Math 170: Brief notes

Svetlana Makarova

Truong-Son Van

Contents

Chapter 1. Logic 4
1. What can logic be about? 4
2. Mathematical logic 6
3. Symbolic logic 8
4. Variables and quantifiers 10

Chapter 2. Sets and functions 16
1. Sets 16
2. Functions 19

Chapter 3. Mathematical induction 23
1. All natural numbers are interesting 23
2. Proof by induction 25
3. Structure of proofs by induction 25
4. Exercises 25

Chapter 4. Playing games with infinity 28
1. Games 28
2. Infinity 28

Chapter 5. Graph theory 30
1. Warm-up 30
2. Graph theory 34
3. Coloring maps and what graphs have to do with it 35
4. Chromatic polynomial of a graph 43
5. Euler characteristics 47

Chapter 6. Discrete probability 50
1. The basics 50
2. Probability from counting 53
3. Conditional probability 57
4. Independence 58
5. Random Variable and Expectation 60

Bibliography 62

CHAPTER 1

Logic

At its core, mathematics is a way of reasoning and is very similar to philosphy. The first part of this chapter wil reflect this basic observation. However, what sets apart mathematics from general philosophy is that the language of mathematics requires precision. There should be no ambiguity in a mathematical statement. The main goal of this chapter is to give the students a taste of what it is like to be mathematically precise.

1. What can logic be about?

We follow [Sai91] for this part.

Most broadly, logic is about reasons and reasoning. There are reasons for *acting*: you may avoid sugary desserts for the reason of wanting to keep thin or lose weight. There are reasons for *believing*: you may think that the potatoes are ready to eat for the reason that they have been boiling for twenty minutes. Historically, logic has concerned itself with reasons for believing. But even this question can be answered in various ways. For example, I asked an Indian friend of mine why she believes that she should not eat meat. Her answer was that this belief was instilled in her by her family at an early age. This explains the origin of this belief, but does not give a *reason* for it. But then she continued her answer and said that now she doesn't like the smell of meat, and eating food that smells bad is usually a bad idea. This also justifies her belief. Some other people may say that killing anything is wrong, and eating meat requires killing, thus reasoning why they shouldn't eat meat.

The way it works is that one deduces the reasons for a certain belief by making it a consequence of a "higher", more abstract, belief. Logic as a discipline of "good reasoning" was first considered as early as the 6th century B.C. and independently in India, Greek and China.

Example 1.1. Consider the following chain of sentences.

"I believe that humans breath oxygen to live. I believe that I am a human. Therefore, I believe that I breath oxygen to live."

"All humans breath oxygen" is a higher, more abstract, fact than " I breath oxygen", as I am just a particular member of the human race.

Warning 1.2. *Not every belief has its reasons. Every logical system has its core "beliefs" (called* axioms*) that are taken for granted, and they have*

no further explanation. These are the most abstract beliefs that are used to deduce every other belief in the same logical system. To have reasons, one needs to take a leap of faith at some point.

Logic is a *normative* discipline. It sets out standard for *good* and *bad* arguments. These are technical terms and should not be confused by subjective opinions. However, the these technical terms are somewhat inspired by daily commonsense distinction between good and bad reasons. In our daily conversation, to make good reasons for something is to create premises so that the something *follows*.

EXAMPLE 1.3. "James is a banker and all bankers are rich" is a good reason for "James is rich."

QUESTION. *If James is not a banker, can we conclude that he is not rich?*

EXAMPLE 1.4. "James likes expensive cars" is not a good reason for "James is rich."

It is important to note that one can discern good and bad reasons without having to believe in the reasons themselves. In fact, Einstein himself did not believe in (even reject) quantum physics while being one of the founding fathers of the theory. A lot of modern mathematics revolves around physics and biology but a lot of mathematicians barely know any physics or biology (confession time).

1.1. Inductive and deductive logic. The result of assembling premises and conclusions together is called an *argument*. An argument is *valid*, or *true*, or *good*, if the conclusions follow from the premises. The two most common forms of logic are inductive logic and deductive logic.

An example of inductive logic is the following.

EXAMPLE 1.5. The sun has risen every morning so far; therefore it (probably) will rise tomorrow.

EXERCISE 1. *Contrast the previous example with the following sentence:*

" The sun has risen every morning so far; therefore it (probably) will NOT rise tomorrow."

Is one of these sentences more true than the other? How do you know?

An example of deductive logic is the following.

EXAMPLE 1.6. All men are mortal. Socrates is a man. Therefore, Socrates is mortal.

Thus, a way to distinguish between inductive and deductive logic is: for deductive logic, it is impossible for the conclusion to be false if the premises

are true. For inductive logic, this is not the case as the conclusion in this case may be false despite the premises being true. We can see that it is only in deductive logic, one can talk about the validity of an argument. In inductive logic, there are degrees in strength of an argument but inductive reasoning can *never be valid* by our definition of validity[1]. However, an inductive argument can be stronger than another inductive argument (just make sure one is talking about the same thing – comparing apples to apples and not to pears).

Mathematics is all about deductive logic whereas science must involve both inductive and deductive logic. The combination of inductive and deductive logic in science gives birth to the need of probability and statistics (whose theories are all mathematical and deductive), and in modern day data science and machine learning that are based on statistics.

WARNING 1.7. *Do not confuse inductive logic with the method of induction, which is a method in deductive logic. Although, there are resemblance between the two. The difference is that in the method of induction, one is given the super power in theory to transcend time to "go off to infinity" whereas inductive logic is limited by physical evidence, where time is a major road block...*

EXERCISE 2. *Make a table of comparison between inductive and deductive logic.*

Watch this lecture about inductive logic: https://youtu.be/DRx-3jvC918

EXERCISE 3. *What's wrong with the following?*
Statement: You have horns.
"Proof": What you haven't lost, you have. You haven't lost your horns. Therefore you have horns.

EXERCISE 4. *What's wrong with the following?*
Statement: You don't know your father.
"Proof": I show you a photo of someone, the photo is covered by a cloth. Do you know who's in the photo? You can't see, so you don't. But it's a photo of your father. Therefore you don't know your father.

2. Mathematical logic

We follow [New] for this part.

Mathematical logic is the study of logic restricted to mathematics. Its existence is to address the biggest problem in the foundation of mathematics: are theories of mathematics consistent with each others? That said, many

[1]Be careful here that validity is a technical term and should not be confused with the daily use of the word

working mathematicians do not pay attention to the question of foundations, which might be a worrisome fact. I can only have my fingers crossed that mathematics will not fall apart one day (which, it did for a period of time, when Georg Cantor discovered different infinities in the 19th century)...

A **mathematical statement** is a statement that at least the statement maker has to be able to assign a **truth value** ('true' or 'false') to it. The truth assignment could be the result of an immediate observation or a long chain of difficult reasoning. To make the truth assignment valid, every single argument in the chain of reasoning must be valid. A **proof** of a mathematical statement is a chain of valid arguments that make the mathematical statement *true*.

Sveta: ?

SV Aug 13: fixed- thanks

There are many names for a *true* mathematical statement, depending on the use:

- **Theorem:** a particularly important mathematical statement given the context.
- **Proposition:** general term that can be used anytime.
- **Lemma:** a mathematical statement that will be used as a stepping stone to prove a theorem.
- **Corollary:** a mathematical statement whose truth value could be deduced from a theorem without much effort.

A statement that are believed to be verifiable but no human has seen or discovered its proof yet is called a **conjecture**.

2.1. Structure. Every mathematical statement has the following structure:

Assumptions + Goals

EXAMPLE 2.1. Suppose Philadelphia is in Massachusetts and Penn is in Philadelphia, then Penn is in Massachusetts.

WARNING 2.2. *Assumptions themselves need not to be true. We will talk more about this later when we talk about truth table.*

EXERCISE 5. *Find an example of a famous mathematical statement that its assumptions are not yet verified.*

Sveta: What are your examples?

SV aug 13: I have one example about the hypothesis that particles in a room behave as if they are independent identically distributed. This is the key assumption to derive the Boltzmann equation. Something that stem from Riemann hypothesis?

EXAMPLE 2.3. For example, abc conjecture implies Fermat's last theorem. But as of now, the status of the abc conjecture is subject to debate. A Fields medalist Peter Scholze and Jakob Stix found a gap in Mochizuki's proof.

Now a quote from Wikipedia: "Scholze and Stix wrote a report asserting and explaining an error in the logic of the proof and claiming that the resulting gap was "so severe that ... small modifications will not rescue the

proof strategy"; Mochizuki claimed that they misunderstood vital aspects of the theory and made invalid simplifications.

On April 3, 2020, two Japanese mathematicians announced that Mochizuki's claimed proof would be published in Publications of the Research Institute for Mathematical Sciences (RIMS), a journal of which Mochizuki is chief editor. In March 2021, Mochizuki's proof was published in RIMS."

We will end this section by discussing a few logical axioms that look obvious but people use all the time in mathematics without realizing it.

3. Symbolic logic

One of the main goals of mathematics is to reduce complicated statements/observations to simple abstract structures that are more tractable to human minds and still keep the essential features of the things one would like to study. This is as much of an art as anything else because too much abstraction would lead to triviality, which may not be very interesting.

Symbolic logic is a system of logic that can be used to reduce a mathematical statement into "agreed" formulas that are easier to determine its truth value.[2]

Let us consider a simple example from [New]:

ex:divides

EXAMPLE 3.1. If c divides b and b divides a, then c divides a.

We see that each of the statements "c divides b", "b divides a", and "c divides a" is a proposition if they stand alone by themselves. Thus, abstractly, each of them could be assigned a symbol

- $P = c$ divides b
- $Q = b$ divides a
- $R = c$ divides a

Then you can write

$$\text{If } P \text{ and } Q, \text{ then } R.$$

The above form of the statement looks easier to follow (at least to the mind of a non-English speaker) since at least we don't need to know what "divides" means. While it is not too useful in terms of conveying knowledge, it is extremely useful when it comes to determining the validity of the statement itself. This leads us to the next question: *What makes a statement true?*

[2]Gottfried Leibniz (another inventor of Calculus) was among the first people to realize the importance of having a system of logic that is universal and calculatable but couldn't actualize this dream. The goal was to reduce confusions and disputes among philosophers and arguers. (Just turn on the Presidential debates and you will understand why we need such a system...) The first well-known work that successfully made symbolic logic a mathematical field was by George Boole in 1854 [Boo09]. One of the earliest work that started the modern account of logic and foundation of mathematics was by Russel and Whitehead [WR97] (there is a comic book about it [DP09]!).

We will need a few new words.

DEFINITION 3.2. A **propositional variable** is a symbol that represents a proposition.

As we said earlier, propositions are just mathematical statements, which are required to have truth values ('true' or 'false').

DEFINITION 3.3. A **logical operator** is a symbol (or collection of words) that turn one or more propositional variables to a *single* new statement.

Basic logical operators are:

- Conjuction ('and', $\wedge$)
- Disjunction ('or', $\vee$)
- Implication ('if...then...', $\implies$)
- Negation ('not', $\neg$)

As simple as they look, these four operators build most of mathematics and anything that require reasoning (philosophy, law, computer science, etc.).

DEFINITION 3.4. A **propositional formula** is an expression that is either a propositional variable, or is built up from simpler propositional formulae using logical operators.

REMARK 3.5. When I ask, "What is the variable for the proposition?", I am more interested in what the symbol you give to the mathematical statement, not so much the content of it. Similarly, when I ask, "What is the formula for the proposition?" I am more interested in the way the proposition is written up, not so much what the proposition conveys.

The simplest kind of propositions is one that only contains one single propositional variable that already explicitly has the truth value ('true' or 'false') known (whether it is a *proven statement*, ~~a common knowledge~~ or an assumption). From these atomic propositions, we could build more complicated kinds of propositions with more complicated propositional formula by obeying certain logical rules of the logical operators[3]. This process is entirely "calculatable". Here are the rules for the basic logical operators listed above.

SV 2021-08-17: I emphasize that common knowledge may not be trusted always

Sveta 08-26: We can ask to give examples of "common knowledge" about some topic, e.g. time (less divisive), and find contradictions. * Time is money. * Life is a marathon, not a sprint. * The two most powerful warriors are patience and time. * Time waits for no one. * Better three hours too soon than a minute too late. * The key is in not spending time, but in investing it. * Punctuality is the thief of time.

SV 08-28: Agreed.

Conjunction ('and', $\wedge$). The propositional formula for a proposition made by a conjuction has the following form

$$P \wedge Q\,.$$

RULE. *The proposition $P \wedge Q$ (we say "P and Q") is true if* **both** *P and Q are true. Otherwise, if either (or both) P or Q is false, $P \wedge Q$ is false.*

[3]Even though we call them rules, they follow an intuitive model of our daily reasoning. The advantage of defining explicit rules is to make the reasoning more consistent by the abstraction.

Disjunction ('or', $\vee$). The propositional formula for a proposition made by a disjunction has the following form

$$P \vee Q\,.$$

RULE. *The proposition $P \vee Q$ (we say "P or Q") is true if* **either one** *(or both) of P or Q is true. $P \vee Q$ is flase if* **both** *P and Q are false.*

Implication ('if...then...', $\implies$). The propositional formula for a proposition made by a disjunction has the following form

$$P \implies Q\,.$$

RULE. *The proposition $P \implies Q$ (we say "P implies Q") is true if one of the following cases holds:*

- *P is true and Q is true.*
- *P is false.*

EXERCISE 6. *This is one of the most confusing rules in logic. Meditate on the rule of implication.*

Negation ('not', $\neg$). The propositional formula for a proposition made by a disjunction has the following form

$$\neg P\,.$$

RULE. *The proposition $\neg P$ (we say "not P") is true if P is false.*

EXERCISE 7. *Mess with your friends/parents/siblings with 'and', 'or', 'if not ... then...'*

AXIOM 1 (Law of excluded middle). Let P be a propositional formula. Then $P \vee (\neg P)$ is true. In plain English, this says that every proposition is either true or false.

AXIOM 2 (Principle of explosion). If a contradiction is assumed, any consequence may be derived.

4. Variables and quantifiers

4.1. Variables. It is nice to know basic rules of logic and how propositions work in sequences in order to produce proofs (arguments). It is unfortunate, however, that if we only work with propositions, our reasoning would be fairly limited. Consider the following sentence:

"x is divisible by 7."

QUESTION. *Is this a proposition?*

ANSWER. This is not a proposition as one cannot assign a truth value to it– we don't know what x is. □

Sveta 08-26: We can include a topic on most common logical fallacies and advertise blog LessWrong and the long fanfic "Harry Potter and the Methods of Rationality".

SV 08-28: That sounds like fun.

If we know a specific value of x, we would be able to determine the truth value of the sentence above and it would become a proper proposition. For example, if $x = 49$, the sentence would be true and if $x = 42$, it would be false. If we suppose that x should belong the set of natural number, $\mathbb{N}$, then The symbol x is called a *free variable* the set $\mathbb{N}$.

DEFINITION 4.1. Let x be a variable that is understood to refer to an element of a set X. In a statement involving x, we say it is *free* if it makes sense to substitute particular elements of X in the sentence; otherwise we say x is *bound*.

So if the sentence in the above question is not a proposition, what do we call it? Glad you ask– statements like those, which depend on free variables (hence abstract the notion of proposition) are called *predicate*. More formally, we have the following definition.

DEFINITION 4.2. A *predicate* is a symbol P with a specified list of free variables $x_1, x_2, \ldots, x_n$ and, for each variable x_i, a specification of a set X_i (called the *domain of disclosure* of x_i).

Notation: we will typically write $P(x_1, \ldots, x_n)$ in order to make the variable explicit.

EXAMPLE 4.3. We can represent the sentence 'x is divisible by 7' by $P(x)$, where $x \in \mathbb{N}$. $P(49)$ is true and $P(10)$ is false.

e:exists

EXAMPLE 4.4. The sentence "there exist integers a, b such that $x = a^2+b^2$" has free variable x and bound variables a^2+b^2, and can be represented by a predicate $R(x)$, where the domain of disclosure can be chosen to be $\mathbb{Z}$.

REMARK 4.5. A predicate with no free variables is precisely a propositional variable.

EXERCISE 8. *How would you represent the sentence "$x - y$ is rational" as a predicate.*

e:every

EXERCISE 9. *How would you represent the sentence "every even natural number $n \geqslant 2$ is divisible by k"?*

4.2. Quantifiers. Looking back to Example 4.4 and Exercise 9, we notice that the bound variables come with either "there exist" or "every". Without those terms, those variables will be come free variables.

Expressions that refer to *how many* elements of a set make a statement true, such as "there exists" and "every" turn free variables into bound variables. We represent such expression using symbols called *quantifiers*.

In mathematics, there are two *universal quantifiers*, $\forall$ (every) and $\exists$ (there exists).

EXAMPLE 4.6. intuitively speaking,

- The expression "$\exists x \in X$" denotes "there exists $x \in X$".
- The expression "$\forall x \in X$" denotes "for every $x \in X$".

Just like we can build propositional formulae from propositions and logical operators, we can build something out of predicates and logical operators.

DEFINITION 4.7 (**Logical formula**). A *logical formula* is an expression that is built from predicates using logical operators and quantifiers; it may have both free and boundary variables. The truth value of a logical formula depends on the free variables according to the rules for logical operators and quantifiers.

It is an important skill to translate from human languages into purely symbolic logical formulae and vice versa.

Formally, we have the following definitions of quantifiers.

DEFINITION 4.8 (**The universal quantifier** $\forall$). If $p(x)$ is a logical formula with free variable with free variable x with domain X, then $\forall x \in X, p(x)$ is the logical formula defined according to the following rules:

- If $p(x)$ can be derrived from the assumption that x is an arbitrary element of X, then $\forall x \in X, p(x)$;
- If $a \in X$ and $\forall x \in X, p(x)$ is true, then $p(a)$ is true.

DEFINITION 4.9 (**The universal quantifier** $\exists$). If $p(x)$ is a logical formula with free variable x with domain X, then $\exists x \in X, p(x)$ is the logical formula defined according to the following rules:

- If $a \in X$ and $p(a)$ is true, then $\exists x \in X, p(x)$;
- If $\exists x \in X, p(x)$ is true, and q can be derived from the assumption that $p(a)$ is true for some fixed $a \in X$, then q is true.

There are more quantifiers out there in the wild world of mathematics but they depend on specific fields of study. The above two quantifiers are used in every field of mathemtics, however.

One can combine quantifiers in a logical formula and the order of the quantifiers matter.

EXERCISE 10. *Translate the following expressions and convince yourself that they are different.*

(1) $\forall x \in \mathbb{Z}, \exists y \in \mathbb{Z}, \exists z \in \mathbb{Z}, x = y^2 + z^2$.
(2) $\exists y \in \mathbb{Z}, \exists z \in \mathbb{Z}, \forall x \in \mathbb{Z}, x = y^2 + z^2$.
(3) $\exists y \in \mathbb{Z}, \forall x \in \mathbb{Z}, \exists z \in \mathbb{Z}, x = y^2 + z^2$.

Are those propositions?

Are the following statements different?

(1) $\forall x \in \mathbb{Z}, \exists y \in \mathbb{Z}, \exists z \in \mathbb{Z}, x = y^2 + z^2$.
(2) $\forall x \in \mathbb{Z}, \exists z \in \mathbb{Z}, \exists y \in \mathbb{Z}, x = y^2 + z^2$.

4.3. Logical equivalence. We start out with a question

QUESTION. *Let P be the set of all prime numbers. Are these two logical formulae the same?*

(1) $\forall n \in P, (n > 2 \implies (\exists k \in \mathbb{Z}, n = 2k + 1))$.
(2) $\neg\exists n \in P, (n > 2 \wedge (\exists k \in \mathbb{Z}, n = 2k))$.

In plain English, the two logical formulae read as follows.

(1) Every prime number greater than two is odd.
(2) There does not exists an even prime number greater than two.

Because of the way they are framed, one would go on to prove these statements by using two completely different techniques.

(1) Fix a prime number n, assume that $n > 2$, and then prove that $n = 2k + 1$ for some $k \in \mathbb{Z}$.
(2) Assume that there is some prime number n such that $n > 2$ and $n = 2k$ for some $k \in \mathbb{Z}$ and derive a contradiction.

QUESTION. *Which strategy is easier to follow/prove?*

One can see that knowing more ways to rephrase a statement gives us more ways to prove/disprove it. The notion of *logical equivalence* tells us exactly when two statements have the same logical meaning, hence gives us confidence to think about one statement in the form of a different statement without worrying about being led astray by irrelevant thoughts and discover later that all the work we have done was in vain (even though this happens all the time).

DEFINITION 4.10. Let P and Q be logical formulae. We say that P and Q are logically equivalent and write $P \equiv Q$ if Q can be derived from P and P can be derived from Q.

ex:equiv1

EXAMPLE 4.11. We claim that

$$P \wedge (Q \vee R) \equiv (P \wedge Q) \vee (P \wedge R)\,,$$

where P, Q, R are propositional variables.

PROOF. First, assume that $P \wedge (Q \vee R)$ is true. Then P is true and $Q \vee R$ is true by definition of conjunction. By definition of disjunction, either Q is true or R is true. So, we divide our reasoning into two cases:

- If Q is true, then $P \wedge Q$ is true by definition of conjunction.
- If R is true, then $P \wedge R$ is true by definition of conjunction.

In both cases, we have that $(P \wedge Q) \vee (P \wedge R)$ is true by definition of disjunction.

Second, assume that $(P \wedge Q) \vee (P \wedge R)$ is true. Then, either $P \wedge Q$ is true or $P \wedge R$ is true by definition of disjunction. Again, we divide our reasoning into two cases:

- If $P \wedge Q$ is true, then P is true and Q is true by definition of conjunction.
- If $P \wedge R$ is true, then P is true and R is true by definition of conjunction.

In both cases, we have P is true and $Q \vee R$ is true by definition of disjunction. Therefore, $P \wedge (Q \vee R)$ is true.

Thus, we have derived that $(P \wedge Q) \vee (P \wedge R)$ is true from $P \wedge (Q \vee R)$ being true and vice versa. This proves our claim by definition of logical equivalence. □

Now we turn to everyone's favorite tool, the *truth table*.

4.4. Truth table.

Definition 4.12. The truth table of a propositional formula is the table with one row for each possible assignment of truth values to its constituent propositional variables, and one column for each sub-formula (incluidng the propositional variables and the propositional formula itself). The entries of the truth table are the truth values of the sub-formulae.

There are many ways that one could employ to prove the logical equivalences of propositional formulae. The most fundamental way is to use the definition. Another way that is one of the all-time favorites is to use the truth table: in order to prove that propositional formulae are logically equivalent, it suffices to show that they have identical columns in a truth table.

Example 4.13. We will prove the what's claimed in Example 4.11 using the truth table.

P	Q	R	$Q \vee R$	$P \wedge (Q \vee R)$	$P \wedge Q$	$Q \wedge R$	$(P \wedge Q) \vee (P \wedge R)$
T	T	T	T	T	T	T	T
T	T	F	T	T	T	F	T
T	F	T	T	T	F	T	T
T	F	F	F	F	F	F	F
F	T	T	T	F	F	F	F
F	T	F	T	F	F	F	F
F	F	T	T	F	F	F	F
F	F	F	F	F	F	F	F

Example 4.14. Use the truth table to show that

$$P \implies Q \equiv (\neg P) \vee Q\,.$$

Theorem 4.15 (De Morgan's laws for logical operators). *Let P, Q be propositional variables. Then,*

(1) $\neg(P \vee Q) \equiv (\neg P) \wedge (\neg Q)$,
(2) $\neg(P \wedge Q) \equiv (\neg P) \vee (\neg Q)$.

THEOREM 4.16 (De Morgan's laws for quantifiers). *Let $P(x)$ be a logical predicate and X be a set. Then,*

(1) $\neg(\forall x \in X, P(x)) \equiv \exists x \in X, \neg P(x)$,
(2) $\neg(\exists x \in X, P(x)) \equiv \forall x \in X, \neg P(x)$.

CHAPTER 2

Sets and functions

1. Sets

The notion of a *set* is extremely fundamental (as we already used it in previous discussions) yet, if not defined carefully, could lead to paradoxes and break mathematics from its core. Perhaps the most infamous paradox in naive set theory is the so-call Russell's paradox. However, for the sake of sanity in this course, let's stick with the naive notion and trust that mathematicians already fixed this issue and unbroke mathematics...

DEFINITION 1.1. A set is a collection of objects. The objects in the sets are called *elements* of the set. If x is an element in the set X then we write $x \in X$. We write $x \notin X$ to mean $\neg(x \in X)$.

The way that Bertrand Russell broke naive set theory is via the following chain of reasoning, taken directly from Wikipedia:

> Let R be the set of all sets that are not members of themselves. If R is not a member of itself, then its definition entails that it is a member of itself; if it is a member of itself, then it is not a member of itself, since it is the set of all sets that are not members of themselves.

We will need to use the so-called *set-builder notation* to describe sets in general. Given a set X, set set of elements of X satisfying some property $P(x)$ is denoted by

$$\left\{x \in X \mid P(x)\right\}.$$

From now on, we define the set of all rationals to be

$$\mathbb{Q} = \left\{p/q \mid p \in \mathbb{Z}, q \in \mathbb{Z} \text{ and } q \neq 0\right\},$$

and the set of real numbers to be

$$\mathbb{R} = \left\{\text{rationals and irrational numbers}\right\}.$$

We are cheating in the definition of real numbers above but that is too technical for the moment. Let's just go with what you imagine it to be from high school. We will also use the usual open and closed set notations.

$$[a, b] = \left\{x \in \mathbb{R} \mid a \leqslant x \leqslant b\right\},$$
$$[a, b) = \left\{x \in \mathbb{R} \mid a \leqslant x < b\right\},$$

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

See Newstead [New] (Chapter 2) for representations of intervals on the number line.

DEFINITION 1.2. Let X be a set. A *subset* of X is a set U such that

$$\forall a, (a \in U \implies a \in X).$$

We write $U \subseteq X$ for the assertion that U is a subset of X. The notation $\subsetneqq$ means that U is a *proper subset* of X, that is a subset of X that is not equal to X.

In order to prove that U is a subset of X, it is sufficient to take an arbitrary element $a \in U$ and prove that $a \in X$.

EXAMPLE 1.3. Prove that $\mathbb{Z} \subset \mathbb{Q}$.

We want to be able to say when two sets are the same (equal) with each other. There are different ways to go about this. One can say that two sets are equal if they have the exact same definition or there definition are somehow logically equivalent to each other. In real practice, these are not very useful. However, there are those who believe that "we are what we are made of." This sounds like a reasonable description and is a criteria to distinguish two different people. Surely, even though our bodies might be made of atoms but my atoms are different than your atoms. This inspires the following axiom about set equality.

AXIOM 3 (Axiom of extentionality). Let X and Y be sets. Then $X = Y$ if and only if $X \subseteq Y$ and $Y \subseteq X$.

EXAMPLE 1.4. Prove that,

$$\{x \in \mathbb{R} \mid x^2 \leqslant 1\} = [-1, 1].$$

One question may arise when we define intevals. Consider, (a, b), for example. Some people might object this definition because if $a > b$ we have a contradiction

$$b < a < x < b.$$

How can this be? A careful look at this definition, we realized that it says, if $x \in \mathbb{R}$ and $a < x < b$, then we admit x into the set. If the description itself does not make sense, we can't even start to consider anything, let alone the admission. When this situation, the set (a, b) simply does not contain any element, and we call that an empty set.

DEFINITION 1.5. A set is *non-empty* if it contains at least one element. Otherwise, it is *empty*.

QUESTION. *How many empty sets are there?*

ANSWER. There is only one empty set. That is if E' and E are both empty set, then

$$E' = E\,.$$

To see this, we want to show $E \subseteq E'$ and $E' \subseteq E$ (axiom of extentionality). By definition, Let a be an element in the universe that E belongs to. $a \in E$ is always false because E is empty. Therefore, the statement

$$a \in E \implies a \in E'$$

is always true. By definition of subsets, $E \subseteq E'$. Likewise, $E' \subseteq E$, showing our claim. □

1.1. Set operations. We will introduce some basic operations on sets. There are many more but the interested reader could find them in fuller details in Newstead's book.

DEFINITION 1.6 (Pairwise intersection). Let X and Y be sets. The *pairwise intersection* of X and Y, denoted $X \cap Y$ is defined by

$$X \cap Y \stackrel{\text{def}}{=} \{a \mid a \in X \wedge a \in Y\}\,.$$

EXAMPLE 1.7. Let $X = \{1, 2, 3, 4\}$ and $Y = \{2, 4, 7\}$, then

$$X \cap Y = \{2, 4\}\,.$$

DEFINITION 1.8 (Pairwise union). Let X and Y be sets. The *pairwise union* of X and Y, denoted $X \cup Y$ is defined by

$$X \cup Y \stackrel{\text{def}}{=} \{a \mid a \in X \vee a \in Y\}\,.$$

EXAMPLE 1.9. Let $X = \{1, 2, 3, 4\}$ and $Y = \{2, 4, 7\}$, then

$$X \cup Y = \{1, 2, 3, 4, 7\}\,.$$

DEFINITION 1.10 (Relative complement). Let X and Y be sets. The *relative complement* of Y and X, denoted $Y \setminus X$ is defined by

$$Y \setminus X \stackrel{\text{def}}{=} \{a \mid a \in Y \wedge a \notin X\}\,.$$

EXAMPLE 1.11. Let $X = \{1, 2, 3, 4\}$ and $Y = \{2, 4, 7\}$, then

$$Y \setminus X = \{7\}\,,$$
$$X \setminus Y = \{1, 3\}\,.$$

1.2. Venn diagram. For this part, please refer to in-class lecture. You can prove the following using Venn diagram.

THEOREM 1.12 (de Morgan's laws for sets). *Let X, Y, Z be sets. We have*

(1) $X \setminus (Y \cup Z) = (X \setminus Y) \cap (X \setminus Z)$,
(2) $X \setminus (Y \cap Z) = (X \setminus Y) \cup (X \setminus Z)$.

DEFINITION 1.13 (Ordered pair). For any two objects, x and y, an ordered pair (x, y) is the notation for the two objects being arranged in that particular order.

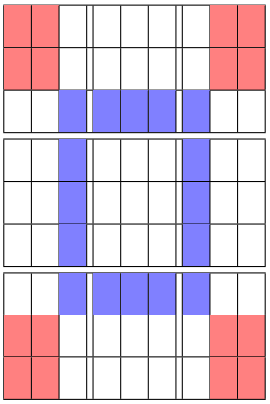
Thus, $(x, y) \neq (y, x)$ unless $x = y$.

DEFINITION 1.14 (Cartesian product). Let X and Y be sets. The Cartesian product of X and Y, denoted by $X \times Y$ is the set of all ordered pairs (x, y) such that $x \in X$ and $y \in Y$. In set-builder notation

$$X \times Y \stackrel{\text{def}}{=} \{(x, y) \mid x \in X, y \in Y\}.$$

1.3. Application of set theory thinking to sudoku.

THEOREM 1.15 (Phistomefel ring). *In a completed sudoku grid, the set of numbers (with multiplicities) in the blue region is exactly the same as in the red region.*



2. Functions

We all talk about functions. So much in math classes that we almost think they are synonyms. Believe it or not, they are not synonyms and a function has its own definition.

DEFINITION 2.1 (Function). A *function* f from a set X to a set Y is a specification of elements $f(x) \in Y$ for $x \in X$, such that

$$\forall x \in X, \exists! y \in Y, y = f(x).$$

The symbol $\exists!$ represents the phrase"there exists a *unique*". The unique element $f(x) \in Y$ is called the *value* of f at $x \in X$.

X is called the *domain* of f and Y is called the *codomain* of f.

We next discuss how to specify a function so that it satisfies the above definition.

(1) *Totality.* A value $f(x)$ should be specified for each $x \in X$ – this corresponds to the quantifier $\forall x$.
(2) *Existence.* For each X, the specified value $f(x)$ should exist, and should be an element in Y.
(3) *Uniqueness.* For each x the specified value $f(x)$ should refer to only one element in Y.

(2) and (3) correspond to the quantifier $\exists! y$.

EXAMPLE 2.2. The following are functions:

(1) $f : X \to X$, where $f(x) = x$ for any set X. This function is called the *identity function*.
(2) $f : \emptyset \to X$ is called the *empty function*. It has no values since there is no element in its domain.
(3) $f : \{1, 2, 3\} \to \{\text{red, yellow, green, blue}\}$ where $f(1) = red$, $f(2) = blue$, $f(3) = blue$.
(4) $g : \mathbb{R} \to \mathbb{R}$, where $g(x) = 2x$.

The following are NOT functions:

(1) $f : \{1, 2, 3\} \to \{\text{red, yellow, green, blue}\}$ where $f(1) = red$, $f(3) = blue$.
(2) $g : \mathbb{R} \to \mathbb{R}$, where
$$g(x) = \frac{1}{x}\,.$$

DEFINITION 2.3 (Graph of a function). Let $f : X \to Y$ be a functin. The *graph* of f is the subset $\mathrm{Gr}(f) \subseteq X \times Y$ defined by
$$\mathrm{Gr}(f) \stackrel{\text{def}}{=} \{(x, f(x)) \mid x \in X\} = \{(x, y) \in X \times Y \mid y = f(x)\}\,.$$

The graph of a function is perhaps the most important idea in modern mathematics as one can graphically draw functions on paper or turn graphs on papers into mathematical equations so that computations can be done. It is this idea that bridges geometry and calculus together. It is not an understatement to say science (and pseudo-science!) would not reach the its height today without this simple idea of Decartes.

EXAMPLE 2.4. Graph of the function $f : \mathbb{R} \to \mathbb{R}$, $f(x) = x/2$ is
$$\mathrm{Gr}(f) = \left\{ \left(x, \frac{x}{2}\right) \,\middle|\, x \in \mathbb{R} \right\}.$$

DEFINITION 2.5 (Composition of functions). Given two functions $f : X \to Y$ and $g : Y \to Z$. Their *composite* $g \circ f$ (read g composed with f) is the function $g \circ f : X \to Z$, defined by
$$(g \circ f)(x) = g(f(x)) \quad \text{for all } x \in X\,.$$

EXAMPLE 2.6. Let $f : [0,\infty) \to [0,\infty)$ be that $f(x) = x^3$ and $g : [0,\infty) \to [0,\infty)$ be that $g(x) = \frac{1}{1+x}$. Then, $f \circ g : [0,\infty) \to [0,1]$ is given by

$$(f \circ g)(x) = f(g(x)) = (g(x))^3 = \frac{1}{(1+x)^3}\,.$$

What happens if I keep the above formula and change $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \setminus \{-1\} \to [0,\infty)$?

EXAMPLE 2.7. We can write the function $M : \mathbb{Q} \to \mathbb{Q}$, $M(x) = \frac{(2x+5)^2}{6}$ as a composition as follows.

$$M = ((k \circ h) \circ g) \circ f\,,$$

where

- $f : \mathbb{Q} \to \mathbb{Q}$ is defined by $f(x) = 2x$,
- $g : \mathbb{Q} \to Q$ is defined by $g(x) = x + 5$,
- $h : \mathbb{Q} \to \mathbb{Q}$ is defined by $h(x) = x^2$,
- $k : \mathbb{Q} \to \mathbb{Q}$ is defined by $k(x) = \frac{x}{6}$.

2.1. Injections and surjections. The concepts of injections and surjections play very crucial roles in mathematics. Among other use, they are the tools for mathematicians to compare sizes of sets. It is with these concepts that one could talk about different sizes of infinities! We will just list definitions here, for a full reading, please read [New]. He does a fantastic job there discussing the concepts so there's no need to copy his text to here.

DEFINITION 2.8 (Injection). A function $f : X \to Y$ is *injective* (or *one-to-one*) if

$$\forall a \in X, \forall b \in X, f(a) = f(b) \Rightarrow a = b\,.$$

An injective function is said to be an *injection*.

EXAMPLE 2.9. Let $f : \mathbb{Z} \to \mathbb{Z}$ be a function that $f(n) = 2n + 1$. We will show that f is injective. Fix some $m, n \in \mathbb{Z}$ and suppose that $f(m) = f(n)$. By definition, we have

$$2m + 1 = 2n + 1 \iff m = n\,.$$

Therefore, f is injective.

EXAMPLE 2.10. Let $f : \mathbb{R} \to [0,\infty)$ be a function that $f(x) = x^2$. f is not injective since $f(-1) = f(1) = 1$, for example.

However, if we change the domain of f so that $f : [0,\infty) \to [0,\infty)$, it would be injective (why?).

PROPOSITION 2.11. *Let $f : X \to Y$ and $g : Y \to Z$ be functions. If f and g are injective, then $f \circ g$ is injective.*

PROOF. Let $a, b \in X$ and suppose that $(f \circ g)(a) = (f \circ g)(b)$. By definition of composition,

$$f(g(a)) = f(g(b)) .$$

Because f is injective, $g(a) = g(b)$; and because g is injective, $a = b$. Because a, b are arbitrary in X, $(f \circ g)$ is injective. □

DEFINITION 2.12 (Surjection). A function $f : X \to Y$ is *surjective* (or *onto*) if

$$\forall y \in Y, \exists x \in X, f(x) = y .$$

An surjective function is said to be an *surjection*.

EXAMPLE 2.13. Fix $n \in \mathbb{N}$ with $n > 0$ and define a function $r : \mathbb{Z} \to \{0, 1, \ldots, n-1\}$ by letting $r(a)$ be the remainder of a when divided by n. This function is surjective since for each $k \in \{0, 1, \ldots, n-1\}$, we have $r(k) = k$.

This function is not injective, however (why?).

DEFINITION 2.14 (Bijection). A function $f : X \to Y$ is bejective if it is both injective and surjective. A bijective function is called a bijection.

EXAMPLE 2.15. Let $f : \mathbb{Z} \to \mathbb{Z}$ be a function that $f(n) = n + 10$. Then f is a bijective.

When we change the domain of f so that $f : \mathbb{N} \to \mathbb{Z}$, it is no longer a bijective. Which fails – injectivity or surjectivity?

CHAPTER 3

Mathematical induction

Mathematical induction is a way to prove a sequence of statements by scaffolding. In a way, it can be compared to inductive logic, because in both cases we start by considering "small examples" and from those, we deduce that all of the examples have some property. However, there is an important distinction: mathematical induction is a part of *deductive reasoning*, because it provides a *formal proof* that yields *correct statements*, and does not just show that these statements are plausible. Here is the formal statement:

THEOREM 0.1 (The Induction Principle). *Suppose that we have a sequence of statements $P(n)$ labeled by the natural numbers $0, 1, 2, \ldots$ such that we know that*

(1) $P(0)$ is true, and
(2) $\big(P(0) \wedge P(1) \wedge \cdots \wedge P(n)\big) \Rightarrow P(n+1)$.

Then all the statements $P(0), P(1), P(2), \ldots$ are true.

1. All natural numbers are interesting

As a warm-up, let us prove that all natural numbers are interesting.

DEFINITION 1.1. A natural number n is called *interesting* if it has some special property that no other natural number has.

THEOREM 1.2. *All natural numbers are interesting.*

Let us first do a survey of the first few natural numbers:

- 0 is interesting because it is the only number that yields itself when you multiply it by another number.
- 1 is interesting because it is the only number that doesn't change the other number when multiplied.
- 2 is interesting because it is the first prime number, i.e. the number that has exactly two positive divisors (1 and itself).
- 3 is interesting because it is the first odd prime number.
- 4 is interesting because it is the first nontrivial square: $4 = 2^2$.
- 5...

We could say that 5 is a prime number, but it is not the first one of those. However, it is recognizable as a member of a certain sequence that is well-known to mathematicians and even in popular culture – Fibonacci

sequence, which appears in Indian mathematics in connection with Sanskrit prosody (study of poetic metres and verse in Sanskrit).

EXERCISE 11. *Watch this really nice introduction to Fibonacci numbers in nature by Vi Hart (total duration of all combined is 18 minutes):*

Part 1: https://www.youtube.com/watch?v=ahXIMUkSXX0

Part 2: https://www.youtube.com/watch?v=lOIP_Z_-0Hs

Now create an angle 137.5°*, draw approximately* 30 *petals, each* 137.5° *from the previous one, and when you are done, mark the spirals and count the number of them.*

Part 3: https://www.youtube.com/watch?v=14-NdQwKz9w

After watching the third video, can you explain why we almost always get a Fibonacci number of spirals in plants?

And so we can continue finding interesting properties for numbers:

- 5 is the first Fibonacci number for which we didn't find another property.
- 6 is the first *perfect number*, which by definition means that it is the sum of its positive divisors excluding itself: $6 = 1+2+3$. Incidentally, 6 is also a *triangular number*, i.e. a sum of consecutive numbers starting from 1. The next perfect number is $28 = 1+2+4+7+14$.
- 7 is the first *Mersenne prime number* for which we didn't find another property. Mersenne prime numbers are prime numbers that can be written as $2^k - 1$, e.g. $7 = 2^3 - 1$, and the next one is $31 = 2^5 - 1$.

In fact, even perfect numbers and Mersenne prime numbers are connected by this beautiful theorem:

THEOREM 1.3 (Euclid-Euler theorem). *If* $2^k - 1$ *is a Mersenne prime number, then* $2^{k-1} \cdot (2^k - 1)$ *is a perfect number, and all even perfect numbers are of this form.*

We omit the proof, but the interested reader will be able to read and understand the proof here: https://primes.utm.edu/notes/proofs/EvenPerfect.html. Before you read, note that $\sigma(n)$ is defined as the *sum of divisors function*, e.g. $\sigma(6) = 1+2+3+6 = 2 \cdot 6$, so a number n is perfect exactly when $\sigma(n) = 2n$.

It is still an open question whether there are infinitely many Mersenne prime numbers or perfect numbers. Additionally, it is not even known if there are odd perfect numbers! It has been proved however that there are no odd perfect numbers that have 1500 digits or less, or in math terms, a lower bound for the odd perfect numbers is 10^{1500}.

So far, we have seen three sequences of natural numbers: Fibonacci, Mersenne primes, perfect number. For even more, visit this page: https://oeis.org.

EXERCISE 12. *Imagine the sequence that starts with* $1, 1, 1, 1$. *What would be the next term? Go to OEIS and see what the encyclopedia shows.*

2. Proof by induction

We can talk about numbers, their properties and sequences of numbers all day long, but since there are infinitely many of them, we will never stop the case by case analysis. So let us do all cases at once by induction!

PROOF. Recall the Induction Principle: we first need a list of statements labeled by natural numbers. We have a natural candidate for this:

$$P(n) = \text{“}n \text{ is interesting”}.$$

Then we need to prove that $P(0)$ is true, but we have already observed a unique property of 0.

Now, let us assume that $P(0) \wedge P(1) \wedge \cdots \wedge P(n)$ is true, or in plain English, that all natural numbers up to n are interesting. We prove that $n+1$ is interesting by contradiction: if it wasn't, then it will have the special property that it is the smallest natural number that is not interesting. Isn't that interesting?! So $P(n+1)$ is true.

Finally, we can apply the Principle of Induction and get that all $P(n)$ are true, i.e. that all natural numbers are interesting. □

3. Structure of proofs by induction

By analyzing the proof above, we can divide proofs by induction into several steps:

(1) Identify the statements $P(n)$.
(2) Step 2 is also called *base of induction*: prove $P(0)$ or $P(1)$ (sometimes it doesn't make sense to talk about $P(0)$).
(3) Assume that all $P(k)$ with $k \leqslant n$ are true – this is called the *induction hypothesis*. Now perform the *step of induction*: prove that $P(n+1)$ is true.
(4) Finally, conclude by the Principle of Induction that all $P(n)$ are true.

4. Exercises

EXAMPLE 4.1 (Triangular numbers). Prove that the sum of the first n positive integers is $\frac{n(n+1)}{2}$.

(1) Here $P(n)$ means "$1 + 2 + \cdots + n = \frac{n(n+1)}{2}$", and we will avoid talking about $P(0)$. Although, strictly speaking, it makes sense, because $P(0)$ simply states that $0 = 0$.
(2) $P(1)$ states $1 = \frac{1 \cdot 2}{2}$, which can be seen is true.
(3) Now assume that $P(n)$ is true, i.e. $1 + \cdots + n = \frac{n(n+1)}{2}$. Therefore, by induction hypothesis,

$$1 + \cdots + n + (n+1) = \frac{n(n+1)}{2} + (n+1).$$

We can calculate the sum and get:

$$1 + \cdots + n + (n+1) = \frac{n(n+1) + 2(n+1)}{2} = \frac{(n+2)(n+1)}{2}.$$

But now the equality that we get is exactly the statement $P(n+1)$.
(4) So by induction, the sum $1 + \cdots + n$ is equal to $\frac{n(n+1)}{2}$ for all natural numbers n.

EXAMPLE 4.2. Prove that the sum of the first n odd positive integers is n^2.

Let's first recall that odd integers look like $2k + 1$ where $k \in \mathbb{N}$. Then, the sum of the first n odd positive integers should be

$$1 + 3 + 5 + \cdots + 2n + 1\,.$$

So, our statement $P(n)$ is

(1) $P(n)$ says

$$1 + 3 + 5 + \cdots + 2n + 1 = (n+1)^2\,.$$

(2) $P(0)$ says "$0 = 0^2$" and $P(1)$ says "$1 = 1^2$".
(3) Assume that for $k > 1$, $P(k)$ is true. We want to show that $P(k+1)$ is true. By the induction hypothesis,

$$\begin{aligned} &1 + 3 + \cdots + (2k+1) + (2(k+1)+1) \\ &= (k+1)^2 + 2(k+1) + 1 = (k+2)^2\,. \end{aligned}$$

The last equality is the quadratic formula $(a+b)^2 = a^2 + 2ab + b^2$.
(4) By the induction theorem, $P(k)$ is true for all $k \in \mathbb{N}$.

We also have a picture proof.

EXAMPLE 4.3. Prove that the natural number $n^3 - n$ is divisible by 3.

(1) $P(n)$ says

$$3 \mid (n^3 - n)\,.$$

(2) $P(0)$ says "$3 \mid 0$", which is true.

(3) Assume that for $k > 0$, $P(k)$ is true. This means that there exists a number l such that

$$k^3 - k = 3l\,.$$

We want to show that $P(k+1)$ is true. We write

$$\begin{aligned}(k+1)^3 - (k+1) &= k^3 + 3k^2 + 3k + 1 - k - 1\\ &= k^3 - k + 3k^2 + 3k = 3(l + k^2 + k)\,.\end{aligned}$$

By definition of divisibility, $P(k+1)$ is true.

(4) By the induction theorem, $P(k)$ is true for all $k \in \mathbb{N}$.

EXAMPLE 4.4. Show that for every $n \in \mathbb{N}$,

$$5 \mid F_{5n}\,.$$

EXAMPLE 4.5. Recall that the Fibonacci sequence is a sequence of number with the following pattern

$$F_0 = 0\,, F_1 = 1\,, \ldots, F_n = F_{n-1} + F_{n-2}\,.$$

The golden ratio is the following number

$$\varphi = \frac{1+\sqrt{5}}{2}\,.$$

The conjugate of the golden ratio is the following number

$$\phi = \frac{1-\sqrt{5}}{2}\,.$$

Show by induction that

$$F_n = \frac{\varphi^n - \phi^n}{\sqrt{5}}\,.$$

CHAPTER 4

Playing games with infinity

1. Games

Consider the following two-player game, which we will call Cantor's Game.

Player 1 begins by writing a sequence of X's and O's in the top row of the grid below. Player 2 then writes either an X or an O in the first box on the bottom. Player 1 then writes a sequence of X's and O's in the second row of the grid. Player 2 writes an X or O in the second bottom box. The players continue until all boxes are filled. Player 1 wins if the sequence on the bottom exactly matches any of the sequences Player 1 has written in the grid. Player 2 wins otherwise.

There is a grid below. Play the game a few times, then answer the following question: which player has a winning strategy, and why?

2. Infinity

Definition 2.1 (Countable sets). A set X is *countably infinite* if there exists a bijection $f : \mathbb{N} \to X$. The bijection f is called an *enumeration* of X. We say X is *countable* if it is finite or countably infinite. If there is no such bijection and the set is not finite, then X is said to be *uncountably infinite*.

EXAMPLE 2.2. The set of even natural numbers, denoted by E, is countably infinite. To see this, we consider the function $f : \mathbb{N} \to E$.

$$f(n) = 2n\,.$$

A more interesting example is the following.

EXAMPLE 2.3. The set of non-negative rational numbers is countably infinite. We see this by the snake argument.

THEOREM 2.4. $\mathbb{R}$ *is uncountably infinite.*

CHAPTER 5

Graph theory

1. Warm-up

In this chapter, we will learn how to represent interactions via graphs. Let us do so by talking about a question about a city whose existence is of major importance in the history of mathematics – Königsberg, now part of Russia and known as Kaliningrad.

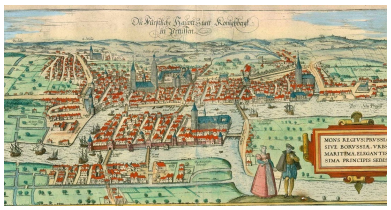


FIGURE 1. The map of Königsberg

fig:konigsberg

Let's highlight the number of bridges in the above map. Let me give you another picture (copied from the internet) of the bridges so that you don't have to suffer through the pixelated image.

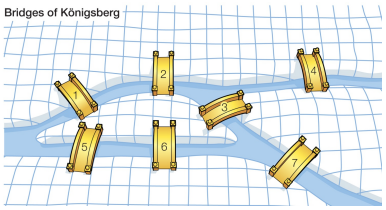


FIGURE 2. Bridges of Königsberg

fig:konigsberg-2

So you see there are 7 bridges. According to lore, people in this city would spend Sunday afternoons to walk around the city and to keep their thoughts occupied, they invented a game: find a path to walk around the city, crossing each of the bridges only once. It was an insanely difficult problem as perhaps none of the citizens was able to find a way to do it.

QUESTION. *Try for 10 minutes to an hour to find a path around the city that crosses each bridge only once.*

Leonhard Euler, perhaps the most prolific mathematician of all time, was asked this problem and thought it had nothing to do with math. But

the more he thought about it, the more he was intrigued. As a result, he invented two fields of mathematics, which are very fundamental to modern world: graph theory and topology.

Before going to talk about the problem, let me give an excerpt from an article on the Mathematical Association of America. [1]

> Why would Euler concern himself with a problem so unrelated to the field of mathematics? Why would such a great mathematician spend a great deal of time with a trivial problem like the Königsberg Bridge Problem? Euler was obviously a busy man, publishing more than 500 books and papers during his lifetime. In 1775 alone, he wrote an average of one mathematical paper per week, and during his lifetime he wrote on a variety of topics besides mathematics including mechanics, optics, astronomy, navigation, and hydrodynamics. It is not surprising that Euler felt this problem was trivial, stating in a 1736 letter to Carl Leonhard Gottlieb Ehler, mayor of Danzig, who asked him for a solution to the problem:
>
> ". . . Thus you see, most noble Sir, how this type of solution bears little relationship to mathematics, and I do not understand why you expect a mathematician to produce it, rather than anyone else, for the solution is based on reason alone, and its discovery does not depend on any mathematical principle. Because of this, I do not know why even questions which bear so little relationship to mathematics are solved more quickly by mathematicians than by others."
>
> Even though Euler found the problem trivial, he was still intrigued by it. In a letter written the same year to Giovanni Marinoni, an Italian mathematician and engineer, Euler said,
>
> "This question is so banal, but seemed to me worthy of attention in that [neither] geometry, nor algebra, nor even the art of counting was sufficient to solve it."
>
> Euler believed this problem was related to a topic that Gottfried Wilhelm Leibniz had once discussed and longed to work with, something Leibniz referred to as geometria situs, or geometry of position. This so-called geometry of

[1] https://www.maa.org/press/periodicals/convergence/leonard-eulers-solution-to-the-konigsberg-bridge-problem

position is what is now called graph theory, which Euler introduces and utilizes while solving this famous problem.

I hope you tried the problem out for a good ten minutes. It's worth trying to think about the problem as you will have a feel of what's going on and experience the historic walk in a city with changed name. It is not a conincidence that the citizens of Königsberg tried for a long time and did not succeed. It is impossible to find such a path!

Euler, as clever as he was, realized that usual ways of mathematical thinking will not solve this problem. He then, out of thin air, invented an entirely new way to think about it and proved that it is impossible to solve this problem, settling the game for the citizens (which means they will have to invent a new game to play...). Let's see how he did it!

The following solution is copied from the book [Bur13] as the argument there is a direct application to Euler's abstract thinking without any abstract theorem.

Solution to the Königsberg problem [Bur13]. The shapes of the landmasses don't matter either, so we can simply replace each landmass by a dot. In fact, the only relevant features are the connections between the landmasses. That is, we focus on which landmass is connected to which other one by a bridge. Each bridge creates a connection between two landmasses.

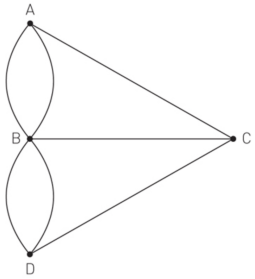


Figure 3. The map of Königsberg in graph form

fig:konigsberg-graph

Now we are well on our way to isolating the essential features of the Königsberg Bridge Puzzle. We have a collection of places (landmasses now denoted as dots) and some pairs of them are connected (bridges now denoted as arcs between pairs of dots)—and these places and connections are all that matter for the challenge at hand. We could name the places A, B, C, and D. And we could describe the connecting bridges by writing which pair of places each bridge connects. So the seven bridges of Königsberg can now be denoted as AB, AB, AC, BC, BD, BD, and CD.

After we've stripped away all the unnecessary diversions, we could restate the Königsberg Bridge Puzzle as follows: Suppose we have four dots named A, B, C, and D, and we have seven connections among them, namely, AB, AB,

AC, BC, BD, BD, and CD. Can we start at some dot and choose connections to take us from dot to dot in such a way that we use every connection exactly once and we return to the dot at which we started? Notice that the starting place does not matter; that is, if we could solve the puzzle starting at some dot we could also solve it starting from any other dot. Why is this true?

It might, at first, appear that we have not really made any progress toward solving the puzzle, but we have isolated the essential ingredients, and that is an enormous step forward. In fact, those essential ingredients—dots and connections—are the essential ingredients that comprise a modern mathematical area called graph theory. A graph is simply a set of vertices (that is, the dots) together with a collection of connections of pairs of the vertices (that is, the lines or arcs connecting pairs of dots). The connections are called edges. So the graph associated with the Königsberg Bridge Puzzle has vertices A, B, C, and D and edges AB, AB, AC, BC, BD, BD, and CD. Our simple picture that just shows the vertices and edges has all the information we require to tackle the Königsberg Bridge Puzzle. Now we can restate the Königsberg Bridge Puzzle as follows: Can we start at some vertex in the Königsberg graph, then choose edges to take us from vertex to vertex in such a way that we use every edge exactly once and we return to the vertex at which we started?

The Königsberg graph shows us that there are four landmasses (represented by the vertices) and it shows which landmasses are connected by bridges (represented by the edges). By the way, the order of the vertices that describe an edge does not matter; in other words, CD and DC both mean the same thing, namely, a "bridge" (or edge) connecting landmasses C and D. Since two bridges connect A to B and two bridges connect B to D, we simply list AB twice and BD twice, but we don't make any attempt to distinguish the two AB bridges from one another. For example, we don't care which AB is the bridge nearer D. We make no distinction between the two AB edges, because an edge is just a connection between its vertices. No other feature of an edge matters, such as how or where we draw it.

If we take a walk over the edges of the Königsberg graph, we can represent that walk by the ordered sequence (list) of edges that we traversed. For example, if we just go from A to B to D to C to B and back to A, we could represent that walk by listing the edges (AB)(BD)(DC)(CB)(BA) in the given order. Notice that it doesn't matter which of the two BD edges we choose, as each one accomplishes the same task of getting us from B to D.

Suppose we take a walk around the Königsberg graph and return back to where we started. If we write down the ordered sequence of edges involved, notice that each edge that we traverse must start at the point at which the previous edge ended. So in our example (AB)(BD)(DC)(CB)(BA), after we traveled along one of the edges from A to B, the next edge of course started

at B and ended at some other point, in this case D. Then we took an edge from D to C, then journeyed on the edge from C to B, and then went from B to A using the edge we didn't use the first time. Since we returned to the place where we started, the first letter and the last letter will coincide.

Let's forget the parentheses and just look at the letters that are written down in our list of the edge-sequence, namely, ABBDDCCBBA. Notice that every letter appears an even number of times, because every internal letter is the end of one edge and therefore, in turn, the beginning of the next edge, so every internal letter appears in pairs, while the first and last letters are the same (because we return to where we started), so that letter appears in pairs as well. The list of edges of such a circuit is the Noah's Ark of graph theory: Every letter appearing comes in pairs, in this example: BB, DD, CC, BB, and of course our starting and ending location AA.

This observation about letters appearing an even number of times lets us solve the Königsberg Graph Puzzle. Why? Well, let's look at the list of all seven edges of Königsberg. They are AB, AB, AC, BC, BD, BD, CD. If we walked over all the Königsberg edges each exactly once in any order at all, those seven pairs of letters would be the ones describing our walk. It might start (AB)(BC)(CD)... and so on. But if each edge in the Königsberg graph (that is, each given pair of letters such as AC) appeared exactly once in our walk, then the total number of A's would be 3 (an odd number), because we can see that there are exactly three A's in our seven edges—namely one A each in the edges AB, AB, and AC, and no other A's in any of the other edges. Similarly, the total number of B's would be 5, the number of C's would be 3, and the number of D's would be 3.

But we saw that if we were able to take a walk over the edges—traversing each edge exactly once—and return to where we started, then when we recorded the edges in the order that we walked over them, each letter would appear an even number of times. But this even number of appearances of each letter is impossible for the Königsberg graph because we just noticed that each letter appears an odd number of times on our list of bridges. Thus we conclude that it is impossible to start at one location, traverse each and every edge exactly once, and return to our starting point.

This observation settles the Königsberg Graph Puzzle and thus settles the Königsberg Bridge Puzzle, definitely proving that it is impossible to walk over each bridge just one time. Please think through the reasoning and "bridge" the ideas of the argument together for yourself until every step makes sense. □

2. Graph theory

Let us recap some essential ideas in the above argument and turn them into mathematical symbols.

DEFINITION 2.1 (Graph). A *graph* $G = (V, E)$ consists of a set V of *vertices* (or *nodes*) and a set E of *edges*.

An *edge* in a graph is a curve that has endpoints connecting a vertex to another vertex or a vertex to itself.

A *loop* in a graph is a curve that connects a vertex to itself.

There are various ways to represent vertices and edges. One way is just to draw out the graph as in Figure 3. Another way is to just write down the list of vertices in a set such as

$$\{A, B, C, \ldots\},$$

and a list of edges and loops connecting the vertices by writing the vertices next to each other such as

$$\{AB, AA, BC, BA, \ldots\}.$$

DEFINITION 2.2 (Path). A *path* is a succession of edges of the form $v_1v_2, v_2v_3, \ldots, v_nv_{n+1}$ so that all the edges and vertices only appear once when drawn out.

DEFINITION 2.3 (Connected graph). A graph $G = (V, E)$ is called *connected* if, for any pair of vertices $A, B \in V$, there is a path from A to B.

DEFINITION 2.4 (Eulerian path). An *Eulerian path* is a path through a graph which traverses each edge exactly once.

DEFINITION 2.5 (Eulerian cycle). An *Eulerian cycle* is an Eulerian path that ends in the same vertex where it started.

DEFINITION 2.6 (Degree). In a graph $G = (V, E)$, the *degree* of a vertex A, denoted by $\deg(A)$ is the number of edges that connect to A, and loops (i.e. edges AA) are counted twice.

THEOREM 2.7. *Let G be a finite connected graph. Then an Eulerian path exists if and only if there are at most two vertices with odd degree, and if there are, vertices of odd degrees must be the start or the end of the path. An Eulerian cycle exists if and only if all vertices have even degrees.*

3. Coloring maps and what graphs have to do with it

3.1. Motivation.

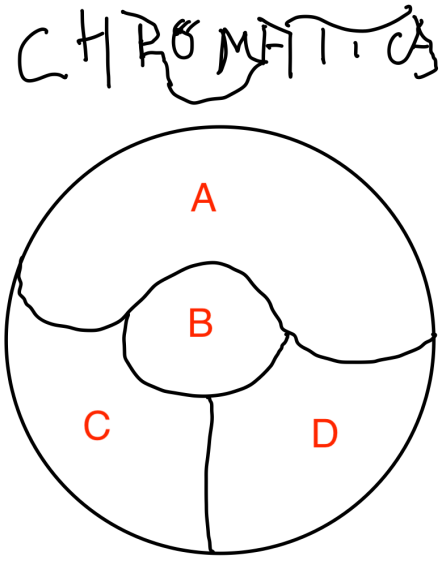
Acknowledgement: This lecture is based on material shared by Renee Bell and Patrick Shields.

QUESTION. *If you take a world atlas and pick a region, what is the minimal number of colors you would need to color every country in a way that each pair of countries that share a borderline has different colors?*

You can try, unsuccessfully, to color Central America with two colors. Try it and notice how you run into trouble around the point where three countries come together, for example Mexico, Belize and Guatemala:

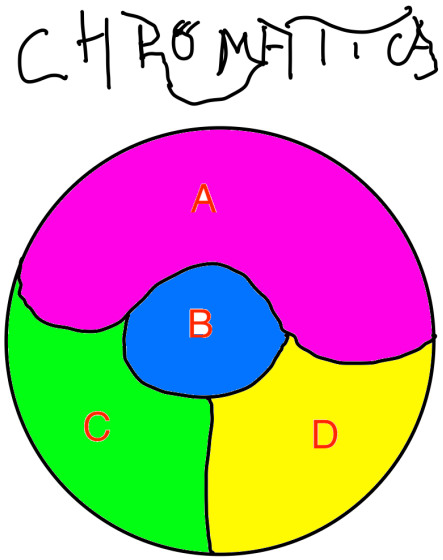
However, it *is* possible to color this map using three colors:

Alright, maybe we can do this for all maps using 3 crayons. Let's consider the following map of the planet of Chromatica, which is in another galaxy, and has countries A,B,C, and D.



QUESTION. *Can you find a coloring of Chromatica which uses only three colors?*

In this case, even three is not enough! If we color A pink, then neither C nor D can be pink, since they're touching A. But C and D cannot be the same color, since C is touching D. So, let's say C is green and D is yellow. Then B is left touching a pink state, a green state, and a yellow state. So it can't be any of those colors. We will need a fourth color, let's say blue.

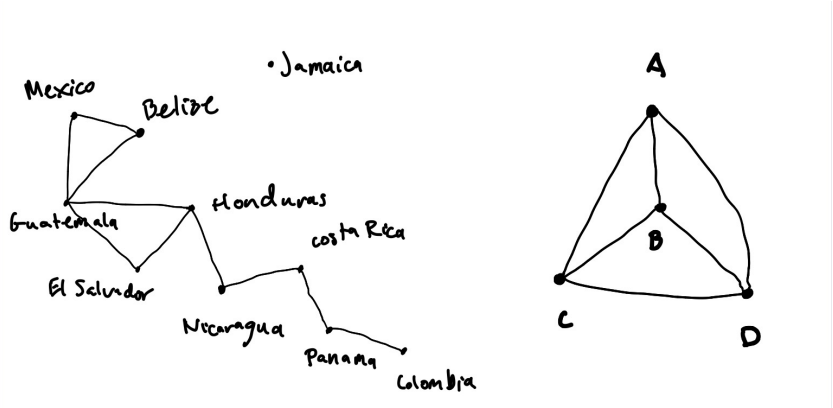


Could there be a map where even four colors isn't enough? It turns out the answer is no, but it took a long time to prove that decisively. The conclusive answer to this question was given by Kenneth Appel and Wolfgang Haken in 1976, in the form of the following theorem, called the Four Color Theorem.

THEOREM 3.1 (Four Color Theorem for Maps). *Every map can be colored such that no two regions which share a boundary are the same color, using four colors.*

3.2. Graphic content. Now let's try to abstract and simplify this problem, especially for people who are not great at drawing countries. The only information we needed was which countries share a boundary! So let's record that information using a graph.

We create a graph associated to this map, with a vertex for each country and an edge between two vertices if the associated countries share a boundary. Here are the graphs for Central America and Chromatica:

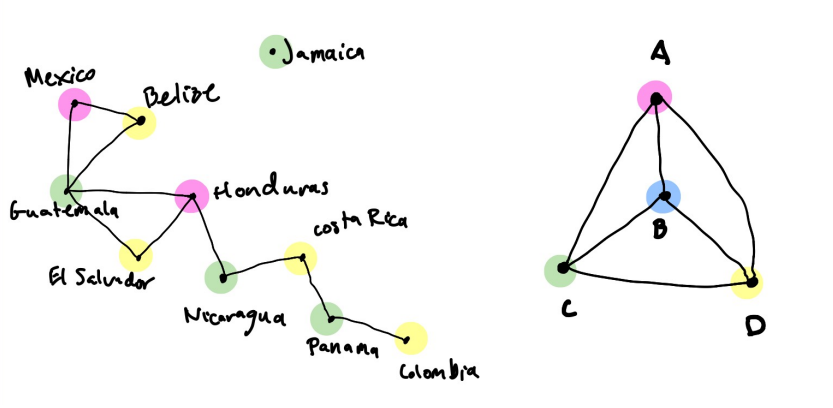


Now the question reduces to coloring the *vertices* of these graphs so that no two vertices which share an edge are the same color. Let's formalize this in a definition:

DEFINITION 3.2. A **coloring** of a graph is a labeling of the graph's vertices with colors such that no two vertices sharing the same edge have the same color.

If you don't have a lot of colors at your immediate disposal, you can label colors by numbers, symbols, letters of Japanese alphabets, whatever suits your tastes better.

Here are the colorings of the graphs associated to Central America and Chromatica corresponding to the map colorings we just did:

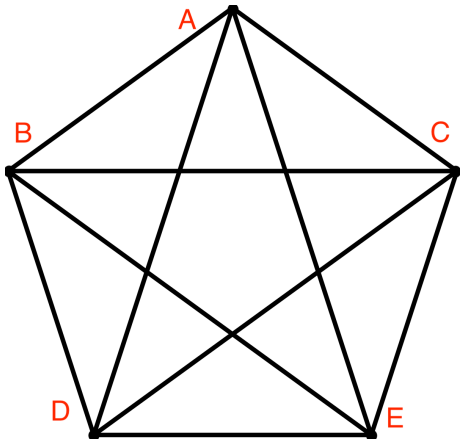


We're also trying to figure out the number of colors we need to do this. This motivates the following definition:

DEFINITION 3.3. A coloring using k colors is called a k-**coloring**. The smallest number of colors needed to color a graph G is called its **chromatic number**, and is denoted $\chi(G)$.

In the examples we just saw, $\chi(\text{Chromatica}) = 4$, since we can color it using four colors but NOT using fewer. We also calculated that $\chi(\text{Central America}) = 3$.

The definitions we just gave could be for any kind of graph. For example, the pentagram, which we will call P:



Let's calculate its chromatic number. Of course, we could 5-color it by just making every single vertex a different color, so we know $\chi(P) \leqslant 5$. So, can we color it using four colors? Well, let's say we color A "color 1". Since B shares an edge with A, it must be a new color, "color 2". Since C shares an edge with both A and B, it must be a new color, "color 3". D must be a fourth color, "color 4". And finally, E touches every other vertex, so it must be a fifth color. P cannot be colored using 4 colors, so $\chi(P) = 5$.

But wait, didn't we just say that four colors is enough? Well, that was only for *maps* which are "2-dimensional" in some sense, not *all* graphs. The type of graph that we made is called **planar**.

DEFINITION 3.4. A graph is **planar** if it can be drawn on the plane (that is, in two dimensions) in such a way that its edges intersect only at their endpoints.

We can now give a different version of the Four Color Theorem:

THEOREM 3.5 (Four Color Theorem for Graphs). *Every planar graph can be 4-colored.*

Now we see that the problem with the pentagram is that it's not planar; it can't be "untangled" into a nice 2D graph that lies flat in the plane.

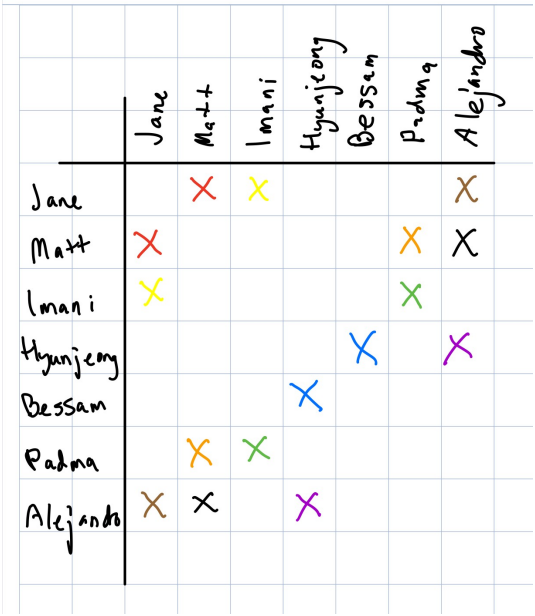
3.3. Conflict Resolution. I realize that not all of you are planning to become cartographers, but most of us have to deal with conflict sometimes, so let's look at a more likely application. Suppose you are planning a small dinner, now that the worst of the lockdown is over, and want to determine a seating chart, seating people at different tables. You are inviting your friends Jane, Matt, Imani, Hyunjeong, Bessam, Padma, and Alejandro. You love

them dearly, but they are very... passionate people, and you know that if you seat two of them who disagree at the same table, the dinner is going to go very badly. Here are their conflicts:

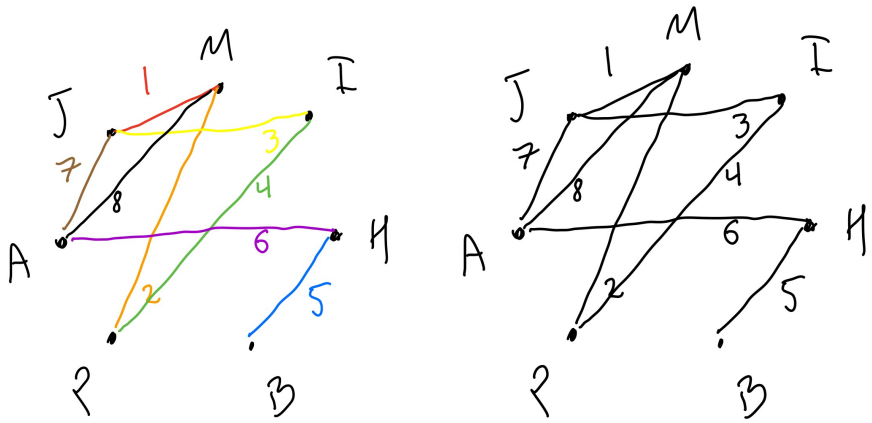
Matt and Imani are Patriots fans. Jane and Padma are Eagles fans. Do not sit them next to each other! Also, Hyunjeong likes Cardi B but Alejandro and Bessam are hardcore Nicki Minaj stans (barbz), so she won't sit with either of them. Alejandro dated both Matt and Jane so he doesn't want to sit with them, because that would be awkward. Other than that, everyone is cool. Let's list the conflicts:

(1) Matt can't sit with Jane
(2) Matt can't sit with Padma
(3) Imani can't sit with Jane
(4) Imani can't sit with Padma
(5) Hyunjeong can't sit with Bessam
(6) Hyunjeong can't sit with Alejandro
(7) Alejandro can't sit with Jane
(8) Alejandro can't sit with Matt

We can record this information in the following chart:



This is a very neat, excel-spreadsheet style way to record the information. But we can also record it in a graph. The only thing that matters for us is whether or not any two people have conflict. So we can create a graph where the vertices represent people, and there is an edge between two people if there is conflict between them. We show this with the colors corresponding to the conflicts on the left, and in black and white on the right. We will not really need to color the edges, but we will have to color the vertices.

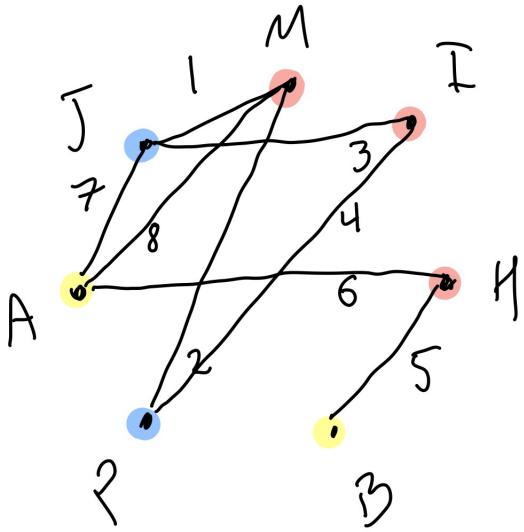


QUESTION. *How can you figure out how many tables you will need for your dinner using this graph?*

We just learned a new thing we can do with graphs, which is color them. You can think of a coloring using k-colors as a way of putting the vertices into k groups, where no group has any two vertices which share an edge.

Conversely, if we put the vertices into k groups, where no group has any two vertices which share an edge, we can use that to make a k-coloring for the graph! We just pick a color for each group, and then color the vertices in each group accordingly.

Back to the dinner; we can find a 3-coloring of the graph using red, yellow, and blue as follows:



This allows us to seat everyone using 3 tables! You can refer to tables as the red table, the yellow table, and the blue table. The graph above corresponds to the following seating:

Tables

Color	Vertices of that color
Red	Matt, Imani, Hyunjeong
Yellow	Alejandro, Bessam
Blue	Jane, Padma

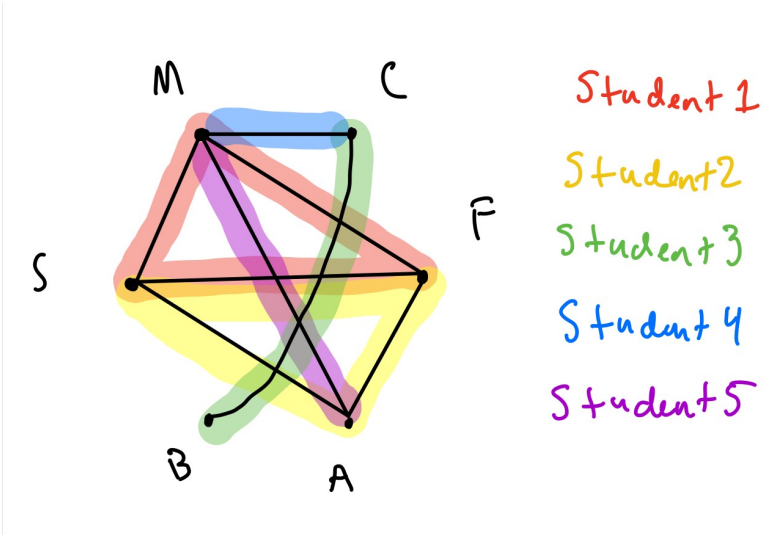
You can check that this graph can't be 2-colored. This means there's no way of seating everyone at 2 tables without there being some conflict. Better bring enough tables!

3.4. Scheduling. Let's look at one *final* application: scheduling final exams. Suppose we're scheduling exams for Math, Computer science, French, Sociology, Biology, and Anthropology. We'll make a graph where the vertices are subjects, and there is an edge between two vertices if the corresponding subjects have a conflict.

Suppose that Student 1 is in Math, French, and Sociology. This means that Math and French are in conflict, French and Sociology are in conflict, and Math and Sociology are in conflict (since this student can't attend both exams at the same time). So there is an edge between Math and French, an edge between French and Sociology, and an edge between Sociology and Math.

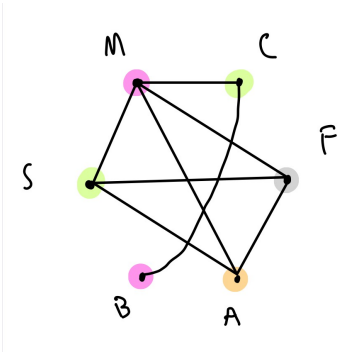
Suppose also that Student 2 is in Anthropology, French, and Sociology, Student 3 is in Biology and Computer science, Student 4 is in Computer science and Mathematics, and Student 5 is in Math and Anthropology.

This gives rise to the following graph, where the conflict edges are colored according to which Student is the cause of them:



QUESTION. *How many timeslots do we need for these final exams? How can we pick which exams go in which timeslot?*

Again, colorings of this graph correspond to conflict-free exam schedules. You can check that it's impossible to 3-color the graph (so we need more than 3 timeslots), but we can 4-color the graph as follows:



Timeslots

Color	Vertices of that color
Hot pink	Math, Biology
Lime green	Computer science, Sociology
Grey	French
Orange	Anthropology

Hence we can have Math and Biology in the first timeslot, Computer science and Sociology in the second timeslot, French in the third timeslot, and Anthropology in the fourth timeslot.

4. Chromatic polynomial of a graph

The previous section was motivated by a more applied thinking about graphs. In this section, we would like to discuss what mathematicians care about.

In general, when mathematicians define a certain kind of object – geometric shape, function, graph, set – we then ask a question about how to classify these objects. For example, for sets, we can count the number of elements (which may be infinite), and if two sets have the same number of elements, then there is a bijection between them, in other words, we can identify their elements. After suitable identifications, we can start asking what are significant distinctions between the objects. For example, to each we can assign a number, like the chromatic number for a graph. These numbers are called *invariants*.

For graphs, one invariant that is closely related to the chromatic number is the *chromatic function*.

DEFINITION 4.1. Let G be a graph. Then we define the *chromatic function* $P_G : \mathbb{N} \to \mathbb{N}$ by the following rule: $P_G(n) =$ the number of colorings of the vertices of G into n colors.

Let us compute some examples.

EXAMPLE 4.2. Let first G be the graph with one vertex and no edges, so $V = \{A\}$ and $E = \varnothing$. It looks like a dot. Given n colors, we can color this dot in exactly n ways, one for each choice of color, so $P_G(n) = n$.

EXAMPLE 4.3. Let us go a few steps further and set G to be the graph with 3 vertices and no edges, so $V = \{A, B, C\}$ and $E = \varnothing$. It looks like three dots. Given n colors, we can color the vertex A in n ways. Then we are free to choose any color for B, since it is not connected to any other vertices, so for each of the first n variants, we have n variants in turn, which brings us to n^2 variants. Finally, the third vertex C is not connected to any others, so we can color it in n ways independently of the previous choices. In total, it brings us to $P_G(n) = n^3$.

EXERCISE 13. *Generalize the previous two examples as follows: if G is a graph with v vertices and no edges, prove by induction that $P_G(n) = n^v$.*

Let us now consider a more involved example – graphs that have not just vertices, but also edges. You may have already guessed that graphs with loops (edges that begin and end in the same vertex) cannot be colored at all. Indeed, if you have a loop AA, definition of coloring tells us that the color of the vertex on the one side of AA cannot be the same as color of the vertex on the other side... but oh no, it is the same vertex! So we can ignore graphs with loops because for them the problem of counting is trivial. Furthermore, if you have multiple edges that connect the two vertices, say A and B, then each of the edges gives the same condition – that A and B should be colored differently. So we can also ignore all but one edge between A and B. To sum up, from now on we can only consider graphs that don't have loops, and each pair of vertices shares at most one edge.

DEFINITION 4.4. A *simple graph* is one in which there is at most one edge joining a given pair of vertices and there are no loops (i.e. edges joining a given vertex with itself).

EXAMPLE 4.5. Let us consider a graph with v vertices that looks like a string, we will call this graph A_v.

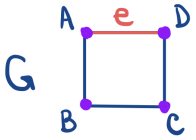
Let us start with $v = 5$ vertices and count the number of colorings of this graph into n colors, that is we will calculate the value $P_{A_5}(n)$. We'll first color the leftmost vertex – for this, we have n choices of colors. But when we pick a color for the second vertex, we notice that one color – the color of the first vertex – is no longer available. So we now have $n - 1$ choices. Similarly, the third vertex is connected to one vertex which is already colored, yielding $n - 1$ choices again, and same for the fourth and fifth vertex. So altogether, we get the following result:

$$P_{A_5}(n) = n(n-1)^4 .$$

EXERCISE 14. *Guess a formula for $P_{A_v}(n)$ and prove it by induction.*

oloring_square

EXAMPLE 4.6. Now let us consider the graph that can be drawn as a square, let's call it G:



Its set of vertices is

$$V = \{A, B, C, D\},$$

and the set of edges is

$$E = \{AB, BC, CD, AD\}.$$

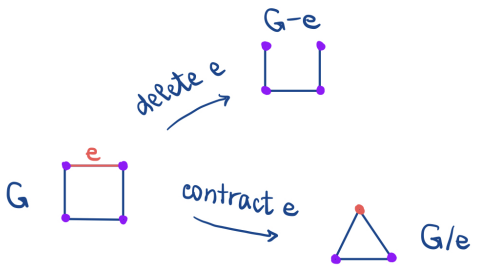
We denote by e the edge AD.

Now we can see how we can color this graph using, let's say, 4 colors. We will be coloring the vertices one after another and see which cases we encounter. There are 4 variants for the vertex A. For the next vertex, B, there are now three variants, because it is connected to one vertex that is already colored. So far, we have $4 \cdot 3 = 12$ variants to color two first vertices.

Similarly, after A and B are colored, there are three variants to color C. However, we run into trouble now with D, because it is connected to both A and C, and we don't know if they are the same color or different. Out of the three variants to color C, there is one case when A and C are of the same color – in this case, we can choose out of three colors for D; and there are two cases when A and C have different colors – then in each we can choose out of 2 colors to color D. In total, we have the following number of colorings:

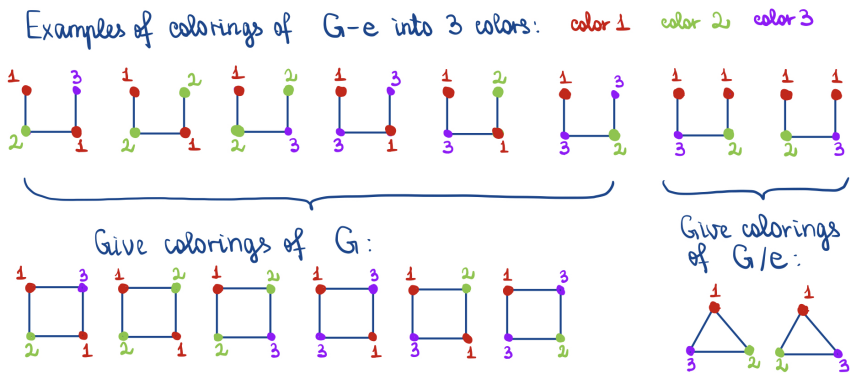
$$P_G(4) = 4 \cdot 3 \cdot (1 \cdot 3 + 2 \cdot 2) = 84.$$

There is another approach to the same problem where you may choose to ignore the edge $e = AD$. If you delete this edge from G, you will get the graph A_4, and there we have less restrictions for coloring. In general, when you *delete* e from G, the result is denoted by $G - e$.



In fact, for each coloring of $G - e$, there are two cases:

(1) Vertices of e (in our case A and D) have different colors. In this case, you can restore the edge e and get a coloring of the original graph G.
(2) Vertices of e have the same color. Then we can no longer draw an edge between them, because we cannot have an edge that connects vertices of the same color; however, we can glue these two vertices instead. This is called *contracting* the edge e of G.



The following is a cute theorem that generalizes the above observation. To prove it, you may use the same case-by-case analysis as we did for the square, as you may have already noticed that it didn't use the structure of G.

THEOREM 4.7 (Deletion-contraction formula). *Let G be a graph, and let $G-e$ and G/e, respectively, be the graphs obtained from G by deleting and contracting the edge e. Then*

$$P_G(n) = P_{G-e}(n) - P_{G/e}(n)\,.$$

EXAMPLE 4.8. Returning to the example with the square, denoted again by G, we can now compute $P_G(n)$ with the help of the new formula:

$$P_G(n) = P_{A_4}(n) - P_{G/e}(n) = n(n-1)^3 - P_{G/e}(n).$$

Looking at the picture of G/e, one can compute $P_{G/e}(n)$: when we first color the vertex A, we have n choices; for B, since it is connected to A, we have $n-1$ choices; and vertex C is connected to B as well as to A (because we contracted AD, so A and D are now one vertex) – so we have $n-2$ choices for C, because A and B are two different colors. In total, we have

$$P_{G/e}(n) = n(n-1)(n-2)$$

colorings of G/e.

Finally, we plug in this result into $P_G(n)$:

$$P_G(n) = n(n-1)^3 - n(n-1)(n-2) = n(n-1)\left((n-1)^2 - (n-2)\right) =$$
$$= n(n-1)(n^2-3n+3).$$

We can check that this result recovers our calculation in Example 4.6 for $n = 4$:

$$P_G(4) = 4 \cdot 3 \cdot (16 - 12 + 3) = 84.$$

5. Euler characteristics

Here is a little bit of magic. Close your eyes and start doodling without lifting your pencils. Don't try too hard because we are going to do some math with your masterpiece. Here's an example of what a doodling looks like.

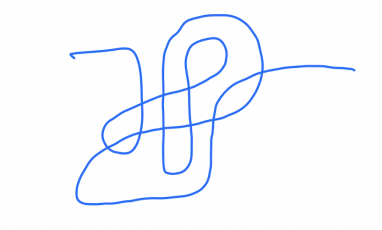


FIGURE 4. Doodling

fig:doodle1

Now, in order to do some math, please use a red pen to color all the intersections as well as the starting and ending points in your masterpiece like the Figure 5.

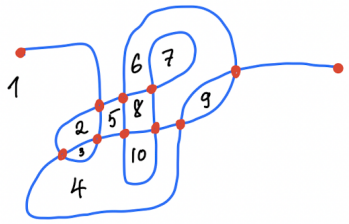


FIGURE 5. Color the intersections

fig:doodle2

Now, we have turned our doodling into a graph! Let us count the number of vertices, number of edges and the number of regions this graph has divided our paper into (each region is the blank spaced enclosed by some vertices and edges, just like countries and borders; we also count the area not enclosed by anything one big region of its own).

Here are the counts for this graph.

- Regions: 10
- Vertices: 11
- Edges: 19

Whatever you do, you should have the following relationship:

eq:euler

$$F + V - E = 2\,, \tag{5.1}$$

where F stands for face (the regions), V stands for vertices, E stands for edges.

DEFINITION 5.1. The number $\mathcal{C}(G) = F + V - E$ is called the Euler characteristics of a graph G.

DEFINITION 5.2 (Spanning Tree). A spanning tree of a graph G is a connected subgraph of G so that every pair of vertices is connected by exactly one path.

t:spanning-tree

PROPOSITION 5.3. *Every connected graph G contains a spanning tree.*

We will not prove this proposition but use it to prove the following theorem.

THEOREM 5.4. *The Euler characteristics of a connected planar graph is always 2.*

PROOF. Let G be a connected graph. Let T be a spanning tree of G given by Theorem 5.3. For a tree T, the number of edges is always one less than the number of vertices, i.e., $V = E + 1$. Because there are no cycles in a tree, the number of faces is $F = 1$. So $\mathcal{C}(T) = 2$. To obtain G from T, we add more edges. Each time we add an edge, we add a new face. Thus, the Euler characteristics doesn't change after we add the edges to make T become G. As a consequence, $\mathcal{C}(G) = 2$. □

We will have to define what a region means for non-planar graphs (which is out of the scope of this class) but once we define that, we can prove that the Euler characteristics of non-planar graphs are not 2.

Let us use the Euler characteristics to show a very surprising fact about Platonic solids.

DEFINITION 5.5 (Regular polygon). A *regular polygon* is a polygon that has all angles equal and all sides equal.

DEFINITION 5.6 (Platnoic solids). A *Platonic solid* is a 3-dimensional shape such that

- Each face is the same regular polygon.
- The same number of polygons meet at each vertex.

THEOREM 5.7. *There are only 5 Platonic solids.*

PROOF. Suppose we have a Platonic solid. Let F be the number of faces, E be the number of edges, V be the number of vertices of the solid.

Let us observe a few things. First, on each face, the number of edges and vertices is the same. Let N be the number of edge and vertices on *each* face. Let D be the degree of each vertex.

Second, since each edge belongs to two different faces and each edge also joins two vertices,

$$NF = 2E = DV\,.$$

Note that we can redraw the faces, vertices and edges of a Platonic solid into a plane graph (you should try a few!). By the Euler's characteristics theorem,

$$2 = V + F - E = V + \frac{DV}{N} - \frac{DV}{2}\,.$$

Rearrange terms, we have

$$V(2N + 2D - ND) = 4N\,.$$

It must be true then that $2N + 2D - ND > 0$ as $V, N > 0$. This means $(N-2)(D-2) < 4$ and there are only 5 possibilities for $N, D > 0$,

$$(N, D) = \begin{cases} (3,3) & \text{tetrahedron}\,, \\ (3,4) & \text{cube}\,, \\ (4,3) & \text{octahedron}\,, \\ (3,5) & \text{dodecahedron}\,, \\ (5,3) & \text{icosahedron}\,. \end{cases}$$

This is our claim! □

CHAPTER 6

Discrete probability

We will follow the books [New; DS12] in this chapter. Some examples are just copied directly from these books., just as we distinguished the three coins in Example 1.6.3.

We now get to the final chapter of the course. Probability is perhaps one of the most important concepts in human history. Since the very ancient time, humans entertain themselves with probability by the game of chance: gambling... This game is almost universal: as far as I can tell, every culture has a version of gambling that ruins people's lives. It is very interesting how this game is a universal phenomenon.

At its heart, probability is a little bit paradoxical: it tells us how to predict randomness! How can one predict anything if everything is random?! The main point is that probability does not tell you what is going to happen to one particular sequence of events. However, it will tell you how likely a sequence of events will happen with some "confidence". When one uses the language of probability, one needs to be careful not to treat it as an absolute way to predict something but one needs to allow the possibility that the sequence of events under consideration may not happen.

1. The basics

In daily language, we talk about the probability of some event to happen to mean the following

$$\text{Probability of an event happening } = \frac{\text{Number of times the event appear}}{\text{Total number of outcomes}}.$$

Most of the time, we unconsciously think that all the events are equally likely such as the probability for a coin to be head or tail or the probability for a die to be $1, 2, \ldots, 6$. This is because we *assume* that there is no reason for the head to appear more often than the tail or 1 to appear more often than 6.

Example 1.1. In daily life, we assume

- The probability for a fair coin to be head is $1/2$.
- The probability for a fair die to be 4 is $1/6$.

However, this assumptioin is not entirely well thought out as there are ways to make a coin land on its head more often than its tail. The coin manufacturer could play with the physics to do it!

DEFINITION 1.2. Let A be a set. The *power set* of A, denoted by $\mathcal{P}(A)$, is the set of all subsets of the set A.

EXAMPLE 1.3. $\mathcal{P}(\{1,2\}) = \{\emptyset, \{1\}, \{2\}, \{1,2\}\}$.

DEFINITION 1.4. A *finite discrete probability space* is a pair $(\Omega, \boldsymbol{P})$ where Ω is a countable set and $\boldsymbol{P} : \mathcal{P}(\Omega) \to [0,1]$ is a function such that

(1) $\boldsymbol{P}(\Omega) = 1$,
(2) $\boldsymbol{P}(\bigcup_{i=1}^{\infty} A_i) = \boldsymbol{P}(A_1) + \boldsymbol{P}(A_2) + \dots$ when $A_i \cap A_j = \emptyset$ and $i \neq j$.

The set Ω is called the *sample space*, an element $\omega \in \Omega$ is called an *outcome*, a subset $A \subseteq \Omega$ is called an *event*, Given A, $\boldsymbol{P}(A)$ is called the *probability of* A.

One can think of the sample space as the set of all possible outcomes. An observation based on the above definition is that a sample space is also an event.

EXAMPLE 1.5. Let us model a coin toss.

The outcomes of the toss are head or tail. So, we can take $\Omega = \{H, T\}$.

The events correspond to the subsets of $\{H, T\}$:

(1) $\boldsymbol{P}(\{H\}) = 1/2 = \boldsymbol{P}(\{T\})$,
(2) $\boldsymbol{P}(\emptyset) = 0$,
(3) $\boldsymbol{P}(\{H, T\}) = 1$.

EXAMPLE 1.6. A slightly more interesting thing to talk about is modeling two coin tosses. In this case, $\Omega = \{HH, HT, TT, TH\}$. The first letter represents the outcome of the first toss and the second letter the outcome of the second toss. Here's a few probabilities:

(1) $\boldsymbol{P}(\{HT\}) = \boldsymbol{P}(\{HH\}) = \boldsymbol{P}(\{TT\}) = \boldsymbol{P}(\{TT\}) = 1/4$
(2) $\boldsymbol{P}(\{H\}) = \boldsymbol{P}(\{HT, HH\}) = 1/2$.

EXERCISE 15. *Assume you have a fair coin. Write down the probability space for three coin tosses.*

EXAMPLE 1.7. Assume you have a fair die. Write down the probability space for one die toss. What is the probability for the die to be odd?

A good way to visualize this is to draw a pie, color the parts that you're looking at and take the area of the colored part and divide it by the total pie.

PROPOSITION 1.8 (Some properties of probability). *The followings are true.*

(1) $\boldsymbol{P}(A^C) = 1 - \boldsymbol{P}(A)$,
(2) $\boldsymbol{P}(\emptyset) = 0$,
(3) If $A \subseteq B$, *then* $\boldsymbol{P}(A) \leqslant \boldsymbol{P}(B)$,

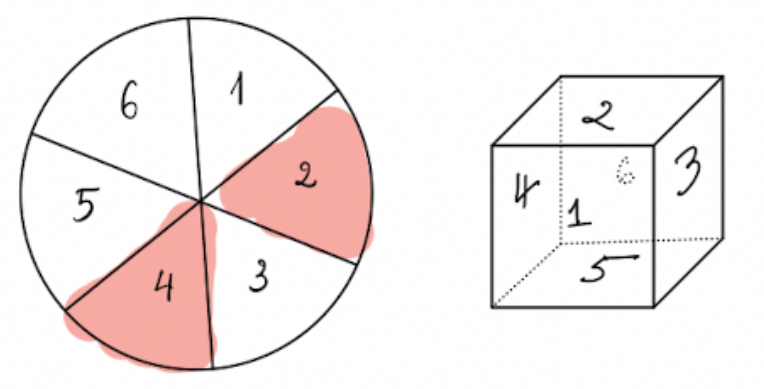


FIGURE 1. The ratio of the color part divided by the total number of equally divided parts gives the probability of the event that either the die turns out to be either 2 or 4. fig:Pie

(4) $\boldsymbol{P}(B \cap A^C) = \boldsymbol{P}(B) - \boldsymbol{P}(A \cap B)$,

(5) Inclusion-Exclusion principle:

$$\boldsymbol{P}(A \cup B) = \boldsymbol{P}(A) + \boldsymbol{P}(B) - \boldsymbol{P}(A \cap B)\,.$$

PROOF. Let's go through the list.

(1) We know that $\Omega = A \cup A^C$ and that $\boldsymbol{P}(\Omega) = 1$. Thus,

$$1 = \boldsymbol{P}(\Omega) = \boldsymbol{P}(A) + \boldsymbol{P}(A^C)\,.$$

Rearrange this, we have

$$\boldsymbol{P}(A^C) = 1 - \boldsymbol{P}(A)\,.$$

(2) $\emptyset = \Omega^C$. Applying (1), we get

$$\boldsymbol{P}(\emptyset) = 1 - \boldsymbol{P}(\Omega) = 1 - 1 = 0\,.$$

(3) $B = (B \setminus A) \cup A$. Thus,

$$\boldsymbol{P}(B) = \boldsymbol{P}(B \setminus A) + \boldsymbol{P}(A) \geqslant \boldsymbol{P}(A)\,.$$

(4) This follows by the following equation.

$$\boldsymbol{P}(B) = \boldsymbol{P}(B \cap A^C) + \boldsymbol{P}(B \cap A)\,.$$

Rearranging this, we get

$$\boldsymbol{P}(B \cap A^C) = \boldsymbol{P}(B) - \boldsymbol{P}(A \cap B)\,.$$

(5) We have that

$$A \cup B = A \cup (B \cap A^C)\,.$$

Therefore, using (4),

$$\boldsymbol{P}(A \cup B) = \boldsymbol{P}(A) + \boldsymbol{P}(B \cap A^C) = \boldsymbol{P}(A) + \boldsymbol{P}(B) - \boldsymbol{P}(A \cup B)\,.$$

□

EXAMPLE 1.9. A patient arrives at a doctor's office with a sore throat and low- grade fever. After an exam, the doctor decides that the patient has either a bacterial infection or a viral infection or both. The doctor decides that there is a probability of 0.7 that the patient has a bacterial infection and a probability of 0.4 that the person has a viral infection. What is the probability that the patient has both infections?

EXAMPLE 1.10. Inherited traits in humans are determined by material in specific locations on chromosomes. Each normal human receives 23 chromosomes from each parent, and these chromosomes are naturally paired, with one chromosome in each pair coming from each parent. For the purposes of this text, it is safe to think of a gene as a portion of each chromosome in a pair. The genes, either one at a time or in combination, determine the inherited traits, such as blood type and hair color. The material in the two locations that make up a gene on the pair of chromosomes comes in forms called alleles. Each distinct combination of alleles (one on each chromosome) is called a genotype.

Consider a gene with only two different alleles A and a. Suppose that both parents have genotype Aa, that is, each parent has allele A on one chromosome and allele a on the other. (We do not distinguish the same alleles in a different order as a different genotype. For example, aA would be the same genotype as Aa. But it can be convenient to distinguish the two chromosomes during intermediate steps in probability calculations.) What are the possible genotypes of an offspring of these two parents? If all possible results of the parents contributing pairs of alleles are equally likely, what are the probabilities of the different genotypes?

2. Probability from counting

Let us learn how to come up with probability law for some"simple" situations, where the number of outcomes is finite. We call this finite probability (as opposed to countably inifinite probability). We will assume that every outcome is equally likely to another and, hence, in order to determine the probability of an event, one needs to be able to count accurately. The general principle is

eq:gen-principle

(2.1)
$$\text{Probability of an event happening } = \frac{\text{Number of ways it can appear}}{\text{Total number of outcomes}}.$$

EXAMPLE 2.1. Let A_3 be the graph in homework 8. Suppose we have 4 colors: red, blue, yellow, green. Assume that each coloring is equally likely. Write down the probability space for the chance for a particular way of coloring is chosen.

EXAMPLE 2.2. Let K_3 be the complete graph with 3 vertices. Suppose we have 4 colors: red, blue, yellow, green. Assume that each coloring is equally likely. Write down the probability space for the chance for a particular way of coloring is chosen.

Thus, it is important to learn how to count.

2.1. Counting. In order to apply the general principle 2.1, we need to know the total number of outcomes. One could list them all out. However, when the number of outcomes gets large, this is very inefficient and often leads to errors. It is nice to have a way (formula) to find the total number of outcomes efficiently. This section will discuss several ways, which we can use to do just that.

EXAMPLE 2.3. Consider an experiment that has the following two characteristics:

(1) The experiment is performed in two parts.
(2) The first part has m possible outcomes and regardless of which outcome in part 1 we end up with, the second part always has n possible outcomes.

How many possible outcomes can one have? Find a way to write down the sample space in mathematical langauge.

One can show the following general theorem using induction.

THEOREM 2.4 (Multiplication rule). *Suppose that an experiment has k parts ($k \geqslant 2$), that the i^{th} part of the experiment can have n_i possible outcomes ($i = 1, ..., k$), and that all of the outcomes in each part can occur regardless of which specific outcomes have occurred in the other parts. Then the sample space Ω will have $n_1 n_2 \dots n_k$ outcomes.*

In the above theorem, it is assumed that the number of outcomes in each part of the experiment does not depend on that of the other parts. This is not the case in general as one might imagine that when one does a survey, one does not ask the same person to fill out the survey multiple times. This process is called *sampling without replacement*.

ex:n-permute-3

EXAMPLE 2.5. Consider an experiment in which a card is selected and removed from a deck of n different cards, a second card is then selected and removed from the remaining $n - 1$ cards, and finally a third card is selected from the remaining $n - 2$ cards. Each outcome consists of the three cards in the order selected. A process of this kind is called sampling without replacement, since a card that is drawn is not replaced in the deck before the next card is selected. In this experiment, any one of the n cards could be selected first. Once this card has been removed, any one of the other

$n-1$ cards could be selected second. Therefore, there are $n(n-1)$ possible outcomes for the first two selections. Finally, for every given outcome of the first two selections, there are $n-2$ other cards that could possibly be selected third. Therefore, the total number of possible outcomes for all three selections is $n(n-1)(n-2)$.

DEFINITION 2.6 (Permutation). Suppose that a set has n elements. Suppose that an experiment consists of selecting k of the elements one at a time without replacement. Let each outcome consist of the k elements in the order selected. Each such outcome is called a *permutation* of n elements taken k at a time. We denote the number of distinct such permutations by the symbol $P_{n,k}$.

We say "n permute k" to talk about $P_{n,k}$.

EXAMPLE 2.7. We see from the Example 2.5 that n permute 3 is $P_{n,3} = n(n-1)(n-2)$. An astute observer would see that

$$P_{n,n} = n(n-1)(n-2)\cdot \ldots \cdot 2 \cdot 1\,.$$

DEFINITION 2.8 (Factorial). We write $n! = n(n-1)\cdot\cdots\cdot 2\cdot 1$ and call it "n factorial". Thus,

$$n! \stackrel{\text{def}}{=} P_{n,n}\,.$$

EXAMPLE 2.9. From the above definition, $0! = 1$ as there is one way to order an empty set – you don't do anything! We have that for $n \geqslant k \geqslant 0$,

$$P_{n,k} = \frac{n!}{k!}\,.$$

Another way to think about permutation is the way of selecting (without replacement) k out of n elements where the order of the elements matter. Sometimes, one does NOT care about the order of the selection. This leads to the concept of *combination*.

DEFINITION 2.10. Consider a set of n elements. Each subset of size k chosen from this set is called a *combination of n elements taken k at a time*. We denote the number of distinct such combinations by the symbol $C_{n,k}$, or often enough $\binom{n}{k}$.

THEOREM 2.11. *For $n \geqslant k \geqslant 0$, we have that*

$$C_{n,k} = \frac{P_{n,k}}{k!} = \frac{n!}{k!(n-k)!}\,.$$

THEOREM 2.12 (Binomial theorem). *For all numbers x and y,*

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

Because of this formula, $\binom{n}{k}$ is often referred to as a binomial coefficient.

ex:blood

EXAMPLE 2.13. The gene for human blood type consists of a pair of alleles chosen from the three alleles commonly called O, A and B. For example, a possible combination of alleles is AA. There is no distiction between the orders of the combination, OA would be the same with AO. There are 3 pairs where both alleles are the same and $\binom{3}{2}$ pairs where the alleles are different. Thus, the number of possible blood type is

$$3 + \binom{3}{2} = 3 + \frac{3!}{2!}1! = 3 + 3 = 6\,.$$

Can you think of another way to derrive this number?

QUESTION. *What happen if there is an alien that has n alleles but the blood type is still the combination of the alleles? How many possible blood type does this alien have?*

2.2. Generating probability from counting. Recall the principle

$$\text{Probability of an event happening } = \frac{\text{Number of ways it can appear}}{\text{Total number of outcomes}}\,.$$

We will have some examples to get used to using this to find out the probability of something happening.

EXAMPLE 2.14. Suppose a fair coin is tossed 10 times. What is the probability that

(1) exactly 3 heads appear.
(2) 3 or fewer heads appear.

In total, there are 2^{10} possibilities.

(1) The number of ways three heads can appear is $\binom{10}{3}$. So the probability for exactly 3 heads to appear is

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

(2) The number of ways three or fewer heads appear would be the sum of the numbers of ways exactly $0, 1, 2$ or 3 heads appear

$$\binom{10}{0} + \binom{10}{1} + \binom{10}{2} + \binom{10}{3}\,.$$

Thus, the probability for three or fewer heads to appear is

$$\frac{\binom{10}{0} + \binom{10}{1} + \binom{10}{2} + \binom{10}{3}}{2^{10}} = 0.1719\,.$$

EXAMPLE 2.15. Suppose a class contains 15 men and 30 women and that 10 students are selected at random with equal probability. What is the probability that exactly 3 men are selected?

There are $\binom{45}{10}$ ways to select 10 students out of 45.

Because there are three men to be selected, there are seven women that would be selected. The number of ways for this to happen is

$$\binom{15}{3}\binom{30}{7}.$$

So the probability to select exactly three men out of the 45 students is

$$\frac{\binom{15}{3}\binom{30}{7}}{\binom{45}{10}} = 0.2904\,.$$

EXAMPLE 2.16. There is a deck of 52 cards that have been shuffled thoroughly. There are four players and each player receives 13 cards. What is the probability that each player receives exactly one ace?

If each player were to receive one ace, there are 13^4 possible positions that the aces could appear.

Without this restriction that each player were to receive one ace, there are $\binom{52}{4}$ positions that the aces could appear.

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

3. Conditional probability

The idea of probability is based on the likelihood of something to happen within an idealized universe. Say, within the universe of MATH 170, 50% of the students is female. However, if one expand the universe to the entire human population on earth, according to the Worldbank, the number is 49.585%. This opens up the idea of conditional probability, where we "zoom" into one universe to study the odds of certain events to happen without worrying about the odds of the same events to happen in the bigger universe.

DEFINITION 3.1 (Conditional probability). Given a probability space $(\Omega, \boldsymbol{P})$ and events A and B. *The conditional probability of the event A given that the event B has occured*, denoted by $\boldsymbol{P}(A \mid B)$, is given by

$$\boldsymbol{P}(A \mid B) \stackrel{\text{def}}{=} \frac{\boldsymbol{P}(A \cap B)}{\boldsymbol{P}(B)}\,.$$

Typically, we read the above as "probability of A given B".

It is sometimes good to draw the Venn diagram to see what's going on with probability.

EXAMPLE 3.2. Consider the following figure The unconditional probability $\boldsymbol{P}(A) = 0.30 + 0.10 + 0.12 = 0.52$. However, the conditional probability $P(A \mid B_1) = 1$, $P(A \mid B_2) = 0.12\ddot{o}(0.12 + 0.04) = 0.75$, and $P(A \mid B_3) = 0$.

EXAMPLE 3.3. Here's something from Wikipedia. Even if 100% of patients with pancreatic cancer have a certain symptom, when someone has the same symptom, it does not mean that this person has a 100% chance of

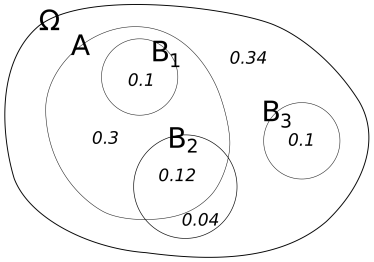


FIGURE 2. Sets in Ω (from Wikipedia)

fig:conditional-prob

getting pancreatic cancer. Assume the incidence rate of pancreatic cancer is 1/100000, while 10/100000 healthy individuals have the same symptoms worldwide, the probability of having pancreatic cancer given the symptoms is only 9.1%, and the other 90.9% could be "false positives" (that is, falsely said to have cancer; "positive" is a confusing term when, as here, the test gives bad news). Based on incidence rate, the following table presents the corresponding numbers per 100,000 people.

Symptom \ Cancer	Yes	No	Total
Yes	1	10	11
No	0	99989	99989
Total	1	99999	100000

fig:cancer

Which can then be used to calculate the probability of having cancer when you have the symptoms:

$$
\begin{aligned}
&P(\text{Cancer}|\text{Symptoms}) \\
&= \frac{P(\text{Symptoms}|\text{Cancer})P(\text{Cancer})}{P(\text{Symptoms})} \\
&= \frac{P(\text{Symptoms}|\text{Cancer})P(\text{Cancer})}{P(\text{Symptoms}|\text{Cancer})P(\text{Cancer}) + P(\text{Symptoms}|\text{Non-Cancer})P(\text{Non-Cancer})} \\
&= \frac{1 \times 0.00001}{1 \times 0.00001 + (10/99999) \times 0.99999} = \frac{1}{11} \approx 9.1\%
\end{aligned}
$$

4. Independence

EXAMPLE 4.1. Suppose that a fair coin is tossed twice. The experiment has four outcomes, HH, HT, TH, and TT, that tell us how the coin landed on each of the two tosses. We can assume that this sample space is simple so that each outcome has probability 1/4. Suppose that we are interested in the second toss. In particular, we want to calculate the probability of the event $A = \{H \text{ on second toss}\}$. We see that $A = \{HH, TH\}$, so that $\boldsymbol{P}(A) = 2/4 = 1/2$. If we learn that the first coin landed T, we might wish

to compute the conditional probability $\boldsymbol{P}(A|B)$ where $B = \{T \text{ on first toss}\}$. Using the definition of conditional probability, we easily compute

$$\boldsymbol{P}(A \mid B) = \frac{\boldsymbol{P}(A \cap B)}{\boldsymbol{P}(B)} = \frac{1/4}{1/2} = \frac{1}{2} = \boldsymbol{P}(A)\,.$$

So, we can see that

$$\boldsymbol{P}(A \cap B) = \boldsymbol{P}(A)\boldsymbol{P}(B)\,.$$

Another way to view this is that A doesn't live under the world of B. What happens to B doesn NOT affect what happens in A.

This inspires the following definition.

DEFINITION 4.2 (Independence). Two events A and B are independent if

$$\boldsymbol{P}(A \cap B) = \boldsymbol{P}(A)\boldsymbol{P}(B)\,.$$

The example with the coin does not seem to have any consequence in our daily life. However, understanding probability (or anything) could mean life or death. There are a lot of tragedies when people misuse probability and deduce wrong conclusions that lead to deaths of others.

EXAMPLE 4.3. In this example, we will write SIDS for Sudden Infant Death Syndrome. Sally Clark[1] (1964-2007) was an English woman who was wrongly accused of murdering her own children in 1999. In the trial, certain "expert" Roy Meadow, calculated the following probaility:

(1) $\boldsymbol{P}(1 \text{ SIDS in a relatively well-off family}) \approx 1/8500$,
(2) $\boldsymbol{P}(2 \text{ SIDS in a relatively well-off family}) \approx (1/8500)^2 = 1/73\text{million}$

From this, he concluded that this event happens once in 100 years and it is likely that Sally was guilty of killing her own children. The court ruled accordingly. This is WRONG! For a few reasons.

(1) You need to look at the probability of SIDS in a typical family, not a well-off family. This probability turns out to be $1/1300$.
(2) Two SIDS in the same family are not two independent events.
(3) A chance for a mother to kill her own baby is *incredibly low*.

Let's do a conservative calculation to see the lowest chance for Sally Clark to be guilty, given the evidence provided in court. The key words here are "given the evidence".

Let us have a few statistics up. Let us set up some notations.

(1) $\boldsymbol{P}(1 \text{ SIDS }) \approx 1/1,300$,
(2) $\boldsymbol{P}(\text{SIDS in a family given there's already 1 SIDS}) \approx 1/130$,
(3) $\boldsymbol{P}(\text{ two children in the same household dies not from SIDS}) \approx 3/650,000$

[1] https://en.wikipedia.org/wiki/Sally_Clark
https://blogs.cornell.edu/info2040/2018/11/28/bayes-theorem-in-the-court-the-prosecutors-fallacy/

So, from (1) and (2) already, we have

$$\boldsymbol{P}(\text{2 SIDS in a family}) \approx \frac{1}{1300}\frac{1}{130} = \frac{1}{169,000}\,.$$

This is still about 1/1million. But this still doesn't tell us about the chance that Clark murderred her children. Let us use the knowledge of conditional probability to proceed. Let D denote the event of two death babies, H be the event of 2 SIDS in a family, Note that $H \subseteq D$. Therefore,

$$\begin{aligned}\boldsymbol{P}(H \mid D) &= \frac{\boldsymbol{P}(H \cap D)}{\boldsymbol{P}(D)} = \frac{\boldsymbol{P}(H)}{\boldsymbol{P}(D \cap H) + \boldsymbol{P}(D \cap H^C)} = \frac{\boldsymbol{P}(H)}{\boldsymbol{P}(H) + \boldsymbol{P}(D \mid H^C)\boldsymbol{P}(H^C)} \\ &\approx \frac{1/169,000}{1/169,000 + (1 - 1/169,000) * 3/650,000} \approx 0.56\,.\end{aligned}$$

There is more chance for Sally Clark to not kill both babies than to kill. This is, of course, not enough to overturn the juridiction but it is not a death sentence as the previous outrageous calculation.

Denote the notation

$$\sum_{i=1}^{k} a_i \stackrel{\text{def}}{=} a_1 + \cdots + a_k\,.$$

THEOREM 4.4 (Bayes Theorem). *Let $(\Omega, \boldsymbol{P})$ be a probability space. Let $B_1, \ldots, B_k$ be disjoint events such that $\boldsymbol{P}(B_i) > 0$ for $i = 1, \ldots, k$ and*

$$B_1 \cup B_2 \cup \cdots \cup B_k = \Omega\,.$$

Then

$$\boldsymbol{P}(B_i \mid A) = \frac{\boldsymbol{P}(B_i \cap A)}{\sum_{i=1}^{k} \boldsymbol{P}(B_i)\boldsymbol{P}(A \mid B_i)}\,.$$

5. Random Variable and Expectation

DEFINITION 5.1. Given a sample space Ω. A discrete random variable X is a function $X : \Omega \to \mathbb{N}$.

With this definition, we can ask what is the likelihood of $X = 1$? Denote $\{X = n\}$ the event that consists of all the outcomes $\omega \in \Omega$ so that $X(\omega) = n$.

So,

$$\{X = n\} \stackrel{\text{def}}{=} \{\omega \in \Omega \mid X(\omega) = n\}\,.$$

We can certainly compute

$$\boldsymbol{P}(\{X = n\})$$

if we know the probability function $\boldsymbol{P}$.

DEFINITION 5.2 (Expectation). Let $(\Omega, \boldsymbol{P})$ be a probability space and X be a discrete random variable. Let $p_n = \boldsymbol{P}(X = n)$. The *expectation* of X is defined by

$$\boldsymbol{E}(X) \stackrel{\text{def}}{=} \sum_{n=0}^{\infty} np_n\,.$$

Sometimes we use μ to be the notation for $\boldsymbol{E}(X)$.

DEFINITION 5.3 (Standard deviation). The *variance* of X is defined by

$$Var(X) \stackrel{\text{def}}{=} \boldsymbol{E}((X-\mu)^2) = \sum_{n=0}^{\infty} (n-\mu)^2 p_n \,.$$

The *standard deviation* of X is

$$\sigma_X = \sqrt{Var(X)}\,.$$

Bibliography

Boole2009 [Boo09] George Boole. *An investigation of the laws of thought*. Cambridge Library Collection. On which are founded the mathematical theories of logic and probabilities, Reprint of the 1854 original, Previously published by Dover Publications, Inc., New York, 1957 [MR0085180]; Prometheus Books, Amherst, NY, 2003 [MR1994936]. Cambridge University Press, Cambridge, 2009, pp. ii+viii+425. ISBN: 978-1-108-00153-3. DOI: 10.1017/CBO9780511693090.024.

Burger2013 [Bur13] Edward Burger. *The heart of mathematics : an invitation to effective thinking*. Hoboken, NJ: Wiley, 2013. ISBN: 9781118156599.

adimitriou2009 [DP09] Apostolos Doxiadis and Christos H. Papadimitriou. *Logicomix*. An epic search for truth, Character design and drawings by Alecos Papadatos, color by Annie Di Donna. Bloomsbury Press, New York, 2009, p. 347. ISBN: 978-1-59691-452-0; 1-59691-452-1.

DeGroot2012 [DS12] Morris DeGroot and Mark J. Schervish. *Probability and statistics*. Boston: Addison-Wesley, 2012. ISBN: 9780321500465.

Newstead [New] Clive Newstead. *Infinite Descent*. URL: https://infinitedescent.xyz/dl/infdesc.pdf.

Sainsbury1991 [Sai91] R. M. Sainsbury. *Logical forms : an introduction to philosophical logic*. Oxford, UK Cambridge, Mass: B. Blackwell, 1991. ISBN: 0631177787.

eadRussell1997 [WR97] Alfred North Whitehead and Bertrand Russell. *Principia mathematica to *56*. Cambridge Mathematical Library. Reprint of the second (1927) edition. Cambridge University Press, Cambridge, 1997, pp. xlvi+410. ISBN: 0-521-62606-4. DOI: 10.1017/CBO9780511623585