Think like the pros

Emanuele Viola

September 7, 2011

$$\exists = \bigcup$$

Disclaimer. This is a draft that is not intended for wide dissemination. Parts of it are taken from various sources that are not yet acknowledged. Still, any comment is highly appreciated!

Contents

1 Claims

Uniquely to our field, we make claims that are either true or false, but not both.

Equivalent ways of saying true and false:
$$\begin{cases} \text{true} & \text{right} & \text{holds} & 1 & \top & \checkmark \\ \text{false} & \text{wrong} & \text{does not hold} & 0 & \bot & X. \end{cases}$$

To jump-start things, I'll assume that you know simple facts about numbers, for example "$1+1=2$," "$17 \leq 23$," and "$2^3 = 8$." All these are claims. We are going to represent claims using capital letters A, B, and so on. So for example A can stand for "$1+1=2$." In which case A is true. Similarly, B can stand for a false claim such as "$2+3>9$." In this case B is false.

One may fantasize about "doing everything from scratch:" defining even basic concepts such as numbers etc. Such enterprizes have indeed been attempted. But they are all disappointing, because in "doing everything from scratch" one appears to need to use the very concepts one is trying to define. So let's just assume we all know what an integer number is and move on.

Simple claims are combined with each other to make more interesting claims. The combination is achieved via the use of *logical connectives and quantifiers*. The logical connectives are: Implication , Or, And, Not. The quantifiers are There exists and For all. All these are absolutely crucial. We discuss these in turn.

1.1 Implication

Implication is a logical connective that is often misunderstood. To understand its meaning, we form a *truth table*. This shows when the claim "A implies B" is True depending on the truth of A and B.

A	B	A implies B
False	False	True
False	True	True
True	False	False
True	True	True

For example, consider the claim "the earth is flat implies $3>2$." Let's let A denote "the earth is flat" and B denote "$3>2$." A is false, while B is true. What is the truth of "A implies B?" Let's look at the table. The second line says that in this case the implication is true. So that's it: "the earth is flat implies $3>2$" is a true statement.

You should construct examples by yourself corresponding of the other 3 possibilities.

Looking at the table, we see that there is **exactly one case in which implication is false**: when A is true and B is false. **In every other case the implication is true!**

Equivalent ways of saying A implies B: $A \Rightarrow B$, $A \to B$, if A then B, A only if B, let/assume/suppose A then B, $(\neg A) \vee B$,

The last thing, $(\neg A) \vee B$, we will see later.

The reverse implication could be obtained by swapping A and B. But it's often useful to turn the arrow instead:

Equivalent ways of saying A is implied by B: $A \Leftarrow B$, $A \leftarrow B$, if B then A, A if B, $A \vee (\neg B)$.

When both the forward and the reverse implications hold, the two claims are said to be *equivalent*.

Equivalent ways of saying "A is equivalent to B": $A \Rightarrow B$ and $A \Leftarrow$, $A \Leftrightarrow B$, $A \leftrightarrow B$, A if and only if B, A iff B, A when and only when B.

Common mistakes: (1) confusing $\Rightarrow$ and $\Leftarrow$, (2) confusing $\Rightarrow$ and $\Leftrightarrow$ People often confuse the meanings of $\Rightarrow$, $\Leftarrow$, and $\Leftrightarrow$. The confusion derives from the fact that in everyday language, the word "if" is sometimes used to express a meaning that is different from its meaning in mathematics. Consider for example the informal claim:

You go out if you finish your homework.

If A stands for "You go out" and B for "finish your homework, one is tempted to think of the claim as A if B, that is $A \Leftarrow B$. But this is not the intended meaning. $A \Leftarrow B$ means that if you finish your homework then you can go out. All this is saying is that if you finish your homework then you must go out, but it is perfectly allowed for you to go out without having finished the homework!

The intended meaning is precisely opposite: if you go out then you must have finished your homework. The intended meaning is A **only if** B, $A \Rightarrow B$.

When a pro hears "if" it asks herself if this is really "if" or rather "only if" or "if and only if" and then translates accordingly. In this case:

You go out ***only if*** *you finish your homework.*

A similar confusion arises between $\Rightarrow$ and $\Leftrightarrow$.

Do you understand implication? Let's now test your understanding of implication via a quiz. Here are four cards:

You know for true that each card has a letter on one side and a number on the other. Right now, you are looking at only one side of the cards, and you do not know what is written on the other.

Quiz: Suppose I claim that

If a card has a vowel on one side, then it has an even number on the other side.

You task is to determine if my claim is right or wrong by turning over the **minimum** number of cards. That is: which of the cards you **need** to turn over to see if the statement is true or false?

1.2 Not

Equivalent ways of saying "not A": A is false, A is wrong, $1 - A$, $\neg A$, $\bar{A}$, $\not{A}$.

Writing $1 - A$ for "not A" makes sense if you think of true as 1 and false as 0. Note that in this case things work beautifully! If A is 1 then $1 - A$ is 0, and if A is 0 then $1 - A$ is 1.

The important thing of Not is that Not (Not A) is equivalent to A. In fact, $(1-(1-A)) = A$.

1.3 Or

Truth table.

We see Or is *commutative*.

Common mistake: confusing Or with Exclusive Or. Note that if $p \vee q$ is true, it means that at least one of p and q must be true, and it may well be that both are true. For example, if p is "3 is odd" and q is "2 is even" then p and q are both true and $p \vee q$ is true as well.

This apparently innocuous feature is actually source of great confusion. The confusion derives from the fact that in everyday language, the word "or" is often used to express a different meaning. Consider for example the informal claim:

We shall triumph or perish.

Here it is implied that we shall not both triumph and perish. We will do one or the other, but not both. Observe how different this is from the meaning above! This different type of "or" is called exclusive or, and is written xor. A pro would mentally translate the above claim into

*We shall triumph **xor** perish.*

When a pro hears "or" it asks herself if this is really "or" or rather "xor" and then translates accordingly. In mathematics, "or" usually stands for the non-exclusive type, where it's OK if both parts are true. However, sometimes people confuse it, so watch out for danger!

1.4 And

Equivalent ways of saying "A and B": $A \wedge B$, "A, B", $A\&B$, $A \cdot B$.

For the last one, think of true as 1 and false and 0, and note the truth of "A and B" indeed corresponds to the product of the truths of A and B.

Exercise 1. Look again at the four cards in Section 1.1. Which cards do you **need** to turn over to see if the next statement are true or false?

A card has a vowel on one side **if** *it has an even number on the other side.*
A card has a vowel on one side **if and only if** *it has an even number on the other side.*
A card has a vowel on one side **or** *it has an even number on the other side.*
A card has a vowel on one side **xor** *it has an even number on the other side.*

1.5 Very important relationships

Implication using Or and Not: $A \Rightarrow B$	is equivalent to	$(\neg A) \vee B$.

This equivalence is evident looking at the truth tables. The expression $(\neg A) \vee B$ also makes clear that implication is false exactly in one case, when A is true and B is false, because that's the only case in which $(\neg A) \vee B$ is false. When in doubt about the meaning implication, go back to $(\neg A) \vee B$.

Contrapositive: $A \Rightarrow B$	is equivalent to	$(\neg B) \Rightarrow (\neg A)$

Looks weird? If so let's just look at the previous rule. What we are claiming is that $(\neg A) \vee B$ is equivalent to $(\neg(\neg B)) \vee (\neg A)$. Indeed, this is true because $(\neg(\neg B)) \Leftrightarrow B$, and so $(\neg(\neg B)) \vee (\neg A) \Leftrightarrow B \vee (\neg A) \Leftrightarrow (\neg A) \vee B \Leftrightarrow A \Rightarrow B$.

De Morgan's laws:	$\neg(A \vee B)$ is equivalent to $(\neg A) \wedge (\neg B)$.
	$\neg(A \wedge B)$ is equivalent to $(\neg A) \vee (\neg B)$.

De Morgan's laws also give us an algebraic way to express

$$A \vee B = \neg(\neg A \wedge \neg B) = 1 - ((1 - A) \cdot (1 - B)).$$

2 Quantifiers

There are two quantifiers in mathematics, exists ($\exists$) and for all ($\forall$). The $\exists$ quantifier is nothing but a different way of saying Or, and the $\forall$ quantifier is nothing but a different way of saying And.

$$\exists = \bigvee \qquad\qquad \forall = \bigwedge$$

Typically, quantifiers are used to range over *many*, usually infinitely many claims, while logical connectives range over *few* claims things, usually 1 or 2. We now illustrate this. To range over many claims, it is convenient to think of claims as having variables.

Claims with variables. So far we considered claims such as "2 is even" and we represented them by symbols such as A. We now consider claims with some variables. For example "n is even." The truth of this claim depends on the value of n, so it is not defined until we define n. We represent such an "open-ended" claim by

$$A(n)$$

to indicate that has a variable n. We then can say $A(2)$ is true, $A(3)$ is false, and so on.

Quantifiers are typically used to range over the variable n:

$$\exists n : P(n) \qquad \Leftrightarrow \qquad P(1) \vee P(2) \vee P(3) \vee \ldots \qquad \Leftrightarrow \qquad \bigvee_n P(n),$$

$$\forall n : P(n) \qquad \Leftrightarrow \qquad P(1) \wedge P(2) \wedge P(3) \wedge \ldots \qquad \Leftrightarrow \qquad \bigwedge_n P(n).$$

Warning The pros often drop quantifiers when they are "clear from the context." This is often done because expressions with too many quantifiers are hard to read. However, you can be sure that the pros, when faced with some new definition of theorem where the quantifiers have been dropped, the first thing they do is to put back all quantifiers!

Note to take the negation you use De Morgan's law, which swap the quantifiers, so you'll end up moving at what was previously your opponent's turn.

De Morgan's laws for quantifiers:	$\neg\exists x : A(x)$ is equivalent to $\forall x \neg A(x)$. $\neg\forall x : A(x)$ is equivalent to $\exists x \neg A(x)$.

Thus you can move negation across quantifiers swapping them.

2.1 So, what can I do?

If you know that certain claims are true, you can infer others. We also say, deduce, derive, etc. How? It's simple, you can infer any claim that *must be true*.

Example: Suppose you know that A is false, and that $A \vee B$ is true. Then you can deduce B.

From $A \wedge B$ you can deduce A, and also B.

If you know that both A and $A \Rightarrow B$ are true you can deduce B. This case is so important that it has its own name: modus ponens.

End of the example.

Example: Suppose you know that $\forall$ integer $x, A(x)$. Then you can infer $A(5)$.

End of the example.

Equivalent ways of saying define A to mean B: Say A if B, write A if B.

3 Proofs by contradiction

We see a couple of proofs by contradiction. The first is the use of the contrapositive. The second deduces a contradiction (not exactly a proof by contrapositive).

3.1 Using the contrapositive

Claim 1 (Markov's inequality)**.** Suppose you have n positive numbers $a_1, a_2, \ldots, a_n$. Let t be an integer. If $a_1 + a_2 + \cdots + a_n \leq t$ then there are at most $t/2$ integers that are bigger than 2.

Proof: The first thing we note is that all those "fancy" words like "Suppose you have" and "Let" are immaterial from a logical point of view: they all collapse to the $\forall$ quantifier. The claim is:

$$\forall n, a_1, \ldots, a_n, t : A \Rightarrow B,$$

where A stands for "$a_1 + a_2 + \cdots + a_n \leq t$" and B for there are at most $t/2$ integers that are bigger than 2.

To prove $A \Rightarrow B$, we prove the contrapositive: $\neg B \Rightarrow \neg A$. The negation of A, $\neg A$, is simply "$a_1 + a_2 + \cdots + a_n > t$." The negation of B, $\neg B$, slightly more complicated, is: "there are more than $t/2$ integers that are bigger than 2."

Now you can finish the proof

End of the proof.

Is this result *tight*? Can we replace $t/2$ by $t/2.0001$? With such a substitution, the claim becomes false. How do we prove a claim false? Simple, you take the negation and you prove it true. In this case, the negation is

$$\exists \text{ integers } n, a_1, a_2, \ldots, a_n, t \neg(A \Rightarrow B).$$

Internally, we have the negation of an implication. Using the previous rules, we rewrite it as

$$\exists \text{ integers } n, a_1, a_2, \ldots, a_n, t : A \wedge \neg B.$$

So that's what we need to show. We need to exhibit numbers... that's easily done. Pick $n = 3, a_1 = a_2 = a_3 = 2, t = 6$. Their sum equals t and there are $3 > 6.2.0001$ numbers (all of them) that are bigger than 2.

The pros always ask if a claim is tight:

If you don't ask if a claim is tight, you are not doing math

Summary of the things the pros ask: is it if, only if, or iff? Is it or or xor? Is it tight?

3.2 Irrationality of $\sqrt{2}$

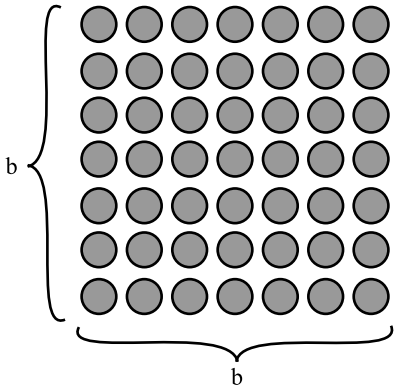
Let us make further practice with a beautiful proof by contradiction. We are going to prove the following claim.

Claim 2. There are no positive integers a and b such that $2a^2 = b^2$.

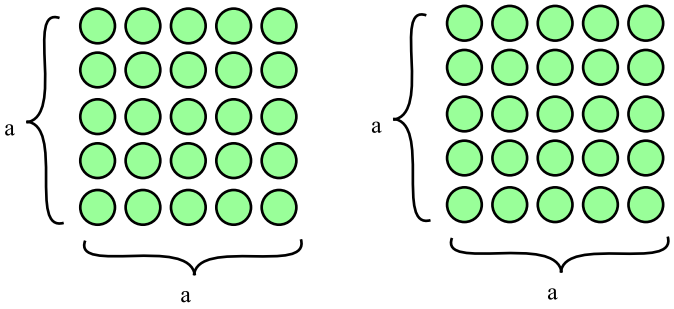
The pros call it the "irrationality of $\sqrt{2}$." And the legend goes that the Pythagorean who proved the claim – this was quite a while back – was then drowned at sea.

Proof: Assume, towards a contradiction, that there are positive integers a and b such that $2a^2 = b^2$. What are we going to do with this a and b? With the benefit of hindsight, let us make the following move: We further require that a *is as small possible*. Why can we do this? Well, we assumed that there are integers such that. While for all we know at this point there could be multiple choices for a and b, surely there has to be a smallest possible a for which this holds.

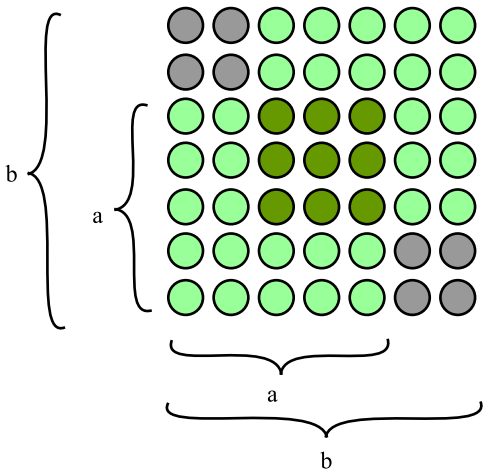
OK, that was boring. Let's get to the action then. What does it mean that $2a^2 = b^2$? It means that the number of discs in this square with side b



is equal to twice the number of discs in a square of side a:



Now let's place the two copies of squares of side a at opposite corners of the bigger square:



They must overlap. By symmetry, the overlap is a square . The two corners uncovered are also squares. So we have found a smaller configuration. This is a contradiction.

End of the proof.

This proof seems to use some "geometric intuition." Is it really correct, or are the pictures misleading? Also, what is exactly the contradiction obtained?

Proof: We start the same.

Our assumptions: (1) $2a^2 = b^2$, (2) among all pairs yielding (1), (a, b) is one with the smallest possible a.

Now consider the numbers $a' := (b - a)$, and $b' := (2a - b)$ (the idea of considering these two numbers comes from looking at the previous picture).

We want to prove two things: (i) $a' < a$ and (ii) $2a'^2 = b'^2$. (i) + (ii) contradicts (2).

To prove (i), note that if $a' > a$ then $b \geq 2a$ and so $b^2 \geq 4a^2 > 2a^2$. This contradicts (1). So if this is the case we already reach a contradiction and we are done. It's a "mini" proof by contradiction inside a large proof by contradiction. So now we know (i) is true.

For (ii), first note that

$$2a'^2 = 2(b^2 + a^2 - 2ab),$$
$$b'^2 = 4a^2 + b^2 - 4ab.$$

Simplifying, we see that

$$2a'^2 = b'^2 \Leftrightarrow 2a^2 = b^2.$$

This is true by (2). Hence we contradicted (1). This concludes the proof.

Let's have a look at the steps structure of the proof.

(a) Assume that there is a pair $a, b : 2a^2 = b^2 \Rightarrow$

(b) There is a pair $a, b : 2a^2 = b^2$ and among all such pairs, a is smallest. $\Rightarrow$

(c) Define a', b'.

(d) Note that $a' < a$. This is because if $a' > a$ we contradict (a) and we are done. So we can assume $a' < a$.

(e) Using (a) prove that $2a'^2 = b'^2$.

(d)+(e) contradict (b).
This contradicts (b).

End of the proof.

4 Quantifiers as games

The vast majority of claims in mathematics starts with a sequence of nested quantifiers, typically 4 or 5 looking like:

$$\exists a\, \forall b\, \exists c\, \forall d \ldots \text{Something(a,b,c,d,}\ldots\text{)}.$$

Such a format is often not evident, because quantifiers have been dropped (see above) or are hidden under various definitions. But that's what the claim "really looks like," and when in doubt the pros write it that way.

It is of **crucial importance that you think of such a claim with nested quantifiers as a game between two players**:

$$P_\exists \text{ and } P_\forall.$$

The players alternate making moves as described in the claim.

4.1 Tic-tac-toe

To explain, we consider the tic-tac-toe game. This is played by two players, X and O, on a 3x3 grid. For convenience we number the cells:

1	2	3
4	5	6
7	8	9

The players alternate in placing X or O on one of the empty cells. The first to get 3 straight in a row, column, or diagonal, wins.

Some board configurations are a win for player X. That is X has a winning strategy: no matter how O plays, X can win. This type of reasoning is exactly what goes on with quantifiers. Let's make a formal claim.

Claim 3. Suppose the board is

	X	
	X	O

and it is O's turn to move. Then X has a winning strategy.

Proof: The proof is a series of cases which analyze all possible moves of O and show that, in any case, X can win.

First, if O does not play 8, X wins by playing 8. Hence the claim is proved in this case. All that is left is to prove the claim in the case in which O does not play 8. So we are going to assume that O plays 8 and continue. The board at this point is:

	X	
	X	O
	O	

.

X plays 1. The board at this point is:

X	X	
	X	O
	O	

.

Now if O plays 3, X plays 9 and wins. If O does not play 3, then X plays 3 and wins. This covers all cases. So X has a winning strategy.

End of the proof.

The previous proof exactly corresponds to analyzing quantifiers. We have shown that X has a winning strategy by showing that the following claim is true:

$$\forall \text{ move of } O \ \exists \text{ move of } X \ \forall \text{ move of } O \ \exists \text{ move of } X \ : \ X \text{ wins the game.}$$

Note how each quantifier has a corresponding part in the proof. The first $\forall$ is the first move of O. If that is not 8, then in the next $\exists$ we can pick 8 and X wins. If it is 8, then we pick 1. Then we analyze the next $\forall$ quantifier, etc.

99% of mathematics can be viewed as analyzing boards of games that are only slightly more complicated than tic-tac-toe.

4.2 Back to numbers

If you want to show that a claim is true, you should think of impersonating the player $P_\exists$. The claim is true if and only if you can always win *no matter how* $P_\forall$ *plays*, that is if you have a winning strategy. Important note: when proving the claim, you do **not** get to pick the moves corresponding to $\forall$. Those belong to your opponent. The moves you **do** get to pick are those of the $\exists$ quantifier.

Example: Consider the claim "$\forall n, n = 2 \vee n = 5 \exists m, m = 3 \vee m = 4 : n + m$ is even ."

In the corresponding game, $P_\forall$ plays first, picking either $n = 2$ or $n = 5$. Then it's $P_\exists$ turn to pick either $m = 3$ or $m = 4$. $P_\exists$ wins if $n + m$ is even, and loses otherwise.

Let's prove this claim true. So you play $P_\exists$. First, our opponent picks n. We have no control over this. Now it's our turn to pick m. We must show that no matter how n is picked, you can pick m so that $n + m$ is true. Indeed, if $n = 2$ you can pick $m = 4$ and

$n + m = 6$ is even. While if $n = 5$ you can pick $m = 3$ and $n + m = 8$ which is also even. We have covered every possible move of the opponent $P_\forall$. In any case you win. Therefore you have just proved this claim correct.

End of the example.

Want to prove that some claim is wrong? No problem. Let's take the negation of the claim and prove that is right. For this, you again impersonate $P_\exists$.

Example:

$$\neg\exists x\forall y\exists z : A(x, y, z) \Leftrightarrow \forall x\exists y\forall z : \neg A(x, y, z).$$

End of the example.

4.3 Examples with order of growth

Big-Oh:

Claim 4. $\exists c, n_0 : \forall n \in \mathbb{N}, n \geq n_0 : 2n^2 + n + 1000 \leq cn^2$.

We write $2n^2 + n + 1000 = O(n^2)$.
Let's make this more interesting:

Claim 5. $\forall a \geq 0 : an^2 + n + 1000 = O(n^2)$.

Note that we suppressed quite a bit of quantifiers using the $O(\cdot)$ notation. What this really means is:

$$\forall a \geq 0\exists c, n_0 : \forall n \in \mathbb{N}, n \geq n_0 : an^2 + n + 1000 \leq cn^2.$$

You see why the pros drop quantifiers in their claims, it makes them more readable.

Proof: Adversary: pick a.
You pick $c := a + 1$.

End of the proof.

Little-oh:

Claim 6. $\forall c > 0\exists n_0 : \forall n \in \mathbb{N}, n \geq n_0 : 15n^2 \leq (n^2 \lg n)/c$.

Proof: Adversary move: chooses c.
Your move: Pick $n_0 := 2^{15c}$.
You win because for $n \geq n_0$, $(n^2 \lg n)/c \geq (n^2 2^{15c})/c \geq 15n^2$.
Since we consider an arbitrary move c of the adversary, and we showed how you win in any case, the claim is true.

End of the proof.

This is an example of "little-oh." We write $15n^2 = o(n^2 \lg n)$. Similarly we can write $n^{10} = o(2^n)$ etc.
We also write $n^2 + 15n = n^2 + O(n)$ etc.
Ω, ω are like O, o with inequalities swapped and c instead of $1/c$.

Claim 7. $n^{\sqrt{\lg n}} = n^{\omega(1)}$.

Example: The pros hardly ever prove claims they know are true. Instead, they *attempt* to prove a claim *in the hope* it is true. Since the truth of the claim is unknown, the pros often try both ways, they try to prove it true, and try to prove it wrong. Consider for example the claim:

$$10n = \omega(n).$$

This means

$$\forall c \geq 0 \exists n_0 \forall n \geq n_0 : 10n \geq cn.$$

Let's say we try to prove it true. Again, we play the game. If our opponent $P_\forall$ picks $c = 1, 2, \ldots, 10$ we are good, we can win. But if the opponent picks $c = 11$ things seem to fail.

So, let's try to prove the claim *false*. We take the negation of the claim:

$$\exists c \geq 0 \forall n_0 \exists n \geq n_0 : 10n < cn.$$

Now life is easy. It's our move to pick c. We pick $c = 11$ and it is easy to see we win: $10n < 11n$ no matter what n_0 is chosen (for example we can pick $n = n_0$).

Hence the claim is false.

End of the example.

We say Θ if both O and Ω.

Arbitrarily large. A claim $P(\cdot)$ holds for arbitrarily large n if, no matter how an adversary picks a threshold t, you can always find a value $n > t$ for which $P(n)$ is true. This is written as

$$\forall t \exists n > t : P(n).$$

An equivalent way of saying arbitrarily large is saying "infinitely many."

For all sufficiently large. A claim $P(\cdot)$ holds for all sufficiently large n if there exists some "threshold" t such that for $n > t$ the claim is true:

$$\exists t \forall n > t : P(n).$$

Claim 8. For all sufficiently large $\Rightarrow$ for arbitrarily large. Specifically, let $P(\cdot)$ be a claim. If P holds for all sufficiently large n, then P holds for arbitrarily large n.

The proof is a basic example of a very important concept: how to use a winning strategy in a game to win in another game.

Proof: Our assumption is that $\exists t \forall n \geq t P(n)$. This means that we can win this game. That is, we have a move t^* such that no matter how the opponent plays $n \geq t^*$, $P(n)$ is true.

We need to prove that $\forall t \exists n > t P(n)$.

The opponent moves first and picks t.

We need to show that we can pick $n \geq t$ so that $P(n)$ is true. Our move is $n := \max\{t, t^*\}$, that is the bigger between t and t^*. This guarantees that $n > t$, which is the kind of move we must make. It also guarantees that $n > t^*$. Since t^* was a winning move in the first game, $P(n)$ is true.

End of the proof.

We now show that the reverse implication does not hold. To show this, we exhibit a claim that holds for arbitrarily large n, but it does not hold for all sufficiently large n.

Let $P(n)$ mean "n is even." Obviously, $P(n)$ holds for arbitrarily large n. However, it does not hold for all sufficiently large n, because if n is even then $n + 1$ is not.

You may want to work out the quantifiers in detail.

Note that "NOT ($P(n)$ holds for all sufficiently large n)" is equivalent to "for arbitrarily large n, NOT $P(n)$." Note how sufficiently large becomes arbitrarily large, when negated.

4.4 Sets, functions

Sets are just different notation to express the same claims we constructed using logical connectives and quantifiers. This redundant notation turns out to be useful.

$$\begin{array}{lcr} x = 1 \vee x = 16 \vee x = 23 & \Leftrightarrow & x \in \{1, 16, 23\} \\ x \text{ is even} & \Leftrightarrow & x \in \{x | x \text{ is even } \} \\ A(x) & \Leftrightarrow & x \in \{x | A(x)\} \end{array}$$

With this in mind, sets become straightforward. When are two sets equal? When the defining claims are equivalent.

$$A(x) \Leftrightarrow B(x) \qquad \Leftrightarrow \qquad \{x|A(x)\} = \{x|B(x)\}$$

This shows that order and repetitions do not matter, for example $\{b, a, a\} = \{a, b\}$, because the claims $x = b \vee x = a \vee x = a$ and $x = a \vee x = b$ are equivalent.

When is a set contained in an another, when its defining claim implies the defining claim of the latter.

$$\begin{array}{lcl} A(x) \Rightarrow B(x) & \Leftrightarrow & \{x|A(x)\} \subseteq \{x|B(x)\} \\ A(x) \Leftarrow B(x) & \Leftrightarrow & \{x|A(x)\} \supseteq \{x|B(x)\} \end{array}$$

We also write

$$A \subsetneq B \Leftrightarrow A \subseteq B \wedge A \neq B.$$

We construct new sets using set operations in entirely the same way as we construct new claims using logical connectives:

$$\{x|A(x)\} \bigcup \{x|B(x)\} = \{x|A(x) \vee B(x)\}$$
$$\{x|A(x)\} \bigcap \{x|B(x)\} = \{x|A(x) \wedge B(x)\}$$
$$\overline{\{x|A(x)\}} = \{x|\neg A(x)\}$$

We also take unions and intersections of infinite families:

$$\bigcup_i \{x|A_i(x)\} = \{x|\exists i : A_i(x)\}$$
$$\bigcap_i \{x|A_i(x)\} = \{x|\forall i : A_i(x)\}$$

The set that contains no elements is denoted $\emptyset$. One can think of it as the set of x's for which something *false* holds (which is never the case of course) such as $\{x|1+1=5\}$. This also shows that $\emptyset$ is a subset of any set

- De Morgan's laws for sets
- Math primer. The size of a set is denoted by $|A|$. The number of subsets of size k of an n-element set is $\binom{n}{k}$. It holds that for every integers $n, k > 0$: $(n/k)^k \leq \binom{n}{k} \leq (en/k)^k$, where $e < 2.7183$.

  The powerset is the set of all subsets. Its size is 2^n if the set has size n. So we get: $2^n = \sum_{i=0}^{n} \binom{n}{i}$.
- Tuples. $(a,b) = \{\{a,b\},\{a\}\}$. $A \times B = \{(a,b)|a \in A, b \in B\}$. We also write A^k for $A \times A \times \cdots \times A$ (k times). A 2-tuple is called pair, a 3-tuple triple.
- Strings. A string is a tuple written without brackets and commas. The base set is called alphabet. The empty string is ϵ. Concatenation is denoted $\circ$. Note $\epsilon x = x\epsilon = x$. A set of strings is also called language. x is a substring of y if $y = z \circ x \circ z'$ for possibly empty strings z, z'. So $\{0,1\}^n$ = binary strings of length n. The length of a string is denoted $|x|$.
- The Kleene $*$. For a set of strings, $A^* = \{\epsilon\} \bigcup \{x_1 x_2 \ldots x_k | k > 0, x_i \in A \forall i\}$. So $\{0,1\}^*$ = all binary strings of any length.
- Functions: $f : A \to B$ is a way to associate to every element $a \in B$ exactly one element $b \in B$. 1-1, injective, onto, bijections.

Ways to think of functions. A function $f : A \to B$ can be thought of in several equivalent ways:

(1) A "dynamic" process/transformation that "sends" any element $a \in A$ to some element $b = f(a) \in B$.

(2) A partition or grouping of the elements of A, depending on which elements of b they map to. For example if $A = \{1, 2, 3\}$ and $B = \{0, 1\}$, and $f(a) =$ "is a odd?" the function partitions A in the sets $\{1, 3\}$ whose elements map to 1 and the set $\{2\}$ whose element maps to 0. These two sets are denotes $f^{-1}(1)$ and $f^{-1}(0)$, respectively.

(3) A sequence (or tuple or string) of length $|A|$ over the alphabet B. For this, we fix some ordering of A, and list the images of the elements in order. Using the definitions in (2), the function f can be written as $101 \in B^{|A|}$. If $A = B = \mathbb{N}$, and f is squaring, we can write $f = 1 \circ 4 \circ 9 \circ \ldots$.

Indeed, there are $|B|^{|A|}$ functions from A to B, and the set of these functions is conveniently denoted by B^A.

Sometimes the subscript notation f_a is used instead of $f(a)$. In the example where f is squaring, we can write $f_i = i^2$.

(4) A swap of quantifiers: For any sentence $P(a, b)$ we have

$$\forall a \exists b P(a, b) \Leftrightarrow \exists f : A \to B \forall a P(a, f(a)).$$

Here the function f "commits" us to a choice for every $a \in A$.

(5) A set of tuples: $f \subseteq A \times B$ where $(a, b), (a, b') \in f \Rightarrow b = b'$.

(6) A set: if the range of the function is $\{0, 1\}$, we can think of $f : A \to \{0, 1\}$ as the set $S \subseteq A$ defined as

$$S = \{x | f(x) = 1\}.$$

That is, $f(x) = 1 \Leftrightarrow x \in S$.

Russell's paradox After playing a bit with sets, one (and even the pros...) is tempted to define sets with any combination of symbols, basically

$$\{x | \text{Anything you can think of } (x)\}.$$

Actually, one cannot do that. For example, consider the expression

$$\{x | x \text{ is a set and } x \notin x\}.$$

This may look like a definition of set that is valid, albeit strange. Actually, it is not valid. This set cannot be formed for it leads to a logical contradiction. To see this, suppose this set exists and call it A. Now let us ask the question, does $A \in A$ hold?

If it does, then A should have the defining property of elements of A, which is precisely the property of not being a member of itself, so $A \notin A$, contradiction.

If it does not, then again by definition of A it should be the case that $A \in A$, which is a contradiction again.

Since we have a contradiction in any case, this is not a valid definition.

The problem with it is that x is supposed to range over "all sets." But there is no such thing as "all sets." Figuratively, the concept of "all sets" is so huge a concept that makes our definitions explode.

But to "think like the pros" you don't really need these huge things. As long as you make your variables range over "small things," such as integers, strings, etc., it is believed that you are free from such contradictions. To find out more, consider a career in metamathematics.

Exercise 2 (Binomials and asymptotics)**.** Prove that for every integer $k > 0$, $\binom{n}{k} = \Theta(n^k)$.

Let $k : \mathbb{N} \to \mathbb{N}$ be a function. Suppose that $k(n)$ is $\omega(1)$. Prove that $\binom{n}{k(n)} = o(n^{k(n)})$.

Let $k : \mathbb{N} \to \mathbb{N}$ be a function. Suppose $k(n)$ is both $\omega(1)$ and $O(\sqrt{n})$. Prove that $\binom{n}{k(n)} = n^{\omega(1)}$.

Hint: use that for every integers $n, k > 0$: $(n/k)^k \leq \binom{n}{k} \leq (en/k)^k$, where $e < 2.7183$.

Exercise 3. Let $a = a_0, a_1, a_2, \ldots$ be a sequence of integers. Let us write $a \to \infty$ if $\forall t$, for sufficiently large i, $a_i \geq t$ (this is a definition).

Prove that, for any two sequences a and b, if $a \to \infty$ and $b \to \infty$, then the sequence $c = c_0, c_1, c_2, \ldots$ defined as $c_i := a_{b_i}$ also $\to \infty$.

Exercise 4. Let $A = \{x | a(x)\}, B = \{x | b(x)\}$ be sets. Prove that

$$A = B \Leftrightarrow \left(A \bigcap \overline{B}\right) \bigcup \left(\overline{A} \bigcap B\right) = \emptyset.$$

5 Practice: Regular sets (a.k.a. regular expressions)

Regular sets are meant to capture some "very simple" subsets of $\{0,1\}^*$.

Definition 9 (Regular sets)**.** The regular sets over alphabet $\Sigma = \{0,1\}$ are:

- $\emptyset, \{0\}, \{1\}$, and $\{\epsilon\}$ are regular sets.
- If R_1, R_2 are regular sets, then $R_1 \bigcup R_2$, $R_1 \circ R_2$, and R_1^* are regular sets.

This is how you find this definition in books. What this definition really means is that regular sets are those that can be obtained applying a *finite number* of the rules in the definition. This point is important, for if we were to allow an infinite number of applications of the rules then *any* subset of $\{0,1\}^*$ would be regular (prove this!). Let us see a more precise definition:

Definition 10 (Regular sets, more precise definition)**.** A set S is regular if $\exists$ an integer $k \geq 1$ and a sequence $(S_1, S_2, \ldots, S_k)$ of sets such that:

(i) $S_k = S$, and

(ii) $\forall i \leq k$, set S_i is either

(ii.a) one of $\emptyset, \{0\}, \{1\}$, and $\{\epsilon\}$ or

(ii.b) $\exists p, q < i$ such that $S_i = S_p \bigcup S_q$ or $S_i = S_p \circ S_q$ or $S_i = S_p^*$.

For conciseness we write S_pS_q instead of S_qS_q.

Example: $S := \{0\}^* \bigcup (\{0\}^*\{1\}\{0\}^*\{1\}\{0\}^*)^*$ is a regular set. Here's the sequence of sets in the more formal definition:

$$\Big(\{0\}, \{1\}, \{0\}^*, \{0\}^*\{1\}, \{0\}^*\{1\}\{0\}^*, \{0\}^*\{1\}\{0\}^*\{1\}, \{0\}^*\{1\}\{0\}^*\{1\}\{0\}^*,$$
$$(\{0\}^*\{1\}\{0\}^*\{1\}\{0\}^*)^*, \{0\}^* \bigcup (\{0\}^*\{1\}\{0\}^*\{1\}\{0\}^*)^* = S\Big)$$

Note how each coordinate in the sequence indeed is either one of the sets in (ii.a) or follows from previous coordinates by applying the rules in (ii.b).

After playing a bit with the elements of this set, you guess that this is the set of strings with an even number of ones. You should prove this formally to verify you understand.

Is this the only way to obtain this set? Of course not. Here's a more compact expression:

$$\Big(\{0\} \bigcup \{1\}\{0\}^*\{1\}\Big)^*.$$

Your should prove the equivalence formally to verify you understand.

End of the example.

5.1 The pumping lemma as a game

The *pumping lemma* is a useful tool to show that certain sets are not regular. Here is the statement:

Lemma 11 (Pumping lemma)**.** Let L be a regular language. Then there exists an integer $p \geq 0$ such that for any $w \in L$, $|w| \geq p$, there are strings x, y, z such that $xyz = w$, $|xy| \leq p$, and $|y| > 0$ and for every $i \geq 0$, $xy^iz \in L$.

We are not going to prove this lemma now, we are just going to *use it*. Let's just accept the above is true and see what can we do with it, and how.

The pumping lemma is *not* useful to prove that a language is regular, but the pumping lemma *is* useful to prove that a language is *not* regular. To do this, let's first rewrite the pumping lemma in a starker way:

$$\forall L$$
$$\underbrace{L \text{ regular}}_{A} \Rightarrow \underbrace{\exists p \geq 0 \forall w \in L, |w| \geq p, \exists x, y, z, w = xyz, |xy| \leq p, |y| > 0, \forall i \geq 0 : xy^iz \in L}_{B}.$$

So the pumping lemma is of the form $\forall L(A \Rightarrow B)$.

By the contrapositive, the pumping lemma is equivalent to $\forall L(\neg B \Rightarrow \neg A)$.

Note "$\neg A$" is "L is not regular."

Therefore, **to prove L not regular it is sufficient to prove $\neg B$.** What is "$\neg B$"? Using De Morgan's laws we see that:

$$\neg B \Leftrightarrow$$
$$\forall p \geq 0 \exists w \in L, |w| \geq p, \forall x, y, z, w = xyz, |xy| \leq p, |y| > 0, \exists i \geq 0 : xy^i z \notin L.$$

To recap: to prove L not regular it is enough to prove $\neg B$, and to prove $\neg B$ means to win the above game, where recall we play $\exists$.

Let's see what this game is all about. Our opponent picks p. Then it's our turn to pick a word w in L that has length $\geq p$. Now the opponent decomposes w as $w = xyz$ with certain restrictions on the decomposition, and finally it's our turn to pick an integer i, and we win the game if we can pick i so that $xy^i z$ is not in the language.

That's the game you must play and win, in order to prove that L is not regular using the pumping lemma.

5.2 examples

6 Induction

6.1 Sums

Induction is a rule to prove claims of the type $\forall n, P(n)$. The basic formulation says that if

- *The base case* $P(t)$ and,
- *the induction step* $\forall i \geq t, P(i) \Rightarrow P(i+1)$,

then you can infer $\forall n, P(n)$.

Claim 12. $\forall$ integer $n \geq 0$, $1 + 2 + 3 + \cdots + n = n(n+1)/2$

Proof: Let $Q(n)$ be the claim "$1 + 2 + 3 + \cdots + n = n(n+1)/2$." We need to prove that $\forall n, Q(n)$. We proceed by induction on n.

Base case $n = 0$: In this case we get $0 = 0$ which is true. So $Q(0)$ is true.

Induction step. Assume $Q(n-1)$ is true. So $1 + 2 + 3 + \cdots + n - 1 = (n-1)n/2$. Hence $1 + 2 + 3 + \cdots + n - 1 + n = (n-1)n/2 + n = n(n+1)/2$. Thus $Q(n)$ is true.

End of the proof.

Exercise 5. Prove by induction that $1^3 + 2^3 + 3^3 + \cdots + n^3 = (1 + 2 + 3 + \cdots + n)^2$.

6.2 More general forms of induction. Section 1.4.2 from HMU.

6.3 Structural induction. Sec. 1.4.3 HMU

6.4 The reverse of a set is regular

What these proofs really mean? In a proof by structural inductions such as those pointed to above, it's not exactly transparent what you induct on. Once again, to make things transparent we can use a more transparent definition such as Definition 10. If you work with such a definition, you prove by induction on integer i that every object in your sequence satisfies the claim.

6.5 Ramsey theory

Ramsey theory has been interpreted as "Total chaos is not possible."

Claim 13 (Ramsey's theorem)**.** $\forall k \geq 2 \exists R(k)$ such that any graph on $n \geq R(k)$ vertices has k nodes every two of which are connected (a.k.a. a k-clique) or has k nodes every two of which are not connected (a.k.a. an independent set of size k).

We would like to prove this theorem by induction on k. However, some difficult arises if we try to come up with the numbers $R(k)$ directly. The way to go is to consider a more refined quantity $R(s,t)$, prove its existence, and infer $R(k)$.

Claim 14 (Ramsey's theorem)**.** $\forall$ integers $s \geq 2, t \geq 2$ $\exists$ an integer $R(s,t)$ such that any graph on $n \geq R(s,t)$ vertices has s nodes every two of which are connected or has t nodes every two of which are not connected.

Observe that Claim 14 implies Claim 13 simply by letting $R(k) := R(k,k)$.

Proof: [Proof of Claim 14] We prove this claim by induction. For the induction, consider the claim $Q(w)$ "$\forall$ integers $s \geq 2, t \geq 2$ $s + t = w \exists$ an integer $R(s,t)$ such that any graph on $n \geq R(s,t)$ vertices has s nodes every two of which are connected or has t nodes every two of which are not connected."

Our claim is equivalent to the claim that $\forall w \geq 4, Q(w)$ is true. We prove the latter by induction on w.

- *Base case:* $w = 4$. In this case $s = t = 2$. We can pick $R(s,t) := 2$. This is the trivial statement that in any graph with at least two nodes, there are two nodes connected by an edge or there are two nodes not connected by an edge. Hence $Q(4)$ holds.
- *Induction step:* Assume $w > 4$, and that $Q(w-1)$ is true. Our goal is to show that $Q(w)$ is true. There are several cases.
  - If $s = 2$, then pick $R(s,t) := t$. In any graph with at least $R(s,t)$ nodes, consider the first t nodes. If there is an edge between two of them, $Q(w)$ holds because $s = 2$. Otherwise, there is no edge between two of them, and since there are t such nodes, $Q(w)$ holds.

- If $t = 2$, reason similarly to the case $s = 2$.
- If $s > 2$ and $t > 2$. Pick $R(s,t) := R(s-1,t)+R(s,t-1)+1$. Consider any graph with at least $R(s,t)$ nodes. Let x be the first node. Say x has exactly g neighbors and h non-neighbors. So the number of nodes in this graph is $g + h + 1$, which we know is at least $R(s-1,t) + R(s,t-1) + 1$. This means that $g \geq R(s-1,t)$ or $h \geq R(s,t-1)$. We now show that in either case, $Q(w)$ is true.
  * $g \geq R(s-1,t)$*:* Apply the induction hypothesis to the graph of the neighbors of x. This has $g \geq R(s-1,t)$ nodes. Hence either it has $s-1$ nodes all connected to each other – in which case by adding x we obtain s nodes all connected to each other – or t nodes no two of which are connected. In either case, $Q(w)$ is true.
  * $h \geq R(s,t-1)$*:* Reason similarly to the case $g \geq R(s-1,t)$.

End of the proof.

It is a fascinating open problem to compute the (smallest possible) values of $R(k)$. It is known that $R(3) = 6$ and $R(4) = 18$, and that's about it:

Imagine an alien force, vastly more powerful than us, demanding the value of $R(5)$ *or they will destroy our planet. In that case, we should marshal all our computers and all our mathematicians and attempt to find the value.*
But suppose, instead, that they ask for $R(6)$*.*
Then we should attempt to destroy the aliens.
Paul Erdös.

Exercise 6. Ramsey theorem can be equivalently stated in terms of colors. Let K_n be the graph on n nodes with an edge between any two nodes. Ramsey theorem states that for any integers $s \geq 2, t \geq 2$ there exists a number $R(s,t)$ such that if we color the edges of K_n, where $n \geq R(s,t)$ with two colors Red and Blue, there are either s nodes such that all edges between them are Red, or t nodes such that all edges between them are Blue.

The exercise asks you to prove an extension to three colors: for any integers $s \geq 2, t \geq 2, u \geq 2$ there exists a number $R(s,t,u)$ such that if we color the edges of K_n, where $n \geq R(s,t,u)$ with three colors Red, Blue, and Green, there are either s nodes such that all edges between them are Red, or t nodes such that all edges between them are Blue, or u nodes such that all edges between them are Green.

7 Counting

Not everything that counts can be counted,
not everything that can be counted counts.

Often you argue by counting. For example, if you have a bag with 10 objects, and you are told that at least 6 are *cool* and at most 4 are *green*, then you can deduce that there are at least $6 - 4 = 2$ objects that are cool and not green. This is obvious. Less obvious is the fact that such innocent-looking arguments are the cornerstone of many fundamental results.

A particularly useful type of counting principle is the following:

The pigeonhole principle: If you put s sweaters in d drawers then some drawer must contains at least s/d sweaters. In particular, if $s > d$ then some drawer contains at least 2 sweaters.

Again, the principle is obvious.

In this chapter we present several wide-ranging results proved using counting arguments.

7.1 Geometric-looking problems

Claim 15. Put $n+1$ points on a segment of length 1. Then two must be at distance $\leq 1/n$.

Proof: Think of the segment divided up into n subsegments (or buckets) of length $1/n$. By the pigeonhole principle, two points fall in the same bucket. Hence they are at distance $\leq 1/n$.

End of the proof.

Claim 16. Put $n^2 + 1$ points on a square of side length 1. Then two must be at distance $\leq \sqrt{2}/n$.

Proof: Divide up the square into n^2 subsquares of side length $1/n$. Two points must fall in the same subsquare. Hence they'll be at distance at most the length of the diagonal of the subsquare, which is $\sqrt{2}/n$.

End of the proof.

Proofs often require a bit more ingenuity, like in Exercise 7.

Note: The distance between two points x, y can be written as $|x - y|$. Try this out on the real line!

7.2 Fractions

Consider fractions such as

$$\frac{15}{1}, \frac{50}{2} = 25, \frac{886537}{17452} = 50.7985..., \frac{49}{27} = 1.8148..., \frac{1}{3} = 0.3333...$$

The symbol to denote such numbers is $\mathbb{Q}$.

When given a fraction involving large numbers such as $\frac{886537}{17452}$, it is natural to ask if it can be written using smaller numbers. In particular we would like to reduce the denominators,

since small denominators are easier to deal with. For certain fractions, you cannot reduce the denominator. For example $\frac{57}{2}$ obviously cannot be written using small denominators. But what if we are just interested in *approximating* the fraction? For example, suppose given a fraction α such as $\alpha = \frac{886537}{17452}$ we want to write $\alpha = \frac{p}{q} + \epsilon$ for some small error ϵ that is at most $1/n$. How small can we pick q depending on n?

The following tradeoff is simple.

Claim 17. For every positive fraction $\alpha \in \mathbb{Q}$, for every integer $q > 0$ there is an integer p such that

$$\left|\alpha - \frac{p}{q}\right| \leq \frac{1}{q}.$$

Proof: The proof of this fact is straightforward. Consider the fractions $\frac{0}{q} = 0, \frac{1}{q}, \frac{2}{q}, \ldots$. For some integer i, it must be the case that $\frac{i}{q} \leq \alpha \leq \frac{i+1}{q}$. Hence, we can pick $p := i$. This gives the desired conclusion since

$$\left|\alpha - \frac{i}{q}\right| \leq \left|\frac{i+1}{q} - \frac{i}{q}\right| = \frac{1}{q}.$$

End of the proof.

Can we improve on this claim? The proof looks so trivial that one may think the claim cannot be improved. In fact, it is possible to improve the claim somewhat, by showing that for any n, we can get an error of $1/q \cdot n$ for some denominator $q \leq n$. For the q (which depends on n) given by the next claim, the previous claim could only guarantee error $1/q$. Now we guarantee $1/q \cdot n$. This improvement relies on the pigeonhole principle.

Claim 18 (Dirichlet's approximation theorem)**.** For every fraction $\alpha \in \mathbb{Q}$, for every $n \geq 0$, there is an integer p, and an integer q, $0 < q \leq n$ such that

$$\left|\alpha - \frac{p}{q}\right| \leq \frac{1}{q \cdot n}.$$

Proof: First, we setup some convenient notation. Imagine multiplying α by an integer i, and consider the *decimal part* of the product, denoted by $\mathrm{DP}(i \cdot \alpha)$. For example, if $\alpha = 13/7$ and $i = 3$, then $i \cdot \alpha = 39/7 = 5.5714...$ and thus $\mathrm{DecimalPart}(i \cdot \alpha) = 0.5714....$ Note that, by definition, the decimal part is always between 0 and 1, and any fraction α can always be written as $\alpha = \mathrm{DP}(\alpha) + t$, for some integer t. We will exploit this later.

Now, divide the interval from 0 to 1 into n subintervals $I_1 := [0, 1/n), I_2 := [1/n, 2/n), \ldots$. Thus, I_i is the interval between $(i-1)/n$ and i/n.

Now consider the $n+1$ decimal parts $\mathrm{DP}(0 \cdot \alpha), \mathrm{DP}(1 \cdot \alpha), \ldots \mathrm{DP}(n \cdot \alpha)$.

Since there are $n+1$ decimal parts, but only n intervals, by the pigeonhole principle two decimal parts are in the same interval.

Let us call $i < j$ the two corresponding integers such that $\mathrm{DP}(i \cdot \alpha)$ and $\mathrm{DP}(j \cdot \alpha)$ are in the same interval. This means that there are integers t_i and t_j such that

$$i \cdot \alpha = t_i + \epsilon_i, \tag{1}$$
$$j \cdot \alpha = t_j + \epsilon_j, \text{and} \tag{2}$$
$$|\epsilon_i - \epsilon_j| \leq 1/n. \tag{3}$$

This last equation (3) is the key; it holds because ϵ_i and ϵ_j are in the same interval.

We get by subtracting (1) to (2) that

$$(j-i)\alpha = t_j - t_i + \epsilon_j - \epsilon_i,$$

and so

$$|(j-i)\alpha - (t_j - t_i)| = |\epsilon_j - \epsilon_i| \leq 1/n.$$

We now let $p := (t_j - t_i)$ and $q := (j - i)$. Note that $0 < q \leq n$, because $0 \leq i < j \leq n$. So dividing by $q > 0$ we obtain

$$\left|\alpha - \frac{p}{q}\right| \leq \frac{1}{q \cdot n},$$

as desired.

End of the proof.

7.3 Increasing subsequences (Erdös-Szekeres' theorem)

A sequence is a subsequence of another if can be obtained by dropping some elements. For example $(2, 7)$ is a subsequence of $(1, 2, 3, 4, 5, 6, 7, 8), (2, 13, 7)$, and $(2, 7)$. Formally, a sequence $(a_1, \ldots, a_k)$ is a *subsequence* of a sequence $(b_1, \ldots, b_n)$ if there are increasing indices $i_1 < i_2 < \cdots < i_k$ such that $a_j = b_{i_j}$ for every j.

A sequence $(a_1, \ldots, a_k)$ is increasing (decreasing) if $i < j \Rightarrow a_i < a_j$ $(a_i > a_j)$.

Claim 19. Let $(a_1, \ldots, a_n)$ be a sequence of n distinct integers. Then either there exists an increasing subsequence of length $\geq \sqrt{n}$ or there exists a decreasing subsequence of length $\geq \sqrt{n}$.

The proof of this cool result amounts to a clever definition and the pigeonhole principle.

Proof: For every i define $x_i \geq 1$ as the length of the longest increasing subsequence ending with a_i, and $y_i \geq 1$ as the length of the longest decreasing subsequence starting with a_i.

Note that $(x_i, y_i) \neq (x_j, y_j)$ for $i \neq j$. Indeed, if $a_i < a_j$ then $x_j > x_i$, while if $a_i > a_j$ then $y_i > y_j$.

So each pair (x_i, y_i) can only be assigned to one i. But there are less than n pairs (x_i, y_i) with both $1 \leq x_i < \sqrt{n}$ and $1 \leq y_i < \sqrt{n}$. So some pair with one component larger must be assigned to some i.

End of the proof.

7.4 Hard functions

We now use the pigeonhole principle to prove the existence of functions that are very hard to compute in your favorite programming language, say C.

Consider a function $f : \{0,1\}^n \to \{0,1\}$. You can think of an input x to this function as being an array of n bits, which is given as input to a C program. The C program is supposed to return $f(x)$. We say a C program P computes f if

$$\forall x \in \{0,1\}^n, P(x) = f(x).$$

Some functions have short, compact C programs. All those you ever coded up have reasonable-length C programs. We are going to prove that there exist functions however that require an astronomical amount of C code to be computed. A leading goal of theoretical computer science is to "get our hands" on such functions. This goal has proved elusive. Oversimplifying, the best technique we have available is the following argument, based on the pigeonhole principle, which shows the *existence* of such functions without actually telling us which these functions are.

Claim 20. For every n, there is a function $f : \{0,1\}^n \to \{0,1\}$ which requires at least $2^n/8$ characters in any C code computing it.

Even for n as small as 300, the bound in the claim is an astronomical 2^{297}. This is larger than the estimated number of atoms in the observable universe, which is 10^{80}.[1] Even if you managed to store an ASCII character in an atom, you still couldn't store the C program for some function on 300 bits.

Note this claim is not far from being tight, because any function $f : \{0,1\}^n \to \{0,1\}$ can be computed by a C code of size $O(2^n)$, which has an array of size 2^n with all outputs of the function.

Proof: Each ASCII character is one of at most 256. So the number of C programs of length t is at most the number of strings of length t, which is $(256)^t = 2^{8t}$. Most of these strings won't even be syntactically correct, or will crash when you execute them. We don't care. All we need is an upper bound on the number of programs of length t, and counting strings of length t does it.

Now observe that the number of functions from $\{0,1\}^n$ to $\{0,1\}$ is $2^{(2^n)}$.

Now suppose by contradiction that every such function can be computed by some C program. If $2^{(2^n)} \geq 2^{8t}$ then by the pigeonhole principle two functions must be computed by the same program. This is a contradiction, for each program computes exactly one function. This last statement is clear: two different functions f, f' must differ on some input y: $f(y) \neq f'(y)$. But $P(y)$ is either 0 or 1, so it can't agree with both.

Thus we reach a contradiction as soon as $t < 2^n/8$.

End of the proof.

Of course, we are not using anything specific to the C programming language – the same proof would apply to any other language.

[1] `http://en.wikipedia.org/wiki/Observable_universe`.

Exercise 7. Consider an equilateral triangle of side length 1.

Prove that no matter how you place 5 points inside it, two must be at distance $\leq 1/2$.

Now generalize this to prove that for every integer $k \geq 1$, if you place at least $4^k + 1$ points inside the triangle, two must be at distance $\leq 1/2^k$.

Exercise 8. Prove that in any graph with $2n$ nodes and at least $n^2 + 1$ edges, there is a triangle (a.k.a. a 3-clique).

Hint: Induction + Pigeonhole principle.

Exercise 9. In this exercise you will show the existence of sets that are not regular without using the pumping lemma. Unlike the pumping lemma, the argument you will be using applies to a large number of definitions of "simple" sets. From a pedagogical point of view, this exercise is beneficial because lets you practice the pigeonhole principle and quantifiers and the distinction between finite and infinite.

Say that a set is t-regular if it is obtained as in Definition 10 where the integer k in that definition is constrained to be equal to t.

For $D \subseteq \{0,1\}^*$, say that a function $f : D \to \{0,1\}$ is (t-)regular if there exists a (t-)regular set S such that $\forall x \in D, f(x) = 1 \Leftrightarrow x \in S$.

(1) Prove that the number of t-regular sets is $2^{O(t \lg t)}$.

(2) Use (1) and the pigeonhole principle to prove that for sufficiently large t $\exists f_t : \{0,1\}^t \to \{0,1\}$ that is not t-regular.

(3) use (2) to prove that there exists a set $S \subseteq \{0,1\}^*$ that is not regular.

Exercise 10. Prove that $\{x | x = wtw$ for some $w, t \in \{0,1\}(\{0,1\}^*)\}$ is not regular (this is [S problem 1.46 d]).

Exercise 11. Prove or disprove the claim: There exists a function $f(n) = \omega(n)$, with range the positive integers, such that $\{1^{f(n)} | n > 0$ is an integer$\}$ is regular.

8 The probabilistic method

We saw in the previous section that counting is key. Probability is "just" a convenient language for counting. You can think of probability as a "normalized" way of counting, where you always remember to normalize your quantities with respect to a common measure. Everything you do with probability you can do via direct counting without ever writing the symbol Pr.

8.1 Basics

To start, we have an experiment in mind. For example, tossing a die. The *sample space* (S) is the set of possible outcomes of the experiment. For example, in the tossing a die experiment, the sample space $S = \{1,2,3,4,5,6\}$.

A *probability distribution* is a map from $S \to \mathbb{R}^+$ (so $\forall s \in S$, $\Pr[s] \geq 0$) such that $\sum_{s \in S} Pr[s] = 1$. An *event* is a subset $E \subseteq S$.

For example, the event "the die is even" is given by $E = \{2, 4, 6\}$.

The probability of an event is defined as

$$Pr\left[E\right] = \sum_{s \in E} Pr\left[s\right].$$

Note $\Pr[\neg E] = 1 - \Pr[E]$.

Fact 21 (Union bound)**.** $\forall E_1, E_2,\ Pr\left[E_1 \cup E_2\right] \leq Pr\left[E_1\right] + Pr\left[E_2\right]$. If $E_1 \cap E_2 = \emptyset$, then this is tight equality.

Two events E_1 and E_2 are *independent* if $Pr\left[E_1 \cap E_2\right] = Pr\left[E_1\right] \cdot Pr\left[E_2\right]$. Note this implies $\neg E_1$ and E_2 are independent as well (try this out). To generalize to k events, we ask that *for every* subset of at most k events, the probability that all of them hold is the product of the probabilities that each holds.

The way to think of independent events is that you have an experiment in mind (e.g., tossing a coin) and then another experiment in mind (e.g., tossing a coin) that is *unrelated*. Then events talking about different experiments will be independent.

Example: Toss n coins. The sample space is $\{0,1\}^n$. Since the coins are independent, the probability of getting n heads is

End of the example.

Probabilistic Method Suppose that your task it prove that there is an object x that satisfies property A. One way to do that is to prove that $Pr_x[x \in A] > 0$, since this implies that $\exists \bar{x}$ such that $\bar{x} \in A$. This is the basic form of the probabilistic method.

8.2 Summary

$$\begin{array}{rll} & \text{if } A \Rightarrow B \text{ then} & \Pr[A] \leq \Pr[B] \\ & \Pr[\neg A] & = 1 - \Pr[A] \\ & \Pr[A \bigcup B] & \leq \Pr[A] + \Pr[B] \\ A, B \text{ independent } & \Rightarrow \Pr[A \bigcap B] & = \Pr[A] \cdot \Pr[B] \\ & \Pr_x[A(x) = 1] > 0 & \Rightarrow \exists x : A(x) = 1 \end{array}$$

8.3 Hard functions, again

The bound in Section 7.4 can be proved via the probabilistic method as follows. Pick $f : \{0,1\}^n \to \{0,1\}$ at random. What does this mean? Toss a coin independently for each output. Recall a function it's like a string of length 2^n, so we pick one such string at random. Now suppose $t < 2^n/8$.

Our goal is to show that there is a function such that there does not exist a code P of length t computing it. Following the probabilistic method, we pick f at random and we attempt to show that

$$\Pr_f[\text{NOT } \exists \text{ code } P \text{ of length } t \text{ computing } f] > 0.$$

To prove this bound, we actually consider the negation of the event we are interested in, and prove its probability to be < 1. This is sufficient because

$$\Pr_f[\text{NOT } \exists \text{ code } P \text{ of length } t \text{ computing } f] = 1 - \Pr_f[\exists \text{ code } P \text{ of length } t \text{ computing } f].$$

Here is the derivation:

$$\begin{aligned}
&\Pr_f[\exists \text{ code } P \text{ of length } t \text{ computing } f] \\
&\leq \sum_{\text{code } P \text{ of length } t} \Pr_f[P \text{ computes } f] && \text{(union bound)} \\
&\leq \sum_{\text{code } P \text{ of length } t} 1/2^{2^n} && (1/2^{2^n} \text{ is the probability that } f = P \text{ (on every input)}) \\
&\leq (256)^t/2^{2^n} < 1 && (t < 2^n/8).
\end{aligned}$$

Therefore,

$$\Pr_f[\text{NOT } \exists \text{ code } P \text{ of length } t \text{ computing } f] = 1 - \Pr_f[\exists \text{ code } P \text{ of length } t \text{ computing } f] > 0.$$

In particular, there exists some function f such that

$$\text{NOT } \exists \text{ code } P \text{ of length } t \text{ computing } f,$$

which is what we wanted to prove.

8.4 Lower bound for Ramsey numbers

Claim 22. There exists a constant c such that for all sufficiently large n, there exists a graph on n nodes with no clique or independent set of size $\geq c \lg n =: k$.

Similarly, there exists a constant $\epsilon > 0$ such that for all large enough r we have $R(r) \geq 2^{er}$. Proof:

Proof: Pick a graph G on n nodes at random: toss a coin for each of its $\binom{n}{2}$ edges. Again, we want to show that

$$\Pr_G[G \text{ has } k\text{-clique or } k\text{-independent set}] < 1.$$

We write the probability as follows, then apply a union bound:

$$\begin{aligned}
&\Pr_G\Big[\exists v_1 < v_2 < \ldots, v_k : (\forall i < j \le k : v_i \text{ and } v_j \text{ are connected }) \bigvee \\
&\qquad\qquad (\forall i < j \le k : v_i \text{ and } v_j \text{ are not connected })\Big] \\
\le &\sum_{v_1<v_2<\ldots,v_k} \Pr_G\Big[(\forall i < j \le k : v_i \text{ and } v_j \text{ are connected }) \bigvee (\forall i < j \le k : v_i \text{ and } v_j \text{ are not connected })\Big] \\
\le &\sum_{v_1<v_2<\ldots,v_k} \Pr_G\Big[(\forall i < j \le k : v_i \text{ and } v_j \text{ are connected })] + \Pr[(\forall i < j \le k : v_i \text{ and } v_j \text{ are not connected} \\
= &\sum_{v_1<v_2<\ldots,v_k} \frac{1}{2^{\binom{n}{k}}} + \frac{1}{2^{\binom{n}{k}}} \\
\le &\binom{n}{k} 2 \cdot \frac{1}{2}^{\binom{k}{2}} \\
\le &2^{k\lg(en/k)+1-k(k-1)/2} \qquad (\text{Using } \tbinom{n}{k} \le (en/k)^k)
\end{aligned}$$

which is < 1 as soon as $(k-1)/2 > \lg(en/k)+1$, which is true for $k = c\lg n$ for a sufficiently large c independent of n.

End of the proof.

Once again, the probabilistic argument is nothing but counting. What the proof establishes is that the number of n-node graphs having a "large" clique or an independent set is smaller than the total number of graphs on n nodes. So obviously there exists a graph without "large" cliques or independent sets (cf. example at the beginning of the Counting Section 7).

Proof: [Proof by "direct counting"] Let k be a parameter for the size of the clique or independent set. How many graphs are there with a clique of size k? We need to specify the position of the clique, $\binom{n}{k}$. When we decided that, we need to decide for $\binom{n}{2} - \binom{k}{2}$ edges, where note we subtracted the edges we already know are in the clique.

Doubling the quantity to count for both independent sets and cliques, we get

$$2\binom{n}{k}2^{\binom{n}{2}-\binom{k}{2}}.$$

Whenever this quantity is less than the total number of graphs, which is

$$2^{\binom{n}{2}},$$

we have the desired conclusion by the pigeonhole principle.

You can verify this indeed happens for k as small as $k = O(\lg n)$. This concludes the proof.

End of the proof.

The proof by the probabilistic method exactly parallels the above proof: in the probabilistic method, all quantities are normalized by $2^{\binom{n}{2}}$. So you can think of the probabilistic method as "normalized quantity." Basically you set things so that the "number of all graphs is 1," and prove that the "bad" graphs are "less than 1 in number." Also, the counting in the above "direct-counting" proof is very loose. We count more than once many graphs. For example the complete graphs is counted multiple times. This overcounting corresponds to the slackness of the union bound.

There's yet another language in which this proof can be cast. The type goes under the name of *incompressibility*.

8.5 Random variables, expectation, variance, and all that

The real power of the probabilistic method comes with using random variables. Random variables are nothing but a useful way to define events. Formally, a *random variable* is a mappings $X : S \to \mathbb{R}$. If X is a random variable, we have the expected value of X: $E[X] := \sum_{a \in \mathbb{R}} a \cdot Pr[X = a]$.

If X is a random variable, it makes sense to define the random variable X^2, $X - 17$, and so on. You can think of X^2 as: Just perform your experiment, get your outcome, obtain X, and then square.

- $\mu = E[X]$, $\sigma^2 = E[(X - \mu)^2] = E[X^2] - E^2[X]$.
- Linearity of expectation: $E[\sum_i c_i X_i] = \sum_i c_i E[X_i]$.

Note: If X takes values in $0, 1$, then $E[X] = \Pr[X = 1]$.

Example: You toss n coins. X_i is the $0/1$ outcome of the i-th coin. $S := \sum_{i \leq n} X_i$.

$$\begin{aligned}
\mu = E[S] &= E[\sum_{i\le n} X_i] \\
&= \sum_{i\le n} E[X_i] \qquad \text{(Linearity of expectation)} \\
&= \sum_{i\le n} \Pr[X_i = 1] \qquad \text{(Because } X \in \{0,1\}\text{)} \\
&= \sum_{i\le n} \frac{1}{2} = n/2.
\end{aligned}$$

$$\begin{aligned}
E[S^2] &= E\left[\sum_i x_i^2 + 2\sum_{i<j} x_i x_j\right] \\
&= \frac{n}{2} + 2\binom{n}{2}\frac{1}{4} \\
&= \frac{n^2}{4} + \frac{n}{4}.
\end{aligned}$$

$$\begin{aligned}
\sigma^2 = E[S^2] - E[S]^2 &= \frac{n^2}{4} + \frac{n}{4} - \left(\frac{n}{2}\right)^2 \\
&= \frac{n}{4}.
\end{aligned}$$

End of the example.

Independence Two random variables X, Y are *independent* if for every a, b, $\Pr[X = a \wedge Y = b] = \Pr[X = a] \cdot \Pr[Y = b]$. Independent variables satisfy

$$E[X \cdot Y] = E[X] \cdot E[Y].$$

8.6 Concentration of measure

Concentration of measure is the general phenomenon that random variables often are close to their expectation. That is, with high probability $|X - E[X]| \le \epsilon$ for a certain ϵ. Note that $E[X]$ is just a number.

Markov's inequality

Claim 23. Let X be a non-negative random variable with mean μ. Then for every $k > 0$:

$$\Pr[X \ge k\mu] \le \frac{1}{k}.$$

Equivalently,

$$\Pr[X \ge t] \le \frac{\mu}{t}.$$

Proof:

$$\begin{aligned}
\mu &= \sum_a a \Pr[X = a] \\
&= \sum_{a \geq t} a \Pr[X = a] + \sum_{a < t} a \Pr[X = a] \\
&\geq \sum_{a \geq t} a \Pr[X = a] \\
&\geq t \left(\sum_{a \geq t} \Pr[X = a] \right) \\
&= t \Pr[X \geq t].
\end{aligned}$$

End of the proof.

Example: You toss n independent 0/1 coins. The probability of getting $\geq (3/4)n$ heads is the probability of deviating by a factor $2 \cdot 3/4 = 3/2$ from the expectation, hance it is at most $2/3$.

You have just proved this scary-looking inequality:

$$\frac{\sum_{i \leq (3/4)n} \binom{n}{i}}{2^n} \leq \frac{2}{3}.$$

Doesn't it look easier using probability?

End of the example.

Example: Reproving the first version of Markov's inequality we proved earlier. If you have n numbers $a_1, a_2, \ldots, a_n$, positive, that sum to at most t, then at most $t/2$ of them are bigger than 2.

To prove, consider the random variable X that's uniformly distributed over $\{a_1, a_2, \ldots, a_n\}$, i.e., $\Pr[X = a_i] = 1/n$ for every i. Note $E[X] = t/n$. Hence $\Pr[X \geq 2] = \frac{|\{a_i | a_i \geq 2\}|}{n} \leq \frac{t}{2n}$. Hence

$$|\{a_i | a_i \geq 2\}| \leq \frac{t}{2}.$$

End of the example.

Chebychev's inequality

Claim 24. Let X be a random variable with mean μ and standard deviation σ. Then for every $k > 0$:

$$\Pr[X \geq \mu + k\sigma] \leq \frac{1}{k^2}.$$

Proof: Set $Y := X - \mu$. Since $k\sigma > 0$, $\Pr[Y \geq k\sigma] \leq \Pr[Y^2 \geq k^2\sigma^2]$. By Markov's inequality, $\Pr[Y \geq k\sigma] \leq \frac{E[Y^2]}{k^2\sigma^2} = \frac{1}{k^2}$.

End of the proof.

Example: You toss n independent 0/1 coins. The probability of getting $\geq (3/4)n$ heads is the probability of deviating by $\Omega(\sqrt{n})$ standard deviations from the expectation, and so it is $O(1/n)$.

Again, you have just improved the scary-looking inequality above to:

$$\frac{\sum_{i \leq (3/4)n} \binom{n}{i}}{2^n} \leq O(1/n).$$

Doesn't it look easier using probability?

End of the example.

Chernoff bound

Claim 25. Let $X_1, X_2, \ldots, X_n$ be n independent random variables taking values in $\{0, 1\}$. Suppose $E[X_i] = \Pr[X_i = 1] = p$ for every i. Then for every $0 \leq \epsilon \leq 1-p$ (so that $p+\epsilon \leq 1$):

$$\Pr[\sum_i X_i \geq pn + \epsilon n] \leq 1/2^{\epsilon^2 n}.$$

Unlike for Markov and Chebychev's inequality, for Chernoff's bound there is a vast number of inequalities providing tight bounds under various conditions. The above is just a particularly useful type.

Example: You toss n independent 0/1 coins. The probability of getting $\geq (3/4)n$ heads can be bounded from above using Chernoff's bound with $p = 1/2$ and $\epsilon = 1/4$. Thus we get the bound $1/2^{\Omega(n)}$ – an exponential bound which wipes out the weaker bounds obtained using Markov and Chebychev's inequalities.

End of the example.

Cantelli's one-sided version of Chebychev's inequality Chebychev's inequality is useful to bound deviation by k times the standard deviation, when $k > 1$. It does not give anything for $k \in [0, 1]$. The next inequality proves a bound in this regime, which is often useful.

Claim 26. Let X be a r.v. with mean μ and standard deviation σ. Then for any $k > 0$:

$$\Pr[X \geq \mu + k\sigma] \leq \frac{1}{1+k^2}, \tag{4}$$

$$\Pr[X \leq \mu - k\sigma] \leq \frac{1}{1+k^2}. \tag{5}$$

Proof: First we prove (1). With the benefit of hindsight, define $t := -\mu + \sigma/k$. One can derive this choice for t by minimizing a certain expression, but I don't find remembering the correct choice of t less painful than remembering the rest.

We now have:

$$\begin{aligned}
\Pr[X \geq \mu + k\sigma] &= \Pr[X + t \geq \mu + k\sigma + t] \\
&= \Pr[X + t \geq \sigma(k + 1/k)] \\
&\leq \Pr[(X + t)^2 \geq \sigma^2(k + 1/k)^2] \qquad \text{(because } \sigma(k + 1/k) \geq 0\text{)} \\
&\leq \frac{E[(X + t)^2]}{\sigma^2(k + 1/k)^2} \qquad \text{(by Markov's inequality).}
\end{aligned}$$

We now compute $E[(X + t)^2]$. First, note that

$$(X + t)^2 = X^2 + t(t + 2X).$$

So $E[(X+t)^2] = E[X^2] + (-\mu+\sigma/k)(-\mu+\sigma/k+2\mu) = E[X^2] + \sigma^2/k^2 - \mu^2 = \sigma^2(1+1/k^2)$.

Hence we get

$$\Pr[X \geq \mu + k\sigma] \leq \frac{\sigma^2(1 + 1/k^2)}{\sigma^2(k + 1/k)^2} = \frac{1}{1 + k^2}.$$

The proof of (2) is symmetrical. Specifically, now choose $t := -\mu - \sigma/k$. Going through the same derivation, this time using that $\mu - k\sigma + t < 0$, you again obtain the bound

$$\Pr[X \leq \mu - k\sigma] \leq \frac{E[(X + t)^2]}{\sigma^2(k + 1/k)^2}.$$

You also again obtain that $E[(X+t)^2] = \sigma^2(1+1/k^2)$. So the proof is concluded as before.

End of the proof.

Exercise 12. In this exercise you will use concentration inequalities to prove anti-concentration inequalities, that is, the result that a certain random variable does deviate from its expectation.

It is convenient to think of the experiment of tossing n $\{-1, 1\}$ coins (instead of $\{0, 1\}$ coins). Let S be the sum. Prove:

(1) $E[S] = 0$.
(2) $E[S^2] = n$.
(3) $E[S^4] = 3n^2 - 2n$. (This takes some patience.)
(4) The standard deviation of S^2 is at most $\sqrt{2} \cdot n$.

Now the idea is that if it was the case that S is often very close to its expectation 0, then also S^2 would be often very close to 0. But since S^2 has a much larger expectation (n) and a standard deviations about the expectation ($\sqrt{2} \cdot n$), this would violate a concentration inequality. Specifically:

(5) Use Cantelli's one-sided Chebychev's inequality to show this anti-concentration inequality:

$$\Pr[|S| \geq \sqrt{n}/2] \geq \Omega(1).$$

Finally, answer this question:

(6) Could you have used Chebychev's inequality to infer (5) from (1)-(4)?

Note: Similar bounds can be obtained by approximations of the binomial coefficients such as $\binom{n}{n/2} = \Theta(2^n/\sqrt{n})$ for n even. One benefit of the proof in this exercise is that it applies in other contexts as well. For example, if v is vector in $\mathbb{R}$ of length 1, with probability $\Omega(1)$ you have that $|\sum_i x_i v_i| \geq 1/2$.

8.7 Error-correcting codes

You can send a string of n bits, but an adversary can change at most $n/6$ locations arbitrarily. How many different messages can you send so that the receiver will be able to decode?

The question is, how big C a subset of $\{0,1\}^n$ can you construct so that every two strings there are at hamming distance $> n/3$? This is because then the received is guaranteed that there is at most one string at distance $\leq n/6$, for if there were two you would have two strings in C at distance $\leq n/3$.

You can have $C = \{1^n, 0^n\}$.

With the chernoff bound we can construct C of size $2^{\Omega(n)}$.

Claim 27. There are sets $C \subseteq \{0,1\}^n$ of size $2^{\Omega(n)}$ such that any two strings in C have hamming distance $\geq n/3$.

Proof: Let C consist of k strings of n bits independently at random.

$$\begin{aligned} &\Pr[\exists x, y \in C, x \neq y \text{ at distance } < n/3\ ]] \\ \leq &k^2 \Pr[x, y \text{ at distance } < n/3] \qquad \text{(Union bound)}. \end{aligned}$$

To bound $\Pr[x, y \text{ at distance } < n/3]$, define z_i to be a $\{0,1\}$ random variable that is 1 if and only if $x_i = y_i$. Note $\Pr[z_i = 1] = 1/2$. Now the probability $\Pr[x, y \text{ at distance } < n/3]$ is the probability that the sum of the variables z_i is at least $2n/3$. By a chernoff bound, this probability is at most $1/2^{\Omega(n)}$.

Hence in the above derivation we can set $k = 2^{\Omega(n)}$ and get the probability to be < 1, which guarantees the existence of such a set.

End of the proof.

Exercise 13. Recall Section 8.7. It is possible to show that $n/6$ cannot be changed to $0.49n$, when sending strings of *bits*. However, suppose that instead of sending a string of n bits you send a string of n symbols in $[t] = \{1, 2, \ldots, t\}$. Suppose that the adversary changes at most $0.49n$ symbols. Show that $\exists t$ such that for sufficiently large n you can still send $2^{\Omega(n)}$ messages.

Exercise 14. In this exercise you will see another example where probabilistic reasoning greatly simplifies counting. Suppose you toss n biased coins. The coins are independent, and each comes up heads with probability p.

What is the probability that you get an even number of heads?

Hint: Think of the outcome of each toss as a number in $\{-1, 1\}$, and use $E[X \cdot Y] = E[X] \cdot E[Y]$ for independent random variables.

9 Conditional probability

Conditional Probability: if E_1 and E_2 are events (and $Pr[E_2] > 0$), then the probability of E_1 conditional to E_2 is given by

$$\Pr[E_1|E_2] = \frac{\Pr[E_1 \cap E_2]}{\Pr[E_2]}.$$

Note if E_1 and E_2 are independent, then $\Pr[E_1|E_2] = \Pr[E_1] \cdot \Pr[E_2]$.

We have, for events A_i with non-zero probability:

$$\Pr[E] = \sum_i \Pr[E|A_i] \cdot \Pr[A_i].$$

For example, in the die tossing experiment, if $E_1 = \{2\}$ and $E_2 = \{2, 4, 6\}$, then $Pr[E_1|E_2] = \frac{1/6}{3/6} = \frac{1}{3}$.

For random variable X and event A, $X|A$ means X conditioned to event A. We again require that $Pr[A] > 0$.

$$Pr[(X|A) = a] = Pr[X = a|A] = \frac{Pr[X = a \cap A]}{Pr[A]}$$

Similarly, $E[X|A]$ is the expectation of random variable X conditioned to event A.

Conditional probability is a difficult notion whose mastery requires long experience and is sometimes elusive.

Claim 28. Let X, Y be independent random variables, and f a function $f : X, Y \to \mathbb{R}$. If $E_{X,Y}[f(X, Y)] \geq \epsilon$, then $\exists$ fixed $\bar{x}$ such that $E_Y[f(\bar{x}, Y)] \geq \epsilon$.

Proof: Suppose for the sake of contradiction that $\forall \bar{x}$, $E_Y[f(\bar{x}, y)] < \epsilon$. Then,

$$E_{X,Y}[f(X, Y)] = \sum_{\bar{x}} E_{X,Y}[f(X, Y)|X = \bar{x}] \cdot Pr[X = \bar{x}].$$

But X and Y are independent, so this is equal to

$$\sum_{\bar{x}} E_Y[f(\bar{x}, Y)] \cdot Pr[X = \bar{x}] < \epsilon \cdot \sum_{\bar{x}} Pr[X = \bar{x}] = \epsilon$$

This contradicts our hypothesis on f in the claim.

End of the proof.

A Summary of equivalences

connectives, quantifiers, sets, pigeonhole, counting, prob., strings, sequences,

B Examples of bad proofs

Proof: [Example of what is not a proof]
Here is the algorithm

$$\begin{aligned}
\delta(a,0) &= (c,1,R)\\
\delta(a,1) &= (b,2,L)\\
\delta(b,0) &= (a,\#,R)\\
\delta(b,1) &= (b,1,L)\\
\delta(c,0) &= (d,1,L)\\
\delta(c,1) &= (c,2,R)\\
\delta(d,0) &= (b,1,L)\\
\delta(d,1) &= (c,2,R).
\end{aligned}$$

End of the proof.

What is wrong with the above is that it does not explain anything! *Why* is the algorithm supposed to be correct? What are the ideas behind it? If the author of the "proof" cannot explain, chances are things are not clear for them either.

Proofs should have an *explanation in English* of *why* things are true. We are allowed to say that things like $2+2=4$ are obvious. But the correctness of the above algorithm is not obvious, and must be explained.