



# Chapter 1 Basic Ideas in Probability

- 1.1 Basic Concepts
- 1.2 Computing Probabilities
- 1.3 Conditional Probability and Independence



## 1.1 Basic Concepts

- What's your understanding of probability?
- In daily life, people usually interpret probability as **the measure of an individual's degree of belief** in the statement that he or she is making.

### Example 1.1

Probability in daily life:

- It is 40% probable that 吴承恩 actually wrote the novel *Journey to the West* (西游记).
- The probability that Oswald acted alone in assassinating Kennedy is 0.8.
- You are 90% sure that you will receive at least an A if you work hard in this course.

- Using the concept of probability in this way is acceptable in daily life, but in academic contexts, probability has a more rigorous and precise definition.
- Probability theory studies **random events**, but whether the novel *Journey to the West* was written by 吴承恩 is not random, it is just unknown.



## 1.1 Basic Concepts

---

- What exactly does "random" mean?
- The concept of "random" refers to events or outcomes that cannot be predicted with certainty, even though the set of possible outcomes is known.
- Therefore, randomness implies uncertainty, but it is not entirely uncertain.
- The essence of probability theory is to transform individual randomness into overall certainty.
- Based on the understanding of "random," we introduce basic concepts such as random experiments, sample space, and random events.
- With these basic concepts, we can describe and study random phenomena using mathematical methods.



# 1.1 Basic Concepts

## Random Experiment

A **random experiment** (随机试验), often simply called an **experiment**, represents the realization or observation of a random phenomenon and has the following characteristics:

- it can be repeated under the same conditions;
- all possible outcomes are clearly known;
- exactly one of these possible outcomes occurs each time, but it cannot be determined in advance which outcome will occur.

## Sample Space and Random Event

- Each possible **fundamental outcome** (基本结果) of a random experiment is called a **sample point** (样本点), usually denoted as  $\omega$ .
- A set that includes all sample points of the experiment is called the **sample space** (样本空间), usually denoted as  $\Omega$ .
- From a set theory perspective, a **random event** (随机事件), or **event**, is a subset of the sample space  $\Omega$ , typically denoted by uppercase letters (e.g.,  $A, B, C$ , etc.)
- If the outcome is a sample point in event  $A$ , then we say that event  $A$  happens.



## 1.1 Basic Concepts

### Example 1.2

A person driving to work needs to pass through three traffic lights. At each traffic light, he either comes to a red light (denoted as 0) or a green light (denoted as 1). Then:

- The sample space is  $\Omega = \{000, 001, 010, 100, 011, 101, 110, 111\}$ ;
- The event that “just one of the three traffic lights is red” is  $A = \{011, 101, 110\} \subset \Omega$ .

### Example 1.3

A couple decides to have children and plans to have children until a boy is born. Write B for boy and G for girl, then:

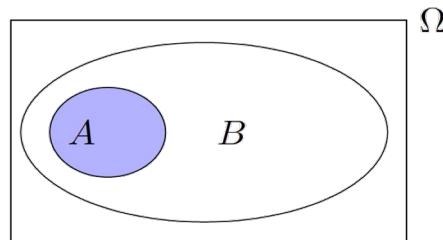
- The sample space is  $\Omega = \{B, GB, GGB, GGGB, \dots\}$ ;
- This example shows that the sample space is not necessarily finite.



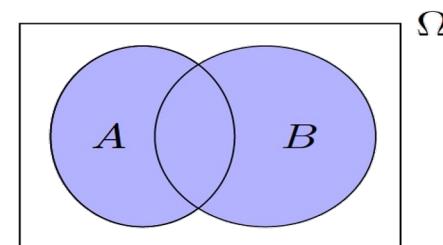
# 1.1 Basic Concepts

- From a set theory perspective, the relationships and operations of events become very clear:

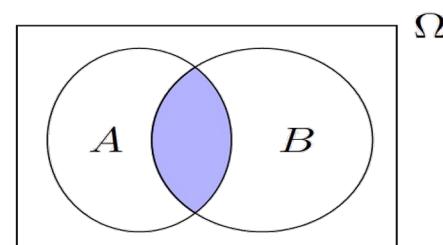
- **Inclusion  $A \subset B$ :** All sample points in event  $A$  are also in  $B$ , i.e.,  $B$  happens when  $A$  happens. (包含)



- **Sum/union  $A \cup B(A + B)$ :**  $= \{\omega | \omega \in A \text{ or } \omega \in B\}$ , at least one of  $A$  and  $B$  happens. (和/并)



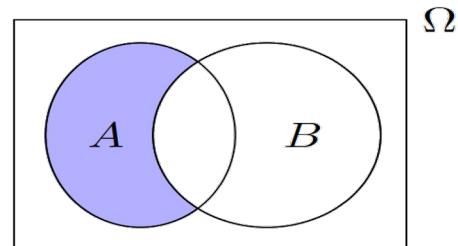
- **Product/intersection  $A \cap B(AB)$ :**  $= \{\omega | \omega \in A \text{ and } \omega \in B\}$ ,  $A$  and  $B$  both happen. (积/交)



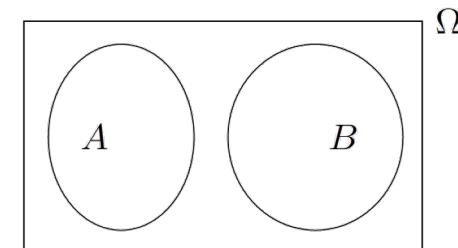
## 1.1 Basic Concepts

- From a set theory perspective, the relationships and operations of events become very clear:

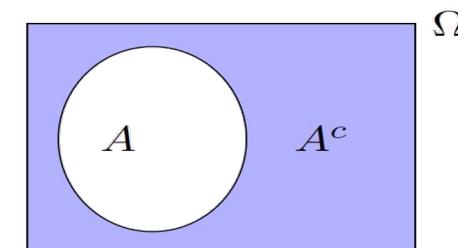
- Difference  $A - B$  ( $A \setminus B$ ):**  $= \{\omega | \omega \in A \text{ and } \omega \notin B\}$ ,  
 $A$  happens and  $B$  does not happen. (差)



- Mutually exclusive/disjoint**  $A \cap B = \emptyset$ :  $A$  and  $B$  cannot happen at the same time. (互斥 / 互不相容)



- Complement**  $A \cap B = \emptyset \text{ & } A \cup B = \Omega$ : either  $A$  or  $B$  happens, denoted as  $B = A^c$  or  $\bar{A}$ . (对立 / 互补)



# 1.1 Basic Concepts

- The operations of events obey certain rules similar to the rules of sets:

- Communicative laws (交换律):**

$$A \cup B = B \cup A, \quad A \cap B = B \cap A$$

- Associative laws (结合律):**

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

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

- Distributive law (分配律):**

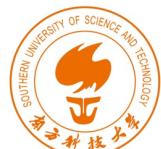
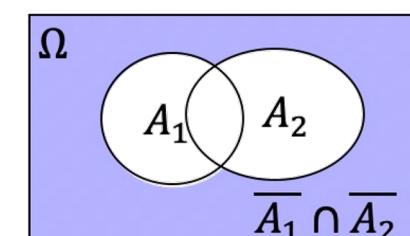
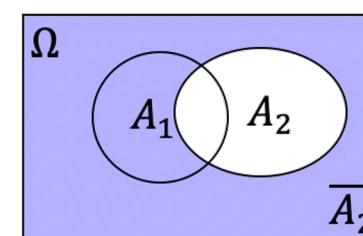
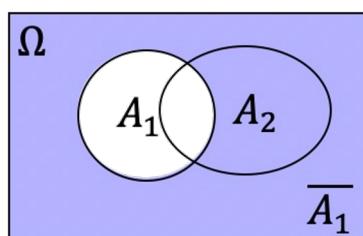
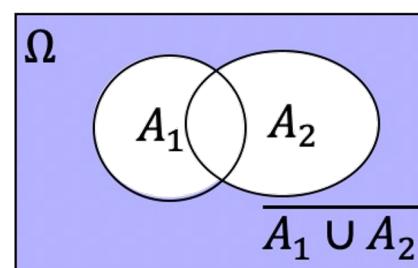
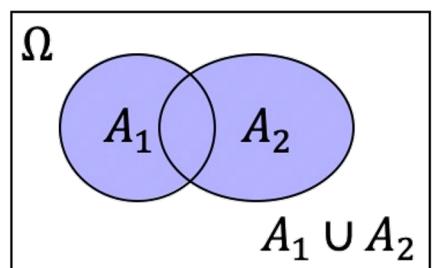
$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

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

- De Morgan's laws (德·摩根律):**

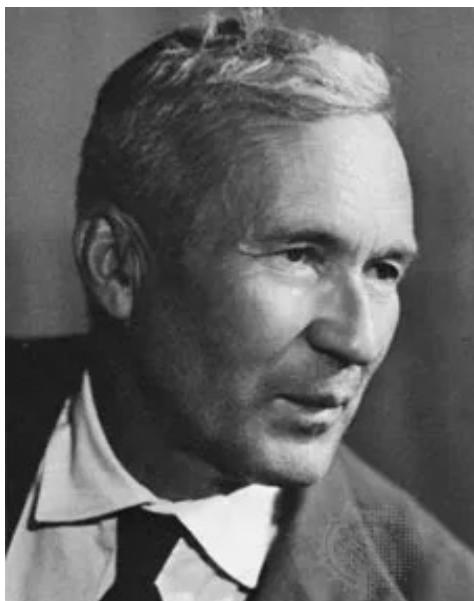
$$\overline{\bigcup_{i=1}^{\infty} A_i} = \bigcap_{i=1}^{\infty} \overline{A_i}, \quad \overline{\bigcap_{i=1}^{\infty} A_i} = \bigcup_{i=1}^{\infty} \overline{A_i}$$

A brief illustration of the De Morgan's laws



## 1.1 Basic Concepts

- Having defined the sample space and random event, we can now discuss the probability of events.
- It is agreed that probability is that it is a quantitative description of the likelihood of a random event occurring. But what's the formal [mathematical definition](#) of probability?



Andrey Kolmogorov  
(1903-1987)

### Probability

**Probability measure** (概率测度), or simply **probability**, is a real-valued function defined on subsets of the sample space  $\Omega$ , satisfying the following three axioms:

- **Non-negativity** (非负性): for any event  $A \subseteq \Omega$ , we have  $P(A) \geq 0$ .
- **Normalization** (规范性):  $P(\Omega) = 1$ .
- **Additivity** (可加性): for any **mutually exclusive** events  $A_1, A_2, \dots$ , we have

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



## 1.1 Basic Concepts

- Following the three axioms, many useful rules for computing probability can be derived.

### Properties of Probability

- $P(\emptyset) = 0$ .

Proof: Consider a sequence of events  $A_1 = \Omega, A_2 = A_3 = \dots = \emptyset$ . Then, these events are mutually exclusive and  $\Omega = A_1 \cup A_2 \cup \dots$ . Therefore, by the third axiom (i.e., additivity):

$$P(\Omega) = P(\Omega) + \sum_{i=1}^{\infty} P(\emptyset),$$

which implies that  $P(\emptyset) = 0$ .

- Finite additivity (有限可加性):** for any finite sequence of mutually exclusive events  $A_1, A_2, \dots, A_n$ ,

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



# 1.1 Basic Concepts

## Properties of Probability

- The complement rule:  $P(\bar{A}) = 1 - P(A)$ .

Proof: Let  $A_1 = A$  and  $A_2 = \bar{A}$ , then by the finite additivity with  $n = 2$ ,

$$1 = P(\Omega) = P(A \cup \bar{A}) = P(A) + P(\bar{A}) \Rightarrow P(\bar{A}) = 1 - P(A).$$

- The numeric bound:  $0 \leq P(A) \leq 1$ .
- Monotonicity (单调性): if  $A \subseteq B$ , then  $P(A) \leq P(B)$  and  $P(B - A) = P(B) - P(A)$ .
- The addition law (加法定律):  $P(A \cup B) = P(A) + P(B) - P(AB)$ .
- The inclusion-exclusion principle (容斥原理): ( $A_1, A_2, \dots, A_n$  not necessarily mutually exclusive)

$$\begin{aligned} P(A_1 \cup A_2 \cup A_3 \cup \dots \cup A_n) &= \sum_{i=1}^n P(A_i) - \sum_{1 \leq i < j \leq n} P(A_i A_j) + \sum_{1 \leq i < j < k \leq n} P(A_i A_j A_k) \\ &\quad - \sum_{1 \leq i < j < k < l \leq n} P(A_i A_j A_k A_l) + \dots + (-1)^{n-1} P(A_1 A_2 \dots A_n) \end{aligned}$$

+ for odd counts  
- for even counts



## 1.1 Basic Concepts

### Example 1.4

Let  $A, B, C$  be three random events, and  $P(A) = P(B) = P(C) = 0.25$ ,  $P(AB) = P(BC) = 0$ ,  $P(AC) = 0.125$ . Please find the probability that at least one of  $A, B, C$  happens.

### Solution

By the monotonicity of probability, since  $ABC \subseteq AB$ , then  $P(ABC) \leq P(AB) = 0$ .

Following the non-negativity axiom,  $P(ABC) = 0$ .

The probability that at least one of  $A, B, C$  happens is expressed as  $P(A \cup B \cup C)$ . Finally, by the inclusion-exclusion principle:

$$\begin{aligned}P(A \cup B \cup C) &= P(A) + P(B) + P(C) - P(AB) - P(BC) - P(AC) + P(ABC) \\&= 0.25 + 0.25 + 0.25 - 0 - 0 - 0.125 + 0 = 0.625.\end{aligned}$$



# Chapter 1 Basic Ideas in Probability

- 1.1 Basic Concepts
- 1.2 Computing Probabilities
- 1.3 Conditional Probability and Independence



## 1.2 Computing Probabilities

---

- Probability is a precise description of the likelihood of a random event occurring. Is it necessary to precisely compute probability?
- In everyday life, it may not be necessary, as it is difficult for us to distinguish between events with probabilities of 0.3 and 0.4.
- In professional fields, however, precise probability measurement becomes very important. E.g.,
  - In a **casino**, through precise probability measurement and design, the house only needs to have a slightly higher probability of winning than the players to make a profit from an overall perspective.
  - **Insurance companies** also operate a probability-based business. They design and price insurance products by accurately calculating the probability of claims, ensuring that they make a profit overall.



## 1.2 Computing Probabilities

- The probability computations in early days were mostly based on a relatively simple model known as the **classical model of probability** (古典概型).

### Classical Model of Probability

If a random experiment satisfies:

- there are only finite number of sample points in the sample space  $\Omega$ , i.e.,

$$\Omega = \{\omega_1, \omega_2, \dots, \omega_n\},$$

- each sample point is equally likely to occur, i.e.,

$$P(\{\omega_1\}) = P(\{\omega_2\}) = \dots = P(\{\omega_n\}) = \frac{1}{n},$$

then the probability model of the experiment is called a **classical model of probability**.

### Probability Computation Under the Classical Model of Probability

Assume that there are  $k$  sample points in event  $A$ , then the probability of event  $A$  is

$$P(A) = \frac{\text{Number of sample points in } A}{\text{Total number of sample points in } \Omega} = \frac{k}{n}.$$

Hence, under the classical model of probability, probability computation only involve counting the number of sample points in the sample space and the event of interest.

- The probability of a random event is the proportion of that event, a subset of the sample space, within the sample space.



## 1.2 Computing Probabilities

### Example 1.5

If two boxes each contain a certain number of red and green balls, and you are allowed to first choose a box and then randomly pick a ball from the chosen box. If a red ball is picked, you win a prize. Then which box should you choose from?

5red, 6green



$$\frac{5}{11}$$



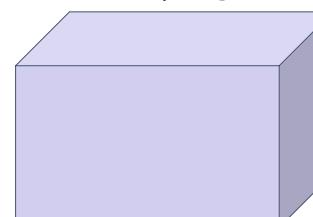
6red, 3green



$$\frac{6}{9}$$

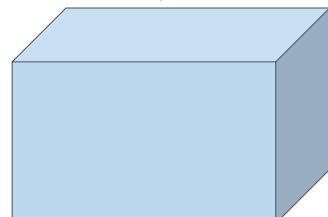


11red, 9green



$$\frac{11}{20}$$

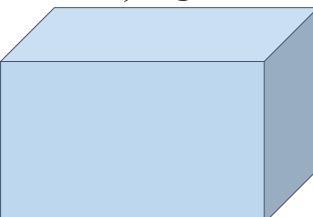
3red, 4green



$$\frac{3}{7}$$



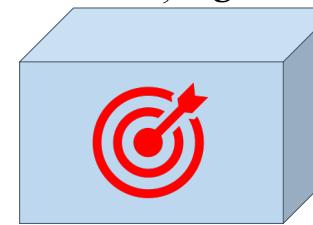
9red, 5green



$$\frac{9}{14}$$



12red, 9green



$$\frac{12}{21}$$

Simpson's  
Paradox  
(辛普森悖论)

(加州大学伯克利分校在1970年代的性别歧视)



## 1.2 Computing Probabilities

- However, more often, the number of sample points is not easy to determine and requires more systematic counting methods.
- We introduce/review two commonly used counting methods: the addition and multiplication principles, and permutations and combinations.

### Addition and Multiplication Principles

- **Addition principle (加法原理)**: If there are  $n$  types of methods to complete a task, with  $m_1$  specific methods in the first type,  $m_2$  specific methods in the second type, ..., and  $m_n$  specific methods in the  $n$ th type, then the total number of specific methods to complete this task is:

$$N = m_1 + m_2 + \cdots + m_n.$$

- **Multiplication principle (乘法原理)**: If there are  $n$  steps to complete a task, with  $m_1$  possible methods for the first step,  $m_2$  possible methods for the second step, ...,  $m_n$  methods for the  $n$ th step, then the total number of methods to complete this task is:

$$N = m_1 \times m_2 \times \cdots \times m_n.$$



## 1.2 Computing Probabilities

### Permutation and Combination

- **Permutation (排列):** If  $k$  elements are randomly selected **without replacement** (无重复随机抽取) from  $n$  distinct elements ( $k \leq n$ ) and **placed in order**, then the number of different permutations is

$$A_n^k = n(n - 1) \dots (n - k + 1) = \frac{n!}{(n - k)!}.$$

- **Combination (组合):** The number of combinations of  $k$  elements randomly selected **without replacement** (无重复随机抽取) from  $n$  distinct elements ( $k \leq n$ ), where the **order does not matter**, is given by

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

- **Think:** what's the permutation and combination under the case of **randomly selecting with replacement** (有重复随机抽取)?



## 1.2 Computing Probabilities

### Example 1.6

Suppose a class of  $n$  students has been allocated  $m$  ( $< n$ ) concert tickets by the university. The teacher decides to distribute the tickets by drawing lots.

The teacher prepares a hat containing  $n$  slips of paper, with  $m$  slips marked with a "1" and the remaining slips marked with a "0".

The students take turns drawing slips from the hat, and those who draw a slip marked with a "1" will get a concert ticket.

If you are one of the students in the class and you really want to get a ticket, would you choose to draw early or late?



## 1.2 Computing Probabilities

### Solution

- The essential question is: Suppose you are the  $l$ th to draw a slip, does the probability of you getting a ticket depend on  $l$ ?
- To compute probability, we first need to figure out what's the sample space?
- Considering the ticket drawing of the class as a random experiment, then each possible outcome of the drawing process is a sample point (i.e., an  $n$ -digit number with  $m$  digits being 1 and the remaining being 0), the sample space is the set of all these sample points.
- Then, determine the number of sample points in the sample space:  $C_n^m$ .
- Let  $A$  denote the event that you get a ticket, then the number of sample points in event  $A$  is:  $C_{n-1}^{m-1}$ .
- Put these together, the probability of event  $A$  is

$$P(A) = \frac{\text{Number of sample points in } A}{\text{Total number of sample points in } \Omega} = \frac{C_{n-1}^{m-1}}{C_n^m} = \frac{m}{n}.$$

- Therefore, the probability of getting a ticket **does not** depend on the order in which you draw a slip!



## 1.2 Computing Probabilities

### Example 1.7

(Birthday problem) There are  $n$  ( $n < 365$ ) students in a class, what's the probability that at least two students have the same birthday? (leap years are not taken into account)

### Solution

- Each sequence of  $n$  birthdays is a sample point, and the sample space is the set of all possible sequences of  $n$  birthdays.
- Then, the number of sample points in the sample space is:  $365^n$ .
- Let  $A$  be the event that at least two students have the same birthday, then  $\bar{A}$  is the event that every student has a different birthday.
- It is easier to determine the number of sample points in  $\bar{A}$ :

$$A_{365}^n = 365 \times 364 \times \cdots \times (365 - n + 1)$$

- Therefore, the probability of event  $A$  is computed as:

$$P(A) = 1 - P(\bar{A}) = 1 - \frac{A_{365}^n}{365^n}.$$



## 1.2 Computing Probabilities

Solution



Results							
$n$	20	25	30	40	50	55	100
$P(A)$	0.41	0.57	0.71	0.89	0.97	0.99	0.9999997

- In a class of 50 students, the probability that at least two people share the same birthday is as high as 97%.
- In a class of 100 students, it is almost certain that at least two people will share the same birthday.
- Does this surprise you?



## 1.2 Computing Probabilities

### Example 1.8

(Matching problem 配对问题)

- There are  $n$  students in a class with student IDs  $1, 2, \dots, n$ .
- Before the last Chinese New Year, everyone prepared a gift which was numbered with his/her student ID.
- Then all the gifts were put into a bag, and everyone randomly select a gift from the bag.
- What's the probability that at least one student get his/her own gift?



### Solution

- Each permutation of  $\{1, 2, \dots, n\}$  is a sample point, representing the gift numbers corresponding to students with IDs  $1, 2, \dots, n$ .
- So, the number of sample points in the sample space is:  $A_n^n = n!$ .
- Let  $A$  be the event that at least one student get his/her own gift, then directly determining the number of sample points in  $A$  is not so straightforward.



## 1.2 Computing Probabilities

### Solution

- The idea is to break down vague and complex problems into clear and simple ones.
- Let  $A_i$  be the event that the student with ID  $i$  get his/her own gift, then

$$\begin{aligned} P(A) &= P(A_1 \cup A_2 \cup \dots \cup A_n) \\ &= \sum_{i=1}^n P(A_i) - \sum_{i < j} P(A_i A_j) + \dots + (-1)^{n-1} P(A_1 A_2 \dots A_n) \end{aligned}$$

- The events  $A_i, A_i A_j, A_1 A_2 \dots A_n$  are very clear and their probability computation is simple:

$$P(A_i) = \frac{(n-1)!}{n!} = \frac{1}{C_n^1}, \quad P(A_i A_j) = \frac{(n-2)!}{n!} = \frac{1}{2! C_n^2}, \dots$$

- Put these together, we have:

$$\begin{aligned} P(A) &= \frac{1}{C_n^1} \times C_n^1 - \frac{1}{2! C_n^2} \times C_n^2 + \frac{1}{3! C_n^3} \times C_n^3 + \dots + (-1)^{n-1} \frac{1}{n!} \\ &= 1 - \frac{1}{2!} + \frac{1}{3!} + \dots + (-1)^{n-1} \frac{1}{n!} \approx 1 - e^{-1} \approx 0.632. \end{aligned}$$



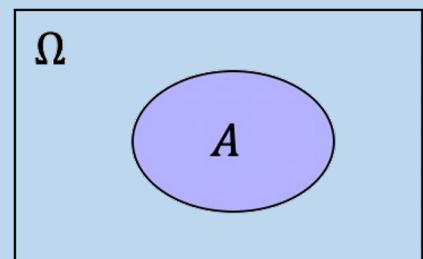
## 1.2 Computing Probabilities

- The classical model of probability assumes finite number of sample points and equal likelihood.
- Another model, called **the geometric model of probability (几何概率)**, is an extension of the classical model to an infinite number of sample points while maintaining equal likelihood.

### Geometric Model of Probability and its Probability Computation

If a random experiment can be represented as randomly throwing a point onto a bounded region  $\Omega$ , where the point is **equally likely** to land at any position within the region, then the probability model of the experiment is called a **geometric model of probability**. Let  $A$  be the event that the point lands at a subregion  $A$  of  $\Omega$ , then the probability of event  $A$  is computed as

$$P(A) = \frac{\text{The length/area/volume of } A}{\text{Total length/area/volume of } \Omega}.$$



- No matter which model is used, probability is essentially the proportion of the random event, a subsect of the sample space, within the sample space.



## 1.2 Computing Probabilities

### Example 1.9

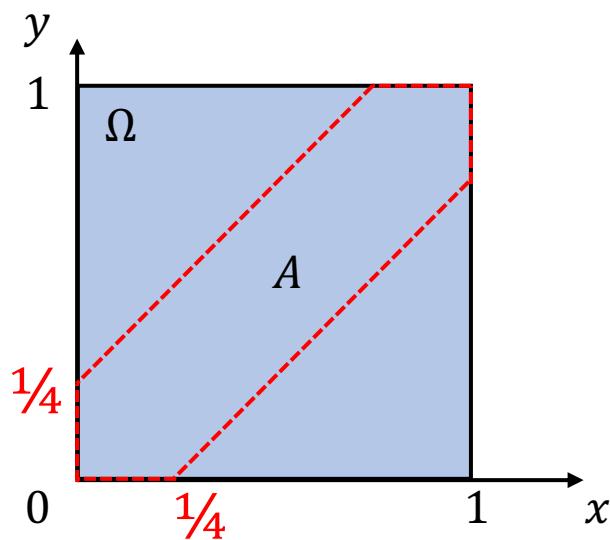
- Romeo and Juliet plan to meet in an evening, and each will arrive at the garden between 7pm and 8pm, with all pairs of arrival time equally likely.
- The first to arrive will wait for 15 minutes and will leave if the other has not yet arrived.
- What is the probability that they will meet?



### Solution

- Let 7pm be the origin,  $x$ -axis and  $y$ -axis be the arrival time of Romeo and Juliet, respectively.
- Then the sample space is  $\Omega = \{(x, y) \in [0, 1] \times [0, 1]\}$ .
- For Romeo and Juliet to meet, we need  $|x - y| \leq 1/4, 0 \leq x, y \leq 1$ .

$$P(A) = \frac{\text{The area of } A}{\text{Total area of } \Omega} = 1 - \frac{3}{4} \times \frac{3}{4} = \frac{7}{16}.$$



## 1.2 Computing Probabilities

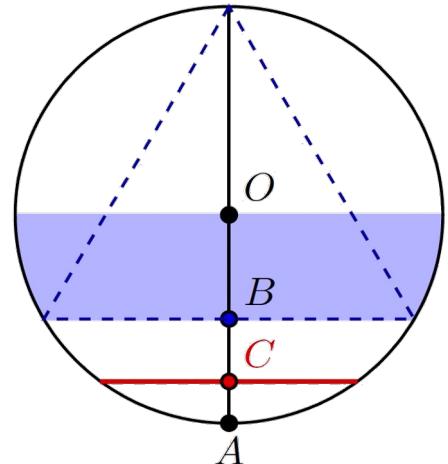
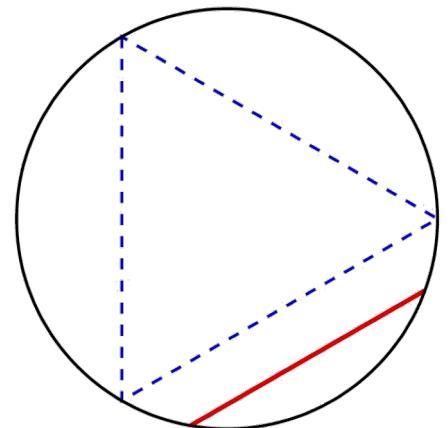
### Example 1.10

#### Bertrand's paradox (贝特朗悖论)

Consider a circle with radius (半径) 1. What is the probability that a randomly draw chord (弦) of the circle is longer than the side of the inscribed equilateral triangle (内切等边三角形) of the circle?

#### Solution 1

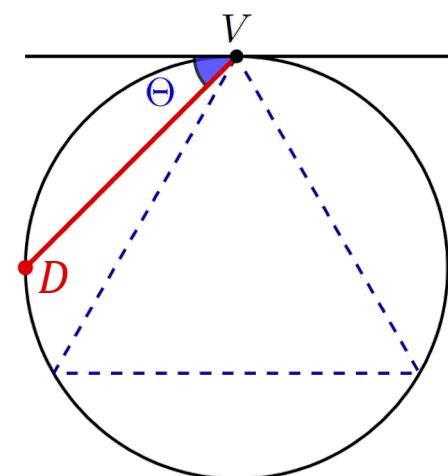
- Take a radius of the circle  $OA$ , and randomly choose a point  $C$  on the radius (all points being equally likely), then draw a chord through  $C$  orthogonal to  $OA$ .
- By elementary geometry,  $OA$  intersects the triangle at the midpoint of  $OA$ , say  $B$ .
- Then the sample space is all points on  $OA$
- For the chord to be longer than the side of the triangle,  $C$  must fall on  $OB$ .
- Therefore, the probability is  $\frac{\text{The length of } OB}{\text{The length of } OA} = \frac{1}{2}$ .



## 1.2 Computing Probabilities

### Solution 2

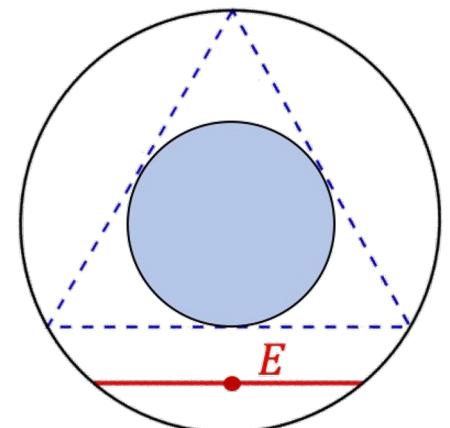
- Take a point on the circle, e.g., the vertex  $V$ , draw the tangent (切线) to the circle through  $V$ .
- Then randomly choose another point  $D$  on the circle and draw a chord by connecting  $V$  and  $D$  that forms a random angle  $\Theta$  with the tangent.
- Then the sample space is  $\Omega = \{\Theta \in [0, \pi]\}$ , and for the chord to be longer than the side of the inscribed equilateral triangle, we need  $\pi/3 < \Theta < 2\pi/3$ .
- Therefore, the probability is  $\frac{\pi/3}{\pi} = \frac{1}{3}$ .



## 1.2 Computing Probabilities

### Solution 3

- Randomly choose a point  $E$  within the circle and draw a chord with  $E$  as the midpoint.
- For the chord to be longer than the side of the inscribed equilateral triangle,  $E$  must fall within the inscribed circle (内切圆) of the triangle, whose radius is  $1/2$ .
- Then the sample space is the original circle, and the event of interest is the inscribed circle.
- Therefore, the probability is  $\frac{\pi/4}{\pi} = \frac{1}{4}$ .



## 1.2 Computing Probabilities

### Discussion

- Which solution is correct?
- The three solutions are all correct!
- The reason why the same question have three different answers is that the “randomly draw chord” is not clearly defined in the problem.
- Different assumptions of equal likelihood lead to **different sample spaces**:
  - In Solution 1, the midpoint of the chord is assumed to be equally likely chosen on a radius, so the sample space consists of all points on the radius.
  - In Solution 2, the other endpoint of the chord is assumed to be equally likely chosen on the circle, so the sample space consists of all points on the circle.
  - In Solution 3, the midpoint of the chord is assumed to be equally likely chosen within the circle, so the sample space consists of all points within the circle.
- Therefore, when computing probability, it is crucial to clearly define the sample space first.



# Chapter 1 Basic Ideas in Probability

- 1.1 Basic Concepts
- 1.2 Computing Probabilities
- 1.3 Conditional Probability and Independence

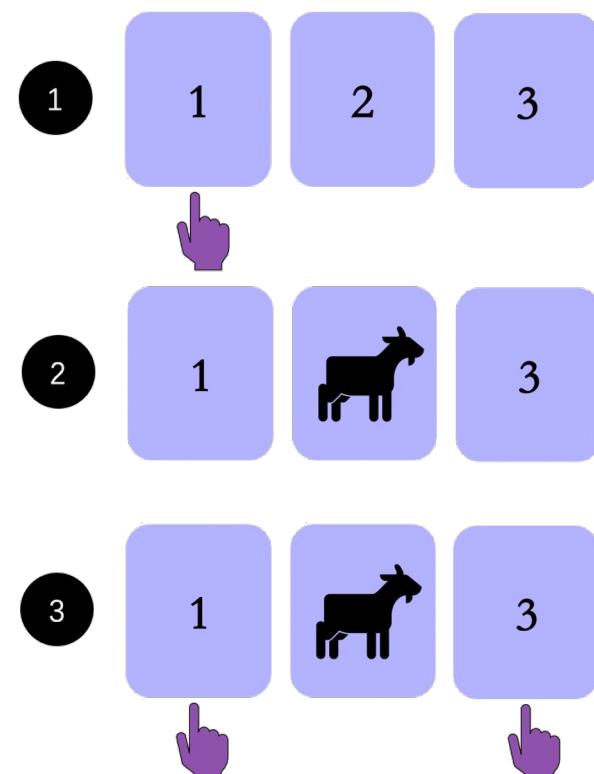


## 1.3 Conditional Probability and Independence

- Let's first look at a problem that once sparked widespread discussion in the United States—the **Monty Hall problem (三门问题)**.

### Example 1.11

- The problem originates from the American TV show *Let's Make a Deal*, named after the show's host, Monty Hall.
- The show came into the public eye in 1975. Contestants would see three closed doors, with a car behind one of them and a goat behind each of the other two. If they successfully pick the door with the car, they will win the car.
- You are first asked to choose one of the doors, say Door 1.
- Then, the host (who knows where the car is) opens one of the remaining two doors, revealing a goat, say behind Door 2.
- Finally, the host gives you a chance to change your choice—either stick with your original choice (Door 1) or switch to Door 3.
- What's your choice?



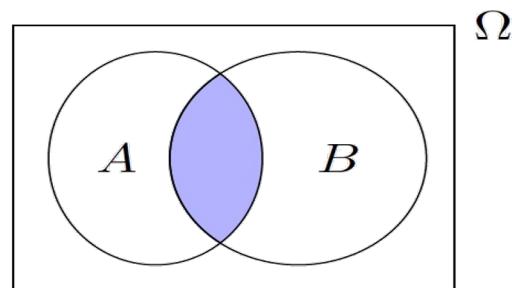
## 1.3 Conditional Probability and Independence

- After learning about conditional probability, you will find a logically rigorous way to solve the Monty Hall problem.
- Conditional probability refers to the idea that the probability of a random event will change given the occurrence of another event.

### Conditional Probability

Let  $A$  and  $B$  be two random events and  $P(B) > 0$ . Then the conditional probability (条件概率) of event  $A$  given that event  $B$  occurs is defined as

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



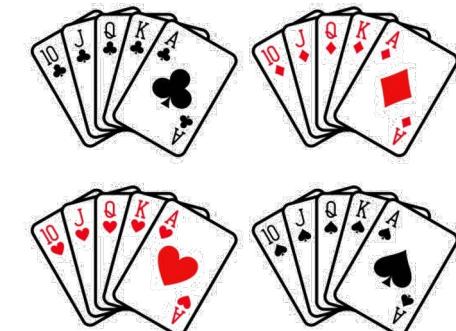
- The intuitive understanding of  $P(A|B)$  is that the occurrence of event  $B$  provides new information that updates our belief on the likelihood of event  $A$ .
- The idea behind this definition is that if event  $B$  has occurred, the sample space becomes  $B$  instead of  $\Omega$ .



## 1.3 Conditional Probability and Independence

### Example 1.12

- You are playing a poker game where you are dealt 5 cards face down.
- A royal flush (皇家同花顺) is a hand of AKQJ10 all in one suit.
- 1. What is the probability that you are dealt a royal flush?
- 2. If one of the cards that you are dealt lands face up, showing the Ace of spades (黑桃A), what is the probability now?



### Solution

- Let  $A$  be the event of a royal flush, then the number of sample points in  $A$  is 4 (one for each suit).
- The sample space is all possible combinations of 5 cards with total number of sample points  $C_{52}^5$ .
- So, the probability of  $A$  is:

$$P(A) = \frac{4}{C_{52}^5} = \frac{1}{649,740}.$$



## 1.3 Conditional Probability and Independence

### Solution

- Let  $B$  be the event that one of the 5 cards is the Ace of spades, then the number of sample points in  $B$  is  $C_{51}^4$ .
- $AB = A \cap B$  is the event of a royal flush of spades, so the number of sample points in  $AB$  is 1. Consequently, the probability of  $A$  given  $B$  is

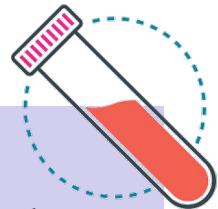
$$P(A|B) = \frac{P(AB)}{P(B)} = \frac{1}{C_{52}^4} = \frac{1}{249,900}.$$



## 1.3 Conditional Probability and Independence

### Example 1.13

- You have a blood test for a rare disease that occurs by chance in 1 person in 100,000.
- If you have the disease, the test will report that you do with probability 0.95 (and that you do not with probability 0.05).
- If you do not have the disease, the test will report a false positive with probability 0.001.
- If the test says you have the disease, what is the probability that you actually have the disease?



### Thinking

- First, we express the given conditions in the example using probability language.
- Let  $A$  be the event that the test reports a positive result for a randomly chosen person,  $B$  be the event that a randomly chosen person has the disease, then:

$$P(A|B) = 0.95, P(A|\bar{B}) = 0.001, P(B) = 0.00001.$$

- The probability to be computed is expressed as  $P(B|A) = P(AB)/P(A)$ .
- How to compute  $P(AB)$  and  $P(A)$  based on the given conditions?



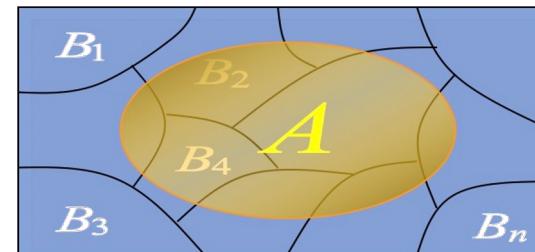
## 1.3 Conditional Probability and Independence

- To solve the problem, we introduce the multiplication law (乘法定律, used to compute  $P(AB)$ ) and the law of total probability (全概率定律, used to compute  $P(A)$ ).

### Multiplication Law

Let  $A$  and  $B$  be two random events and  $P(B) > 0$ . Then

$$P(AB) = P(A|B)P(B).$$



### Law of Total Probability

Let  $A$  and  $B$  be two random events, then (assume that  $P(A|B) = 0$  if  $P(B) = 0$ )

$$P(A) = P(AB) + P(A\bar{B}) = P(A|B)P(B) + P(A|\bar{B})P(\bar{B}).$$

More generally, if  $B_1, B_2, \dots, B_n$  are  $n$  mutually exclusive random events, and  $B_1 \cup B_2 \cup \dots \cup B_n = \Omega$ , then we call  $B_1, B_2, \dots, B_n$  a partition (分割/分划) of the sample space  $\Omega$ , and

$$\begin{aligned} P(A) &= P(A \cap \Omega) = P(AB_1 \cup AB_2 \cup \dots \cup AB_n) = P(AB_1) + P(AB_2) + \dots + P(AB_n) \\ &= P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \dots + P(A|B_n)P(B_n). \end{aligned}$$



## 1.3 Conditional Probability and Independence

- Like the addition law which has a general version extended to multiple events, i.e., the inclusion-exclusion principle, the multiplication law also has a general version: ( $P(A_1A_2 \cdots A_{n-1}) > 0$ )

$$\begin{aligned}P(A_1A_2 \cdots A_n) &= P(A_n|A_1A_2 \cdots A_{n-1})P(A_1A_2 \cdots A_{n-1}) = \cdots \\&= P(A_n|A_1A_2 \cdots A_{n-1})P(A_{n-1}|A_1A_2 \cdots A_{n-2}) \cdots P(A_2|A_1)P(A_1),\end{aligned}$$

- which is called the chain rule (链式法则) for random events.



### Solution of Example 1.13

- By the multiplication law and the law of total probability:

$$P(B|A) = \frac{P(AB)}{P(A)} = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|\bar{B})P(\bar{B})} = \frac{0.95 \times 0.00001}{0.95 \times 0.00001 + 0.001 \times 0.99999} \approx 0.0094.$$

- Even if the test accuracy is very high, the probability that a positive result is correct is very small.
- Does this contradict your experience or intuition? Why does this happen?
- The reason is that incidence of the disease in the population is very low.



## 1.3 Conditional Probability and Independence

- Actually, a famous theorem has been used in the solution of the blood test example, which is the Bayes' theorem (贝叶斯定理).

### Bayes' Theorem / Bayes' Rule

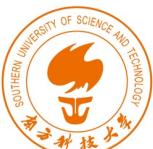
Let  $B_1, B_2, \dots, B_n$  be random events and  $B_1, B_2, \dots, B_n$  is a partition of the sample space. Then for any event  $A$  such that  $P(A) > 0$ , we have

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

- $P(B_i|A)$  is the probability of  $B_i$  given that  $A$  occurs, also termed the **posterior probability** (后验概率) of  $B_i$ .
- $P(B_i)$  is called the **prior probability** (先验概率) or marginal probability (边缘概率) of  $B_i$ , which refers to the probability value in the absence of any other prior information.
- The Bayes' rule is widely used across various fields, especially in scenarios involving uncertainty and decision-making under incomplete information.



Thomas Bayes 1702-1761  
British Mathematician



## 1.3 Conditional Probability and Independence

8/24/2024, 11:05:24 PM



Is the Bayes' rule widely used?



8/24/2024, 11:05:39 PM

Yes, Bayes' rule is widely used across various fields, especially in scenarios involving uncertainty and decision-making under incomplete information. Here are some key applications:

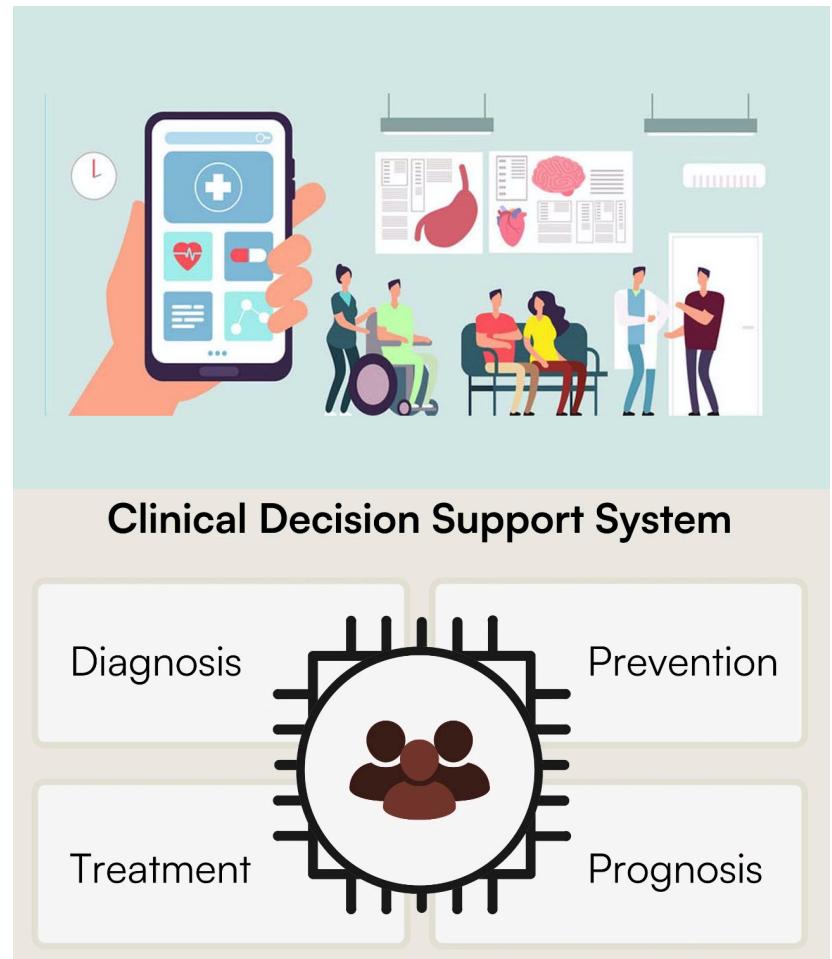
- 1. Medical Diagnosis:** In healthcare, Bayes' theorem helps calculate the probability that a patient has a particular disease given a positive test result, considering both the test's reliability and the disease's prevalence.
- 2. Machine Learning and AI:** Many algorithms, like the Naive Bayes classifier, rely on Bayes' rule to make predictions. It's used in spam detection, sentiment analysis, and document classification.
- 3. Finance:** Bayes' theorem is used in risk assessment, portfolio management, and market predictions, updating investment strategies as new information becomes available.
- 4. Forensics and Legal Reasoning:** In court cases, Bayes' theorem can assist in evaluating evidence, determining the likelihood of guilt based on the available data.
- 5. Cognitive Science:** Bayes' rule models human reasoning and decision-making, particularly how people update their beliefs in light of new evidence.

Overall, Bayes' rule is a powerful tool wherever probabilistic reasoning and inference are needed, from science and engineering to everyday decision-making.



## 1.3 Conditional Probability and Independence

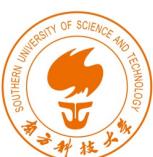
- A real-world example of the application of Bayes' rule – the **Clinical Decision Support System**
  - $B_1, B_2, \dots, B_n$  represent various diseases, and  $A$  represents a certain symptom or indicator (such as the level of transaminase in the blood 血液中的转氨酶含量).
  - The prior probability  $P(B_i)$  can be determined using statistical methods.
  - $P(A|B_i)$  can be determined using medical knowledge.
  - By applying the Bayes' rule, we can calculate the posterior probability  $P(B_i|A)$ .
  - The diseases corresponding to the larger  $P(B_i|A)$  can be provided to the doctor for further clinical diagnosis.



## 1.3 Conditional Probability and Independence

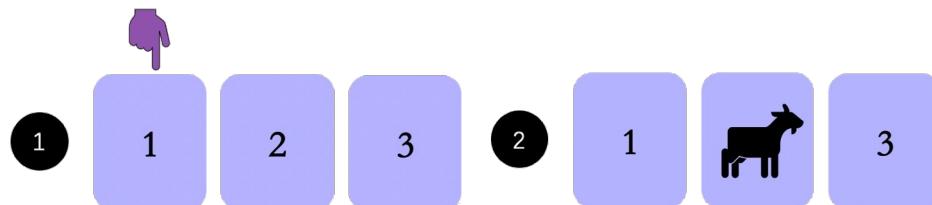
- How to understand the Bayes' rule?

- In reality, we can view event  $A$  as the outcome (结果) and events  $B_1, B_2, \dots, B_n$  as the various causes (原因) leading to this outcome.
- The law of total probability infers the probability of the outcome  $A$  occurring based on the different causes—it's a **reasoning process from cause to effect**.
- However, there is another important scenario in our daily lives: we observe a certain phenomenon and then reason backward to determine the probabilities of various causes that led to it. Simply put, it is **reasoning from effect to cause**.
- The conditional probability  $P(B_i|A)$  obtained by the Bayes' rule helps us infer the likelihood that a specific cause  $B_i$  led to the observed outcome  $A$ , supporting our subsequent decision-making.
- The posterior probability  $P(B_i|A)$  is a revision of the prior probability  $P(B_i)$  after acquiring new information.



## 1.3 Conditional Probability and Independence

- Lastly, let's solve the Monty Hall problem.



### Solution of Example 1.11

- Without loss of generality, suppose that you chose Door 1 and the host opened Door 2.
- Let  $B_1, B_2, B_3$  be event that there is a car behind Door 1, 2, 3 and  $A$  be the event that the host opens Door 2. Then:

$$P(B_1) = P(B_2) = P(B_3) = \frac{1}{3}, P(A|B_1) = \frac{1}{2}, P(A|B_2) = 0, P(A|B_3) = 1.$$

- The probability of interest is  $P(B_1|A)$  and  $P(B_3|A)$ . By the law of total probability, we have

$$P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + P(A|B_3)P(B_3) = \frac{1}{2} \times \frac{1}{3} + 1 \times \frac{1}{3} = \frac{1}{2}.$$

- Finally, by the Bayes' rule:

$$P(B_1|A) = \frac{P(A|B_1)P(B_1)}{P(A)} = \frac{1/2 * 1/3}{1/2} = \frac{1}{3}, \quad P(B_3|A) = \frac{P(A|B_3)P(B_3)}{P(A)} = \frac{1 * 1/3}{1/2} = \frac{2}{3}.$$

- Therefore, you should switch to Door 3!



## 1.3 Conditional Probability and Independence

---

- Finally, let's discuss the independence (独立性) between events. You might have learned about some definitions of independence before, like:
  - The independence between two events means that the occurrence of one event does not affect the probability of another event.
  - If  $P(AB) = P(A)P(B)$ , then we claim that events  $A$  and  $B$  are independent.
- However, the first is not a mathematically rigorous definition, the second seems difficult to understand intuitively.
- With the help of conditional probability, the independence between events can be easily and clearly defined.
- “The occurrence of  $A$  does not affect the probability of  $B$ , and vice versa” can be expressed as

$$P(A|B) = P(A), P(B|A) = P(B).$$

- By the multiplication law, we therefore have

$$P(AB) = P(A|B)P(B) = P(B|A)P(A) = P(A)P(B).$$



## 1.3 Conditional Probability and Independence

### Example 1.14

Let  $A$  and  $B$  be two random events, what is the relationship between the following two statements?

- $A$  and  $B$  are independent.
- $A$  and  $B$  are mutually exclusive.

### Solution

- $A$  and  $B$  are independent  $\Rightarrow P(AB) = P(A)P(B)$ .
- $A$  and  $B$  are mutually exclusive  $\Rightarrow P(AB) = 0$ .
- Therefore, if  $P(A) > 0$  and  $P(B) > 0$ , then “ $A$  and  $B$  are independent” and “ $A$  and  $B$  are mutually exclusive” cannot happen at the same time.
- “ $A$  and  $B$  are mutually exclusive” suggests that  $A$  would never occur if  $B$  occurs, i.e., the occurrence of  $B$  provides new information on the occurrence of  $A$ , so the two events are not independent.



## 1.3 Conditional Probability and Independence

- Let's look at the formal definition of independence.

### Independence

Let  $A, B, C, A_1, A_2, \dots, A_n$  all denote random events.

- $A$  and  $B$  are said to be independent if  $P(AB) = P(A)P(B)$ .
- $A, B, C$  are said to be (mutually) independent (相互独立) if:

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

- $A_1, A_2, \dots, A_n$  are said to be (mutually) independent (相互独立) if for every subset of the events,  $A_{i_1}, A_{i_2}, \dots, A_{i_k}$  ( $1 \leq i_1 < \dots < i_k \leq n, k = 2, \dots, n$ ), we have

$$P(A_{i_1}A_{i_2} \dots A_{i_k}) = P(A_{i_1})P(A_{i_2}) \dots P(A_{i_k}).$$

- The conditions in the definition above can all be derived using conditional probability.
  - You might naturally wonder why the independence of multiple events requires so many conditions. Aren't these conditions redundant?
  - For example, doesn't pairwise independence imply that all three events are mutually independent?



## 1.3 Conditional Probability and Independence



### Example 1.15

- Draw three cards from a properly shuffled standard deck, with replacement and reshuffling (i.e., draw a card, make a note, return to deck, shuffle, draw the next).
- Let  $A$  be the event that “card 1 and 2 have the same suit”,  $B$  be the event that “card 2 and 3 have the same suit”,  $C$  be the event that “card 1 and 3 have the same suit”.
- Are the three events mutually independent?

### Solution

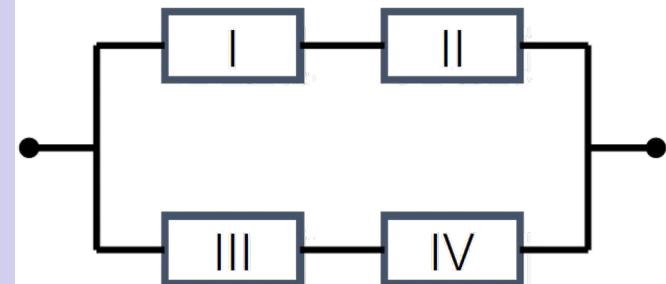
- It is not difficult to obtain that  $P(A) = P(B) = P(C) = 1/4$ .
- Moreover, events  $AB$ ,  $AC$ ,  $BC$ ,  $ABC$  all refer to the event that “card 1, 2, and 3 all have the same suit”, so  $P(AB) = P(AC) = P(BC) = P(ABC) = 1/16$ .
- Therefore, we have  $P(AB) = P(A)P(B)$ ,  $P(AC) = P(A)P(C)$ ,  $P(BC) = P(B)P(C)$ , suggesting that the three events are pairwise independent.
- However, since  $P(ABC) \neq P(A)P(B)P(C)$ , the three events are not mutually independent.



## 1.3 Conditional Probability and Independence

### Example 1.16

- A system has 4 components as shown in the picture.
- Assume the probability of each component working properly is  $p$  and the 4 components are independent of each other.
- What is the probability that the system works properly?



### Solution

- Let  $A$  be the event that the system works properly and  $A_i$  be the event that component  $i$  works properly ( $i = 1, 2, 3, 4$ ). Then  $P(A_1) = P(A_2) = P(A_3) = P(A_4) = p$ .
- From the structure of the system, it is not difficult to see  $A = (A_1A_2) \cup (A_3A_4)$ .
- Therefore, we have the reliability of the system to be

$$P(A) = P(A_1A_2 \cup A_3A_4) = P(A_1A_2) + P(A_3A_4) - P(A_1A_2A_3A_4)$$

By independence  $= P(A_1)P(A_2) + P(A_3)P(A_4) - P(A_1)P(A_2)P(A_3)P(A_4) = p^2(2 - p^2)$ .



## 1.3 Conditional Probability and Independence

- Sometimes, events may not be independent directly, but they are independent conditioned on the occurrence of a specific event. This is the concept of **conditional independence** (条件独立性).

### Conditional Independence

Let  $A_1, A_2, \dots, A_n, B$  all denote random events.

- $A_1, A_2, \dots, A_n$  are said to be **conditionally independent conditioned on  $B$**  if for every subset of the events,  $A_{i_1}, A_{i_2}, \dots, A_{i_k}$  ( $1 \leq i_1 < \dots < i_k \leq n, k = 2, \dots, n$ ), we have

$$P(A_{i_1}A_{i_2} \dots A_{i_k}|B) = P(A_{i_1}|B)P(A_{i_2}|B) \cdots P(A_{i_k}|B).$$

- Conditional independence is a foundational concept widely used in various fields, enabling efficient computation and reasoning about complex systems.
- Some key areas where conditional independence is essential include **graphical models** (e.g., Bayesian networks), **hidden Markov models**, **structural causal models in causal inference**, **topic models in natural language processing** (e.g., Latent Dirichlet Allocation), etc.



## 1.3 Conditional Probability and Independence

### Example 1.17

- Naïve Bayes is a simple yet powerful probabilistic machine learning algorithm used for classification tasks, particularly in text classification, such as spam filtering, sentiment analysis, etc.
- It is based on the **Bayes' theorem** with an assumption of **conditional independence** between the features that are used to perform classification.
- For example, in spam filtering, suppose that there are  $N$  words in an email,  $W_i$  denotes the event that the  $i$ th word is in the email ( $i = 1, 2, \dots, N$ ), and  $S$  denotes the event that the email is a spam, then

$$P(S|W_1 \cap W_2 \cap \dots \cap W_N) = \frac{P(W_1 W_2 \dots W_N | S)P(S)}{P(W_1 W_2 \dots W_N | S)P(S) + P(W_1 W_2 \dots W_N | \bar{S})P(\bar{S})}.$$

- By assuming conditional independence between the words appearing in an email given whether the email is a spam or not, we have

$$P(W_1 W_2 \dots W_N | S) = P(W_1 | S)P(W_2 | S) \dots P(W_N | S),$$

- which simplifies the computation greatly.
- However, the conditional independence assumption may not be realistic, this is why it is called “naïve”.



## 1.3 Conditional Probability and Independence

- Finally, let's discuss the independence in reality.
  - In mathematics, it is not difficult to determine if events are independent since it is clearly defined.
  - However, in real life, it may not be easy to discern whether events are independent.
  - For example, we have heard of the famous butterfly effect, where “the flap of a butterfly’s wings in South America could ultimately cause a tornado in Texas.”
  - So, something seemingly unrelated could bring about a significant change.
  - Many instances of independence in real life are assumptions we make to greatly simplify problems, i.e., independence is often just a mathematical model we use to describe random events.

### Example 1.18

Please compare the Taobao shopping system and the 12306 ticketing system from the perspective of system development difficulty.

