

# **Probability and Mathematical Statistics**

概率与数理统计

唐聿勍

Department of Industrial Engineering & Management

yujietang@pku.edu.cn

# Course Logistics

- Class Schedule
  - Tuesday: 1:00pm–2:50pm
  - Friday: 1:00pm–2:50pm (双周)
  - 理教310
- Instructor
  - 唐聿勘 工业工程与管理系 [yujietang@pku.edu.cn](mailto:yujietang@pku.edu.cn)
  - Office: 新奥工学大楼2046
  - Office hour: Tuesday 3:00pm–5:00pm and by appointments

# Course Logistics

- Teaching assistants
  - 王旭浩 (2301111776@stu.pku.edu.cn)
  - 章斯岚 (2401111821@stu.pku.edu.cn)
  - Contact the TAs if you have questions about the assignments
- The lecture on Week 15 needs to be rescheduled. Details will be provided later.

# Course Contents

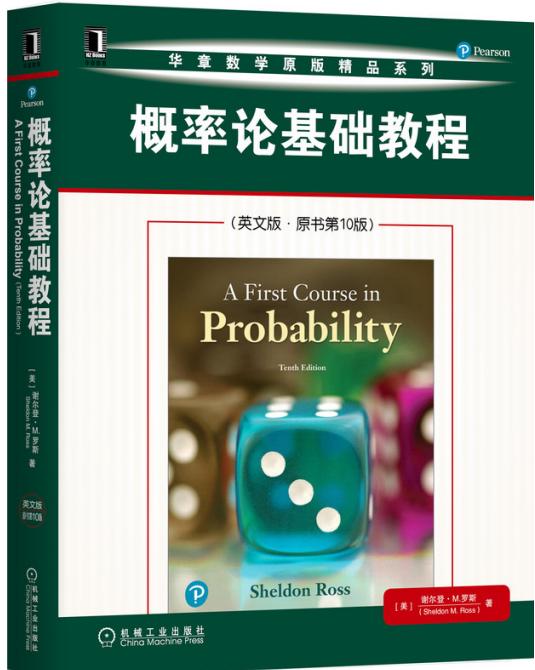
- Probability
  - Fundamentals
  - Discrete random variables
  - Continuous random variables
  - Further topics
  - The law of large numbers and the central limit theorem
- Mathematical Statistics
  - Basic concepts
  - Parameter estimation
  - Hypothesis testing
  - Linear regression analysis

# Prerequisites

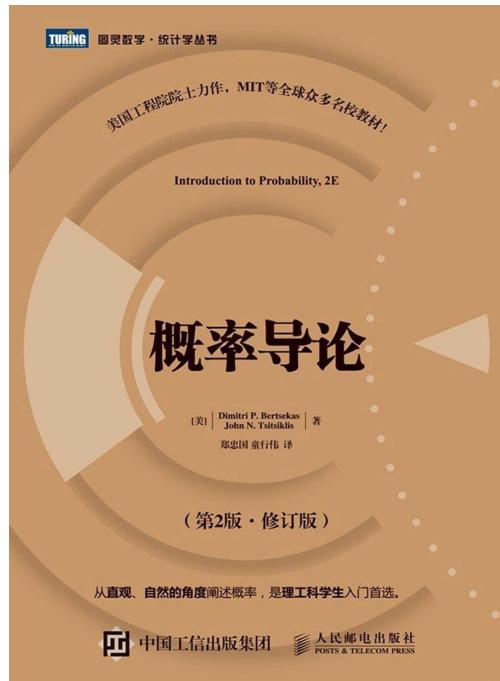
- Univariate and multivariate calculus
  - Basic set theory
  - Limits, continuous functions, derivatives, Taylor expansion, Riemann integrals, Newton–Leibniz theorem, improper integrals
  - Partial derivatives, multiple and iterated integrals
  - Series, absolute convergence, rearrangements, power series
- Linear algebra
  - Vectors, matrices, matrix inverse, determinants
  - Eigenvalue decomposition, real symmetric matrices, positive (semi)definite matrices

# Course Contents

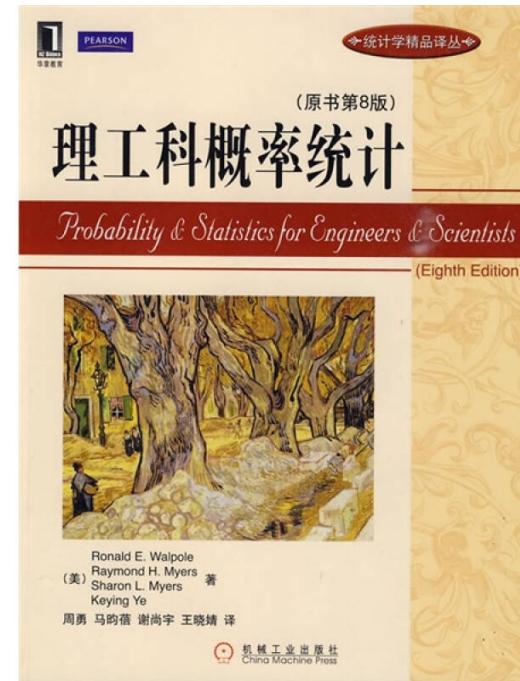
- No official textbook. Some recommended books:



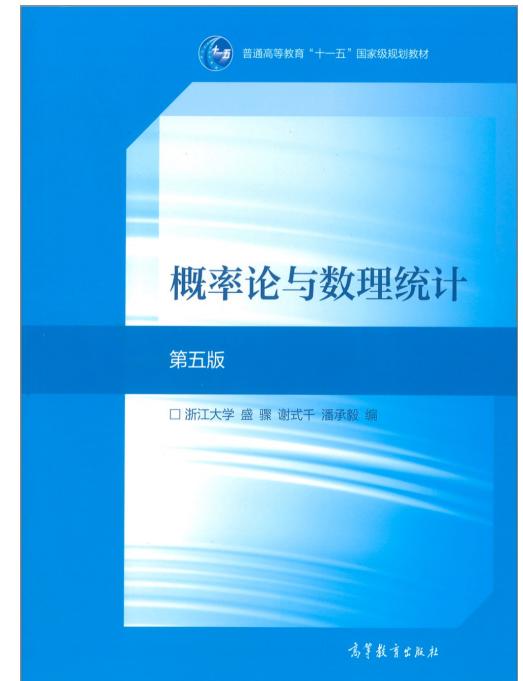
Sheldon Ross



Dimitri P. Bertsekas,  
John N. Tsitsiklis



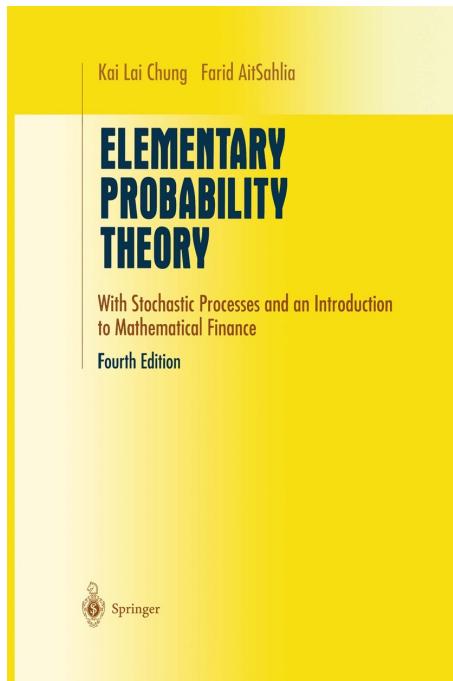
R. E. Walpole, R. H. Myers,  
S. L. Myers, K. Ye



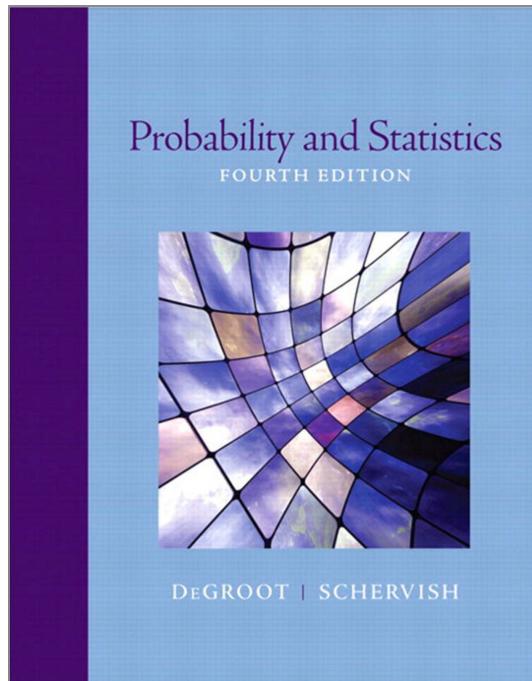
浙江大学  
盛骤 谢式千 潘承毅

# Course Contents

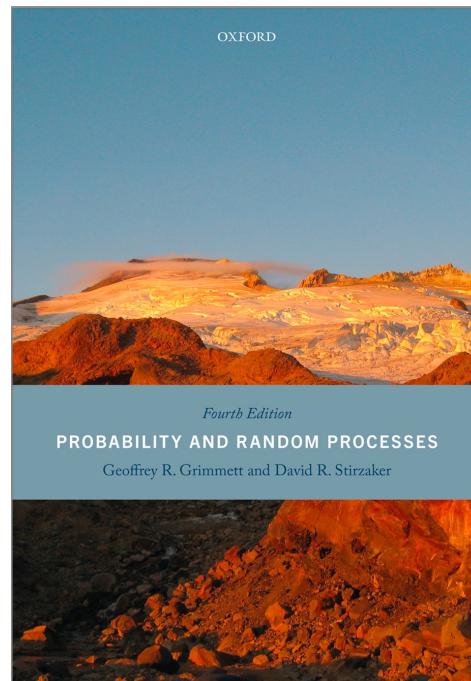
- Some other excellent books that are not in Chinese:



Kai Lai Chung,  
Farid AitSahlia



Morris H. DeGroot,  
Mark J. Schervish



Geoffrey Grimmett,  
David R. Stirzaker

- You don't need to read a whole book.
- Read them when you have confusions and want to see how others explain those concepts.
- Do the exercises.

# Course Contents

- The slides will be mostly in English.
- The slides **DO NOT** include everything taught in the lectures.
- You are encouraged to take your own notes.
- “Official” lecture notes (in Chinese) will be provided on course.pku.edu.cn **AFTER** each chapter is completed.
  - My lectures notes are for you to review the course materials.
  - We strongly recommend that you finish all the exercises in my lecture notes (some of them are not easy), although you don’t need to submit your solutions.

# Grading Policy

- Assignments 25%
  - Submitted by a studying group of up to **2 students**. The group will receive the same grade on each submission.
  - About 5 assignments in total.
  - Late submission will be penalized by a 50% reduction in the grade.
- Course project 15%
  - Carried out in groups. Each group consists of up to **4 students**.
  - Students in the same group will receive the same grade.
  - Will be assigned after the midterm exam.
  - Late submission will be penalized by a 50% reduction in the grade.

# Grading Policy

- Midterm exam 20%
  - In class. Date to be determined (probably early November).
- Final exam 40%
  - Dec. 31<sup>st</sup> afternoon
  - Problems will be similar to or even a little bit harder than assignments and examples in the handouts
- We will **NOT** 划重点 for the exams.
- Grading will be changed only when an error has been made; no negotiation is allowed.

- Task 1:
  - Form your homework team first.
  - Send your names and student ID numbers of homework team to our TA 王旭浩 by Sep. 20th.

# Chapter 1. Fundamentals of Probability Theory

# Randomness and Probability

- Phenomena with uncertainties
  - Their outcomes cannot be predicted accurately beforehand.
  - But some of them seem to exhibit certain statistical regularity (统计规律) when they are repeated under roughly the same conditions for a large number of times.

# Randomness and Probability

- Example: Brownian motion of a pollen particle immersed in water

# Randomness and Probability

- Random experiments 随机试验: An idealized procedure
  - that can be repeated under (approximately) the same conditions;
  - whose outcomes exhibit statistical regularity
- Probability theory: How to model a random experiment mathematically?

# Randomness and Probability

- **Sample space** 样本空间: The set of all possible outcomes
- **Events** 事件: An event is a **subset** of the sample space
- **Probability** 概率: Every event is assigned a number (probability) to quantify how likely this event is to occur.

Ideally:  $\boxed{\mathbb{P}(E)} = \lim_{n \rightarrow \infty} \boxed{\text{Fr}_n(E)}$

Probability of  
the event  $E$

Frequency of the occurrence of  
 $E$  in  $n$  repeated experiments

## 1.1. Review of Set Theory (集合论)

# Review of Set Theory (集合论)

- Given a set  $A$  and an object  $x$ , we can always talk about whether  $x$  is a member of  $A$  or not

$x \in A$ :  $x$  is a member/element of  $A$

- Two sets  $A$  and  $B$  are equal if and only if any member of  $A$  is also a member of  $B$  and vice versa.

$A = B$  iff  $\forall x(x \in A \Leftrightarrow x \in B)$

# Review of Set Theory (集合论)

- Some commonly seen sets:
  - The empty set  $\emptyset$
  - The set of natural numbers  $\mathbb{N} = \{0, 1, 2, 3, \dots\}$
  - The set of positive integers  $\mathbb{N}_+ = \{1, 2, 3, \dots\}$
  - The set of integers  $\mathbb{Z} = \{0, 1, -1, 2, -2, 3, -3, \dots\}$
  - The set of rational numbers  $\mathbb{Q}$
  - The set of real numbers  $\mathbb{R}$

# Review of Set Theory (集合论)

- $A \subseteq B$ : Any member of  $A$  is also a member of  $B$ .
- Power set 幂集  $2^A$ : The set/family/collection of all subsets of  $A$ .
- $\{x \mid \varphi(x)\}$ : The set containing all objects  $x$  satisfying the property  $\varphi(x)$ 
  - $\{x \in A \mid \varphi(x)\}$ : The subset of  $A$  containing all objects  $x$  satisfying  $\varphi(x)$
  - Intervals on the real line:

$$[a, b] = \{x \in \mathbb{R} \mid a \leq x \leq b\}$$

$$(a, b) = \{x \in \mathbb{R} \mid a < x < b\}$$

In my lecture notes, an open interval is denoted by  $]a, b[$  so that you won't confuse it with the order pair

# Countable and Uncountable Sets

- Countably infinite set 可数无限集: A set whose members can be exhaustively enumerated by a sequence.
  - Examples:  $\mathbb{N}$ ,  $\mathbb{Z}$ ,  $\mathbb{Q}$
- Countable set 可数集: Finite or countably infinite
- Uncountable set 不可数集: A set that is not countable
  - Examples:  $\mathbb{R}$  and its open intervals

# Countable and Uncountable Sets

- Subsets of a countable set are countable.
- Given a countable set  $S$  and a mapping  $f$  defined on  $S$ , the **image** (像集/值域)

$$\text{Im } f = \{f(s) \mid s \in S\}$$

is a countable set.

# Set Operations

- Union:  $A \cup B = \{x \mid x \in A \text{ or } x \in B\}$
- Intersection:  $A \cap B = \{x \mid x \in A \text{ and } x \in B\}$
- Commutative and associative laws 交换律与结合律:

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

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

- Venn diagram

# Set Operations

- A more complicated situation:  $\mathcal{I}$  is any nonempty set, and for each  $i \in \mathcal{I}$  there is one corresponding set  $A_i$  assigned to  $i$ .
  - An indexed family of sets;  $i$  is the index (指标) of  $A_i$
  - Union of an indexed family of sets:

$$\begin{aligned}\bigcup_{i \in \mathcal{I}} A_i &= \{x \mid \text{there exists some } i \in \mathcal{I} \text{ such that } x \in A_i\} \\ &= \{x \mid x \text{ belongs to at least one } A_i\}\end{aligned}$$

# Set Operations

- A more complicated situation:  $\mathcal{I}$  is any nonempty set, and for each  $i \in \mathcal{I}$  there is one corresponding set  $A_i$  assigned to  $i$ .
  - An indexed family of sets;  $i$  is the index (指标) of  $A_i$
  - Intersection of an indexed family of sets:

$$\begin{aligned}\bigcap_{i \in \mathcal{I}} A_i &= \{x \mid \text{for all } i \in \mathcal{I} \text{ we have } x \in A_i\} \\ &= \{x \mid x \text{ belongs to all } A_i\}\end{aligned}$$

# Set Operations

- Some examples:

$$\bigcup_{i=1}^{\infty} \left[ 0, \frac{i}{i+1} \right] = \bigcup_{i \in \mathbb{N}_+} \left[ 0, \frac{i}{i+1} \right] = ?$$

$$\bigcap_{i=1}^{\infty} \left( 0, \frac{1}{i} \right) = \bigcap_{i \in \mathbb{N}_+} \left( 0, \frac{1}{i} \right) = ?$$

# Set Operations

- When  $\mathcal{I}$  and each  $A_i, i \in \mathcal{I}$  are all countable, the union and the intersection

$$\bigcup_{i \in \mathcal{I}} A_i \qquad \bigcap_{i \in \mathcal{I}} A_i$$

are countable.

# Set Operations

- Distributive laws 分配律:

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

$$A \cap \left( \bigcup_{i \in \mathcal{I}} B_i \right) = \bigcup_{i \in \mathcal{I}} (A \cap B_i)$$

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

$$A \cup \left( \bigcap_{i \in \mathcal{I}} B_i \right) = \bigcap_{i \in \mathcal{I}} (A \cup B_i)$$

# Set Operations

- Difference of sets 差集:  $A \setminus B = \{x \in A \mid x \notin B\}$
- Complement 补集  $B^c$ : When  $B \subseteq A$  and  $A$  is clear from the context.
- De Morgan's laws:

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

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

$$\left( \bigcup_{i \in \mathcal{I}} A_i \right)^c = \bigcap_{i \in \mathcal{I}} A_i^c$$

$$\left( \bigcap_{i \in \mathcal{I}} A_i \right)^c = \bigcup_{i \in \mathcal{I}} A_i^c$$

# Set Operations

- Cartesian product 笛卡尔积:  $A \times B = \{(\underline{a}, b) \mid a \in A, b \in B\}$   
Ordered pair
- The Cartesian product of two countable sets is countable.

## 1.2. Basics Notions of Probability Theory

# Sample Space, Events and Probability

- Suppose we want to model a random experiment.
- **Sample space** 样本空间: The set containing all possible outcomes of this random experiment.
- In our lectures, we usually use  $\Omega$  to denote the sample space, and use  $\omega$  to denote a possible outcome.
- Some quick examples:
  - Throwing a dice:  $\Omega = \{1, 2, 3, 4, 5, 6\}$
  - Flipping two coins:  $\Omega = \{\underline{(\text{H}, \text{H})}, (\text{H}, \text{T}), (\text{T}, \text{H}), (\text{T}, \text{T})\}$

Ordered pair

# Sample Space, Events and Probability

- **Event 事件:** An event  $E$  is a subset of the sample space.
- We say that the event  $E$  occurred in one trial of the random experiment, if the outcome  $\omega$  of this trial is a member of  $E$ .
- **Probability measure 概率测度:** A mapping/function  $\mathbb{P}$  that assigns each event  $E$  with a real number  $\mathbb{P}(E)$ .
  - Probability measures have to satisfy certain properties.  
We'll talk about it later.

# Quick Examples of Sample Spaces and Events

- Throwing a dice:  $\Omega = \{1, 2, 3, 4, 5, 6\}$ 
  - The outcome is four:  $\{4\}$
  - The outcome is an even number:  $\{2, 4, 6\}$
  - The outcome is a prime number:  $\{2, 3, 5\}$
  - The outcome is less than seven:  $\{1, 2, 3, 4, 5, 6\} = \Omega$  Certain event 必然事件
  - The outcome is not an integer:  $\emptyset$  Impossible event 不可能事件

# Quick Examples of Sample Spaces and Events

- Flipping three coins:  $\Omega = \{(\omega_1, \omega_2, \omega_3) \mid \omega_i \in \{\text{H}, \text{T}\}, i = 1, 2, 3\}$ 
  - The first coin lands on heads:
  - The first and the second coins turn out the same:
  - There is at most one tail:
  - At least two heads appear consecutively:
  - Four heads appear:

*How many elements are there in the samples space?*

*How many events are there in total?*

# More Examples

- Generating a random number in  $[0, 1]$ : Just set  $\Omega = [0, 1]$ 
  - The random number falls in the interval  $I$ :  $I \cap [0, 1]$
- Flipping a coin for (countably) infinitely many times:

$$\Omega = \{(\omega_i)_{i=1}^{\infty} \mid \omega_i \in \{\text{H}, \text{T}\}, \forall i \in \mathbb{N}_+\}$$

- The frequency of heads converges to  $1/2$ :

$$\left\{ (\omega_i)_{i=1}^{\infty} \in \Omega \mid \lim_{n \rightarrow \infty} \frac{q_n(\omega_1, \dots, \omega_n)}{n} = \frac{1}{2} \right\}$$

$q_n(\omega_1, \dots, \omega_n)$  counts  
the number of heads  
in  $\omega_1, \dots, \omega_n$

# Relations and Operations of Events

*Events are sets, and so set relations and operations apply to them.*

- $E \subseteq F$  holds if and only if  $E$  occurs implies  $F$  occurs.

$$\begin{aligned} E \subseteq F &\iff \text{for all } \omega \in E, \text{ we have } \omega \in F \\ &\iff \text{whenever the outcome } \omega \text{ is in } E, \text{ we have } \omega \in F \\ &\iff \text{whenever } E \text{ occurs, we have that } F \text{ occurs} \end{aligned}$$

- Complement of an event:  $E^c = \Omega \setminus E$  occurs if and only if  $E$  does not occur.
  - $E$  and  $E^c$  are called 对立事件.

# Relations and Operations of Events

*Events are sets, and so set relations and operations apply to them.*

- Union of events:  $E \cup F$  occurs if and only if  $E$  or  $F$  (or both) occurs.
- Intersection of events:  $E \cap F$  occurs if and only if both  $E$  and  $F$  occur.
  - $E$  and  $F$  are called disjoint or mutually exclusive 互斥 if  $E \cap F = \emptyset$ .
  - Disjoint events do not occur simultaneously in one trial.

# Relations and Operations of Events

*Events are sets, and so set relations and operations apply to them.*

- More complicated situations: Given a sequence of events  $E_1, E_2, E_3, \dots$

$$\bigcup_{i=1}^{\infty} E_i = \{\omega \in \Omega \mid \omega \text{ belongs to at least one } E_i\} \quad \text{At least one } E_i \text{ occurs}$$

$$\bigcap_{i=1}^{\infty} E_i = \{\omega \in \Omega \mid \omega \text{ belongs to all } E_i\} \quad E_1, E_2, E_3, \dots \text{all occur}$$

# Some Mathematical Intricacy

- Unfortunately, due to some mathematical intricacy (which is well beyond the scope of this course), sometimes we **CANNOT** assign probabilities to **ALL** subsets of the sample space.
- We allow the collection of events to be a **smaller** subset of the power set  $2^{\Omega}$ , but still requires it to satisfy certain properties for theoretical development.

# Event Space: Rigorous Definition

**Definition.** A collection  $\mathcal{F}$  of subsets of  $\Omega$  is called an **event space** 事件域, if

- The sample space  $\Omega$  is in  $\mathcal{F}$ .
- For each  $E \in \mathcal{F}$ , the complement  $E^c$  is in  $\mathcal{F}$ .
- For any sequence  $E_1, E_2, E_3, \dots$  in  $\mathcal{F}$ , the union  $\bigcup_{i=1}^{\infty} E_i$  is in  $\mathcal{F}$ .

\* Uncountable unions of members in  $\mathcal{F}$  may go out of  $\mathcal{F}$ .

# Probability Measure: Rigorous Definition

**Definition.** Given a sample space  $\Omega$  and an event space  $\mathcal{F}$  of  $\Omega$ , a mapping  $\mathbb{P} : \mathcal{F} \rightarrow \mathbb{R}$  is said to be a **probability measure** 概率测度 on  $\mathcal{F}$ , if

- $0 \leq \mathbb{P}(E) \leq 1$  for any  $E$  in the event space  $\mathcal{F}$ .
- $\mathbb{P}(\Omega) = 1$ .
- Countable additivity 可数可加性: For any sequence of mutually exclusive sets  $E_1, E_2, E_3, \dots$  in the event space  $\mathcal{F}$ , we have

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

# Probability Space

- The ordered triple  $(\Omega, \mathcal{F}, \mathbb{P})$  is called a **probability space** 概率空间 if
  - $\mathcal{F}$  is an event space of  $\Omega$ .
  - $\mathbb{P}$  is a probability measure on  $\mathcal{F}$ .
- Given a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ , we call members of  $\mathcal{F}$  **events**, and  $\mathbb{P}(E)$  the **probability** of  $E$  for any event  $E$ .
- In probability theory, we model a random experiment by a probability space.

# Problems Studied by Probability Theory

- Given a random experiment, suppose we know **i)** its sample space, and **ii)** the probabilities of some simple events (or their relations).
  - How to calculate the probabilities of other events we are interested in?
  - How to update/correct the probabilities if we obtain some partial information on the outcome of the random experiment?
- Problems not entirely within the scope of probability theory:
  - How to find those probabilities of simple events (or their relations) from the underlying physics or from data?
  - We need tools from **mathematical statistics**.

# Some Remarks on Event Spaces

- When the sample space  $\Omega$  is countable, we always choose the event space to be the power set  $2^\Omega$ .
- When the sample space is uncountable, things are complicated...
  - Out of the scope of this course.
  - Only theorists focus on this issue. You don't need to worry about it.
- In this course, we'll always assume an appropriate event space  $\mathcal{F}$  is given after the sample space is specified.

# Properties of Probability

- The sample space (必然事件) has probability one (by definition).
- The empty set (不可能事件) has probability zero.
- An event with probability one is not necessarily certain to occur.
  - Such events are called **almost sure** (几乎必然) events.
- An event with probability zero is not necessarily impossible.

# Properties of Probability

- Countable additivity implies **finite additivity**:

$$\mathbb{P}\left(\bigcup_{i=1}^n E_i\right) = \sum_{i=1}^n \mathbb{P}(E_i) \quad \text{if } E_1, \dots, E_n \text{ are } \underline{\text{mutually exclusive}}$$

- $\mathbb{P}(E^c \cap F) = \mathbb{P}(F) - \mathbb{P}(E \cap F)$  . Particularly,  $\mathbb{P}(E^c) = 1 - \mathbb{P}(E)$
- $\mathbb{P}(E) \leq \mathbb{P}(F)$  whenever  $E \subseteq F$ .
- Use Venn diagram to develop their intuitions.

# Properties of Probability

- $\mathbb{P}(E \cup F) = \mathbb{P}(E) + \mathbb{P}(F) - \mathbb{P}(E \cap F)$ 
  - $E$  and  $F$  are arbitrary.
  - Quite intuitive if we look at the Venn diagram.
  - Generalization to finitely many events:

We won't prove it now. You'll learn  
an easier way to prove it after you  
learn expectation (期望).

$$\begin{aligned} & \mathbb{P}(E_1 \cup \dots \cup E_n) \\ &= \sum_{i=1}^n \mathbb{P}(E_i) - \sum_{i_1 < i_2} \mathbb{P}(E_{i_1} \cap E_{i_2}) + \dots \\ & \quad + (-1)^{k-1} \sum_{i_1 < \dots < i_k} \mathbb{P}(E_{i_1} \cap \dots \cap E_{i_k}) + \dots \\ & \quad + (-1)^{n-1} \mathbb{P}(E_1 \cap \dots \cap E_n) \end{aligned}$$

# Properties of Probability

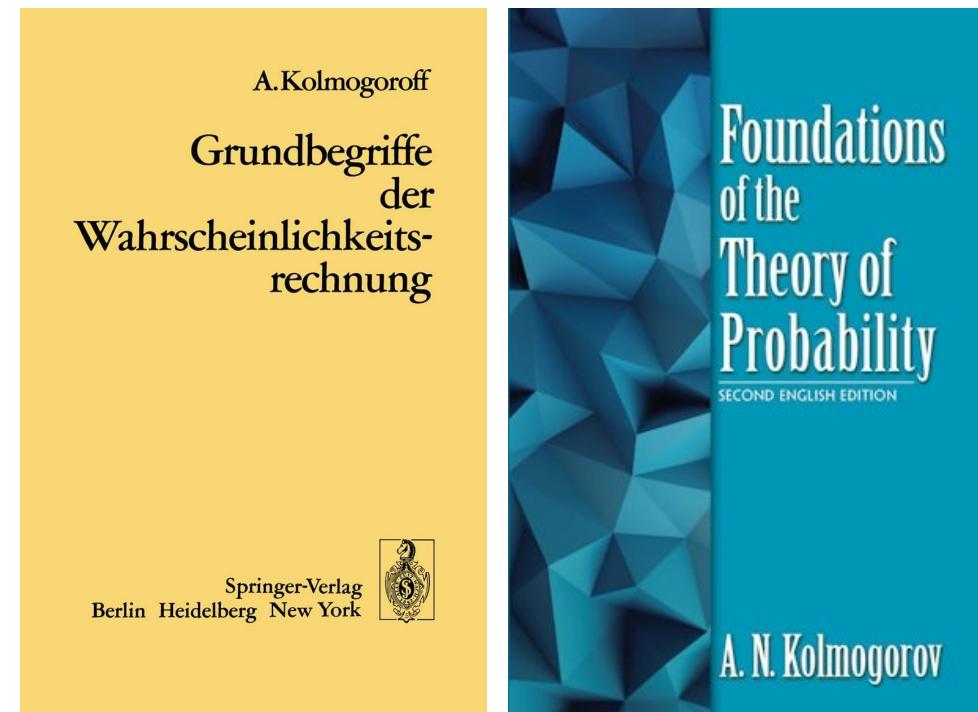
- Boole's inequality:

$$\mathbb{P}(E_1 \cup \dots \cup E_n) \leq \mathbb{P}(E_1) + \dots + \mathbb{P}(E_n)$$

- The proof using mathematical induction in this case is quite clean.

# Remarks on the Definition of Probability

- The definition of probability seems a bit indirect.
  - They do not tell you what probabilities are. They only tell you how probabilities should behave.
  - Some people call it the axiomatic definition of probability (概率的公理化定义).
  - First summarized and proposed by Kolmogorov (柯尔莫哥洛夫) in his 1933 book.
  - The foundation of modern probability theory.



# Remarks on the Definition of Probability

- Why not define probability directly as the limit of frequency?
  - It's going to be complicated...
  - We'll face infinitely repeated trials of random experiments from the start.
  - We'll see that the repeated trials of random experiments need to be **independent**, which is a concept we haven't introduced yet.
- In modern probability theory, it becomes a **theorem** that frequency converges to probability (almost surely).
- Modern probability theory can even apply to problems not related to random experiments with statistical regularity.

### 1.3. Classical Probability (古典概型)

# Classical Probability (古典概型)

- You have learned it in high school.
  - There is really not much more to say in this course.
  - Some problems/quizzes of classical probability are very tricky, though.
- How does classical probability fit in modern probability theory?

# Classical Probability (古典概率型)

- For a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  to be a classical probability model:
  - The sample space  $\Omega$  should be a finite set.
  - The event space should be the power set  $2^\Omega$ .
  - The probability measure should satisfy

$$\mathbb{P}(\{\omega\}) = \frac{1}{|\Omega|}, \quad \forall \omega \in \Omega \qquad \Rightarrow \quad \mathbb{P}(E) = \frac{|E|}{|\Omega|}, \quad \forall E \subseteq \Omega.$$

*All outcomes are equally probable.*

**Just count!**

# Overview of Basic Principles of Counting

- 乘法原理
- 加法原理
- 排列数  $\frac{n!}{(n - k)!}$
- 组合数 binomial coefficient  $\binom{n}{k} = \frac{n!}{k!(n - k)!}$
- 二项式定理 binomial theorem

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

# Overview of Basic Principles of Counting

- 多项式系数 multinomial coefficient: Suppose  $n_1 + n_2 + \cdots + n_r = n$

$$\binom{n}{n_1, n_2, \dots, n_r} = \frac{n!}{n_1! n_2! \cdots n_r!}$$

- The number of ways to partition a set  $A$  with  $n$  elements into  $r$  subsets  $A_1, A_2, \dots, A_r$ , where  $A_i$  has  $n_i$  elements.
- The number of permutations of  $r$  groups of objects, where the  $i$ th group consists of  $n_i$  indistinguishable objects.

# Overview of Basic Principles of Counting

- 多项式定理 multinomial theorem

$$(x_1 + \cdots + x_r)^n = \sum_{\substack{n_1 + \cdots + n_r = n \\ n_1, \dots, n_r \in \mathbb{N}}} \binom{n}{n_1, n_2, \dots, n_r} x_1^{n_1} x_2^{n_2} \cdots x_r^{n_r}$$

- We recommend that you do some exercises in Chapter 1 of *A First Course in Probability* by Sheldon Ross, to get acquainted with basic techniques of counting.

# The Banach Matchbox Problem

- A mathematician carries 2 matchboxes, 1 in left pocket and 1 in right pocket. Each time he needs a match, he is equally likely to pick a box from either pocket. Consider the moment when the mathematician first discovers that the box picked is already empty. Assuming that both matchboxes initially contained  $N$  matches, what is the probability that there are exactly  $k$  matches ( $0 \leq k \leq N$ ) in the other box?

## 1.4. Conditional Probability (条件概率)

# Conditional Probability

- Let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a probability space.
- Assume we observe that some event  $F$  has occurred. How should we update the probability measure?

# Conditional Probability

**Definition.** Given a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ , let  $F$  be an event satisfying  $\mathbb{P}(F) > 0$ . For any event  $E$ , The conditional probability of  $E$  given  $F$  is defined as

$$\mathbb{P}(E | F) = \frac{\mathbb{P}(E \cap F)}{\mathbb{P}(F)}.$$

- If  $\mathbb{P}(F) = 0$ , the conditional probability  $\mathbb{P}(E | F)$  is **UNDEFINED**.

# Conditional Probability

- Example: Suppose  $(\Omega, \mathcal{F}, \mathbb{P})$  is a classical probability model. Let  $F$  be an event that is not empty.
- Then for any event  $E$ , we have

$$\mathbb{P}(E | F) = \frac{|E \cap F|}{|F|} = \sum_{\omega \in E \cap F} \frac{1}{|F|}$$

- Outcomes in  $F$  are equally probable.
- Outcomes not in  $F$  have probability zero.

# Conditional Probability as a Probability Measure

**Proposition.** Given a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ , let  $F$  be an event satisfying  $\mathbb{P}(F) > 0$ . Then the mapping

$$E \in \mathcal{F} \mapsto \mathbb{P}(E|F) \in \mathbb{R}$$

is a probability measure on  $\mathcal{F}$  satisfying

- $\mathbb{P}(F|F) = 1$ .
- $\frac{\mathbb{P}(E_1|F)}{\mathbb{P}(E_2|F)} = \frac{\mathbb{P}(E_1)}{\mathbb{P}(E_2)}$  whenever  $E_1, E_2 \subseteq F$  and  $\mathbb{P}(E_2) > 0$ .

We can adopt  $\mathbb{P}(\cdot | F)$  as the updated probability measure after we observe  $F$  has occurred.

# Multiplication Rule 乘法公式

- We have

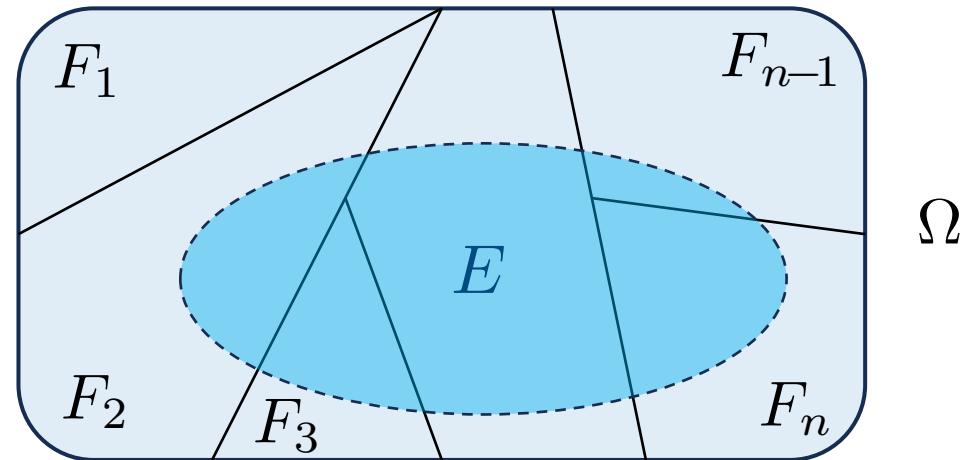
$$\begin{aligned}\mathbb{P}(E_1 \cap \cdots \cap E_n) &= \mathbb{P}(E_1) \cdot \mathbb{P}(E_2 | E_1) \cdot \mathbb{P}(E_3 | E_1 \cap E_2) \cdots \\ &\quad \cdot \mathbb{P}(E_n | E_1 \cap \cdots \cap E_{n-1})\end{aligned}$$

provided that  $\mathbb{P}(E_1 \cap \cdots \cap E_{n-1}) > 0$ .

# Law of Total Probability 全概率公式

- Suppose the events  $F_1, \dots, F_n$  satisfy
  1.  $F_1 \cup \dots \cup F_n = \Omega$
  2.  $F_i \cap F_j = \emptyset, \forall i \neq j$
  3.  $\mathbb{P}(F_i) > 0, \forall i$
- Then for any event  $E$ ,

$$\mathbb{P}(E) = \sum_{i=1}^n \mathbb{P}(E | F_i) \mathbb{P}(F_i)$$

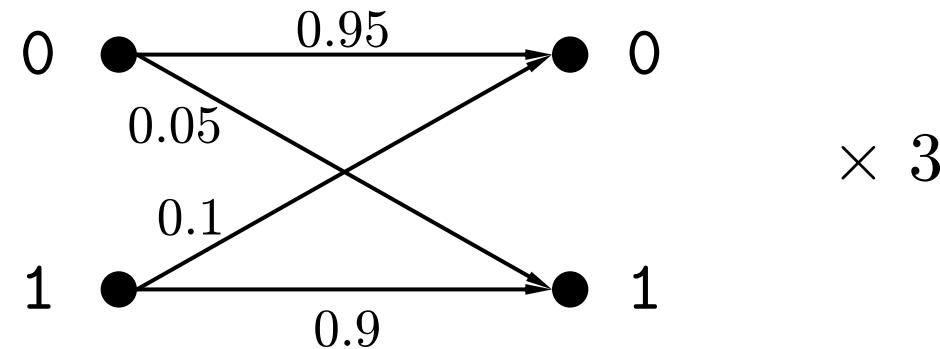


- A special case:

$$\begin{aligned}\mathbb{P}(E) &= \mathbb{P}(E | F) \mathbb{P}(F) \\ &\quad + \mathbb{P}(E | F^c) \mathbb{P}(F^c)\end{aligned}$$

# Exercise: Communication Through Noisy Channels

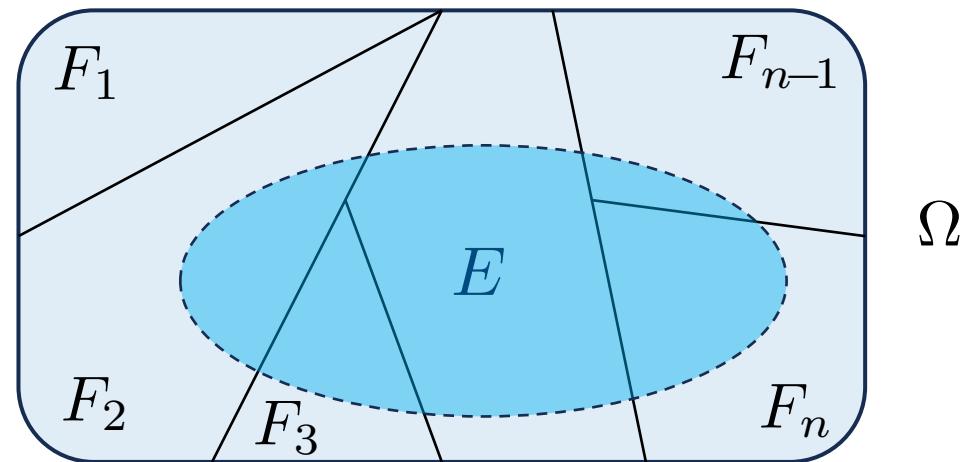
- A sensor sends data (a string of 0/1 symbols) to the server, and each symbol needs to be transmitted through three noisy channels sequentially.
- When 0 is transmitted through each channel, the error probability is 0.05. When 1 is transmitted through each channel, the error probability is 0.1.



- Suppose a symbol 0 is sent from the sensor. What is the probability that the server receives the symbol correctly?

# Bayes' Theorem 贝叶斯定理

- Suppose the events  $F_1, \dots, F_n$  satisfy
  - $F_1 \cup \dots \cup F_n = \Omega$
  - $F_i \cap F_j = \emptyset, \forall i \neq j$
  - $\mathbb{P}(F_i) > 0, \forall i$
- Then for any  $i$  and any event  $E$ ,



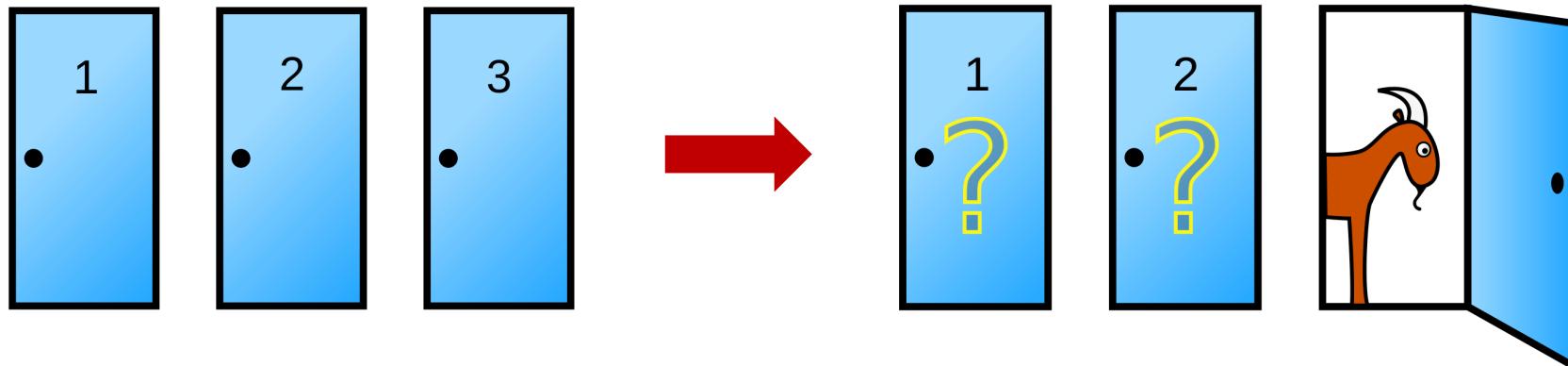
Posterior probability  
后验概率  $\boxed{\mathbb{P}(F_i | E)} = \frac{\mathbb{P}(E | F_i) \boxed{\mathbb{P}(F_i)}}{\sum_{j=1}^n \mathbb{P}(E | F_j) \boxed{\mathbb{P}(F_j)}}$  Prior probability  
先验概率

# Exercise: Diagnosis of A Rare Disease

- A rare disease has an incidence of 1 in  $10^5$  in the population at large. There is a diagnostic test, but it is imperfect. If a person has the disease, the test is positive with probability 99%; otherwise, the test is positive with probability 1%.
- Now suppose a person is randomly picked from the population, and the test result for this person turns out to be positive. What is the probability that this person has the disease?

# Exercise: Monty Hall Problem

- Suppose you're on a television show, and you're given the choice of three doors: Behind one door is a car; behind the others, goats.
- You pick a door, say No. 1.
- The host, who knows what's behind the doors, then opens another door that has a goat (say No. 3), and asks you, "Do you want to pick door No. 2?" Is it to your advantage to switch your choice?



## 1.5. Independence of Events (事件的独立性)

# Independence of Events

**Definition.** We say that two events  $E$  and  $F$  are **independent**, if

$$\mathbb{P}(E \cap F) = \mathbb{P}(E) \mathbb{P}(F).$$

- If  $F$  has a positive probability, then

$$E \text{ and } F \text{ are independent} \iff \mathbb{P}(E | F) = \mathbb{P}(E)$$

- If  $E$  has a positive probability, then

$$E \text{ and } F \text{ are independent} \iff \mathbb{P}(F | E) = \mathbb{P}(F)$$

*Knowing the occurrence of one does not affect the probability of the other.*

# Independence of Events

- An easy exercise: Let  $E$  and  $F$  be any events. The following statements are equivalent:
  - 1)  $E$  and  $F$  are independent.
  - 2)  $E$  and  $F^c$  are independent.
  - 3)  $E^c$  and  $F$  are independent.
  - 4)  $E^c$  and  $F^c$  are independent.

# Independence of Multiple Events

- **Definition.** Let  $\mathcal{I}$  be any nonempty set, and suppose for each  $i \in \mathcal{I}$  there is one corresponding set  $E_i$  assigned to  $i$ . We say that this family of events  $\{E_i, i \in \mathcal{I}\}$  are **mutually independent**, if for any nonempty finite subset  $\mathcal{J} \subseteq \mathcal{I}$ , we have

$$\mathbb{P}\left(\bigcap_{j \in \mathcal{J}} E_j\right) = \prod_{j \in \mathcal{J}} \mathbb{P}(E_j).$$

# Independence of Multiple Events

- Example: For three events  $E$ ,  $F$  and  $G$ , they are mutually independent if and only if all the following identities hold:

$$\mathbb{P}(E \cap F) = \mathbb{P}(E) \mathbb{P}(F)$$

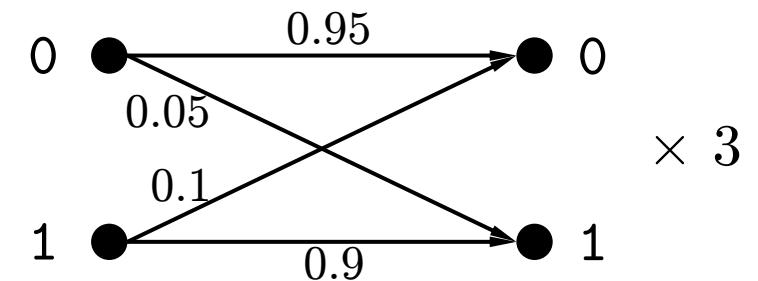
$$\mathbb{P}(E \cap G) = \mathbb{P}(E) \mathbb{P}(G)$$

$$\mathbb{P}(F \cap G) = \mathbb{P}(F) \mathbb{P}(G)$$

$$\mathbb{P}(E \cap F \cap G) = \mathbb{P}(E) \mathbb{P}(F) \mathbb{P}(G)$$

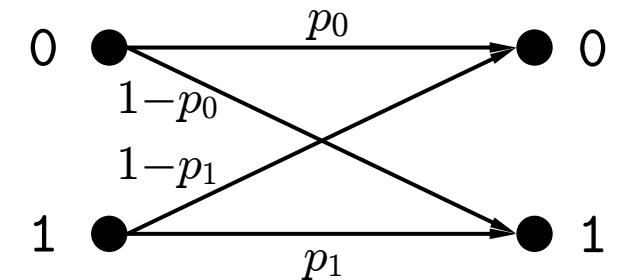
# Exercise: Communication Through Noisy Channels

- Still consider the case where a sensor transmits a string of 0/1 symbols to the server through three noisy channels sequentially.
- Further assume errors in different symbol transmissions are independent.
- To improve reliability, each symbol is transmitted three times and the server decodes the symbol by majority rule.
- What is the probability that 0 can be decoded correctly?



# Exercise: Communication Through Noisy Channels

- Still consider the case where a sensor transmits a string of 0/1 symbols to the server through three noisy channels sequentially.
- Further assume errors in different symbol transmissions are independent.
- To improve reliability, each symbol is transmitted three times and the server decodes the symbol by majority rule.
- What is the probability that 0 can be decoded correctly?



# Summary of Chapter 1

- **Sample space, events and probability measure**
  - Axiomatic definition & basic properties
- **Classical probability**
  - All outcomes are equally probable, so just count!
- **Conditional probability**
  - When you know something has occurred but still don't know the exact outcome
- **Independence of events**
  - More to come when you study random variables

# Assignment 1

- Will be posted on course.pku.edu.cn by today.
- Deadline is **October 8** in class.
- We encourage you to find extra problems from the recommended books for practising.