1) By the total probability formula,

$$\begin{aligned} P(D) &= P(A)P(D|A) + P(B)P(D|B) + P(C)P(D|C) \\ &= 0.25 \times 0.05 + 0.35 \times 0.04 + 0.4 \times 0.02 \\ &= 0.0345. \end{aligned}$$

(2) By the Bayes' formula,

$$\begin{aligned} P(A|D) &= \frac{P(AD)}{P(D)} \\ &= \frac{P(A)P(D|A)}{P(A)P(D|A) + P(B)P(D|B) + P(C)P(D|C)} \\ &= \frac{0.25 \times 0.05}{0.25 \times 0.05 + 0.35 \times 0.04 + 0.4 \times 0.02} \\ &= 0.3623. \end{aligned}$$

Example 1.7.2.

One bag contains 4 white balls and 3 black balls, and second bag contains 3 white balls and 5 black balls. One ball is drawn from the first bag and placed unseen in the second bag.

((1)) What is the probability that a ball now drawn from second bag is black?

(2) If a ball drawn from second bag is black, what is the probability that the ball drawn from the first bag is white?

Solution:

Let A=event that a ball drawn from the second bag is black,
B=event that a white ball is drawn from the first bag
and B^c =event that a black ball is drawn from the first bag.

Solution:

$$P(B)=\tfrac{4}{7},P(B^c)=\tfrac{3}{7},P(A|B)=\tfrac{5}{9},P(A|B^c)=\tfrac{6}{9}$$

(1) By *the total probability formula*,

$$\begin{aligned}P(A) &= P(B)\times P(A|B)+P(B^c)\times P(A|B^c)\\ &= \frac{4}{7}\times\frac{5}{9}+\frac{3}{7}\times\frac{6}{9}\\ &= \frac{38}{63}.\end{aligned}$$

(2) By *the Bayes' formula*,

$$\begin{aligned}P(B|A) &= \frac{P(B\cap A)}{P(A)}\\ &= \frac{P(B)\times P(A|B)}{P(A)}\\ &= \frac{P(B)\times P(A|B)}{P(B)\times P(A|B)+P(B^c)\times P(A|B^c)}\\ &= \frac{\frac{4}{7}\times\frac{5}{9}}{\frac{38}{63}}\\ &= \frac{10}{19}.\end{aligned}$$

Example 1.7.3.

Suppose that two dice are to be rolled repeatedly and the sum T of the two numbers is to be observed for each roll. We shall determine the probability p that the value $T = 7$ will be observed before the value $T = 8$ is observed.

Solution:

We have discussed this problem in Example 1.6.4. Now we want to propose another solution.

Solution:

Let B_1 be the event that the sum for the first roll is 7,
B_2 be the event that the sum for the first roll is 8 ,
B_3 be the event that the sum for the first roll is a value other than 7 or 8.

Let A be the event that the value T = 7 will be observed before the value T = 8 is observed. By the law of total probability,

$$\begin{aligned} P(A) &= P(B_1) \times P(A|B_1) + P(B_2) \times P(A|B_2) + P(B_3) \times P(A|B_3) \\ &= \frac{1}{6} \times 1 + \frac{5}{36} \times 0 + \frac{25}{36} \times P(A) \end{aligned}$$

Thus, P(A) = 6/11.

Example 1.7.4.

Approximately 1% of women aged 40-50 have breast cancer. A woman with breast cancer has a 90% chance of a positive test from a mammogram, while a woman without has a 10% chance of a false positive result. What is the probability a woman has breast cancer given that she just had a positive test?

Solution:

Let B be the event that the woman has breast cancer and A be the event that the test is positive.

We know $P(B) = 0.01, P(A|B) = 0.90, P(A|B^c) = 0.10.$

We need find $P(B|A)$.

Solution: *By the Bayes' formula,*

$$
\begin{aligned}
P(B|A) &= \frac{P(B) \times P(A|B)}{P(B) \times P(A|B) + P(B^c) \times P(A|B^c)} \\
&= \frac{0.01 \times 0.90}{0.01 \times 0.90 + (1 - 0.01) \times 0.10} \\
&= \frac{1}{12}.
\end{aligned}
$$

Remark

This answer is somewhat surprising and interesting. A positive test only means the woman has a $\frac{1}{12} \approx 8.3\%$ chance of cancer, rather than 90%. But it makes sense: the test gives a false positive 10% of the time, so there will be a lot of false positives if the population size is large enough. Most of the positive test results will be incorrect.

1.8 Independent Events

If the outcome of one event does not affect the outcome of another, events are independent.

Definition 1.8.1.

Two events A and B are independent if$P(A \cap B) = P(A) \times P(B)$.

Theorem 1.8.1.

If two events A and B are independent, and $P(B) > 0$, then $P(A|B) = P(A)$.

Proof.

It can be proved by the definition of conditional probability.

Example 1.8.1.

Suppose that a balanced die is rolled twice. Find the probability of getting at least one number which is not less than 5.

Solution:

Let A be the event that the number on the first roll is not less than 5,
B be the event that the number on the second roll is not less than 5.
A and B are physically unrelated and independent.

$$\begin{aligned} P(A\cup B) &= P(A)+P(B)-P(A\cap B) \\ &= P(A)+P(B)-P(A)\times P(B) \\ &= \frac{1}{3}+\frac{1}{3}-\frac{1}{3}\times\frac{1}{3} \\ &= \frac{5}{9}. \end{aligned}$$

Example 1.8.2.

Suppose that a balanced die is rolled. Let A be the event that an even number is obtained, and let B be the event that one of the numbers 1, 2, 3, or 4 is obtained. Show that the events A and B are independent.

Solution:

Sample space S = {1,2,3,4,5,6},A = {2,4,6},B = {1,2,3,4},A ∩ B = {2,4}.

$$P(A) = \frac{1}{2}, P(B) = \frac{2}{3}, P(AB) = \frac{1}{3}$$

$$P(AB) = P(A) \times P(B).$$

It follows that the events A and B are independent events, even though the occurrence of each event depends on the same roll of a die.

Theorem 1.8.2.

If two events A and B are independent, then the events A and B^c, A^c and B, A^c and B^c are also independent.

Proof.

$$
\begin{aligned}
P(AB^c) &= P(A) - P(AB) \\
&= P(A) - P(A)P(B) \\
&= P(A)[1 - P(B)] \\
&= P(A)P(B^c).
\end{aligned}
$$

Therefore, A and B^c are independent. The proofs for the remaining cases are similar.

Theorem 1.8.3.

Suppose that A and B are disjoint events for an experiment, each with positive probability. Then A and B are dependent.

Proof.

A and B are disjoint,

$$A \cap B = \varnothing, P(A \cap B) = 0.$$

$$P(A) \times P(B) > 0, \text{ so } P(A \cap B) \neq P(A) \times P(B).$$

A and B are dependent.

Definition 1.8.2.

Three events A,B, and C are mutually independent if the following four relations are satisfied:

$$
\begin{aligned}
P(AB) &= P(A)P(B) \\
P(AC) &= P(A)P(C) \\
P(BC) &= P(B)P(C) \\
P(ABC) &= P(A)P(B)P(C).
\end{aligned}
$$

Three events A,B, and C are pairwise independent if the first three relations above are satisfied.

Remark

Mutually and pairwise independence are two different concepts. Mutual independence implies pairwise independence.
But the converse is not true.

Example 1.8.3.

Suppose that a fair coin is tossed twice. The sample space
S = {HH,HT,TH,TT}. Consider the following three events:

A = {Head on first toss} = {HH,HT},
B = {Head on second toss} = {HH,TH},
C = {Both tosses the same} = {HH,TT}.

Then AB = AC = BC = ABC = {HH}. Hence,

$$P(AB) = P(AC) = P(BC) = P(ABC) = \frac{1}{4}$$

$$P(AB) = P(A)P(B), P(AC) = P(A)P(C), P(BC) = P(B)P(C)$$

$$P(ABC) \neq P(A)P(B)P(C).$$

So A,B, and C are pairwise independent, but not mutually independent.

Definition 1.8.3.

The n events $A_1, \cdots, A_n$ are independent (or mutually independent) if, for every subset $A_{i1}, \cdots, A_{ik}$ of k of these events ($k = 2, 3, \cdots, n$),

$$P(A_{i1} \cap A_{i2} \cdots \cap A_{ik}) = P(A_{i1})P(A_{i2}) \cdots P(A_{ik}).$$

Thank you!