CMSC-27100 Section 1 (Discrete Math)

Prof. Bill Fefferman

0. Administration/Introduction to the class

About me

- Prof. Bill Fefferman
- I'm a theoretical computer scientist, specializing in the theory of quantum computation
- I too did my undergrad. at UChicago!

Teaching assistants

- Antares Chen (antaresc@uchicago.edu)
- Yuyuan Chen (yuyuanchen0@uchicago.edu)
- The T.As are your first line of communication about all issues concerning grades and late homework. **Please contact me only after you have talked with them first.**
- We also have graders who will help in grading the homework.

Textbook/Syllabus

- Officially our textbook is K. Rosen: Discrete Mathematics and Its Applications, 7th ed.
- This will be useful as a reference, but we won’t adhere to it very closely!
- We will adhere much more closely to the lecture notes of S. Kurtz, (available online, see syllabus) which in turn are derived from Rosen and the lectures of L. Babai (available online)
- Syllabus should be available now on canvas.uchicago.edu.
- Just in case, we’ll review it now!

A note about the format of the class

- I will try to present (at least most) lectures using Powerpoint slides.
- I will circulate the slides at the end of week, after class Friday.
  - They are based on a combination of Prof. Kurtz's and Prof. Babai's notes.
- Not after every class – this is to encourage you to come to class -- so please take notes on the slides!
- Also, I hope that for each lecture you will read the reference I mention at the beginning. This will serve as review for later.

On class participation

- Traditional math classes are not very interactive
- I strongly believe this is not the most effective way to learn math
- My goal is to make this class a *dialogue* – all questions are welcome!
- I will never call on a student who doesn’t raise their hand…

Grading

- Weekly homework assignments: 25% of total grade
  - Assigned Monday, due next Monday at the beginning of class
  - First homework assigned next Monday.
  - Allow for one late homework (due at the beginning of following Friday class) – no need to write us about this – Post on gradescope or send to TA by email
  - LaTeX not necessary but recommended
  - Collaboration is fine but please write who you worked with on the first page!
  - Homeworks will be assigned and handed in through Gradescope only (please make sure you can login properly! If it doesn't work please email TAs
- Midterm (in person) (timing TBA, probably week 5 or 6): 25%
- Participation: 5%
- Final (in person): 45%

Office hours

- My office: Crerar 251
- Office hours: Friday at 10:30-11:30 AM, *starts next week*
- Or by appointment (email me, **always include [27100] in the subject line**)
- T.A. office hours: See discussion sections

Discrete Math discussion sections

- Each student should already have an assigned lab time (my hope: no need to stay with this assignment, subject to capacity constraints)
- The labs will be run by the T.A's and double as T.A office hours.
- The purpose is to review concepts from class and do practice problems!
- **While not mandatory and attendance will not be taken**, I highly recommend attending the labs.
- If you need to change your section, write to Jessica Garza (jdgarza@cs.uchicago.edu) – please do not write to me (I can't do anything about it myself)

Discussion sections (2/2)

- Timing:
  - Wed: 03:30 PM-04:20 PM (in Ryerson 255)
  - Thu: 04:00 PM-04:50 PM (in Rosenwald Hall 329)
- Starts next week!
- These will all follow the material covered in this section (Sections 1). Probably best not to attend other section discussions – these will not follow our class!

Collaboration, etc…

- Although collaboration on homework is permitted, each student should write their own solutions. If you do collaborate please list all collaborators at the beginning of the assignment.
- **If you need special accommodations, please contact me ASAP.**
- **See detailed late homework policy in the syllabus – roughly, one late homework is allowed per student, due following Friday at beginning of class. This late homework should be emailed to the TAs *and* submitted to Gradescope! After that, the policy becomes more rigid.**

Not registered?

- My section is currently filled.
- However if you want to join please write Jessica Garza and she’ll add you to a waitlist

Should I take this class?

So what do we study in “discrete math”?

- It’s the first theoretical computer science in the CS major!
- Themes of the course (and very rough course outline)
    - First week: introduction to logic and set theory
    - Second/Third/Fourth week: Introduction to number theory/Induction
    - Fifth week: Basic combinatorics and counting
    - Sixth/Seventh week: Discrete probability
    - Eighth week: Recurrences/generating functions
    - Ninth week (subject to timing): Graphs, trees, basics of algorithms and complexity theory

Questions?

1. Basic logic [see also, Kurtz, lecture 1]

Proposition

- A **proposition** is a statement that is either true or false.
- Ex. “Washington D.C is the capital of the U.S” is true
- Whereas “Chicago is capital is the U.S” is false
- A proposition is not a question and doesn’t have variables e.g., $x \leq 5$ isn’t one

Complex propositions

- Can join propositions together to make new propositions
- Example: “Prof. Fefferman lives in Illinois ∧ Prof. Fefferman teaches Discrete Math”
  - Where ∧ is logical AND
  - i.e., this is a true statement since both the propositions to AND are true.
- Can also consider the negation of propositions to get a new proposition that is true iff the original proposition is false
- Ex. “Prof. Fefferman lives in Illinois” is true, while the negation, “¬(Prof. Fefferman lives in Illinois)” is false

Truth tables

- We can completely describe the operator like AND with a ”truth table”, which tells us the output value of *all possible* inputs

p	q	$p \wedge q$
True	True	True
True	False	False
False	True	False
False	False	False

- **Question:** these tables describe Boolean functions with two Boolean inputs (p and q). How many values of a general Boolean f which maps n Boolean inputs $x_1, x_2, \dots, x_n \in \{true, false\}$ to $\{true, false\}$?
- What about if we generalize to n k-ary inputs, rather than Boolean inputs (i.e., how many values of a map from $\{1,2,\dots,k\}^n$ to $\{1,\dots,\ell\}$)?
    - k^n
- At most how many bits to describe such a function?
    - $k^n \log_2 \ell$

More operators

- Can also consider OR operator, also known as $\vee$

p	q	$\boldsymbol{p \vee q}$
True	True	True
True	False	True
False	True	True
False	False	False

- Notice that True $\vee$ True = True
- Exclusive OR, also known as $\oplus$ is the same operator except True $\vee$ True = False
- For example, what is “Prof. Fefferman lives in Illinois” $\vee$ “Prof. Fefferman teaches Discrete Math”?
- What is “Prof. Fefferman lives in Illinois”$\oplus$“Prof. Fefferman teaches Discrete Math”?

Implication

- Two propositions p and q can be joined into the conditional “if p, then q”.
- We can express the statement equivalently using the implies operator, denoted “⇒”
- Here’s the truth table:
- Note that (False ⇒ Anything) is True!
  - Ex. False ⇒False is true, as is False ⇒ True
- Ex: “If Bush is the President then Pigs fly” is true!

p	q	$p \Rightarrow q$
True	True	True
True	False	False
False	True	True
False	False	True

Why does implication work this way?

- First it's just a convention (i.e., a definition)
- But it's sometimes convenient.
- E.g., let's consider $(x < 5) \Rightarrow (x < 10)$
  - Is this a proposition?
- This is always true (for any fixed value of x).
- Some examples:
  - TRUE implies TRUE is TRUE
    - E.g, take $x = 3$ then LHS is true, so is the right hand side
  - FALSE implies TRUE is TRUE
    - E.g., take $x = 7$, we have if we assume $(7 < 5)$ then clearly $(7 < 10)$ should also hold.
  - FALSE implies FALSE is TRUE
    - E.g., take $x = 100$, we have if we assume $(100 < 5)$ then clearly $(100 < 10)$.
- But recall TRUE implies FALSE is FALSE – because shouldn't be able to derive something false from something true (this is intuitive)

Announcements

- Homework 1 assigned on Gradescope, due next Monday!
- Office hours this Friday right after class!
- Discussion sections start this week
- I'll send out most current slides this Friday

Converse and contrapositive

- The “**converse**” of $p \Rightarrow q$ is $q \Rightarrow p$
- These statements are not equivalent! Why not?
  - Notice that if p is True and q is False, then $p \Rightarrow q$ is False, but $q \Rightarrow p$ is True!
- However the “contrapositive” of $p \Rightarrow q$ is $\neg q \Rightarrow \neg p$ is equivalent to the original statement!
- Ex: Let’s suppose “If it’s winter then it’s cold” is true, but:
  - It does not imply the converse (“if it’s cold then it’s winter”)
  - But it does imply the contrapositive (“if it’s not cold then it’s not winter”)
- How do we go about formally proving that these statements are equivalent?

De Morgan’s Laws

- De Morgan’s laws regard the negation of the OR, or AND of two propositions
- First: $\neg(p \vee q) = (\neg p \wedge \neg q)$
  - In English: “it is false that either of p or q is true” = “p is not true and q is not true”
  - Where equality here is “logical equivalence” and means the truth table of both propositions are equivalent
- Second: $\neg(p \wedge q) = \neg p \vee \neg q$
  - In English: “it is false that p and q is true” = “p is not true or q is not true”
- How to prove these?
  - We can prove these by checking the truth tables of the LHS and the RHS propositions!

Predicate logic and variables

- A **predicate** is a statement that is true or false if you plug in values for it's variables – i.e., the truth of the statement is a function of the variables' value
- E.g., $\mathrm{P(x)} =$ "$x^3 > 8$" is true if $x = 3$ but false if $x = 1$
- Now we can also add quantification to the variables in the predicate:
- There are two commonly used quantifiers: there exist ("∃") and for all ("∀")
    - Regarding the example above:
        - "$\exists x, P(x) = True$" is True, since as above, $P(x)$ is true for $x = 3$, for example
        - "$\forall x, P(x) = True$" is False, since as above, $P(x)$ is false for $x = 1$, for example

Negating statements with quantifiers

- How do we negate statements with quantifiers?
- Suppose $Q(x) = (\forall x, R(x))$
- Then we can readily verify that $\neg Q(x) = \exists x, \neg R(x)$
  - i.e., if a statement is not universally true, there must exist a counterexample!
- And likewise, if $Q(x) = (\exists x, R(x))$ then $\neg Q(x) = \big(\forall x, \neg R(x)\big)$
  - i.e., if it’s not the case that there exists a setting of x that makes R(x) true, then it must be the case that R(x) is always false!
- Notice that’s this is very much like De Morgan’s laws – we can think of the quantifiers ∃, ∀ like the operators ∨,∧ respectively

2. Basic set theory [see also, Kurtz Lecture 2]

Basics

- Def. A set is an unordered collection of objects, called the members or elements of the set.
    - If A is a set and a is a member of that set, we write $a \in A$
    - Likewise we write $a \notin A$ if a is not a member
- Def. The set which contains no elements is called the empty set, denoted by ∅
- Def. Two sets A and B are equal if and only if they have the same elements, i.e., $A = B \Leftrightarrow \forall x, x \in A \Leftrightarrow x \in B$
- Notice that sets are unordered, so e.g., {1,2,3,4}={4,3,1,2}

More basics

- Def. a set A is a subset of a set B if all elements of A are also elements of B. Using notation, we say $A \subseteq B$
- A is a proper subset of B or equivalently $A \subset B$ if:
  - A is a subset of B
  - A≠B
- Def. The cardinality of a set A, denoted |A|, is the number of elements A contains.
  - For finite sets, it's just the number of elements.
  - For infinite sets we say it's $\aleph_0$ if it's the same cardinality as the natural numbers…
- Def. The power set of a set A, denoted P(A) is the set of all of its subsets
- What is the cardinality of P(A), if A has n elements?
  - 2^n

Set operations

- Def. If A and B are sets, then the union of A and B, denoted $A \cup B = \{x | x \in A \vee x \in B\}$
- Def. The intersection of sets A and B, denoted $A \cap B = \{x | x \in A \wedge x \in B\}$
- Many times we'll be interested in the cardinality of a set created by applying set operations to other sets.
- First (easy) observation: If two sets A, B are disjoint, then $|A \cup B| = |A| + |B|$

Example

- If A=$\{1,2,3,4,5,6\}$, B= {1,2,3}
- What is $A \cup B$?
  - It's $A = \{1,2,3,4,5,6\}$! Since $B \subseteq A$
- What is $A \cap B$?
  - It's $B = \{1,2,3\}$
- Notice that $|A \cup B| = |A| \leq |A| + |B|$
- But notice if instead $A = \{4,5,6\}$, and so A and B are disjoint, then $|A \cup B| = |A| + |B| = 6$

Cartesian products of set

- How can we combine sets to get another set?
- Def. If A, B are two sets, their Cartesian product $(A \times B)$ contains all ordered pairs (x, y) where $x \in A, y \in B$
- Ex. If $A = \{a, b\}$ and $B = \{1,2\}$
  - then $A \times B = \{(a, 1), (a, 2), (b, 1), (b, 2)\}$
  - notice that $B \times A = \{(1, a), (2, a), (1, b), (2, b)\}$
  - These new sets aren’t equal – order matters for cartesian products!
- Notice, however, that $|A \times B| = |B \times A| = |A| \cdot |B| = 4$

A claim about Cartesian products

- **Claim:** If A, B, C are any sets so that $A \times B \subseteq A \times C$ and $A \neq \emptyset$, then $B \subseteq C$
- First, an observation: Why is this false if $A = \emptyset$?
  - Let $B = \{1,2,3\}, C = \{a, b\}$. Notice that $A \times B = \emptyset$ and $A \times C = \emptyset$
  - So $A \times B \subseteq A \times C$
  - But B is not a subset of C!
- **Pf.**
  - Let's choose any element $b \in B$, we need to show $b \in C$
  - Let's take an arbitrary $a \in A$ and consider $(a, b) \in A \times B$ (since A is non-empty by assumption)
  - Now due to the premise, we know that (a,b) $\in A \times C$
  - But now, by definition of Cartesian product, $b \in C$

Set theory is similar to logic

- There are many similarities between set theory and logic
- In some sense ∪ is like OR, and ∩ is like AND
- We can also consider a negation “analogue” called complement
- Let’s suppose that we consider a ”universe” U which is some set that contains all the objects we’ll be considering (e.g., U could be set of all integers)
- Then if we consider a subset $A \subseteq U$, we can consider the complement, $\bar{A} \subseteq U$ to be all elements in U outside A.
- Note that by definition, $A \cup \bar{A} = U$

Other similarities!

- Analogously to De Morgan's laws, not hard to verify:
- $\overline{A \cup B} = \bar{A} \cap \bar{B}$
- What set would be analogous to False?
- What set would be analogous to True?
- i.e., because $\emptyset \cup U = U$ (like F OR T = T) but $\emptyset \cap U = \emptyset$ (like F AND T = F)
- However, there's no obvious equivalent to implication in set theory

Another example, motivating inclusion/exclusion

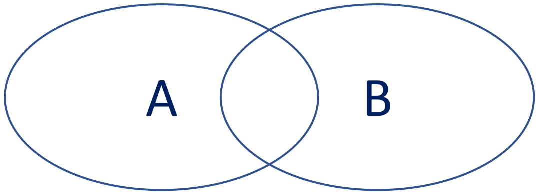
- Def. For two sets A,B the notation A\B means all elements in A not in B
- Here’s a non-obvious fact:
- **Claim:** For any set A,B, if $(A\setminus \mathrm{B}) \cup (B\setminus \mathrm{A}) = A \cup B$ then $A \cap B = \emptyset$
- Pf.
  - Let’s prove the contrapositive that is to say…
    - Let’s assume $A \cap B \neq \emptyset$ we’ll show that in this case $(A\setminus \mathrm{B}) \cup (B\setminus \mathrm{A}) \neq A \cup B$
  - So consider an element an $x \in A \cap B$, which can do since it’s nonempty by assumption
  - Now $x \in A \cup B$ (the RHS of inequality)
  - However, since $x \in B$ $\mathrm{we\ know\ that\ x} \notin A\setminus B$, and similarly since $x \in A, x \notin B\setminus \mathrm{A}$
  - So $x \notin (A\setminus \mathrm{B}) \cup (B\setminus \mathrm{A})$ which is the LHS of the inequality

Announcements

- Slides will be circulated tonight via Canvas
- Homework due Monday on Gradescope before start of class
- Office hours after class, Crerar 251

Basic inclusion/exclusion

- Thm. (basic Incl./Excl.) For all finite sets A and B, $|A \cup B| = |A| + |B| - |A \cap B|$
- Pf. Notice that if we consider the RHS, |A|+|B| we have overcounted– we've counted every element in the intersection, $|A \cap B|$, twice, so we subtract it.



- Def. Two sets A and B are disjoint if $A \cap B = \emptyset$
- Notice if A and B are disjoint, then $|A \cup B| = |A| + |B|$

Alternative, proof of Thm.

- Lemma: For all sets A and B, $A = (A\backslash \mathrm{B}) \cup (A \cap B)$, notice this implies $|A| = |A\backslash \mathrm{B}| + |A \cap B|$
- Pf. We’re taking out from A everything in B and then adding these elements back in.
- Thm. (basic Incl./Excl.) For all finite sets A and B, $|A \cup B| = |A| + |B| - |A \cap B|$
- Pf.
  - $|A \cup B| = |A\backslash \mathrm{B}| + |B\backslash \mathrm{A}| + |A \cap B|$ (because these sets partition $A \cup B$)
  - =$|A\backslash \mathrm{B}| + |B\backslash \mathrm{A}| + |A \cap B| + |A \cap B| - |A \cap B|$ (note I did nothing here!)
  - =$(|A\backslash \mathrm{B}| + |A \cap B|) + (|B\backslash \mathrm{A}| + |A \cap B|) - |A \cap B|$ -- (by reordering terms in sum)
  - = $|A| + |B| - |A \cap B|$ (by the above Lemma!)

Functions

- Functions are one of the most *central concepts* in all of math
- Intuitively a function f maps every element of a set A to an element of B
- Formally, we have definitions:
- Def. A relation R with domain A and co-domain B is a subset of $A \times B$
- Def. A relation $\mathrm{R} \subseteq A \times B$ is total if $\forall a \in A, \exists b \in B\ so\ that\ (a, b) \in R$
- Def. A relation R is single-valued if $\forall a \in A, \forall b_1, b_2 \in B, (a, b_1) \in R \wedge (a, b_2) \in R \Rightarrow b_1 = b_2$
  - I.e., every element a only has one “match” in B
- Def. A function $f: A \rightarrow B$ is a total, single-valued relation with domain A and co-domain B

Properties of functions

- Def. A function $f: A \to B$ is one-to-one (aka injective) if $\forall x, y \in A, f(x) = f(y) \Rightarrow x = y$
- Ex. $f: \mathbb{N} \to \mathbb{N}$ given by $f(x) = x$ is injective but $f(x) = 2$ isn't (since all values of x map to 2)
- Def. A function $f: A \to B$ is onto (aka surjective) if $\forall b \in B, \exists a \in A\ f(a) = b$.
- ?: Is the squaring function $g(x) = x^2$ onto if g: $\mathbb{N} \to \mathbb{N}$?
  - *No!* (e.g., there's no natural number x so that g(x)=2, since $\sqrt{2}$ isn't a natural number)
  - But it would be if $g: \mathbb{R}_+ \to \mathbb{R}_+$

Number theory

a. Divisibility

Definitions and properties

- **First, unless otherwise noted we'll always work over the non-negative integers!**
- **Def.** We say "a divides b", or $a|b \Leftrightarrow \exists d, a \cdot d = b$
  - Also "a is a divisor of b" or "b is a multiple of a"
  - Notice this means, by def., we have $\forall a, a|0$
  - Because $\forall a, a \cdot 0 = 0$
- Note, if a, b are non-negative and $a|b$ and $b < a$ then…
  - b=0
- **Thm:** $\forall\ a, b_1, b_2, a|b_1 \wedge a|b_2 \Rightarrow a|(b_1 + b_2)$
- **Pf:** By definition we know
  - There exits d_1, d_2, so that $a \cdot d_1 = b_1$ and $a \cdot d_2 = b_2$
  - So, plugging in: $b_1 + b_2 = a \cdot d_1 + a \cdot d_2$
  - And this implies that $= b_1 + b_2 = a(d_1 + d_2)$
  - QED (by definition of divides)

Some easy questions about divisibility

- **Question 1:** Does d|a and d|b ⇒ $d|(a+b)$?
- **Question 2:** Does d|(a+b) ⇒ $d|a$ and d|b?
  - Ex. 3|(10+11) but 3 doesn’t divide either 10 or 11

Properties of “divides”

- **Def.** We say a binary relation R is transitive iff:
  - $\forall a, b, c, R(a, b) \land R(b, c) \Rightarrow R(a, c)$
- **Ex:** $=, <, \leq$ (why?)
- **Thm.** Divisibility is transitive
  - i.e., $\forall a, b, c, a|b \land b|c \Rightarrow a|c$
- **Pf.** By def. $\exists d_1, a \cdot d_1 = b$ and $\exists d_2,\ b \cdot d_2 = c$
  - But then by plugging in for b, $a \cdot d_1 \cdot d_2 = c$
  - QED (via $d_1 \cdot d_2$).

Properties of "set of divisors"

- **Def.** The set of divisors of n is denoted $Div(n) = \{a, a|n\}$
- **Ex:** $Div(6) = \{\pm 1, \pm 2, \pm 3, \pm 6\}$
- **Thm.** $a|b \Rightarrow Div(a) \subseteq Div(b)$
- **Pf.** Suffices to show $\forall a' \in Div(a), a'|b$
  - Since $a|b$ we know $\exists d_1, a \cdot d_1 = b$
  - Now since $a' \in Div(a) \Rightarrow \exists d_2, a' \cdot d_2 = a$
  - Plugging in for a, $a' \cdot d_2 \cdot d_1 = b$
  - QED (via $d_2 \cdot d_1$)

Greatest Common Divisor and Division Thm.

- **Def.** If m and n are nonnegative integers, then GCD(m,n) is the largest element of $Div(m) \cap Div(n)$
- **Ex.** GCD(5,15)=5
- **Thm.** (The Division Thm.) $\forall n \text{ and } d > 0, \exists! \, q, r \; so \; that \; n = d \cdot q + r$ and $0 \leq r < d$
  - ∃! is **unique** existence
  - Note: we call q the quotient, d the divisor, r the remainder, n the dividend
- **Pf. 1** (Existence, nonconstructive)
  - Consider the set $A = \{n - d \cdot q \mid q \in \mathbb{Z}\}$
    - "The remainder set"
  - Since $d \neq 0$, A must have a non-negative member
  - Therefore, A must have a *least nonnegative* member, called r
  - Now $r < d$ (why?)
    - Assume for contradiction $r > d$
    - Then, by subtraction, $0 < r - d < r$ would be a non-negative element of A, but this contradicts the min of r
    - i.e., we have $r = n - d \cdot q'$, so $0 < r - d = (n - d \cdot q') - d = n - (q' + 1)d < r$ (contradiction!)
  - So then notice $r = n - d \cdot q' \Rightarrow n = d \cdot q' + r$ where $r < d$. QED

Pf. 2: Uniqueness of division theorem

- Now need uniqueness!
- Suppose, for contradiction, that for fixed (but arbitrary) $n, d > 0\ \exists q_1, q_2, r_1, r_2$ so that:
  - $n = q_1 \cdot d + r_1$
  - $n = q_2 \cdot d + r_2$
- Then, by subtracting both sides:
  - $0 = q_1 \cdot d + r_1 - q_2 \cdot d - r_2 = \mathrm{d}(\mathrm{q}_1 - \mathrm{q}_2) + \mathrm{r}_1 - \mathrm{r}_2$ (*)
- But now we claim: $d|(r_1 - r_2)$ (**)
  - That's because d|LHS and $d|d(q_1 - q_2)$ so must be the case $d|(r_1 - r_2)$
  - This is a bit subtle! Recall that in general $d|(a + b)$ doesn't mean $d|a\ and\ d|b$.
- Claim: Since $0 \leq r_1 - r_2 < d,\ (**)\ \Rightarrow r_1 - r_2 = 0 \Rightarrow r_1 = r_2$
- So now $\mathrm{d} \cdot (q_1 - q_2) = 0$ by (*), but $d > 0$, so must be that $q_1 - q_2 = 0$
- So $q_1 = q_2$. QED

Division theorem definitions

- So division theorem tells us that $\forall n, d > 0, n = q \cdot d + r$ where $r < d$, and this is unique
- This motivates the definitions:
    - $n\ \boldsymbol{div}\ d = q$ and
    - $n\ \boldsymbol{mod}\ d = r$
    - as the q and r whose uniqueness is guaranteed by the division theorem.

Division theorem corollary (proof)

- **Thm.** Suppose $d = \gcd(a, b)$. Then there exist integers x, y so that $d = a \cdot x + b \cdot y$
- (These integers x,y are not necessarily unique!)
- **Pf.**
  - First, if a=b=0 then gcd isn't well defined, so wlog assume at least one of a, $b \neq 0$.
  - Consider set $S = \{a \cdot x + b \cdot y > 0\colon\ x, y \in \mathbb{Z}\}$
  - Because S is a nonempty set of positive integers…
    - It has a minimum element $d = a \cdot s + b \cdot t$
  - Claim that $d = \gcd(a, b)$.
  - First will show that d is a common divisor of a,b. WLOG show this for a (same for b)
    - Divide a by d, by division thm.: $a = dq + r$ with $0 \leq r < d$.
    - Claim: The remainder, r, is in $S \cup \{0\}$
      - because: $r = a - dq = a - (as + bt)q = a(1 - qs) - b(qt)$ (by substitution and algebra)
      - So this means $r \in S$ or $r = 0$
    - But recall d is the smallest positive integer in S *and* r being a remainder, must be less than d. So r=0.
    - But this means that $d|a$ (and by same token $d|b$).
  - Now we'll show d is **greatest** common divisor of a and b
    - Let $c \in div(a) \cap div(b)$. Then there exists u, v so that $a = cu$ and $b = cv$
    - Then $d = as + bt$ (from above) = $cus + cvt$ (by substitution) $= c(us + vt)$
    - But this means that c divides d! Then $c \leq d$. QED

Euclid’s algorithm (preliminaries)

- Euclid’s algorithm for computing GCD(a,b).
  - One of the first algorithms!
  - What is an algorithm?
    - Intuitively, a procedure for solving a problem without going through all possible inputs
- **Starting Thm**: If $d = \gcd(a, b)$, $b \neq 0$ and $r = a \boldsymbol{mod}\ b$ then $d = \gcd(b, r)$.
  - i.e., if d is the gcd of a and b, then d is also the gcd of b and the remainder of a divided by b
- **Pf.** We’ll show $Div(a) \cap Div(b) = Div(b) \cap Div(r)$
  - That is, that the common divisors of both a and b are the same as the common divisors of b and r
  - First, let $q = a\ \boldsymbol{div}\ b$, so say $a = q \cdot b + r$, by definition
  - Now let $z \in Div(a) \cap Div(b)$
    - Then $z|a$ and $z|b \cdot q \Rightarrow z|(a - bq) = r$
    - So $z \in Div(r)$
  - Now let $z' \in Div(b) \cap Div(r)$
    - Then $z'|(q \cdot b + r) = a$, so $z' \in Div(a)$
- Now, recall $d = \gcd(a, b) = \max(Div(a) \cap Div(b))$ so must also be $\gcd(\mathrm{b}, \mathrm{r}) = \max(Div(b) \cap Div(r))$ QED.

Euclid's algorithm

- The **starting thm.** on the last slide gives way immediately to an **algorithm** for computing gcd(a,b)
- Consider, we've just proven:

$$\begin{aligned} gcd(a,b) &= a,\ if\ b = 0 \\ &= gcd(b, a \mod b)\ otherwise \end{aligned}$$

- So can compute by continually replacing the larger argument (here "a") for the remainder: a mod b, until we get 0!
- **Ex:** $\mathrm{gcd}(15{,}9) = \mathrm{gcd}(9{,}15\ mod\ 9 = 6) = \mathrm{gcd}(6{,}3) = \mathrm{gcd}(3{,}0) = 3$

b. Modular arithmetic

Modular arithmetic introduction

- Last time we discussed divisibility, now we'll talk about "equating" integers via their remainders when divided by some fixed integer
- Recall that m|(a-b) doesn't imply m|a and m|b, rather it means…
  - that their remainder is the same when divided by m, this motivates:
- **Def.** If $a, b \in \mathbb{Z}, m > 0$, then we say $a \cong b \boldsymbol{mod}\, m$ or $a \cong_m b$ if:
  1. $m|(a - b)$
  2. Equivalently if $a\ mod\ m = b\ mod\ m$

Properties of Mod function

- **Thm.** Mod is an **equivalence relation**
- **Pf.** 1. Reflexive:
    - $\forall a, a \cong_m a, m|(a-a) = 0$
  - 2. Symmetric:
    - if $a \cong_m b$ then $m|(a-b) \Rightarrow m|(b-a) \Rightarrow b \cong_m a$
  - 3. Transitive:
    - If $a \cong_m b, b \cong_m c$
    - Proof: we have $m|(a-b)$ and $m|(b-c)$ and so $m|\big((a-b)+(b-c)\big) = a-c$, and so $a \cong_m c$.

Equivalence in terms of Mod function

- Consider equivalence classes defined under $\cong_m$
    - i.e., $\forall m > 0$ and $a \in \mathbb{Z}$ define $[a]_m = \{b |\ a \cong_m b\}$
- Let's define operations of these classes:
    - $[a]_m + [b]_m = [a + b]_m$ (where equality is by definition)
    - $[a]_m \cdot [b]_m = [a \cdot b]_m$
- Let's prove these operations are "well-defined":
    - Will prove for addition, the multiplication case will be an (easy) exercise!
    - i.e., we need to show that, if $a_1 \cong_m a_2$ and $b_1 \cong_m b_2$ $\text{then } a_1 + b_1 \cong_m a_2 + b_2$
    - By definition we have $m|(a_1 - a_2)$ and $m|(b_1 - b_2) \Rightarrow m|((a_1 + b_1) - (a_2 + b_2))$ QED

More properties of equiv. classes of mod

- This addition is also associative (and it “inherits this” from integer addition)
  - i.e., $\forall a, b, c$ and for $m > 0$, $[a]_m + ([b]_m + [c]_m)$
  - $= [a]_m + [b + c]_m = [a + (b + c)]_m = [a + b]_m + [c]_m = ([a + b]_m) + [c]_m$
- To recap, there’s an important “meta” phenomenon going on here:
  - We start with a structure, say the integers, $\mathbb{Z}$
  - Then we describe an equivalence relation on $\mathbb{Z}$, such as mod, and define equivalence classes using this equivalence relation
  - Then we notice that certain functions defined on $\mathbb{Z}$ are also well-defined on the equivalence class.
  - Then we can create a new structure named “$\mathbb{Z}/\cong_m$” that consists of the equivalence classes together with certain operations, say $+, \cdot$
- In this case, we define “the integers $mod\ m$” to be the structure $\mathbb{Z}_m = \mathbb{Z}/\cong_m$ together with operations $+, \cdot$
- However, there are **important differences** between the integers mod m and the integers themselves:
  - for instance in the former there’s no meaningful ordering (i.e., no “<“ or “>” operators)

Announcements

- Next week Friday I’m traveling, there will be a guest lecturer!
- My office hours next week are after class on Wednesday
- This week office hours are after class today, as usual

Relatively prime numbers, and Mult. Inverses

- **Def.** Two integers a, b are *relatively prime* (or co-prime), if $\gcd(a, b) = 1$.
- **Def.** Integer a has a *multiplicative inverse* mod m if $\exists b, b \cdot a \cong_m 1$
- **Thm.** If $m > 0$ and a are relatively prime, then a has a multiplicative inverse $mod\ m$
- **Pf.**
    - Recall from last time, we proved Bezout's identity. ("Suppose $d = \gcd(m, a)$. Then there exists n, b so that $d = n \cdot m + b \cdot a$")
    - So now, using this identity, there exits integers n, b so that $n \cdot m + b \cdot a = 1$
    - Then computing equivalence classes mod m, (aka "residue classes") and the fact that addition, multiplication translates over these classes:
    - $[1]_m = [n \cdot m + b \cdot a]_m = [n]_m \cdot [m]_m + [b]_m \cdot [a]_m$
      $= [n]_m \cdot [0]_m + [b]_m \cdot [a]_m = [0]_m + [b]_m \cdot [a]_m = [b \cdot a]_m$ QED

Corollary of mult. Inverse existence

- **Cor.**: if p is prime, and a,b are integers so that $p|(a \cdot b)$ then $p|a$ or $p|b$.
- **Pf.** WLOG suppose p doesn't divide a (i.e., could have just as well considered b)
  - Then we know $\gcd(a, p) = 1.$
  - Now by Bezout's identity, there exist integers x and y so that $x \cdot a + y \cdot p = 1$
  - Multiply both sides on the right by b, so $b = x \cdot a \cdot b + y \cdot p \cdot b$
  - But now p divides the RHS (why?)
    - because it divides both terms of the sum
  - So it must divide the LHS=b. QED

The Chinese Remainder Theorem

- **Thm** (The “Chinese Remainder Thm.”) Suppose $m_1, m_2, \ldots, m_k$ are *pairwise* rel. prime and that $c_1, c_2, \ldots, c_k$ are integers.
    - Then there exists a solution, x, to system of equations:
        - $\{x \cong c_i \; mod \; m_i\}$
- **Proof** (existence, constructive):
    1. Let $n = m_1 \cdot m_2 \cdot \ldots \cdot m_k$
    2. For each $i \in \{1, \ldots, k\}$ define $n_i = n/m_i$
    3. Notice that n_i is rel. prime to m_i (why?)
        - Because if a, b and c are rel. prime then so are a and $c \cdot b$
    4. So let $\{a_i\}$ be set of mult. inverses wrt $\{n_i\}$ – i.e., let $a_i \cong \; n_i^{-1} mod \; m_i$ (why do these exist?)
    5. Recall what this means: it’s the unique a_i so that $a_i n_i \cong 1 \; mod \; m_i$
    6. Define $x_i = a_i n_i$ for all i
    7. Then claim: $x = c_1 x_1 + c_2 x_2 + \cdots + c_k x_k$ is the solution to the system!
    8. To see this, look at any i-th equation in the system: x $\cong c_i \; mod \; m_i$, then:
        - Take i-th term of x: $c_i x_i = c_i \cdot a_i n_i$ (by definition of x_i)
            - **Claim:** $c_i \cdot a_i n_i \cong c_i \; mod \; m_i$ -- why?
            - Since $a_i \cong n_i^{-1} \; mod \; m_i$ so $c_i \cdot a_i n_i \cong c_i \cdot n_i \cdot n_i^{-1} mod \; m_i = c_i \cdot 1 \; mod \; m_i = c_i \; mod \; m_i$
        - But if we consider j-th term, for $j \neq i$, it’s $c_j x_j$
            - Claim: $c_j x_j \cong 0 \; mod \; m_i$
            - Since $c_j x_j = c_j a_j n_j$ (by how we defined x_j)
                - $\cong 0 \; mod \; m_i$ (why??)

                Because by definition of n_j, $m_i | n_j \Rightarrow m_i | c_j a_j n_j$
    - QED

Announcements

- Midterm – next Wednesday, November 1 or Friday, November 3??
- This week only -- office hours on Wednesday after class, Crerar 251
- Homework 3 due and Homework 4 is assigned

Uniqueness

- **Thm.** If a and m are relatively prime then the mult. inverse of a $mod\ m$ is unique
- **Pf.** Suppose, for contradiction, that both b and c are different mult. inverses of a mod m (**note:** equalities in what follows are shorthand for equivalence mod m)
  - Then: $b = b \cdot 1 = b \cdot (c \cdot a)$ (why?)
    - $= (b \cdot c) \cdot a = (c \cdot b) \cdot a = c \cdot (b \cdot a) = c$ – because of how we defined b
- **Notation:** we use a^{-1} to denote the mult. inverse of a, when it exists

c. A brief discussion about induction

Induction review!

- **Goal of induction:** we want to prove a universal statement of the form:
  - $\forall n > 0, \phi(n)\ is\ True.$
- **Steps:**
  - Prove the "base case" $\phi(1)\ is\ True$
  - Prove the "inductive step" $\forall n > 0,\ \phi(n)\ is\ True \Rightarrow \phi(n+1)\ is\ True$

Example: Summing first n numbers

- A very famous example of induction is due to Gauss
- Suppose we want to sum the numbers 1,2,…,100
- Gauss's observation is that if we define the answer, α, to be the sum:
    - $\alpha = 1 + 2 + \cdots + 99 + 100$
    - Then we can, by flipping the terms in the sum write:
    - $\alpha = 100 + 99 + \cdots + 2 + 1$
    - Then notice that $2\alpha = 101 + 101 + \cdots + 101 + 101 = 101(100)$
    - So we have $\alpha = \frac{101(100)}{2} = 101 \cdot 50 = 5050$
- More generally, if we want to sum $1,2, \ldots, n$ we can use the same principle to obtain $\alpha = \sum_{i=1}^{n} i = \frac{n \cdot (n+1)}{2}$

Example: Formal proof

- Gauss's procedure can be formalized by an inductive argument
- **Recall:** want to prove $\sum_{i=1}^{n} i = \frac{n\cdot(n+1)}{2}$
- **Pf:** By induction on n. What is the predicate, ϕ ?
  - $\phi(n) = 1$ if $\sum_{i=1}^{n} i = \frac{n\cdot(n+1)}{2}$
- **Base case:** $\phi(1) = 1, \text{since } 1 = \frac{1\cdot 2}{2}$
- **Inductive step:** Suppose $(\phi(n) = 1)$ which means $\sum_{i=1}^{n} i = \frac{n\cdot(n+1)}{2}$
  - Will show $\phi(n+1)\ is\ true$
  - This is because $\sum_{i=1}^{n+1} i = (\sum_{i=1}^{n} i) + n + 1$
    - $= \frac{n\cdot(n+1)}{2} + n + 1$, by Inductive Hypothesis
    - $= (n+1)(\frac{n}{2} + 1)$, factoring out $n+1$
    - $= \frac{(n+1)(n+2)}{2}$, factoring out ½ (since 1=2/2)
    - QED

Strong induction

- For some problems, helpful to consider “strengthening” inductive hypothesis
- In strong induction we use:
- **Def** (Strong Inductive Principle):
- $(\forall n > 0, \forall k, 0 \leq k < n, (\phi(k) = 1) \Rightarrow (\phi(n) = 1)) \Rightarrow (\forall n > 0, \phi(n) = 1)$
  - *Proof by contrapositive*: Suppose RHS is false, i.e., $\exists n\ \phi(n) = 0$ and so if we look at the set of “counterexamples”: $\{n | \phi(n) = 0\}$ is a non-empty set of natural numbers
  - So it must have a minimum element, n_0. But since that’s the minimum counterexample, we must have $\forall k < n_0,\ \phi(k) = 1$.
  - But notice this contradicts the LHS (since this does not imply that $\phi(n_0) = 1$, because in fact, $\phi(n_0) \neq 1$)
- Will give an example that uses this shortly…

d. Prime numbers

Definitions

- **Def.** A **prime number** is an integer greater than 1 for which the only divisors are 1 and itself.
- **Ex.** 2,3,5,7…
- We'll now show that the primes are the "building blocks" of positive integers
- i.e., the "Fundamental theorem of arithmetic" says that every integer > 1 can be written uniquely as a product of primes
  - **Simpler thm.** Every integer > 1 has a prime divisor i.e., $n > 1 \Rightarrow \exists\ prime\ p\ such\ that\ p|n$
  - The way to prove this uses induction, as with many universally quantified statements in number theory…

But “weak” induction doesn’t work!

- ***(recall)* Simpler thm.** Every positive integer > 1 has a prime divisor i.e., $n > 1 \Rightarrow \exists\ prime\ p\ such\ that\ p|n$
  - But a totally straightforward inductive approach doesn’t work!
    - i.e., in the induction step, we’d need to show $n\ has\ prime\ divisor \Rightarrow \exists p|(n+1)$
    - Consider $n+1$. Either it’s prime or it’s not. If it is, we’re done (since $(n+1)|(n+1)$)
    - But if it’s not, then $n+1 = a \cdot b$ and we’d need to use Inductive Hypothesis (i.e., that a prime divides n) to show that a prime $p|a$ or $p|b$ – this isn’t obvious!

Proof of “simpler theorem”

- (Recall) Simpler thm. Every natural number > 1 has a prime divisor i.e., $n > 1 \Rightarrow \exists\ prime\ p\ such\ that\ p|n$
- **Pf.** Let the statement $\phi(n) = (n > 1) \Rightarrow \exists\ prime\ p$ so that $p|n$
  - Then by strong induction, assume $\phi(k)$ is true whenever $k < n$
  - Proceed as before: if n is prime then $n|n$ and we’re done
  - So assume n is not prime. Then $n = a \cdot b$ for $1 < a, b < n$
  - By inductive hypothesis, $\exists\ prime\ p, so\ that\ p|a$. So this implies $p|n$. QED

Cor. There are infinitely many primes!

- **Pf:** Consider any finite list of prime numbers $p_1, p_2, \dots, p_k$. We'll show it's missing some prime number!
- Let $p = p_1 p_2 \dots p_k$
  - Then consider $n = p + 1 = p_1 \cdot p_2 \cdot \dots \cdot p_k + 1.$
  - Now either n is prime or it's not
  - If it is, then there's at least one prime not on the list!
  - So let's assume n is not prime
  - But then by **simpler theorem**, some prime p'| n.
    - Claim: p' cannot be in our list! (Why?)
      - Because if it's in our list then p'|p
      - But since also p'|n we know…
        - p'|(n-p)=1 (why?)
      - But no number other than 1 can divide 1 !!

A helpful lemma

- **Helpful Lemma**: If p is prime and so are $p_1, p_2, \dots, p_k$ (not necessarily distinct), and $p | p_k \cdot p_{k-1} \cdot \dots \cdot p_1$ then p must equal p_i for some i
- **Pf** (Lemma): how should we proceed?
  - Standard induction on number of primes, k
  - Base case: k=1, then $p | p_1 \Leftrightarrow p = p_1$
  - Inductive Hypothesis: Assume for k, to prove for $k + 1$
  - So consider instance of the $k + 1$ case, i.e., suppose $p | p_1 \cdot p_2 \cdot \dots \cdot p_{k+1}$
  - Then recall that for prime p, if $p | a \cdot b \Rightarrow p | a \text{ or } p | b$.
  - Now suppose $a = p_1 \cdot p_2 \cdot \dots \cdot p_k$ and $b = p_{k+1}$
    - Then either $p | a$, in which case $p = p_i$ for some $i \in \{1, \dots, k\}$ by inductive hypothesis
    - Or $p | b$, in which case $p | p_{k+1}$ and so $p = p_{k+1}$

Fundamental Theorem of Arithmetic

- Recall **Thm** (FTA): Every $n > 1$ can be written uniquely (up to changing the order) as a product of primes.
- **Pf.** (existence): Easy to prove via strong induction (exercise!)
- **Pf.** (uniqueness):
  - Suppose the prime factorizations aren't unique
  - Let *n* **be the least** natural number that doesn't have a unique prime factorization
  - So $n = p_1 \cdot \ldots \cdot p_k = p_1' \cdot \ldots p'_{k'}$
  - Now $p_1 | n$ so by **Helpful Lemma**, we must have $p_1 = p_j'$ for some *j*.
  - But then we can divide both sides by p_1 and obtain a smaller counter-example!
  - This contradicts minimality of n. QED

Announcements

- Midterm – this upcoming Friday!
- No homework this week!
- Office hours: Wednesday after class (10:30 – 11:20 PM) in Crerar 251
- TAs to post practice midterm shortly!
- Last topic on midterm: Fermat's Little Theorem

Factorial function and Wilson's thm.

- **Def.** Factorial function is denoted $n!$ for *n* a non-negative integer the product of all positive integers less than or equal to n
  - *Equivalently* can define inductively:
    - $0! = 1$
    - $(n+1)! = (n+1) \cdot n!$
- **Ex:** $5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$
- **Wilson's Thm.** For all p prime, $(p-1)! \cong -1(mod\ p)$
- **Pf.** Let's consider each number $x \in \{1, \dots, p-1\}$
  - First note that every number x $\in \{1, \dots, p-1\}$ is rel. prime to p
  - This means that every number $x \in \{1, \dots, p-1\}$ has a ***unique mult. Inverse*** mod p
  - Now we **claim:** in this list, only 1 and p-1 are *self-inverse* mod p – why?
    - i.e., $x^2 \cong 1\ (mod\ p) \Rightarrow x^2 - 1 \cong 0\ (mod\ p) \Rightarrow p|(x-1)(x+1)$
    - So this means $x \cong +1(mod\ p)$ or $x \cong -1(mod\ p)$ – which numbers satisfy?
    - Out of the list, only satisfied by $x = 1$ or $x = p - 1$
  - All the other elements in the list have a non-self mult. Inverse
  - Now consider $(p-1)! = (p-1) \cdot (p-2) \cdot \dots \cdot 1$
  - The "middle" terms from $\{2, \dots p-2\}$ can be reordered into distinct pairs of numbers and multiplicative inverses.
  - These pairs each multiply to 1, and disappear, leaving us with 1 and $p-1$. And $1 \cdot p - 1 = p - 1 \cong -1\ (mod\ p)$ QED.

Another very helpful lemma

- **Helpful lemma 2**: Let p be prime, and a be such that $\gcd(a, p) = 1$. Then, as sets, $\{a \cdot 1, a \cdot 2, \ldots, a \cdot (p-1)\} = \{1,2, \ldots, p-1\}$ **(where multiplication here is done mod p)**
- **Pf.**
  - First note that each element in LHS set, is of form $a \cdot j \neq 0 \; mod \; p$ (why?)
    - Because a and j are both rel. prime to p so p can't divide their product
    - So each element mod p is in $\{1,2, \ldots, p-1\}$
  - Now all the elements of the LHS are distinct! (why?)
    - As $\gcd(a, p) = 1$ there must exist a mult. Inverse a^{-1} for a.
    - Let's assume, for contradiction, that $(a \cdot j) mod \; p = (a \cdot j') mod \; p$ for $j \neq j'$
    - Then multiply both sides by a^{-1} and we get $j = j'$ (this is contradiction).

Fermat’s Little Theorem

- **Thm.** For all primes p, and a such that $\gcd(a, p) = 1$, $a^{p-1} \cong 1\ (mod\ p)$
- **Pf.** Consider the product $n = (a \cdot 1) \cdot (a \cdot 2) \cdot \ldots \cdot \big(a \cdot (p-1)\big)\ (mod\ p)$
  - On the one hand, by **helpful lemma 2,** this is just a reordering of the factorial, and so $n = -1$, by Wilson’s Thm. (i.e., this is $= (p-1)!\ = -1\ mod\ p$).
  - On the other hand, by factoring out the a’s (of which there are $p-1$), we have $n \cong a^{p-1} \cdot (p-1)!\ (mod\ p)$
  - So we have $n \cong a^{p-1} \cdot (-1)(mod\ p)$
    - by Wilson’s thm. (used here again!)
  - Equating both expressions for n, we have $-1 \cong a^{p-1} \cdot (-1)(mod\ p)$, and multiplying both sides by -1 gives the Thm. QED

e. Cryptography

Who cares about Fermat’s Little Theorem (or number theory, more generally?)

- Historically, number theorists were proud of studying the “purest” form of math
  - In the sense that it didn’t have any applications to the real world.
- Today, we know that this is not true at all.
- Possibly most interesting connection of number theory to CS: Cryptography!

Announcements

- Office hours: Today after class, 10:30 – 11:30 AM, Crerar 251
- Midterm this Friday, in class!
- If you need accommodations the TAs should have already reached out to you – please let us know ASAP if you haven’t heard from them!

Cryptography!

- **This is a bonus lecture, which will not be on homework or tested**
- So why do we cover this?
- Because this is a discrete mathematics class in the computer science department
- This is one of the best examples of an interesting connection between mathematics and real-world computing

Private key cryptography (used since the Romans)

- Most fundamental task in cryptography: encrypt a message m so that:
  - m can be "encrypted" or made unreadable to an adversary
  - But only intended recipient of the message can somehow decrypt to get back m.
- Most obvious encryption is called "**private key**" or "**symmetric**" cryptography
- In which Alice chooses a private key, k, that she shares only with intended recipient, Bob.
- Alice then encrypts her message with *k* and sends message to Bob
- Bob uses k to decrypt Alice's message

Example of private key encryption – sometimes called the “one-time pad”

- **Setting:** Alice has an n bit message $m \in \{0,1\}^n$ that must be sent to Bob!
- **Encryption**:
  - Alice chooses uniformly random key $k \in_R \{0,1\}^n$ and shares it only with Bob
  - Then Alice encrypts her message m by taking the “bitwise xor of m and k”
    - i.e., xor of two bits a $\oplus b = 1$ iff exactly one of the bits is 1
    - So Alice sends to Bob $y = m \oplus k$ (that is, y is the n bit string so that $y_i = m_i \oplus k_i$ for all i)
- **Decryption**:
  - What can Bob, who holds the encrypted message, y, and k, do to get m?
  - Bob takes the bitwise xor of y and k, and:
    - $\mathrm{m} = y \oplus k = m \oplus k \oplus k$ since $k \oplus k = 0^n$

Why is the one-time pad secure?

- Not hard to show that Eve -- the eavesdropper! -- who doesn't know k, cannot discover k from $y = x \oplus k$. Why?
  - Recall that the key k was chosen uniformly at random over $\{0,1\}^n$
  - So it's not hard to see that the encrypted message, $y = x \oplus k$ is still uniformly distributed over all n bit strings!
    - So she gets as much information about the key, k, from y, as from a random string that she chooses
    - So she may as well guess the key randomly (which isn't helpful!)
- But it's not provably secure if we use the same key again with another message (why?)
  - Because if Eve holds both $y = x \oplus k$ and $\mathrm{y}' = x' \oplus k$ then $y \oplus y' = x \oplus x'$
  - This is not random and also not necessarily secure (e.g., if Eve knows x' but not x she can now discover x) – this is why it's called one-time.

Public key cryptography (1970’s)

- Problem with **private key** cryptography is that in practice…
  - Don’t want a trusted courier to deliver the private key each time we want to encrypt a message!
  - This is generally infeasible (how would Amazon.com work with this?)
- Alternative: **public key**, or “**asymmetric**” cryptography.
- In public-key cryptography there are two different keys
  - One is public (everyone can share!) that is used to encrypt
  - One is private that is used to decrypt
- How does this fix internet e-commerce applications?
  - E.g., if Amazon.com gives you their public key, you can use to encrypt credit card information to send to them.
  - Once you encrypt, no one but Amazon.com can decrypt!

Announcements

- Jesse lecturing on Friday!
- Again office hours are this Wednesday after class
- Homework 5 assigned!

RSA public-key encryption

- Most famous public-key encryption scheme is due to Rivest, Shamir, Adleman from the 70's.
- It uses ideas from number-theory! In particular, a generalization of Fermat's Little theorem called Euler's theorem.
- Euler's theorem:
  - For n any positive integer, let $\phi(n)$ be the *number of numbers* up to n that are rel. prime to n
  - *Not hard to see:* if $n = p \cdot q$ for primes p and q can show $\phi(n) = (p-1)(q-1)$
  - Recall **Fermat's Little Theorem**: For all primes p, and a such that $\gcd(a,p) = 1 \text{ we have } a^{p-1} \cong 1\ (mod\ p)$
  - **Euler's theorem** (generalizes FLT): If n rel. prime to a, then $a^{\phi(n)} \cong 1\ mod\ n$
- **Theorem** [RSA'77]: Assume given $n = p \cdot q$, it's hard to compute $\phi(n) = (p-1) \cdot (q-1)$. Then public-key encryption is possible!
  - This is actually a deep assumption.
  - In particular, it's assuming that factoring is a hard problem!

RSA encryption scheme

- **Theorem** [RSA'77]: **Assume given $\boldsymbol{n = p \cdot q}$, it's hard to compute $\boldsymbol{\phi(n) = (p-1) \cdot (q-1)}$.** Then public-key encryption is possible!
- **Key generation**: Bob chooses two distinct, random primes p, q (very large!!)
  - Compute $n = p \cdot q$.
  - Choose integer e so that $1 < e < \phi(n)$ and $\gcd\big(e, \phi(n)\big) = 1$
    - This will be public!
  - Because e is rel. prime to $\phi(n)$…
    - There exists a mult. inverse, d so that $d \cdot e = 1 \; mod \; \phi(n)$.
  - **Public key:** (n, e)
  - **Private key:** (n, d).
- **Encryption**: Bob shares public key (n, e) with everyone (including Alice and Eve)
  - Say Alice wants to send message m to Bob.
    - Let's say $m \ll n$
  - Alice computes c $= m^e \; mod \; n$ and sends to Bob.
- **Decryption**: Bob receives $c = m^e \; mod \; n$ and holds (n, e, d).
  - Bob computes: $c^d \; mod \; n = (m^e)^d mod \; n = m^{ed} \; mod \; n$ *(claim this equals m)*
  - Recall $ed = 1 \; mod \; \phi(n) \Rightarrow \phi(n) | (ed - 1) \Rightarrow \exists k \; so \; that \; k \cdot \phi(n) + 1 = d \cdot e$ by construction
  - So plugging in from above:
    - $m^{ed} \; mod \; n = m^{k\phi(n)+1} \; mod \; n = m\big(m^{k\phi(n)}\big) \; mod \; n = m(1) \; mod \; n = m$ mod n
  - Where last equality is from Euler's theorem – generalized Fermat's little theorem!!
    - But this requires m to be rel. prime with n – true with high probability!

But is RSA really secure?

- RSA encryption, or similar ideas, are now completely ubiquitous on the internet!!
- But they are based on the assumption that can’t factor efficiently.
- Is this really true?
- It may be true for today’s computers, but it turns out that…
    - The physics of computation really matters here! Computers of the future that could take advantage of quantum mechanics – called quantum computers can factor numbers efficiently.
    - The first quantum computers are being built **now**, although factoring will not be a practical task for these computers for many years to come
- **Open question in CS**: what will replace RSA, that will be secure against futuristic quantum computers?

Counting

a. Permutations and Combinations

Announcements

- Jesse lecturing on Friday!
- Office hours today after class in Crerar 251

Permutations

- Def. A **permutation** of a set is an ***ordered*** arrangement of its elements (a list with the same elements as the set)
- Ex. Suppose we consider the set {1,2,3} – how many permutations?
  - {1,2,3},{1,3,2},{2,1,3},{2,3,1}, {3,1,2},{3,2,1}
- How many permutations of a set of *n* elements?
  - $n!$
- Can also consider **2-permutations** (i.e., ordered subsets of size two) of the same set – this also has size 6
  - {1,2}, {1,3}, {2,1}, {2,3}, {3,1} and {3,2}
- And more generally $k - permutations$, for $k \leq n$
- The fact that we got the same number in both examples is not a coincidence!

Number of r −Permutations

- Def. Let P(n,r) denote the number of r-permutations of an n element set
- Thm. $P(n,r) = n \cdot (n-1) \cdot \ldots \cdot (n-(r-1))$
- Pf. There are n ways to choose the first, n-1 remaining ways to choose the second and so on.
- Cor. If n is positive, and $0 \leq r \leq n$, then $P(n,r) = \frac{n!}{(n-r)!}$
- Pf 2. Start with the number of ways to order an n element set, this is $n!$.
  - But we could have counted differently: first choose r items from the set – there's $P(n,r)$ different ways to do this.
  - Then choose the remaining $n-r$ items in $(n-r)!$ different ways.
  - So we have $n! = P(n,r) \cdot (n-r)! \Rightarrow P(n,r) = \frac{n!}{(n-r)!}$

Permutations example

- Ex: How many permutations of the letters ABCDEFGH contain the letters ABC consecutively?
  - There’s a trick here! We view the underlying set as consisting of 6 elements: ABC, D, E, F, G, and H.
  - Any permutation that has the characters we want amounts to a permutation of these six
  - So the answer is P(6,6)=6!=720.

Combinations

- Closely related problem is how many *distinct* subsets of size *r* can be formed from an *n* element set.
- Each such subset is called an r-**combination**
- Def. An r-combination of a set with n elements is a subset of size r. We denote the number of such subsets as **C(n,r)**
- How are *combinations* different than *permutations*?
  - A permutation refers to ordered arrangements of elements (e.g., {1,2,3} is distinct from {3,2,1}).
  - A combination is just a subset (so ordering within the subset is ignored)
- Are there more **r-permutations** of an *n* element set or **r-combinations**?
  - In general there are more **r-permutations** of an n element set, than **r-combinations** from that set.

Number of combinations

- **Thm.** If n is positive, and $0 \leq r \leq n$, then $C(n, r) = \frac{n!}{r!(n-r)!}$
- **Intuition**: Recall there are $n!/(n-r)!$ r-permutations of a set of size n.
  - But when comparing with r-combinations, we've *overcounted* since many r-permutations correspond to the same r-combination.
  - Easy to see we've overcounted by a factor of $r!$, why?
    - Because there are $r!$ r-permutations per each r-combination
- Pf. Let's just formalize this intuition:
  - $P(n, r) = C(n, r) \cdot P(r, r)$
  - $\Rightarrow C(n, r) = \frac{P(n,r)}{P(r,r)} \Rightarrow C(n, r) = \frac{n!}{r! \cdot (n-r)!}$

Combinations example

- It’s traditional to consider examples with decks of cards
- However, if you’re like me, you don’t play card games very often
- So let’s review the basics:
  - There’s 52 total cards (not counting Jokers)
  - There are 13 “kinds” (i.e., 2,…,9,10,Jack, Queen, King, Ace)
  - There are four “suits” (clubs, diamonds, hearts, spades)
  - A poker hand consists of 5 cards
- Ex. How many distinct poker hands exist?
  - Answer: $C(52,5) = 2{,}598{,}960$

More card counting examples

- **Reminder:** A full house is a poker hand (5 cards) that:
  - consists of a three-of-a-kind (i.e., a triple of the same kind, possibly different suits)
  - together with a pair (two of the same kind) – but can be different kind from the above triple
- **Ex.** How many distinct “full houses” exist?
  - Pick two kinds in order (one for the triple, one for the pair)
    - There’s $P(13,2)$ ways to do this – why not $C(13,2)$?
  - Now we have narrowed our options down to the same kind, but there are four cards of the same kind in each deck (one for each suit)
  - So there’s $C(4,3)$ different ways to choose the triple and $C(4,2)$ ways to choose the pair.
  - So there’s $n = P(13,2) \cdot C(4,3) \cdot C(4,2)$

Even more card examples!

- Recall, a flush is a poker hand (5 cards), in which all cards are the same suit
- Ex. How many flushes exist?
  - First there are $C(4,1) = 4$ different suits
  - Then from that suit there are $C(13,5)$ different hands in that suit
  - So answer $n = C(4,1) \cdot C(13,5) = 5{,}148$

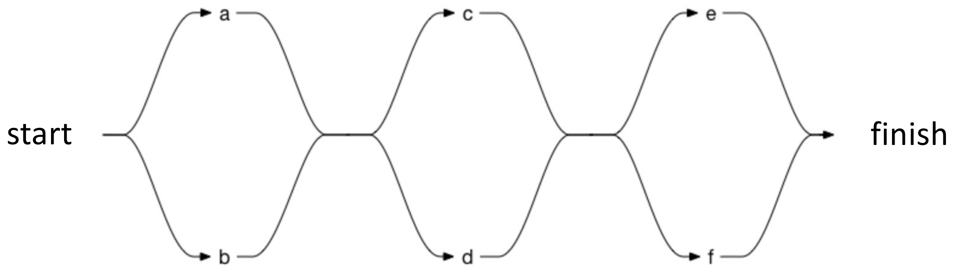
Combinatorial proofs

- Def. A combinatorial proof is a proof based on a counting argument.
  - (Generally these proofs work by counting the same thing two different ways to achieve combinatorial identity)
- Thm. For all $n > 0$ and $0 \leq r \leq n, C(n, r) = C(n, n - r)$
- Pf. Given a set of *n* elements there's two ways to pick r elements.
  - Can enumerate its r elements subsets directly: get $C(n, r)$ or
  - Can enumerate their complements i.e., "throw out" $n - r$ elements, $C(n, n - r)$ different ways – now we're left with r elements!

b. The Binomial Theorem

Introduction to binomial theorem

- Suppose we want to evaluate a ”product of sums” expression like:
  - $(a + b) \cdot (c + d) \cdot (e + f)$
- We can envision this expression graphically like:



- And evaluating this product corresponds to summing the values of all paths through the circuit from start to finish
- For example: could have first gone through b then d then e
- So $b \cdot d \cdot e$ is a term in the expression, along with all other such paths

Introduction to binomial theorem 2

- Of course we can think of a similar diagram to help us compute an expression like $(x+y)^4$.
  - We would have four consecutive “choices” of x or y
- How many (non-distinct) terms in the final sum?
  - There’s $2^4 = 16$ terms in the final sum, corresponding to the total number of paths
- Clearly each path does not correspond to a unique term – and the degree of each term is 4.
- So how many times, for instance, do we get the term x^3y?
  - This is equivalent to either $C(4,3) = 4$ -- if we count the number of ways to get three factors of x
  - Or we count this by $C(4,1) = 4$ – if we count the number of ways to get a single factor of y
- Notice that $C(n,k) = C(n,n-k)$
  - Prove this by the formula definition of combinations

The Binomial theorem

- More generally we note in the product $(x+y)^k$, we have each term $x^l y^{k-l}$ (or "monomial") appearing $C(k,l) = C(k,k-l)$ times.
- Thm. For all natural numbers $n > 0$:
  - $(x+y)^n = \sum_{i=0}^{n} \binom{n}{i} x^{n-i} y^i$
  - Where the notation $\binom{n}{i}$=C(n,i)
  - We've already sketched a proof… how would we prove this formally?
- Ex: What is the coefficient of $x^{11}y^5$ in the expansion of $(x+y)^{16}$?
  - Answer: $\binom{16}{11}$
- Ex. What is the coefficient of $x^7 y^4$ in $(x+2y)^{11}$?
  - Answer: Note without the ”2” we would have had $\binom{11}{7}$, but every time we see a y factor we get an extra factor of 2. So answer is $2^4\binom{11}{7}$

Binomial identities

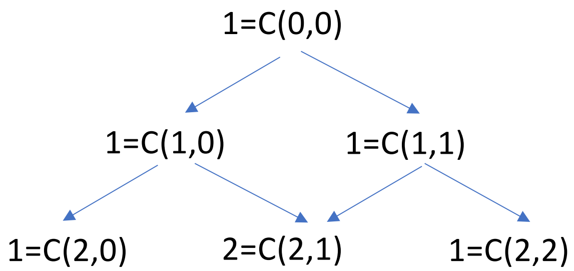
- Recall (binomial theorem): Thm. For all natural numbers $n > 0$:
  - $(x + y)^n = \sum_{i=0}^{n} \binom{n}{i} x^{n-i} y^i$
- Thm. For all natural numbers $n > 0$, $\sum_{i=0}^{n} \binom{n}{i} = ?$
- Pf. LHS is exactly the expansion of $2^{\mathrm{n}} = (1 + 1)^n$
- Thm. For all natural numbers $n > 0$, $\sum_{i=0}^{n} (-1)^i \binom{n}{i} =?$
- Pf. Consider $0 = (1 - 1)^n$. The binomial theorem says this is exactly equivalent to LHS.

Pascal's identity

- Thm. For all natural numbers $n \geq k > 0$, $\binom{n+1}{k} = \binom{n}{k-1} + \binom{n}{k}$
- Pf. The LHS is considering an n+1 element set S and taking the number of k-element subsets. We could equivalently count this by:
  - Picking an arbitrary fixed element $s \in S$ and counting subsets $K \subseteq S$ with k elements
  - Now there are two disjoint possibilities:
    - Either $s \in K$
      - So we know that $k - 1$ elements of K come from set S$-\{s\}$
      - Notice $S - \{s\}$ is a set of size n, and we're interested in a subset of size $k - 1$
      - So $\binom{n}{k-1}$ ways to choose this
    - Or $s \notin K$
      - In which case all k elements of K come from $S - \{s\}$
      - As before $S - \{s\}$ is a set of size n, but now we're interested in a subset of size k
      - So $\binom{n}{k}$ ways to choose

Pascal's triangle

- Recall the identity: $\binom{n+1}{k} = \binom{n}{k-1} + \binom{n}{k}$
- We can visualize the binomial coefficients in the following geometric way:
  - A "triangle" in which the k-th row (starts with 0) has entries C(k,0)…C(k,k)
  - Notice, by Pascal's identity, we have the following relations (pair of arrows indicate sum)



1=C(0,0)

1=C(1,0) 1=C(1,1)

1=C(2,0) 2=C(2,1) 1=C(2,2)

- Where, for example, in third row (i.e., n+1=2, k=1) we have C(2,1)=C(1,0)+C(1,1)
- And so on…

Generalization: Vandermonde's Identity

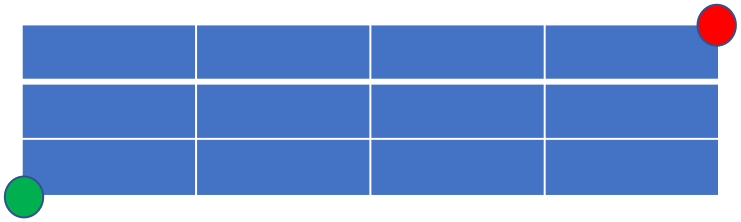
- Thm. For all $, m, n \geq r > 0, \binom{m+n}{r} = \sum_{k=0}^{r} \binom{m}{r-k}\binom{n}{k}$
- Pf. Let A, B be disjoint sets of size m, n.
    - LHS counts the number of subsets of size r of $A \cup B$
    - But there's another way to count this…
        - Can also count this by counting for distinct $0 \leq k \leq r$, the number of ways to select a subset of size $r - k$ from A and a subset of size k from B
- Cor. $\binom{2n}{n} = \sum_{k=0}^{n} \binom{n}{k}\binom{n}{k}$
- Pf. In above equality, consider two equal sized sets, so let $m = r = n$ and appeal to symmetry (recall: $\binom{n}{k} = \binom{n}{n-k}$)

Final binomial identity

- Thm. Let $n \geq r > 0$. Then $\binom{n+1}{r+1} = \sum_{j=r}^{n} \binom{j}{r}$
- Pf. LHS is way to pick r+1 elements from set *S*, of size *n+1*.
  - There's another way to do it!
  - To simplify, let's say the elements of S are {0,1,2,…,n}
  - First question: What is the *smallest* max element a subset of S of size r+1 could have?
    - It's r
    - How many subsets of S of size r+1 have greatest element r?
    - It's C(r,r)=1
  - In general, the number of subsets of S of size $r + 1$ that have the largest element $j \geq r$ is $C(j, r)$ -- why?
    - We're effectively fixing a single element, j, of the subset (the largest element)
    - Since the subset is of size r+1, we have r left to choose from 0,…,j-1

e. Generalized permutations and combinations

Motivating example



- Suppose we have a map, consisting of a grid with 3 north/south blocks and 4 east/west blocks. We want to get from the lower left to the upper right. How many different ”efficient” ways of doing this are there?
  - Efficient means in each step we take steps going only north or east.
  - Notice that each efficient path is of length 7 blocks (which could be north or east), and exactly 3 of these steps needs to going north.
  - So the answer is $\binom{7}{3}$
- Helpful way of thinking about this:
  - Consider an abstract program that has two sorts of instructions: “move north” and “move east”
  - A “successful program” makes 7 such instructions with exactly 3 “move north” instructions
  - So as before, there are $\binom{7}{3}$ such “successful programs”

Another counting example

- Ex: How many different ways can we select five bills from cash drawers that each have at least five each of \$100, \$50, \$20, \$10, \$5, and \$1-dollar bills?
- Answer:
  - Imagine six drawers of cash, from left to right. Program starts at far left pile and must end at the far right
  - The program that has two instructions: “dispense”, to dispense a bill from current drawer, and “shift right”, to move to the next drawer.
  - As before, we consider ”successful programs” that use 10 instructions (of either “dispense” or “shift right”) and use exactly 5 “shift right” instructions
  - So there are are $\binom{10}{5}$ such ”successful programs”
  - Why did we require the program to get to the far right?

Generalizing this: "Combinations with replacement"

- These examples can be generalized by the following scenario:
  - Suppose we have *n* distinct piles of elements (e.g., the six bill denominations).
  - We want to know how many ways we can choose *r* elements
  - But each time we draw a bill, we replace it, so it can be drawn again
  - Formally, we're interested in the number of *r*-combinations that can be formed from a set with n elements, *with replacement* (aka repetition).
- Thm. There are $\binom{n+r-1}{r}$ r-combinations of a set with n distinct elements, with replacement.
- Pf. As before, consider designing a program with two different instructions "dispense" and "next"
  - A "successful program" makes *r* dispense instructions and *n-1* next instructions (because we start on the first element)
  - So there are clearly $\binom{n+r-1}{r}$ different "successful programs"

Another example of “Combinations with replacement”

- Ex. How many natural number solutions to $x_1 + x_2 + x_3 = 11$?
  - This is equivalent to having a set of three distinct elements (one for each variable) and needing to draw 11 of them.
  - So the answer is the number of 11-combinations of 3 elements with repetition.
  - Again, this is $\binom{n+r-1}{r}$, where $r = 11, n = 3,$ so $\binom{11+3-1}{11} = \binom{13}{11} = 78$

Permutations with repetition

- Motivating ex: How many *distinguishable* permutations of the word "SUCCESS" are there?
- Pf. We'll give a combinatorial proof, equating two different ways to count this quantity.
  - First, there are seven letters, so if they were all distinct (which they *are not*!) we would have 7! different permutations of the letters
  - But of course we have to factor in the multiplicities/repetitions of each letter, so accounting for the repeated "C" and "S"'s we have, letting n be the answer:
  - $7! = n \cdot 3! \cdot 1! \cdot 2! \cdot 1! \Rightarrow n = 420$
- This is so important, the concept has it's own notation:
  - $\frac{n!}{n_1!n_2!\ldots n_k!} = \binom{n}{n_1, n_2, \ldots, n_k}$
  - These are called "multinomial coefficients" and can be used to generalize the binomial theorem to n variable – see problem on the homework!

Announcements

- No class next week due to Thanksgiving!
- Correspondingly the current homework is due Monday the 27th
- The following homework will be abridged and due December 1 due to reading period

f. The pigeonhole principle

Basics

- Thm. Let $A_1, A_2, \ldots, A_m$ be disjoint sets so that $|\bigcup_i A_i| > m$. Then there exists an A_i so that $|A_i| \geq 2$.
- Pf. By contrapositive
  - What's the contrapositive of this statement?
  - If all $|A_i| \leq 1$ then $|\bigcup_i A_i| = \sum_i |A_i| \leq \sum_i 1 = m$ (contradiction!)
- This simple theorem is sometimes called the pigeonhole principle (it's unclear exactly why)
- Cor. Suppose A,B are finite sets where $|A| > |B|$ and $f: A \to B$. Then f cannot be one-to-one.
- Pf. Let $b_1, b_2, \ldots, b_k$ be the elements of set B.
  - Define the sets $A_i = f^{-1}(b_i)$
    - i.e., the set of those elements in A that f maps to b_i
    - So there are $|B|$ sets and $|\mathrm{A}| = |\bigcup_i A_i| > |B|$
    - Notice in a one-to-one function these sets would all have…?
  - So we can apply the pigeonhole principle to the sets $\{A_i\}$ (of which there are $|B|$) and conclude there must exist an A_i so that $|A_i| \geq 2$ This means the function f cannot be one-to-one!

Examples

- Ex: “Picking socks!”:
  - Suppose you have 3 colors of socks
  - The socks are all jumbled together in an unsorted sock bag
  - How many socks do we need to pull out to guarantee we get a matching pair?
  - By PHP, 4 – what are the “pigeons” and what are the “holes” in this example?
- Ex. “Hand shaking”
  - Suppose there is a meeting involving n people
  - As the meeting closes there’s a round of hand shaking
  - Show that there exist two people who shook the same number of hands.
    - Let’s map people to number of hands they shook
    - Let’s start by obvious strategy: person i shook i hands (starting with 0, through n-1)
    - But this can’t be! Why?
  - So either someone shook n-1 hands or someone shook no hands, and not both!
  - The point is that there are only $n - 1$ choices of # of hands to shake (i.e., $\{0, \ldots, n-2\}$) and n people and so the pigeonhole principle says someone must have shaken same number of hands!!

Generalized PHP

- Thm. Let $A_1, A_2, \ldots, A_k$ be disjoint sets, such that $|\bigcup_i A_i| = n$. Then there exists an A_i so that $|A_i| \geq \left\lceil \frac{n}{k} \right\rceil$.
  - Where "ceiling" $\lceil r \rceil$ notation means least integer greater than or equal to r.
- Pf. Let's prove the contrapositive.
  - Suppose all i are so that $|A_i| < \left\lceil \frac{n}{k} \right\rceil \Rightarrow |A_i| < \frac{n}{k}$
  - Then $|\bigcup_i A_i| = \sum_i |A_i| < \sum_i \frac{n}{k} = k\left(\frac{n}{k}\right) = n$, so $|\bigcup_i A_i| \neq n$

Generalized PHP example

- Ex. (The sock problem for centipedes!) Let us recall that centipedes are supposed to have 100 legs. Let's say that the jumbled bag of socks still has three different colors. How many socks does the centipede need to draw to ensure there's 100 same color socks?
- Claim: answer is least n so that $\left\lceil \frac{n}{3} \right\rceil = 100$ which is equivalent to requiring $\frac{n}{3} > 99$ so $n > 297$, so answer is 298. Why?
  - By GPHP!
  - Let A_1 be set of picked socks that are red, A_2 blue, A_3 green.
  - Then suppose that we have $\bigcup_i A_i = 298$ (i.e., centipede picked in total 298 socks).
  - Then at least one set (color) $A_i \geq \left\lceil \frac{298}{3} \right\rceil > 99$ (which is 100).

5. Probability theory

a. Basic probability theory

Basics

- **Def.** A **probability space** is a finite set $\Omega \neq \emptyset$ and a function $\Pr: \Omega \to \mathbb{R}$ so that:
  - $\forall \omega \in \Omega, \Pr[\omega] \geq 0$
  - $\sum_{\omega \in \Omega} \Pr[\omega] = 1$
- **Def.** We say Ω is the **sample space**, and Pr is the **probability distribution**.
- **Def.** An **event** is a subset of the sample space $\mathrm{A} \subseteq \Omega$
  - Def. The **atomic** events are singleton sets $\{\omega\}$ for which $\Pr[\{\omega\}] = \Pr[\omega]$
  - The probability of the event A is denoted Pr[A]=$\sum_{\omega \in A} \Pr[\omega]$
  - Def. An event is **trivial** if its probability is 0 or 1
- Notice if $A_1, A_2, \ldots A_k$ are disjoint subsets of Ω (these are events), then $\Pr[A_1 \cup A_2 \ldots \cup A_k] = \Pr[A_1] + \Pr[A_2] + \cdots + \Pr[A_k]$
- **Def.** The **uniform distribution** over a sample space Ω sets $\Pr[\omega] = 1/|\Omega|$ to all $\omega \in \Omega$

More properties of probability…

- **Def.** Events A, B are disjoint if $A \cap B = \emptyset$
- **General thm.** For any two events (*that need not be disjoint*), A, B: $\Pr[A] + \Pr[B]$ = $\Pr[A \cup B] + \Pr[A \cap B]$
- Where have we seen this before?
    - This is effectively the same as set inclusion-exclusion!
- **Lemma:** $\Pr\left[\cup_{i=1}^{k} A_i\right] \leq \sum_{i=1} \Pr[A_i]$
  - When does equality hold?
    - Iff the A_i events are (pairwise) disjoint
  - How would we prove this?

Conditional probability

- **Def.** If A, B are events, then the **conditional probability** of A relative to B is:
  - $\Pr[A|B] = \frac{\Pr[A \cap B]}{\Pr[B]}$
- **Ex.** Consider rolling 3 dice. What is the probability that the *sum* of the dice is 9?
  - How many dice rolls sum to 9? 25
  - How many possible dice rolls?
    - 6^3=216 – this is the size of our sample space
  - Probability is 25/216
- Ex: What's the probability that the first die shows 5 given that sum of the three dice is 9?
  - Let A be event that the first die shows 5, let B be event corresponding to sum of the three dice is 9
  - Then $\Pr[A \cap B] = \Pr[\{5,2,2\}] + \Pr[\{5,3,1\}] + \Pr[\{5,1,3\}] = \frac{3}{216}$
  - $\Pr[A|B] = \frac{Pr|A \cap B|}{\Pr[B]} = \frac{\frac{3}{216}}{\frac{25}{216}} = \frac{3}{25}$

Independence

- *Intuitively*, two events are independent if the occurrence of one does not change the probability of the other occurring
- **Def.** Events A, B are independent if $\Pr[A \cap B] = \Pr[A] \cdot \Pr[B]$
- **Claim:** Events A,B are independent if $\Pr[A|B] = \Pr[A]$ (why?)
  - **Recall:** $\Pr[A|B] = \frac{\Pr[A\cap B]}{\Pr[B]} \Rightarrow \Pr[A \cap B] = \Pr[A|B]\Pr[B]$
    - Since we are assuming $Pr[A|B] = Pr[A] \Rightarrow Pr[A \cap B] = Pr[A]Pr[B]$

Correlated events

- Events A,B are independent if $\Pr[A \cap B] = \Pr[A] \cdot \Pr[B]$
- Def. The events A,B are positively correlated if $\Pr[A \cap B] > \Pr[A] \cdot \Pr[B]$
- Def. The events A,B are negatively correlated if $\Pr[A \cap B] < \Pr[A] \cdot \Pr[B]$
- Ex. Consider the roll of standard dice with 6 sides, and let event A be the “roll is even={2,4,6}”, B be the “roll is prime={2,3,5}”
  - Are A and B independent, or correlated (and if so, how?)
  - $\Pr[A \cap B] = \Pr[\{2\}] = \frac{1}{6} < \Pr[A] \cdot \Pr[B] = \frac{1}{2} \cdot \frac{1}{2}$
  - So A and B are negatively correlated!

Pairwise and mutual independence

- Both terms refer to a collection of (generally) more than 2 events
- Def. Events $A_1, A_2, \dots, A_k$ are pairwise independent if $\forall i, j \ A_i \ and \ A_j$ are independent so $\Pr\left[A_i \cap A_j\right] = \Pr[A_i] \cdot \Pr[A_j]$
- Def. Events $A_1, A_2, \dots, A_k$ are mutually independent if for all subsets of the sets, $\forall \ I \subseteq \{1, \dots, k\}, \Pr[\bigcap_{i \in \mathrm{I}} A_i] = \prod_{i \in I} \Pr[A_i]$
- Which of these is stronger?
- (Homework question): Can you think of a collection of events that are pairwise but *not* mutually independent?

Announcements

- Homework 6 due (now)!
- Homework 7 assigned and due Friday before class! (No late homework!)
- Final exam on Thursday of exam week, December 7, 5:30 PM – 7:30 PM
  - In this room (Stuart 102)!

b. Random variables

Random variables

- A random variable is a function $f\colon \Omega \to \mathbb{R}$ where Ω is the sample space of a probability space.
- Interestingly, random variables are neither random nor variables!
- Def. If X is a random variable, we define $\Pr[X = r] = \Pr[\{\omega \in \Omega | X(\omega) = r\}]$
  - i.e., it's the probability of all atomic events ω that X maps to r.
- Ex. Consider the probability space that considers of flipping a fair coin three times. Let X be the random variable that counts number of heads
  - Can label elements of sample space by a string of three H or T's. e.g., {HTT} or {HHH}
  - What's the Pr[X=3]?
    - Pr[X=3]=Pr[{HHH}]=$\frac{1}{8}$

Bernoulli Trials

- Def. A Bernoulli trial is a random variable B whose codomain is {0,1}
  - We sometimes call the event $\{\omega \in \Omega | B(\omega) = 1\}$ the success event.
  - We sometimes call the event $\{\omega \in \Omega | B(\omega) = 0\}$ the fail event.
- "The point of Bernoulli trials is to repeat them"
- And the key question is to count the number of successes, k, in n trials.
  - Key point: this itself (number of successes of a Bernoulli r.v) is a random variable!

Bernoulli trial example!

- E.g., consider flipping a biased coin which comes up heads with probability p and tails with probability 1-p
  - So consider a random variable $\mathrm{X}(\{H\}) = 1$ and $X(\{T\}) = 0$
  - Then consider flipping n coins, and consider new r.v $Y = \sum_i X_i$ where each $X_i = X$
  - For $k \leq n$, what is the $\Pr[Y = k]$?
    - We consider singleton events in the sample space to be strings of H,T of length n
    - Then how many different atomic events correspond to seeing k H's?
    - It's $\binom{n}{k}$
    - And notice that each sequence of tosses occurs with probability $p^k(1-p)^{n-k}$
    - So answer is $\binom{n}{k} p^k(1-p)^{n-k}$
- We've seen this before… (where?)

c. Expectation

Expectation basics

- Let (Ω, Pr) be a probability space and let X be a random variable. Then expected value of X is defined:
  - $E[X] = \sum_{\omega \in \Omega} X(\omega) \Pr[\omega]$
- So Expectation can be seen as a sort of weighted average (why?)
- Ex: Consider a fair six-sided dice. Let the random variable X denote the value of the roll. Then $E[X]$=?
  - $1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6}$ = 21/6

Linearity of Expectation

- Thm. Let (Ω, Pr) be a probability space and $X_1, X_2, \dots, X_n$ be random variables over Ω, and let $X = \sum_i X_i$, then:
  - $E(X) = \boldsymbol{E}(\boldsymbol{X_1} + \boldsymbol{X_2} + \cdots + \boldsymbol{X_n}) = \sum_{\boldsymbol{i}} \boldsymbol{E}(\boldsymbol{X_i})$
- Pf.
  - $E(X) = \sum_{\omega \in \Omega} X(\omega) \mathrm{Pr}[\omega] = \sum_{\omega \in \Omega} (X_1(\omega) + X_2(\omega) + \cdots + X_n(\omega)) \mathrm{Pr}[\omega]$
    - $= \sum_{\omega \in \Omega} \sum_i X_i(\omega) \mathrm{Pr}[\omega]$
    - $= \sum_i \sum_{\omega \in \Omega} X_i(\omega) \mathrm{Pr}[\omega]$
    - $= \sum_i E[X_i]$

Linearity of Expectation example

- Ex. There is a dinner party where n men check-in their hats. The person at the counter is lazy and so gives back the hats to the men uniformly at random (i.e., gives each man a random hat). What is the expected number of men who get their own hat back?
  - Let R be the random variable corresponding to the number of men that get their hat back
  - Note that it's not obvious how to compute $E[R] = \sum_k k \cdot \Pr[R = k]$ since this probability is quite a mess!
  - So we introduce "Indicator variables"
    - Let R_i be the random variable so that $R_i = 1$ if i-th man gets his own hat and 0 otherwise.
    - We know two things:
      - $R = \sum_i R_i$
      - $E[R_i] = \frac{1}{n}$ why?
        - Because each man is just as likely to get one hat as another and there are n men!
  - Now by linearity of expectation, $E[R] = E[\sum_i R_i] = \sum_i E[R_i] = \sum_i \frac{1}{n} = 1$

Announcements

- Homework 7 due on Friday, before class starts
- Final next Thursday, here in Stuart 102, at 5:30-7:30
- Instructor office hours after class Friday 10:20-11:30
- People who need accommodations for the final – we’ll send an email in the next few days, please email TAs if you don’t hear from us soon

Markov’s inequality

- Thm. [Markov] If X is a non-negative R.V then $\forall a > 0$,
  - $\Pr[X \geq a] \leq \frac{E[X]}{a}$
- Pf.
  - By definition $E[X] = \sum_{\omega \in \Omega} X(\omega) \Pr[\omega] \geq \sum_{\omega \in \Omega, \mathrm{X}(\omega) \geq a} X(\omega) \Pr[\omega]$ (why?)
    - Because we’re summing over fewer non-negative items!
  - $\geq \sum_{\omega, X(\omega) \geq a} a \cdot \Pr[\omega]$ (why?)
    - Because summing smaller things! $(X(\omega) \geq a)$
  - =$a \sum_{\omega, X(\omega) \geq a} \Pr[\omega]$ by linearity
  - $= aPr[X \geq a]$ (this is by definition of $\Pr[X \geq a]$ and then we divide both sides by a)
- Cor [Markov]: If X is a non-negative R.V then $\forall k > 0, \Pr\left[\mathrm{X} \geq kE[X]\right] \leq 1/\mathrm{k}$
- Pf.
  - Let $a = kE[X]$ in the above thm.
  - Then $\Pr[X \geq kE[X]] \leq \frac{E[X]}{kE[X]} = 1/k$
  - What is the intuition here? Hint: think of expectation as average!
  - i.e., if we have some numbers whose mean is k, there can’t be too many numbers much greater than k

Is expectation multiplicative?

- Recall that expectation is additive ("linear")
  - i.e., $E[X+Y] = E[X] + E[Y]$
- Is expectation multiplicative? i.e., is $E[XY] = E[X]E[Y]$?
  - Generally no!
  - Ex. Let $X = 1$ with probability ½ and $X = -1$ with probability ½
  - $E[X]^2 = \left(\frac{1}{2} \cdot -1 + \frac{1}{2} \cdot 1\right)^2 = 0$
  - But $E[X^2] = \frac{1}{2} \cdot (-1)^2 + \frac{1}{2} \cdot 1^2 = 1$
  - So $E[X]^2 \neq E[X^2]$ and so generally $E[XY] \neq E[X]E[Y]$
- However, notice that in this case the "two random variables" are not independent (i.e., say X,Y are independent random variables if $\Pr[(X = x) \cap (Y = y)] = \Pr[X = x] \cdot \Pr[Y = y]$ for all values x,y)
- In fact, not hard to see that X and Y independent implies that $E[XY] = E[X]E[Y]$

d. Variance

Definition of variance

- Def. Let X be a random variable on the probability space $(\Omega, \Pr)$. Then:
  - $Var[X] = E\left[\left(X - E(X)\right)^2\right] = \sum_{\omega\in\Omega}\left(X(\omega) - E(X)\right)^2 \cdot \Pr[\omega]$
- Intuitively, variance tells us how “spread X is from the expectation”
- Def. The standard deviation of X, denoted $\sigma(X) = \sqrt{Var(X)}$.
- Because of this, another notation for $Var[X]$ is $\sigma^2[X]$
- Thm. $Var[X] = E[X^2] - E[X]^2$
- Pf. $Var[X] = \sum_{\omega\in\Omega}\left((X(\omega) - E[X])^2 \cdot \Pr[\omega]\right)$ (by definition)
  - $= \sum_{\omega\in\Omega}\left((X(\omega)^2 - 2 \cdot X(\omega)E[X] + E[X]^2) \cdot \Pr[\omega]\right)$
  - $= \sum_{\omega\in\Omega}(X(\omega)^2 \cdot \Pr[\omega]) - 2E[X] \cdot (\sum_{\omega\in\Omega} X(\omega) \cdot \Pr[\omega]) + E[X]^2 \cdot \sum_{\omega\in\Omega} \Pr[\omega]$
  - $= E[X^2] - 2E[X]^2 + E[X]^2$ (i.e., since there are two $E[X]$ terms in the middle)
  - $= E[X^2] - E[X]^2$ (i.e., just by algebra)

Chebyshev's inequality

- Thm [Chebyshev]. Let X be a random variable (not necessarily non-negative!). Then for all $a > 0$:
  - $\Pr[|X - E[X]| \geq a] \leq \frac{Var[X]}{a^2}$
- Pf.
  - Application of Markov's inequality
  - Let $Y = (X - E[X])^2$
  - So by definition of variance, $E[Y] = Var[X]$
  - Markov's inequality applies to Y (why?) so:
    - $\Pr[Y \geq a^2] \leq \frac{E[Y]}{a^2}$
  - But now: $Y \geq a^2 \Leftrightarrow (X - E[X])^2 \geq a^2$ by definition of Y
    - and this is true iff $|X - E[X]| \geq a$ (by taking square root)
  - So we have $\Pr[|X - E[X]| \geq a] \leq \frac{E[Y]}{a^2} = \frac{Var[X]}{a^2}$ by definition of $E[Y]$

Variance is (sometimes) linear

- If X and Y are independent random variables, then $V(X+Y) = V(X) + V(Y)$
- Pf.
    - $V(X+Y) = E[(X+Y)^2] - E[X+Y]^2$
        - $= E[X^2 + 2XY + Y^2] - (E[X] + E[Y])^2$
        - $= E[X^2] + 2E[XY] + E[Y^2] - E[X]^2 - 2E[X]E[Y] - E[Y]^2$
        - $= (E[X^2] - E[X]^2) + (E[Y^2] - E[Y]^2)$
        - $= V[X] + V[Y]$
    - Which step used independence?
        - The second to the third line (so that $2E[XY] - 2E[X][Y] = 0$)
        - Because X and Y independence implies that $\mathrm{E[XY]} = E[X]E[Y]$

Variance example

- Let random variable X be the sum of rolls of 2 independent fair dice. What are the Expectation, Variance, Standard Deviation?
  - Expectation?
    - $E[X] = E[roll] + E[roll] = 2E[roll] = 2\left(1\left(\frac{1}{6}\right) + 2\left(\frac{1}{6}\right) \ldots + 6\left(\frac{1}{6}\right)\right) = 2\left(\frac{7}{2}\right) = 7$
  - Variance?
    - $Var[roll] = E[roll^2] - E[roll]^2 = (\sum_i \frac{i^2}{6}) - \left(\frac{7}{2}\right)^2 = \frac{35}{12}$
    - And by independence, Var[X]=2 $\cdot \frac{35}{12} = 35/6$
  - Standard deviation?
    - $\sqrt{Var[X]} = \sqrt{35/6}$

6. Asymptotics

Basics of Asymptotics

- What is an algorithm?
  - A sequence of “elementary computation steps” that solves the specified problem on *any* input
- How should we measure the complexity (or difficulty) of an algorithm?
- We’re generally interested in how many steps the algorithm takes as a function of the input length.
  - I.e., as the length of the input grows, how does the number of steps (called the *running time*) of the algorithm scale?
  - E.g., in Euclid’s algorithm for gcd – if we want gcd(m,n) takes log(min(m,n)) steps
- To understand the running time of algorithms we’ll need a language to describe the growth of functions for increasing n.
- E.g., intuitively, it’s clear that 2^n “seems” larger than n^{10}. But it’s not true for sufficiently small n. How do we formalize this?

“Big Oh” notation

- Let $f, g: \mathbb{N} \to \mathbb{R}$. We say that $f(x) = O\big(g(x)\big)$ if there exist constants c and k such that:
  - $\forall x \geq k, |f(x)| \leq c \cdot |g(x)|$
  - i.e., $g(x)$ grows as fast or faster than $f(x)$, past some initial finite portion, and up to a constant multiple
- Example: $x^2 + 2x + 1 = \mathrm{O}(\mathrm{x}^2)$. Why?
  - E.g., Let $k = 1, c = 4$.
  - We know $x^2 + 2x + 1 \leq x^2 + 2x^2 + x^2$ $\text{(since each term on LHS is dominated)}$
  - But $\text{RHS} = 4x^2$

“Big omega” and “Big theta”

- Let $f, g: \mathbb{N} \to \mathbb{R}$. We say that $f(x) = \Omega(g(x))$ if there exist constants c, k so that:
    - $\forall x \geq k, |f(x)| \geq c \ \cdot |g(x)|$
- Let $f, g: \mathbb{N} \to \mathbb{R}$. We say $f(x) = \Theta\big(g(x)\big)$ if $f(x)$ is both $O\big(g(x)\big)$ and $\Omega(g(x))$
    - How can both of these definitions hold simultaneously? Can you give an example?
    - Ex: $f(n) = 2n$ is both $O(n)$ and $\Omega(n)$
    - However, $f(n) = \log n$ is $O(n)$ but not $\Omega(n)$