

Survivor's Manual for CIS 1600

HITMAN7128

May 11, 2023

What are the basic mathematical concepts and techniques needed in computer science? This course provides an introduction to proof principles and logics, functions and relations, induction principles, combinatorics and graph theory, as well as a rigorous grounding in writing and reading mathematical proofs.

— *Description of CIS 1600 in course catalog*

Contents

1	Introduction	3
2	General Course Advice	4
3	Counting	4
3.1	Addition Rule	4
3.2	Multiplication Rule	5
3.3	Applied Counting	6
3.3.1	Subsets	6
3.3.2	Words and Strings	6
3.3.3	Permutation	6
3.4	Complementary Counting	7
3.5	Combinations	8
3.6	Stars and Bars	8
3.7	Anagrams	9
3.8	Pascal's Triangle	10
3.9	Combinatorial Proof	11
3.10	Functions	12
3.11	Principle of Inclusion/Exclusion	12
3.12	Pigeonhole Principle	13
4	Proofs	13
4.1	Odd and even	13
4.2	Divisibility and Primes	14
4.3	Logic Behind Proofs	15
4.4	Predicates and Quantifiers	15
4.5	Negation, Converse, and Contrapositive	16
4.6	Proof by Contradiction	16
4.7	Induction	17
4.8	Strong Induction	20
5	Terminology and Notation	21
5.1	Sets	21
5.1.1	Introductory	21
5.1.2	Intermediate	22
5.1.3	Set Builder Notation	23
5.2	Logical Structure	23

5.3	Truth Tables	24
5.4	Functions	25
5.5	Bijections	25
6	Probability	26
6.1	Probability Space and Events	26
6.2	Uniformity	27
6.3	Bernoulli Trials	28
6.4	Probability Properties	28
6.5	Independence	29
6.6	Conditional Probability	30
6.7	Chain Rule	30
6.8	Random Variables	31
6.9	Expectation	31
6.10	Linearity of Expectation and Indicators	32
6.11	Independent Random Variables	33
6.12	Variance	34
6.13	Binomial Distribution	34
7	Graphs	35
7.1	Basics and Definitions	35
7.1.1	Introductory	35
7.1.2	Intermediate	35
7.2	Special Graphs	36
7.3	Handshaking Lemma	37
7.4	Walks and Paths	38
7.5	Connected Components	38
7.6	Cycles	39
7.7	Trees, Forests, and Leaves	39
7.8	Cut Edge	40
7.9	Spanning Trees	41
7.10	Coloring	41
7.11	Cliques and Independent Sets	42
7.12	Directed Graphs	43
7.13	Reachability and Strong Connectivity	43
7.14	Directed Acyclic Graphs (DAGs)	44
7.15	Rooted Tree	45
7.16	Binary Tree	46
8	Test Strategies	46
8.1	General	46
8.2	Midterm 1	47
8.3	Midterm 2	47
8.4	Final	48

§1 Introduction

Keep in mind that I took this course specifically during Spring 2023 with Val Tannen. Material and conditions might be different depending on the year and if you take it with Rajiv in the fall.

Anyway, this course has a reputation for being one of the hardest at UPenn relative to its level. But it helps to have the preconceptions of why it is so dreaded so that you can tackle the course with the right mindset:

- The thought process behind solving problems in this course is not something that can be easily described with black and white principles.
- The TAs are extremely nitpicky about grading on the homeworks and tests.
- Concepts are not always explained in plain English. The course is not exactly friendly to those who don't have a background in math.
- If you fail to include certain buzzwords they're looking for in the homework, they won't hesitate to take points off. They don't always tell you what buzzwords they're looking for, too.
- The sheer amount of material thrown at you is insane and the pace of the course is unforgivingly fast.
- The midterms and final are closed notes and are fast paced. See more on "Test Strategies."

What this guide is supposed to do:

- Help you understand concepts in Layman's terms
- Show examples of the concepts in action
- Point out when and where a concept is applicable to build a thought process on tackling HW and test problems
- Make you feel more comfortable about what you're learning to give you momentum in the course

That being said, I would greatly appreciate constructive criticism on people who use this guide. If a section is poorly explained, I have no issue in improving it.

Some boxes to keep in mind as you read this guide:

Example 1.1

Examples boxes will have a simple problem that demonstrates the concept at a basic level.

Caveat 1.2. Caveat boxes will have information on what you may need to say in a write-up of a homework problem. Pay attention to these.

Lemma 1.3

Lemma/theorem boxes will have claims we want to prove and take more work or reasoning to show than an example.

Advice 1.4 — Advice boxes will have information on what clues to look for in a problem that may warrant using the concept I explained.

§2 General Course Advice

This goes for the homework and how you could utilize the resources you have available.

- They mean it when they say start the homework early and go to office hours early.
- Don't feel discouraged if you don't see the solution immediately on a HW problem. Sometimes it takes some thought or trial and error to arrive at a complete solution.
- **In general, if you're using a rule, principle, or technique in a solution to a HW or test problem, explicitly namedrop it.**
- You never need to prove anything that is proven in the lectures or recitation.
- When in doubt, ask on Piazza if you need to prove something.
- If you're stuck, try looking back on the problems solved during the lecture. The intuition can give hints on what to consider.
- Be comfortable with converting words to math and vice versa.
- The first third of the course is usually people getting used to the format and circumstances. During the second third, people generally know what to expect. But the last third is a difficulty spike because graph theory problems are abnormally challenging.

§3 Counting

In simplest terms, counting problems involves finding how many of this satisfies some condition. But the "this" can vary. It could be sandwiches, sequences, arrangements, etc.

Counting will primarily be in the first third of the course, where things being counted and methods for counting will be introduced to you.

§3.1 Addition Rule

The Addition Rule is a way to add up things that we sort into different cases.

Example 3.1

Suppose in my collection of cards, I have 20 red cards, 23 blue cards, and 24 green cards. How many cards do I have total?

Intuitively, I have $20 + 23 + 24 = 67$ cards, total right? That's true, but we need to take a closer look on why we can add these. This introduces the concept of disjoint cases and exhaustive cases.

The cases being "disjoint" means there are no overlap between any of the cases. So we have 20 in the sum, we can think of that 20 as the cards in the case where the card is red. Similarly, the 23 comes from the case when the card is blue, and 24 comes from the case when the card is green. But a card isn't more than one color, so no card belongs to more than one case. Imagine if a card belonged to more than one case. A card would be counted excessive times, disrupting the total. That's what it means for the cases to be disjoint: the thing being counted is never included in more than one case.

Sometimes, you'll find no matter what you do, you cannot create cases that are disjoint. We'll see a tool later on that will help with just that.

Now, onto what "exhaustive" means. Well, what if we just did $20 + 23 = 43$, accounting for only the red and blue cards? The case where the card is red is disjoint from the case where the card is blue. There's no overlap. But we missed a possibility! We need to cover every possibility for the correct total. That's what it means for the cases to be exhaustive: every possibility of the thing being counted is covered.

The example above is a simple application of the Addition Rule in action. Usually, it will be harder to establish what the cases are and/or count the number in each case. But the basic idea of adding up each possibility exactly once still applies.

A more general statement is, suppose you have some things in k cases that are disjoint and exhaustive (where k is some integer). You have t_1 things in the first case, t_2 things in the second case, and so on, until t_k things in the k th case. Then, the number of things you have total is $t_1 + t_2 + \cdots + t_k$. In the cards example, $k = 3$ with $t_1 = 20$, $t_2 = 23$, and $t_3 = 24$. **You cannot use the Addition Rule if the cases are not disjoint or exhaustive.**

Caveat 3.2. Always cite the Addition Rule (abbreviated as “AR”) whenever you use it in the course, ever. Also, always say the cases are disjoint and exhaustive when appropriately using it.

§3.2 Multiplication Rule

The Multiplication Rule helps to count things that require a procedure with steps.

Example 3.3

For the dinner, I have to select 1 entree and then 1 dessert. The menu has 9 different entrees and 4 different desserts. How many ways can I select my dinner?

So to select my dinner, I have two steps: choose my entree and then choose my dessert. I have 9 ways to select my entree, and then 4 ways to select my dessert. The important thing to note here is, the number of ways I do the second step does not effect the number of ways I do the first step. Thus, I have $9 \cdot 4 = 36$ ways to choose my dinner.

What allows me to use the Multiplication Rule is that the steps are “independent.” This means that what happens in one step never effects the number of ways to perform any of the other steps. If the steps are not independent, you cannot use the Multiplication Rule. An easy way to check if the steps are independent is, for each rule in the procedure, does the number of ways to perform this step change depending on what happened in prior steps?

Now, keep in mind, the choices available may be effected by prior steps. But that doesn't matter as long as the number of ways to do the step doesn't change.

Example 3.4

I have 10 different cookies. I want to give one to my sister and then one to my brother. How many ways can I give them cookies?

Suppose one of the cookies was a chocolate chip one and another was a peanut butter one. The first step would presumably be to give my sister a cookie. Suppose I give her the chocolate chip cookie. Now, I no longer have the chocolate chip cookie to give to my brother, but I can still give him the peanut butter cookie. Now, suppose I give my sister the peanut butter cookie. Now, I no longer have the peanut butter cookie to give to my brother, but I can still give him the chocolate chip cookie. So it is clear that based on what I give my sister, the choices on what I give my brother are different. However, notice that when I give my sister a cookie, regardless of which one it is, I have 9 cookies left. And one of those 9 cookies goes to my brother.

So even if what I had available to do the second step (give my brother a cookie) changes, the number of ways does not. This is enough for the steps to be independent. In this example, the procedure would be to first give my sister a cookie and then my brother. There are 10 ways to do the first step and then 9 ways to do the second step. Since the steps are independent, by the Multiplication Rule, I have $10 \cdot 9 = 90$ ways total.

A more general statement is, suppose your procedure has k independent steps (where k is some integer). You have w_1 ways to do the first step, w_2 ways to do the second step, and so on, until w_k ways to do the last step. Then, the total number of outcomes is $w_1 \cdot w_2 \cdot \dots \cdot w_k$. In the dinner example, $k = 2$ with $w_1 = 9$ and $w_2 = 4$. **You cannot use the Multiplication Rule if the steps are not independent.**

Caveat 3.5. Like AR, always cite the Multiplication Rule (abbreviated as “MR”) whenever you use it in the course, ever. Also, always say the steps are independent when appropriately using it.

Advice 3.6 — You will often see the AR and MR together in action. Sometimes, you may be creating a procedure and then realize the steps aren't independent partway through. Perhaps divide into cases that address when the steps no longer become independent. Use MR for the procedure in each separate case and add up the cases with AR at the end.

§3.3 Applied Counting

§3.3.1 Subsets

Suppose a set S has n elements. It helps to consider how you would construct a subset. A subset of S can only use elements from S . Intuitively, for each element of S , either we choose to include it in the subset or we don't. Also, our choice for how we handle one element doesn't effect the others. So we can construct a procedure to create a subset:

- 2 ways to handle the first element (include in the subset or we don't)
- 2 ways to handle the second element
- ...
- 2 ways to handle the n th element

The steps are independent, so by MR, we have $\underbrace{2 \cdot 2 \cdot \dots \cdot 2}_{n \text{ twos}} = 2^n$ subsets.

§3.3.2 Words and Strings

Think of a string as a sequence of stuff. A word is essentially a string of letters.

Suppose you have n possible characters to work with. How many strings can you create that are k characters long?

Again, it helps to consider how you make a word. A word is impacted by what is in each position, so maybe consider positions as steps in the procedure. Intuitively, our choice for one position doesn't effect what we do in another position. So we can construct a procedure to create a string:

- n ways to choose the first character (because there are n possibilities)
- n ways to choose the second character
- ...
- n ways to choose the n th character

The steps are independent, so by MR, we have $\underbrace{n \cdot n \cdot \dots \cdot n}_{k \text{ copies of } n} = n^k$ subsets.

§3.3.3 Permutation

Definition 3.7 (Permutation). Given a set of distinct elements, a permutation is simply some ordering of the elements.

Take the set $\{1, 2, 3\}$. Examples of permutations of the set would be $\{1, 2, 3\}$ itself, $\{1, 3, 2\}$, $\{2, 1, 3\}$, and $\{3, 2, 1\}$.

Let's count how many permutations there are of n elements. Again, consider how we would construct it. So we know it is an ordering of n elements, so we can consider n blank slots that we have yet to fill in.

First, we have n choices for the first element. Whatever we chose for that first element, we can no longer use. No matter what we chose, we only have $n - 1$ elements left. So we decide the second element, which can be done in $n - 1$ ways. Whatever we chose here, we can no longer use. Now, we have $n - 2$ elements left, regardless of the prior two choices. Noticing the independence here? Continuing the logic, this inspires the procedure:

- n ways to choose the first element
- $n - 1$ ways to choose the second element
- $n - 2$ ways to choose the third element
- \dots
- 1 way to choose the n th character

The steps are independent, since after each step, you always lose exactly 1 possible element that could have been in the next slot. So by the MR, we have $n \cdot (n - 1) \cdot (n - 2) \cdot \dots \cdot 1$ possible permutations. In particular, this quantity is noted as $n!$ or “ n factorial.”

Advice 3.8 — In permutation counting problems where you're given that some two elements are consecutive, like for example, a and b must be together, it helps to think of a and b as a single block. Of course, if applicable, don't forget to account for ordering a and b within the block.

Definition 3.9 (Partial Permutation). Given a set of distinct elements, a partial permutation is made by using some (not necessarily all) of the elements and ordering them in some way.

Consider a partial permutation of length k from a set of n elements. The procedure is essentially the same as a regular permutation, except we cut early, since we are only using k elements. So once we get to the k th step, we stop.

- n ways to choose the first element
- $n - 1$ ways to choose the second element
- \dots
- $n - (k - 1)$ way to choose the n th character

The steps are independent, since after each step, you always lose exactly 1 possible element that could have been in the next slot. So by the MR, we have $n \cdot (n - 1) \cdot \dots \cdot (n - (k - 1))$ possible permutations. This quantity can be rewritten as $\frac{n!}{(n-k)!}$.

Advice 3.10 — As you can see from these examples, thinking about how you would physically construct the thing from a blank slate can help with coming up with the procedure. Sometimes, the procedure might not always have independent steps, but that's okay. You can use cases to get around where independence breaks in your procedure.

§3.4 Complementary Counting

Sometimes, when counting the number of things that satisfy some condition, it might be easier to count the number of things we don't satisfy the condition instead. So essentially, we can get what we want by taking the number of things total (regardless if the things satisfy the condition) and subtract off the things that don't satisfy the condition.

Example 3.11

How many subsets of $\{1, 2, 3, 4, 5, 6\}$ are not empty?

The number of “things” (subsets in this case) total is $2^6 = 64$. Now, we want to subtract the “things” that don't satisfy the condition. A subset that doesn't satisfy the condition is a subset that is empty, which there is only 1 of. So by complementary counting, our answer is $64 - 1 = 63$.

Advice 3.12 — Whenever you're given a problem where something has to have "at least one of this," that's a good sign to try complementary counting. Because the stuff you don't want will be the ones that have none of the "this" at all. It's also a good thing to at least try if you're ever stuck on a counting problem.

§3.5 Combinations

The number of ways to choose r things from a collection of n different things is denoted as $\binom{n}{r}$. This is read as " n choose r ." The formula is

$$\binom{n}{r} = \frac{n!}{r!(n-r)!},$$

which you should memorize because this will come up a lot.

§3.6 Stars and Bars

First, we need to clear up what "distinguishable" and "indistinguishable" means. Indistinguishable items are those that effectively all function as the same. Suppose we have two indistinguishable balls. Giving one ball to John over there is the same as giving the other ball to John. Distinguishable items are those that effectively function differently. Suppose we have distinguishable balls, like one ball is red and one ball is blue. Giving the red ball to John is different than giving the blue ball to John.

Now, we can comprehend this example:

Example 3.13

Suppose I have 9 indistinguishable balls. I have 4 distinguishable urns that I can put each ball in to. I must put each ball in an urn. How many ways can I distribute the balls?

For this procedure, you just have to take my word on it with how one could come up with it out of thin air. But knowing the procedure will help you generalize it to similar problems.

Lay the indistinguishable balls in a row. So we decide want to put 2 balls in the first urn. Then, we can represent it like this where we have a divider between the 2nd and 3rd balls:

$BB|BBBBBB$

Now, suppose we want to put 3 balls in the second urn. Now, we can represent it like this:

$BB|BBB|BBBB$

Now, suppose we want to put 1 ball in the third urn. Now, we can represent it like this:

$BB|BBB|B|BBB$

Since we must use all the balls, we just give the rest to the last urn. It seems like we needed 3 dividers. Indeed, 3 dividers divides the ball line into 4 parts. Once we lay down the dividers, the balls in the first part go to the first urn, the balls in the second part go to the second urn, and so on.

This trick works because the balls are indistinguishable. If we decide to use 2 balls in the first urn, it doesn't matter which 2 of the 9 balls we use.

Note that consecutive dividers represent an urn getting no balls at all. A divider up front represents the first urn getting no balls.

So distributing the balls is equivalent to laying down 3 dividers within the 9 balls. We have 12 objects total, so imagine 12 slots. We have $\binom{12}{3}$ ways to choose slots for where the dividers go. Then, the balls occupy the rest of the slots. So our answer is $\binom{12}{3}$.

We can consider a more general problem:

Example 3.14

This time, I have n indistinguishable balls and r distinguishable urns. How many ways can I distribute the balls now?

It is essentially the same logic. We have n balls and $r - 1$ dividers. We have $n + r - 1$ objects to line up, and $r - 1$ of those slots go to dividers. The rest go to balls. So there are $\binom{n+r-1}{r-1}$ ways to choose the slots for the dividers, which lays down where the balls go. So our answer is $\binom{n+r-1}{r-1}$.

Advice 3.15 — As you can see, stars and bars works with problems where you have to distribute indistinguishable items to distinguishable receivers. Also, to easier memorize the formula, just think of the n balls and $r - 1$ dividers example.

Stars and bars can also be used to solve problems like find the number of tuples of nonnegative integers (x_1, x_2, \dots, x_5) such that $x_1 + x_2 + \dots + x_5 = 7$.

§3.7 Anagrams

Definition 3.16 (Anagram). A rearrangement of the letters of a word. Similar to a permutation of the letters of the word, except this time, the elements aren't always distinct. The rearrangement can form a sequence of characters that doesn't form a sensical word.

Example 3.17

How many anagrams are there of the word 'science'?

So we have a 7 character word. Normally, there would be $7!$ ways to rearrange 7 distinct elements. But the kicker here is, not all elements of the things we're arranging are distinct. The two 'e's are the same. However, we can pretend we can tell them apart for now, and then reintroduce indistinguishability later. This will help with intuition.

So label the two 'e's as e_1 and e_2 to tell them apart. So we first want to see how many rearrangements there are of $scie_1nce_2$. Obviously, this is $7!$.

But now, we remember to reintroduce indistinguishability. We overcount by a factor of $2!$. Because for every anagram, we can reorder the e_1 and e_2 within that anagram in $2!$ ways and still have the same anagram. So when we consider the e s as distinguishable, we count each anagram $2!$ times. So the total number of anagrams is $\frac{7!}{2!}$.

Here's a generalization:

Example 3.18

Consider a word with k different characters. Character 1 shows up n_1 times, character 2 shows up n_2 times, and so on, until character k shows up n_k times. How many anagrams are there of this word?

We consider the same intuition where we pretend all the characters are distinguishable and then add indistinguishability back later. There are $n_1 + n_2 + \dots + n_k$ characters total. So if we pretend all characters are distinguishable, we have $(n_1 + n_2 + \dots + n_k)!$ ways to rearrange them. But remember for the n_1 times the first character showed up, we can reorder the copies of the first character within the slots they appear in and get the same anagram. So we overcount by a factor of $(n_1)!$. Similarly, by looking at the second character, we realize we overcount by a factor of $(n_2)!$. And so on. So the answer is

$$\frac{(n_1 + n_2 + \dots + n_k)!}{n_1! \cdot n_2! \cdot \dots \cdot n_k!}.$$

In particular, what we do is we say we have a bag. Suppose the m th character is noted by c_m . A bag is a set that allows elements to be repeated. Here's how we would denote the bag:

$$\{n_1 \cdot c_1, n_2 \cdot c_2, \dots, n_k \cdot c_k\}$$

where we have elements in the form $a \cdot b$, where b represents a element from the bag and a represents how many times it shows up.

In particular, the bag $\{n_1 \cdot c_1, n_2 \cdot c_2, \dots, n_k \cdot c_k\}$ has

$$\frac{(n_1 + n_2 + \dots + n_k)!}{n_1! \cdot n_2! \cdot \dots \cdot n_k!}$$

rearrangements.

Advice 3.19 — This bag trick is not just applicable for anagrams. Use it whenever you're rearranging stuff with categories, where stuff in the same category are indistinguishable, but stuff in different categories are distinguishable. In the anagram case, the different letters involved form categories.

§3.8 Pascal's Triangle

This is how you construct Pascal's Triangle. It might help to get a piece of paper and try following the procedure. So the 0th row will have 1 number. The 1st row will have 2 numbers. The 2nd row will have 3 numbers. And so on.

Now, for each row, put a 1 on both ends of the row. Then, to write out the other numbers that are not at either end of some row, we take the two numbers directly above it, add those two numbers, and write the sum in that position. Here are the first few rows of Pascal's triangle:

$$\begin{array}{ccccccc}
 & & & & 1 & & & & \\
 & & & & & 1 & & 1 & \\
 & & & 1 & & 2 & & 1 & \\
 & & 1 & & 3 & & 3 & & 1 \\
 & 1 & & 4 & & 6 & & 4 & & 1
 \end{array}$$

As you can see, each row has a 1 on the either end. The middle 2 in the 2nd row is formed from the two 1s above it and $1 + 1 = 2$. The 3 in the 3rd row is formed from the 1 and 2 directly above it and $1 + 2 = 3$.

Theorem 3.20 (Binomial Theorem)

We have

$$(a + b)^n = \binom{n}{0}a^n + \binom{n}{1}a^{n-1}b + \binom{n}{2}a^{n-2}b^2 + \cdots + \binom{n}{n}b^n$$

for any real numbers a and b and integer n that is at least 0.

This can be written in sigma notation:

$$(a + b)^n = \sum_{i=0}^n \binom{n}{i} a^{n-i} b^i.$$

If you haven't seen sigma notation, which uses \sum , what happens is below the \sum , we have $i = 0$. The variable i represents a "dummy variable" or a variable that will constantly change value. The 0 represents the lower bound. The n above the \sum represents the upper bound. So what we do is have i vary across all integers from the lower bound to upper bound and sum the values up we get for plugging in each such value of i .

In this case, we plug in $i = 0$ into the expression directly to the right of the \sum . Then, we plug in $i = 1$ and then $i = 2$ then $i = 3$, and so on, until $i = n$. Then, we add everything up that we get from plugging in each value of i .

How is this relevant to Pascal's Triangle? What if I told you Pascal's Triangle can be rewritten like this:

$$\begin{array}{ccccccc}
 & & & & \binom{0}{0} & & & \\
 & & & & \binom{1}{0} & & \binom{1}{1} & \\
 & & & \binom{2}{0} & & \binom{2}{1} & & \binom{2}{2} \\
 & & \binom{3}{0} & & \binom{3}{1} & & \binom{3}{2} & & \binom{3}{3} \\
 & \binom{4}{0} & & \binom{4}{1} & & \binom{4}{2} & & \binom{4}{3} & & \binom{4}{4}
 \end{array}$$

where the j th row of the triangle has numbers in the form $\binom{j}{i}$.

Let's take an example for the binomial theorem, say $n = 4$. Then, the theorem says

$$(a + b)^4 = \binom{4}{0}a^4 + \binom{4}{1}a^3b + \binom{4}{2}a^2b^2 + \binom{4}{3}ab^3 + \binom{4}{4}b^4.$$

Notice how the coefficients read off the 4th row in the triangle?

Theorem 3.21 (Pascal's Identity)

We have

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$$

Remember what we said earlier about constructing Pascal's Identity? How an entry was the sum of the two entries above it? The identity should then make sense intuitively, just this time, Pascal's triangle is reformulated in terms of binomial coefficients (that's the name given to terms in the form $\binom{a}{b}$).

§3.9 Combinatorial Proof

This one is TRICKY. Sometimes, we may have to prove some equation holds. We may get lucky and be able to use algebra. But other times, we aren't so lucky. Suppose we need to prove left side = right side, where left side and right side represent sides of some equation.

So what a combinatorial proof does is it conjures and presents some counting problem. It shows that you can solve the problem in some way and get the left hand side as the result. Then, it shows that it turns out that you can solve the problem another way and get the right hand side as the result. Now, the problem's answer shouldn't change just because you changed the procedure, as long as you executed the procedure correctly. So the two alleged answers should be equal. Bear with me through an example so that this all makes sense.

Example 3.22

Show through a combinatorial proof that

$$\binom{2n}{2} = n^2 + 2\binom{n}{2}$$

for all integers $n \geq 2$.

Now, you could use algebra with the choose formulas. But that would get you 0 points when they say show through a combinatorial proof. Consider this counting problem: "A deck of cards has n distinguishable red cards and n distinguishable blue cards. How many ways can I choose 2 cards from this deck?"

So we can get the answer this way: there are simply $\binom{2n}{2}$ ways to choose 2 cards. This gives the left hand side as the answer.

But here's a second way that's slightly more complicated that gives the right hand side as the answer. We have two cases: the cards are the same color or they are different colors.

Case 1: They are different colors

So then we will be choosing a red card and a blue card. This can be done with this procedure:

- n ways to choose the red card
- n ways to choose the blue card

Steps are independent, so by MR, we have $n \cdot n = n^2$ ways for Case 1.

Case 2: They are the same color colors

So the cards are either both red or both blue. Consider this procedure:

- 2 ways to choose which color both cards are
- $\binom{n}{2}$ ways to actually choose the cards of whatever color we picked

Steps are independent, so by MR, we have $2 \cdot \binom{n}{2}$ ways for Case 2.

Coming back, the two cases are clearly disjoint and exhaustive. So by AR, the answer to the counting problem is $n^2 + 2 \cdot \binom{n}{2}$.

The two answers we got are the same, since they are answers to the same problem. So we can say $\binom{2n}{2} = n^2 + 2 \binom{n}{2}$.

Caveat 3.23. Do NOT change the equation in any way in a combinatorial proof or they will take points off. Do not move terms from one side to another. Do not make any simplifications to any term in the equation (like applying the formula for $\binom{2n}{2}$ in the equation).

You are allowed to make simplifications when you're doing the counting for one way like how I did $n \cdot n = n^2$.

But now you might be wondering “how the fuck was I supposed to come up with that problem out of thin air?” This leads me to a lot of advice when conjuring the problem.

Advice 3.24 — You generally have to take a shot in the dark when making the problem. Sometimes, it will work out. Other times, you have to reformulate the problem a bit to get it to work.

Draw your inspiration to make the problem from the less complicated side. Like how the less complicated left hand side in the example earlier inspires to think of choosing 2 items from $2n$ total items.

Advice 3.25 — Here are tip-offs for what you should consider when conjuring your problem:

- Pluses should make you think of AR
- \sum is really just AR
- Times should make you think of MR
- $\binom{n}{k}$ should make you think of choosing k items from a pool of n items
- Minuses should make you think of complementary counting
- n^k should make you think of a sequence of length k with n possibilities for each term in the sequence

§3.10 Functions

Suppose we define a function $f : A \rightarrow B$. Further suppose that $|A| = a$ and $|B| = b$. Then, there are b^a possible functions.

The integer interval $[m..n]$, with $m \leq n$ is a set containing every integer from m to n , including m and n . So $[3..7]$ would be $\{3, 4, 5, 6, 7\}$.

Remember $|[m, n]| = n - m + 1$.

§3.11 Principle of Inclusion/Exclusion

Remember these formula: (where A , B , and C are sets)

$$|A \cup B| = |A| + |B| - |A \cap B|$$

$$|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |B \cap C| - |C \cap A| + |A \cap B \cap C|.$$

To explain the formula for the three sets better, notice how first we add $|A| + |B| + |C|$. Then, we flip the sign and consider intersections of any two sets, giving $-|A \cap B| - |B \cap C| - |C \cap A|$. Then, we flip the sign again and consider the intersection of all three sets.

Remember how with the AR, we could only use it if the cases are disjoint? Now, we no longer have to worry about that. With PIE (abbreviation for Principle of Inclusion/Exclusion), you can break up into cases if you seem to find there's no way getting around overlap between cases.

Advice 3.26 — Going forward, for problems where you want to break up into cases, just divide the cases into the most natural seeming way. Now that you have PIE, you can deal with overlaps between cases. If you have no overlap between any cases, then it is just AR.

Expect to use PIE when it is count the things that satisfy this OR this (key word is “or”).

§3.12 Pigeonhole Principle

Get out a piece of paper and draw 3 non-overlapping circles on your page. Now, I tell you to draw 4 dots, where each dot must be inside one of the 3 circles. You'll notice it seems like no matter what you do, there will a circle containing at least 2 dots. Can we reason this out more formally?

Let's suppose we were drawing the 4 dots while actively trying to avoid giving a circle at least 2 dots. So we draw the first dot in one of the circles. Because of what we are actively avoiding, we don't want to draw another dot in that circle again. Similarly, we draw the second and third dots in the other two circles. Now, each circle has a dot. We have no choice but to draw the fourth dot and force a circle to have at least 2 dots.

We can assert this more formally:

Definition 3.27 (Pigeonhole Principle). If we have n pigeons and m pigeonholes, where $n > m$ and each pigeon must go in one of the pigeonholes, then one of the pigeonholes must contain more than one pigeon.

Of course, when using PHP (pigeonhole principle), it won't always be pigeons that are the objects being placed, and it won't always be pigeonholes that you're placing the objects into. You need to figure that out yourself.

Definition 3.28 (Generalized PHP). If we have n pigeons and m pigeonholes, where $n > m$ and each pigeon must go in one of the pigeonholes, then one of the pigeonholes must contain at least $\lceil \frac{n}{m} \rceil$ pigeons.

The notation $\lceil x \rceil$ is the ceiling function. Basically, if x is an integer, the ceiling function doesn't change x . Otherwise, if x is not an integer, it always rounds x up to the nearest integer (NEVER rounds down). So for example, $\lceil 5 \rceil = 5$, $\lceil \frac{3}{2} \rceil = 2$, and $\lceil 3.7 \rceil = 4$.

Advice 3.29 — PHP is something to consider when you need to prove something exists, or that some two objects have this property.

GPHP is something to consider when tasked to prove something has at least 4 of this, where the “4” can be replaced with any positive integer.

§4 Proofs

This is what this course revolves around. You have to be thorough with proofs and account for every possibility, as well as making sure the steps are valid to get from start to finish. If you don't do this, you may end up “proving” incorrect things that don't make sense.

For the sake of demonstration, consider the general statement “All cars are red.” We may find a red car. But just because we find a red car, that doesn't mean all of cars are red. We would in theory need to check every single car's color.

We'll see general statements all the time that we need to prove, where we can't just use one example and call it a day. We need to figure out how to be general in our proofs, too. Not only that, but we also need to learn different strategies for proving something and when these proofs may be applicable.

§4.1 Odd and even

Definition 4.1 (Odd integer). If n is odd, then there is an integer k such that $n = 2k + 1$. This works the other way around: if there is an integer k such that $n = 2k + 1$, then n is odd.

For instance, since we know $n = 5$ can be written as $2 \cdot 2 + 1$, we can say $k = 2$ is an integer such that $5 = 2k + 1$. So by the definition, we can say 5 is odd.

We can use this definition the other way around. Suppose now that $k = 2023$ but n is yet to be computed out. Then, we can say $n = 2 \cdot 2023 + 1$ is odd.

Definition 4.2 (Even integer). If n is even, then there is an integer k such that $n = 2k$. This works the other way around: if there is an integer k such that $n = 2k$, then n is even.

Similar logic applies here. For instance, since we know $n = 12$ can be written as $2 \cdot 6$, we can say $k = 6$ is an integer such that $6 = 2k$. So by the definition, we can say 12 is even.

And we can use this definition the other way around. Suppose now that $k = 2023$. Then, we can say $n = 2 \cdot 2023$ is even.

In this course, you'll want to use these two definitions whenever you are given something is odd/even, or whenever you want to prove something is odd/even.

Example 4.3

Suppose x is an odd integer and y is an even integer. Prove that $5x + 7y$ is an odd integer.

So suppose we use $x = 1$ and $y = 2$. Then, $5x + 7y = 19$, which is odd. But like the “all cars are red” example at the beginning of Section 4, just because it works for one choice of x and y , that doesn't mean we can call it a day. We need to be more general. We don't know anything specific about x and y beyond the fact that x is odd and y is even.

Instead, we want to use our definitions of odd and even integer because those definitions are general (they hold for any odd integer and any even integer). So we can say $x = 2a + 1$ for some integer a because x is odd. Also, $y = 2b$ for some integer b because y is even. Now, we want to consider the expression we want to prove as odd. Plugging our representations of x and y in, as well as some algebra implies

$$5x + 7y = 5(2a + 1) + 7(2b) = 10a + 5 + 14b = 10a + 14b + 4 + 1 = 2(5a + 7b + 2) + 1.$$

Something that is helpful to remember is that the sum, difference, or product of any two integers is always an integer. So we can use this to reason out that $5a + 7b + 2$ is some integer. As we can see from the algebra, $5x + 7y = 2(\text{some integer}) + 1$. Using our definition, this means $5x + 7y$ is odd.

Advice 4.4 — Essentially, to solve these types of problems, you want to convert the words to math. You want to use the definitions of odd and even with the given information because they are general. Then, you want to use algebra to get the end goal expression using the new variables. Last, you conclude using the definitions of odd and even.

Caveat 4.5. You don't have to prove that the sum, difference, or product of any two integers is always an integer, but you should mention it in homeworks when you're using it.

§4.2 Divisibility and Primes

Definition 4.6 (Divisibility). For integers n and d , we say n is divisible by d if there is an integer k such that $n = d \cdot k$.

For example, suppose we wanted to check if 15 is divisible by 5. Then, we would plug in $n = 15$ and $d = 5$ above so that we have $15 = 5 \cdot k$. Then, $k = 3$, which is an integer. So there is an integer k , meaning yes, 15 is divisible by 5.

Furthermore, suppose n is divisible by d . Then, we can say d is a divisor of n . We can also say n is a multiple of d .

Advice 4.7 — Often times, when given a condition like “this is divisible by this,” you will want to convert the condition to math, similar to the odd and even section. For example, if you are told n is divisible by 3, then you want to write $n = 3k$ for some integer k .

Definition 4.8 (Primes). An integer p is prime if it has exactly 2 positive divisors: itself and 1. Note that 1 is not a prime number.

A property we can use for all primes p (without needing to prove) is $p \geq 2$.

Example 4.9

Suppose p is a prime number and r and s are positive integers such that $p = r \cdot s$. What can we say about r and s ?

Based on our definition of divisibility, p is divisible by r , since r and s are integers. But p 's only divisors (by the definition of a prime) are itself and 1. So either $r = p$ or $r = 1$. This yields the only possibilities of $r = p$ and $s = 1$, or the other way around. Our conclusion is that whenever $p = r \cdot s$, r and s can only be 1 and p in some order. This example on primes is helpful when you have a prime equals the product of two integers because it forces the integers to be specific values.

Lemma 4.10

2 is the only even prime.

This could be a useful fact on the HW, but you would need to prove it. So let p be an even prime. By the definition of even, we can write $p = 2k$ for some integer k . But by the definition of a divisor, 2 would be a divisor of p .

Remember that for a prime p , the full list of divisors of p is only 1 and p . But 2 is a divisor of p , so it must be somewhere in this list. So the only possibility is $p = 2$, proving the lemma.

Also, for proving something is not prime, you generally want to find a divisor of that number that is not equal to 1 or the number itself. Alternatively, show that the number is the product of two integers that are both greater than 1.

§4.3 Logic Behind Proofs

A good amount of the time, you can usually boil down the task in a problem in this course to an implication in the form $p \Rightarrow q$. Basically, you're given some conditions. Then, you have to show if p is true, then it will follow that q is true.

Sometimes though, you'll be tasked with proving a bidirectional: " p holds if and only if q holds." In this case, you need to show $p \Rightarrow q$ and $q \Rightarrow p$.

Either way, you're going to have to work with proving some implication like $p \Rightarrow q$.

How do you go about with the proof? Here's the general pattern. So you assume p is true along with the given information in the problem. Some of the given information and p can imply additional information that you can add to your arsenal. That new information along with the old information can lead to more information. So you keep expanding what you know is true until you reach a point where q is true.

However, sometimes, you may need to break the proof up into cases, which is fine. But you also need to know what is a proper proof that uses cases.

First off, each case has to actually lead to the desired conclusion, or the proof won't hold logically. But the cases have to be **exhaustive**. Remember this word from the Addition Rule? Another way of saying this is, the cases have to cover every single possibility. But unlike the addition rule, the cases can have overlap.

Suppose you need to prove some property holds for all integers x . Dividing into cases where 1 case is $x \geq 1$ and another is $x \leq 0$ is exhaustive because the two cases cover every integer x . Dividing into cases where 1 case is $x \geq -1$ and another is $x \leq 0$ is also exhaustive, even if there is overlap. But dividing into only 2 cases where 1 case is $x \geq 2$ and another is $x \leq 0$ is NOT exhaustive because it misses the possibility of $x = 1$.

§4.4 Predicates and Quantifiers

Definition 4.11 (Predicate). A logical statement containing a variable. We usually don't know the truth value of the statement without specifying the variables.

For example, “ x is prime” is a predicate. Because we don’t know if this is true if we don’t specify what x is. The statement is true if $x = 3$ but not if $x = 4$.

Definition 4.12 (Quantifier). A word/phrase/symbol that specifies what scope we are addressing.

Definition 4.13 (Universal quantifier (\forall)). A quantifier that means “for all.” So $\forall x$ means “for all x .”

Definition 4.14 (Existential quantifier (\exists)). A quantifier that means “there exists.” So $\exists x$ means “there exists an x .”

These quantifiers will come up in problems throughout the course where you may be examining where some property holds. Maybe you need to prove the property holds for all of these things. Maybe you need to prove there exists a thing such that the property holds for it.

§4.5 Negation, Converse, and Contrapositive

DeMorgan’s Laws State:

- The negation of a disjunction is a conjunction: $\neg(p \vee q) = (\neg p) \wedge (\neg q)$
- The negation of a conjunction is a disjunction: $\neg(p \wedge q) = (\neg p) \vee (\neg q)$

To help you better memorize these, think about what $p \vee q$ means: either p or q . The negative would be neither. Which means not p and not q . Also, think about what $p \wedge q$: both p and q . So if is false that both p and q are true, then one of them has to be wrong. This motivates the or.

There are also negations of the quantifiers:

- The negation of universal is existential: $\neg(\forall x, \text{this holds}) = \exists x \text{ such that this does not hold}$
- The negation of existential is universal: $\neg(\exists x, \text{this holds}) = \forall x, \text{ this does not hold}$

These make sense intuitively. If something held for all x , it holds for $x = 1$, $x = 2$, $x = 3.14$, $x = \text{whatever}$, literally every value of x . But if it is NOT true that it holds for all x , then the property slips up somewhere, meaning it is not true for some x out there. That motivates existential.

Likewise, suppose there does not exist an x such that something holds. Then, no matter where we search, we’ll never find an x where it holds. So we can say for all x , it doesn’t hold.

Negative of an implication is trickier, but:

$$\neg(p \Rightarrow q) = p \wedge \neg q$$

The intuition is either p is true or it isn’t. The implication conveys that if p is true, then q has to be true too. If p is not true, it doesn’t matter what q is. So the implication is essentially, either not p or q or $\neg p \wedge q$, since the case in which p is true forces q to be true. Negating this with DeMorgan’s Laws leads to $p \wedge \neg q$.

Definition 4.15 (Converse). Given an implication $p \Rightarrow q$, the converse is $q \Rightarrow p$.

Definition 4.16 (Contrapositive). Given an implication $p \Rightarrow q$, the contrapositive is $\neg q \Rightarrow \neg p$.

Take my word on it for now, but the contrapositive is logically equivalent to the original implication. This means if the contrapositive is true, the implication is true and vice versa. If the contrapositive is false, the implication is false and vice versa. So you can prove an implication by proving the contrapositive is true.

§4.6 Proof by Contradiction

This technique will come up throughout the rest of the course. You may be finding yourself spinning your tires trying to prove something directly. But what if it has to be true because it being false is simply not possible?

This is where proof by contradiction comes in. Suppose you want to prove q . What you do assume q is false. Then, show how q being false gives information that directly contradicts something. The contradiction can be something where you find r is true, but the given information says r is false (for a statement r). Or the contradiction can be a blatantly false statement like $n > n$.

Caveat 4.17. Say you're trying to prove q using proof by contradiction. You have to explicitly say “assume for the sake of contradiction that q is false” with q replaced with the statement you are trying to prove. If you don't make that contradiction statement, they'll take off points.

Advice 4.18 — If you find that your conclusion is hard to prove directly, try proof by contradiction. Proof by contradiction can be useful when you have to prove “there exists this.” Because by the negative of the existential quantifier is the universal quantifier.

§4.7 Induction

This is a proof technique that usually works best in proving something “holds for all integers $n \geq$ [some integer],” where [some integer] is usually 0 or 1. Without an example, it is hard to explain in a way that will make sense, so bear with me here. Yes, this is a long section, but I feel I have to explain this out to show the thought process.

Example 4.19 (Based off 2022 AMC 10B #9)

Prove for all integers $n \geq 1$,

$$\sum_{m=1}^n \frac{m}{(m+1)!} = 1 - \frac{1}{(n+1)!}.$$

Okay, this looks hard to prove directly. But we know if this holds for every single integer $n \geq 1$, it definitely needs to hold for $n = 1$. So let's start simple and show that it holds for $n = 1$.

$$\sum_{m=1}^1 \frac{m}{(m+1)!} = \frac{1}{2!} = \frac{1}{2}.$$

This is equal to $1 - \frac{1}{2!}$. So we know the proposition in the example holds for $n = 1$.

Now, we need to show the proposition holds for $n = 2$. But we already know it holds for $n = 1$. Maybe that can be of some use?

$$\sum_{m=1}^2 \frac{m}{(m+1)!} = \sum_{m=1}^1 \frac{m}{(m+1)!} + \frac{2}{3!}.$$

So what I did here was separated the last term from the sum, so the sum in the equation is more familiar. Now, about that sum where the upper bound is 1. We know that equals $1 - \frac{1}{2!}$, since we proved that earlier. So

$$\sum_{m=1}^2 \frac{m}{(m+1)!} = \sum_{m=1}^1 \frac{m}{(m+1)!} + \frac{2}{3!} = 1 - \frac{1}{2!} + \frac{2}{3!} = \frac{5}{6}.$$

So the sum up to 2 is indeed equal to $1 - \frac{1}{3!}$. So the proposition in the example holds for $n = 2$.

Now, we need to show the proposition holds for $n = 3$. But we already know it holds for $n = 2$. Again, maybe that can be of some use?

$$\sum_{m=1}^3 \frac{m}{(m+1)!} = \sum_{m=1}^2 \frac{m}{(m+1)!} + \frac{3}{4!}.$$

Again, I broke off the last term in the sum, so the sum is something we are familiar with. We know that sum with the upper bound of 2 equals $1 - \frac{1}{3!}$, since we proved that earlier. So

$$\sum_{m=1}^3 \frac{m}{(m+1)!} = \sum_{m=1}^2 \frac{m}{(m+1)!} + \frac{3}{4!} = 1 - \frac{1}{3!} + \frac{3}{4!} = \frac{23}{24}.$$

So the sum up to 3 is indeed equal to $1 - \frac{1}{4!}$. So the proposition in the example holds for $n = 3$.

Okay, it seems like we can reuse what we know about the proposition holding for prior values of n to see if it can hold for future values of n . Let's see if we can generalize this.

So for each integer $n \geq 1$, define $P(n)$ to be the proposition that

$$\sum_{m=1}^n \frac{m}{(m+1)!} = 1 - \frac{1}{(n+1)!}$$

holds. We know $P(1)$ is true, as we checked it earlier. But let's see if we can prove for an arbitrary integer $k \geq 1$ that $P(k)$ being true implies $P(k+1)$ is true. Note the use of the buzzword “arbitrary,” which will be important. I'll get to that later.

So basically, since $k \geq 1$ is an arbitrary integer, what that means is we know nothing about this specific value of k other than it is an integer and it is at least 1. But we have to show $P(k) \Rightarrow P(k+1)$. In other words, our goal is to show $P(k+1)$ is true using the assumption that $P(k)$ is true. We can go about this directly by algebra with the sum with $k+1$ terms:

$$\sum_{m=1}^{k+1} \frac{m}{(m+1)!} = \sum_{m=1}^k \frac{m}{(m+1)!} + \frac{k+1}{(k+2)!}.$$

Again, we broke off the last term, so we have the sum up to k terms. We are assuming $P(k)$ is true, which means, by definition of $P(k)$, we know

$$\sum_{m=1}^k \frac{m}{(m+1)!} = 1 - \frac{1}{(k+1)!}$$

is true. We can use this to our advantage in the algebra, so

$$\begin{aligned} \sum_{m=1}^{k+1} \frac{m}{(m+1)!} &= \sum_{m=1}^k \frac{m}{(m+1)!} + \frac{k+1}{(k+2)!} \\ &= 1 - \frac{1}{(k+1)!} + \frac{k+1}{(k+2)!} \\ &= 1 - \frac{k+2}{(k+2)!} + \frac{k+1}{(k+2)!} \\ &= 1 - \frac{k+2}{(k+2)!} + \frac{k+1}{(k+2)!} \\ &= 1 - \frac{1}{(k+2)!} \end{aligned}$$

And so we have shown $P(k+1)$ is true. But remember that we arbitrarily chose k to be an integer greater than 1. This basically means, we showed the relation $P(k) \Rightarrow P(k+1)$ holds no matter what integer k is, as long as it is at least 1.

So as a recap, we know $P(1)$ is true. With the relation, we know $P(1) \Rightarrow P(2)$, so $P(2)$ is true too. We know $P(2) \Rightarrow P(3)$ with the relation, so $P(3)$ is true because $P(2)$ is true. And you can see where this is going. This will eventually show $P(n)$ is true for all positive integers n to infinity.

Now that you have a crash course at the idea of induction, I can explain things more formally. There are two vital parts to any proof by induction, the base case and inductive step.

Definition 4.20 (Base Case). In the example, this is $P(1)$ being true. A base case in an induction proof is proving the proposition holds for the first integer in your given range.

So our base case in the example was $n = 1$, since we were tasked to prove something held for all integers $n \geq 1$. If it said “prove [this] for all integers $n \geq 0$,” our base case would be $n = 0$ instead.

Definition 4.21 (Inductive Step). In the example, that is basically the part where we find $P(k) \Rightarrow P(k+1)$ for any arbitrary integer $k \geq 1$

The inductive step is basically what allows us to reuse the fact that we know $P(k)$ is true, in order to conclude $P(k+1)$ is true, too. Think of the base case and inductive step together as a chain reaction. $P(1)$ is true, is the base case. Then, the chain reaction is $P(1)$ being true causes $P(2)$ to be true (by the inductive step). Then, $P(2)$ being true causes $P(3)$ to be true. And so on.

I should also explain why you can't have a proof by the induction if you're missing one of those two things.

Say you have the inductive step proven, but you don't show the base case is true. That is basically saying "Okay I know this chain reaction $P(1) \Rightarrow P(2) \Rightarrow P(3) \Rightarrow \dots$ is a thing and it exists, but the chain reaction hasn't even started because the reaction starting relies on $P(1)$ being true!"

Now, we have the base case and the inductive step works everywhere except for one point. Like for example, we don't know for sure if $P(4) \Rightarrow P(5)$. That is basically saying "Okay I have this chain reaction $P(1) \Rightarrow P(2) \Rightarrow P(3) \Rightarrow P(4)$, but it abruptly stops there, with no way to get to $P(5)$. That means the reaction won't reach $P(5), P(6), P(7), \dots$ and beyond, so it won't reach every positive integer."

So basically, we not only have to show the chain reaction is linked all the way for all positive integers (inductive step), we need to show it actually starts in the first place (base case).

Now, I'm going to rewrite our work for the problem into a solution in a format they expect you to use, now that you hopefully understand what is going on. I will highlight the necessary buzzwords.

Solution:

For each integer $n \geq 1$, define $P(n)$ to be the **proposition** that

$$\sum_{m=1}^n \frac{m}{(m+1)!} = 1 - \frac{1}{(n+1)!}$$

holds. We will use induction.

Base Case: We need to make sure $P(1)$ is true:

$$\sum_{m=1}^1 \frac{m}{(m+1)!} = \frac{1}{2!} = \frac{1}{2}.$$

This is indeed equal to $1 - \frac{1}{2!}$. So $P(1)$ is true.

Inductive Step: Now, assume $P(k)$ is true for an **arbitrary** integer $k \geq 1$. We will show this implies $P(k+1)$ is true.

We can break up the sum for $k+1$ terms as follows:

$$\sum_{m=1}^{k+1} \frac{m}{(m+1)!} = \sum_{m=1}^k \frac{m}{(m+1)!} + \frac{k+1}{(k+2)!}.$$

We know

$$\sum_{m=1}^k \frac{m}{(m+1)!} = 1 - \frac{1}{(k+1)!}$$

by the **inductive hypothesis**. Thus,

$$\begin{aligned} \sum_{m=1}^{k+1} \frac{m}{(m+1)!} &= \sum_{m=1}^k \frac{m}{(m+1)!} + \frac{k+1}{(k+2)!} \\ &= 1 - \frac{1}{(k+1)!} + \frac{k+1}{(k+2)!} \\ &= 1 - \frac{k+2}{(k+2)!} + \frac{k+1}{(k+2)!} \\ &= 1 - \frac{k+2}{(k+2)!} + \frac{k+1}{(k+2)!} \\ &= 1 - \frac{1}{(k+2)!} \end{aligned}$$

And so we have shown $P(k+1)$ is true, completing the induction.

Okay, let me point out the mandatory stuff they expect you to put in an induction proof.

- Define the proposition immediately that you want to prove for each integer.
- Label the base case and inductive step.
- The format of the two sentences after “Inductive Step” is what they want to use. You have to say you’re assuming $P(k)$ is true and then saying your goal is to prove $P(k+1)$ is true under the assumption $P(k)$ is true.
- If they say “Prove this for all positive integers n ,” you want to use a different variable other than n (like k in what I did) during the inductive step. If you do not use a different variable, they will take points off.
- Specify k (or whatever variable you’re using for the inductive step) is an arbitrary integer. They’ll take points off if you don’t.
- Put a bound on k like $k \geq 1$ (if they were to ask you to prove something for all integers $n \geq 0$, though, your bound would be $k \geq 0$ instead). Remember that $P(0)$ doesn’t exist in this problem.
- Specifically namedrop “by the inductive hypothesis” during your inductive step when you’re using the information obtained from the assumption that $P(k)$ is true.

This might look like a lot to memorize, but I think once you have typed out several induction proofs yourself, knowing what to write will become more routine.

Advice 4.22 — Induction with sums is usually straightforward once you’ve seen one example. You always to break off the last term in the summation during the inductive step.

§4.8 Strong Induction

Wrapping your head around induction is the hard part. Once you do that, strong induction shouldn’t be too hard to understand, but there are important changes.

- There may be more than one base case necessary now
- Suppose you’re proving something for all integers $n \geq 1$. Instead of showing $P(k) \Rightarrow P(k+1)$ in the inductive step, you’re showing $P(1), P(2), \dots, P(k)$ all being true implies $P(k+1)$ is true. In other words, you’re reusing at least one other past proposition besides $P(k)$.

You don’t have to be explicitly using all past propositions in your inductive step, just more than one besides $P(k)$.

Example 4.23 (Based off 2019 AMC 10A #15)

A sequence of numbers is defined recursively by $a_1 = 1$, $a_2 = \frac{3}{7}$, and

$$a_m = \frac{a_{m-2} \cdot a_{m-1}}{2a_{m-2} - a_{m-1}}$$

for all integers $m \geq 3$. Prove that $a_n = \frac{3}{4n-1}$ for all integers $n \geq 1$.

We’ll define for every positive integer n , $P(n)$ is the proposition that $a_n = \frac{3}{4n-1}$ holds.

So we have to prove this holds for all integers $n \geq 1$. Look at how we find a_m though. That formula given relies on a_{m-1} and a_{m-2} . Meaning it relies on 2 prior terms rather than 1 prior term. This is where multiple base cases come in. It says that formula only works for “all integers $m \geq 3$ ” so we can’t put in $m = 2$. We can plug in $m = 3$ as our first value in that formula. But it relies on knowing a_2 and a_1 , which luckily are given to us.

So instead, we’ll have $n = 1$ AND $n = 2$ be our base cases. The base case of $n = 1$ is not enough on its own to start the chain reaction because our inductive step will use that formula, and it can’t be used if we know only 1 term (the formula requires 2 prior terms).

Then, of course, in the inductive step, we'll have to rely on two prior propositions instead of one. Here's how the solution would be written up:

Solution: For every positive integer n , define $P(n)$ to be the proposition that $a_n = \frac{3}{4n-1}$ holds. We will use strong induction.

Base Case: We need to check $P(1)$ and $P(2)$ hold.

We are given that $a_1 = 1$ and $a_2 = \frac{3}{7}$. We see that $a_1 = \frac{3}{4 \cdot 1 - 1} = 1$ is indeed true, so $P(1)$ is true. We also see that $a_2 = \frac{3}{4 \cdot 2 - 1} = \frac{3}{7}$ is true too, so $P(2)$ is true.

Inductive Step: Now, assume that for an arbitrary integer $k \geq 2$, the propositions $P(1), P(2), \dots, P(k)$ are all true. We will show this implies $P(k+1)$ is true.

Plug in $m = k+1$ into the recursive formula given. This is allowed, since $k \geq 2$ means $m \geq 3$, which is within the given bounds for where the formula accepts a value m . (The inspiration to plug in $m = k+1$ comes from the fact we want to know a_{k+1} from the prior terms). So

$$a_{k+1} = \frac{a_{k-1} \cdot a_k}{2a_{k-1} - a_k}.$$

By the strong induction hypothesis, $P(k-1)$ and $P(k)$ are true, so $a_{k-1} = \frac{3}{4(k-1)-1} = \frac{3}{4k-5}$ and $a_k = \frac{3}{4k-1}$. Plugging this in,

$$\begin{aligned} a_{k+1} &= \frac{a_{k-1} \cdot a_k}{2a_{k-1} - a_k} = \frac{\frac{3}{4k-5} \cdot \frac{3}{4k-1}}{2 \cdot \frac{3}{4k-5} - \frac{3}{4k-1}} = \frac{\left(\frac{3}{4k-5} \cdot \frac{3}{4k-1}\right) (4k-5)(4k-1)}{\left(2 \cdot \frac{3}{4k-5} - \frac{3}{4k-1}\right) (4k-5)(4k-1)} \\ &= \frac{9}{6(4k-1) - 3(4k-5)} = \frac{3}{4(k+1) - 1}. \end{aligned}$$

(Essentially we plugged what we knew in, cleared denominators, and then simplified stuff to get it in the form we want). So $P(k+1)$ is true, completing the strong induction.

Advice 4.24 — Recursive sequences (sequences that rely on prior terms to compute future terms) are the poster child for strong induction. The strong inductive step is usually just using the recursive formula.

Advice 4.25 — If you want to approach a problem with induction, do it with regular induction first. If you find only knowing $P(k)$ is true is not enough to prove $P(k+1)$ is true and that you need to use more propositions before $P(k)$, then switch to strong induction.

Caveat 4.26. Specify you're using strong induction and not just say "I'll use induction" without the "strong." Likewise, don't say you're using "strong induction" if you're not relying on more than one prior proposition for the inductive step. They'll take points off if you do either of those things.

§5 Terminology and Notation

§5.1 Sets

This section will be definition heavy. You will have to get used to notation. Try to understand it in plain English first, then, use whatever method you used for memorizing vocabulary to get accustomed to the notation.

§5.1.1 Introductory

Definition 5.1 (Set). A set is simply a collection of things. The things could be anything. It could be integers, letters, animals, whatever. It could contain a mix of different things too. A set can have finite size or infinite size.

A set is denoted with square brackets and different elements in the set are separated with commas. For example, $\{1, a, L\}$ is a set containing only 1, a , and L .

Definition 5.2 (\in). Let e be a thing and S be a set. We say $e \in S$ (reads e is in S) if e is a member of S . Likewise, we say $e \notin S$ (reads e is NOT in S) if e is NOT a member of S .

For example, let $S = \{1, 2, 3\}$, $e = 2$, and $f = 4$. We can see $e \in S$ and $f \notin S$.

Definition 5.3 (Subset). A subset of a set is a set containing some (maybe all, maybe none) of the elements of the original set. The formal way to think of a subset is, every element of the subset is also an element of the original set. But the intuitive way to think of it is, take the original set. Delete some of its elements (maybe all of them, maybe none of them). Then, boom, you have a subset.

For example, let $S = \{1, 2, 3\}$. We can clearly see $\{1, 2\}$ is a subset of S . But $\{1, 4\}$ wouldn't because 4 isn't in S .

If A is a subset of B , then we use the notation $A \subseteq B$.

Definition 5.4 (Proper Subset). A subset of a set that isn't just the original set.

For example, if $S = \{1, 2, 3\}$, $\{1, 2, 3\}$ is a subset of S but not a proper subset. $\{1, 2\}$ would still be a proper subset of S .

If A is a proper subset of B , then we use the notation $A \subsetneq B$.

Definition 5.5 (Empty Set). The set with no elements. It is noted as \emptyset . It is a subset of any set.

Definition 5.6 (\mathbb{Z}). The set of all integers. This includes the positive integers: $1, 2, 3, \dots$. It also includes 0. It also includes the negative integers: $-1, -2, -3, \dots$.

Definition 5.7 (\mathbb{Z}^+). The set of all positive integers. 0 isn't considered positive, so it's not included.

Definition 5.8 (\mathbb{N}). The set of all natural numbers. These are the positive integers including 0. (Other definitions of natural numbers don't include 0, but for this course, 0 is a natural number)

§5.1.2 Intermediate

Definition 5.9 (Union (\cup)). The union of two sets A and B is the set containing each element that shows up in A or B . The union of A and B is noted as $A \cup B$.

For example, if $A = \{1, 2, 3\}$ and $B = \{1, 3, 5\}$, then $A \cup B = \{1, 2, 3, 5\}$.

Definition 5.10 (Intersection (\cap)). The intersection of two sets A and B is the set containing each element that shows up in both A and B . The intersection of A and B is noted as $A \cap B$.

For example, if $A = \{1, 2, 3\}$ and $B = \{1, 3, 5\}$, then $A \cap B = \{1, 3\}$.

Note that union and intersection can apply to more than two sets. For example, the set $A \cup B \cup C$ means the elements in A or B or C . The set $A \cup B \cup C \cup D$ means the elements in A or B or C or D . And so on.

The set $A \cap B \cap C$ means the elements in A and B and C . The set $A \cap B \cap C \cap D$ means the elements in A and B and C and D . And so on.

Definition 5.11 (Disjoint). Two sets are disjoint if they have no elements in common.

Definition 5.12 (Pairwise disjoint). Suppose we have a collection of sets. We say the sets are pairwise disjoint if any two sets are disjoint.

Definition 5.13 (Set difference (\setminus)). The set difference of two sets A and B is the set containing the elements that are in A but not also in B . The set difference of A and B is noted as $A \setminus B$.

For example, if $A = \{1, 2, 3\}$ and $B = \{1, 3, 5\}$, then $A \setminus B = \{2\}$.

Definition 5.14 (Cardinality). The cardinality of a set is how many elements are in the set. A cardinality of a set A is noted as $|A|$.

For example, if $A = \{1, 4, 6, 9\}$, then $|A| = 4$.

Definition 5.15 (Pair). A list of two elements where the order matters. The pair of a then b would be noted as (a, b) .

Definition 5.16 (Tuple). A list of elements where the order matters.

Definition 5.17 (Powerset). The powerset of a set A is the set containing every subset of A . It is denoted as 2^A .

For example, if $A = \{1, 3\}$, then $2^A = \{\emptyset, \{1\}, \{3\}, \{1, 3\}\}$. Essentially, for $B \in 2^A$ to be true for some set B , we need $B \subseteq A$.

Definition 5.18 (Cross-product). Consider two sets A and B . This is how you construct the cross-product of A and B : consider the pair (a, b) , where $a \in A$ and $b \in B$. Generate every possible pair (a, b) as a varies across all elements of A and b varies across all elements of B . Put them all into a set, and then that set is the cross-product. The cross-product of A and B is denoted as $A \times B$.

For example, if $A = \{1, 3\}$ and $B = \{2, 3, 4\}$, then $A \times B = \{(1, 2), (1, 3), (1, 4), (3, 2), (3, 3), (3, 4)\}$. Essentially, for a pair $(x, y) \in A \times B$, we need $x \in A$ and $y \in B$.

§5.1.3 Set Builder Notation

In plain English, set builder notation goes like: $S = \{\text{thing} \mid \text{some property}\}$. This defines S as the set consisting of all the “things” out there that satisfy the “some property.”

For example, $S = \{x \mid x \text{ is an even integer}\}$ defines S as the set of all x such that x is an even integer.

Advice 5.19 — Suppose they ask you to prove two sets S and T are equal. What you need to do is show S is a subset of T and T is a subset of S . Think about why this is enough to conclude $S = T$.

Advice 5.20 — When they want you to prove S is a subset of T , what they want you to do is: let z be an arbitrary element of S . Then, show that implies z is in T . **Use the word arbitrary or they'll take points off.** Arbitrary in this case means, you don't know anything about this element z other than the fact it is in S . Basically, it is saying when we just randomly choose any element of S , we must show it is guaranteed to also be in T .

Advice 5.21 — With set proofs, it helps to convert stuff in to plain English to make it less intimidating. For example, if $z \in A \cap B$, that is saying z is in A and z is in B .

§5.2 Logical Structure

Definition 5.22 (Statement/Proposition). A statement/proposition is a sentence that is either true or false.

Statements can be given variable names so it is easier to refer to. We can say, for example, $p =$ “This is hard.”

Here are some logical operators:

- \wedge means “and”
- \vee means “or”
- \Rightarrow means “implies”
- \neg means “not”

A lot of problems will usually be in the implication form: “If this is true, prove this is true.”

Definition 5.23 (Biconditional). These are logical statements in the form “if and only if.” Such a statement would be p holds if and only if q holds.

To prove a biconditional, you always need to show $p \Rightarrow q$ AND $q \Rightarrow p$. If you don't do both, the proof isn't complete.

Advice 5.24 — Sometimes, when dealing with problems where you have to write a gnarly sentence in terms of other statements and other logical operators, it helps to define “helper statements.” You'll see in the next example.

Example 5.25

Consider the statement “If I play video games or go to the movies, then I'm not going to the party and I'm staying up all night.” If

- p = “I play video games”
- q = “I go to the movie”
- r = “I'm going to the party”
- s = “I'm staying up all night”

Write the statement in terms of p , q , r , and s and logical operators.

Seems like an ugly statement. But let a = “I play video games or go to the movies.” And let b = “I'm not going to the party and I'm staying up all night.” Then, the statement reads “If a , then b ” which is a much cleaner to work with instead of tripping over ourselves. So it is more clear the statement is $a \Rightarrow b$. But now, we need to write a and b in terms of p , q , r , and s . We can play the same trick again, where we divide a into parts we can manage. We have the exact statement for p , then an or, then the exact statement for q . So $a = p \vee q$. Similar reasoning implies $b = \neg r \wedge s$.

So the statement original statement is $a \Rightarrow b$ or $(p \vee q) \Rightarrow (\neg r \wedge s)$.

§5.3 Truth Tables

Truth tables are tables used in logic to figure out when a logical statement is true or false. The logical statement r comprised of some combination of statements p and q may have a different truth value depending on the truth values of p and q .

Memorize this below:

p	q	$p \wedge q$	$p \vee q$	$p \Rightarrow q$
T	T	T	T	T
T	F	F	T	F
F	T	F	T	T
F	F	F	F	T

Intuitively, the conjunction and disjunction make sense. For “and” to be true, we need p and q to both be true, so it is only true in that one instance. For “or” to be true, we need at least one of p or q to be true. So it should only be false in the instance where both p and q are false.

Implication is tricky. Implication is saying if p is true, it must mean that q is also true. So it makes sense for $p \Rightarrow q$ to be false when p is true and q is false. For the cases where p is false, it is less obvious. The implication only deals with when p is true. So p being false doesn't compromise validity of the implication, so we just mark it as true in those instances.

Advice 5.26 — Like with expressing complicated statements in terms of logical operators, it helps to make intermediate statements if you ever are working on a truth table with a complicated statement. For example, suppose you're making a truth table of $(p \wedge \neg q) \Rightarrow (q \vee p)$. It helps to define $a = p \wedge \neg q$ and $b = q \vee p$, while figuring those out separately. Then, work on $a \Rightarrow b$.

Advice 5.27 — DeMorgan's Laws will come in helpful when dealing with negations.

To prove two statements are logically equivalent, you have to show they have the same truth value always in the truth table.

§5.4 Functions

How you were probably taught what a function is, was through the machines analogy. Think of a function as a machine. You put something (the input) into the machine. Then, you get an output.

Suppose we have a function f . The notation for defining f is $f : A \rightarrow B$, where A and B are sets. Specifically:

- A is called the domain of f . It is the set of all possible inputs of f .
- B is called the codomain of f . It is the set of all candidates for the output of f .

We usually establish what is in A and B before we determine what input becomes what output. Then, to finish creating the function, we take every value in A and establish that putting it into the machine will yield some value in B .

For example, suppose we establish that $A = \{1, 2, 3\}$ and $B = \{1, 2\}$ for a function f . One way we could make f is having $f(1) = 1$, $f(2) = 2$, and $f(3) = 1$. Note that a value in the codomain doesn't have to actually be used for f , but every value in the domain has to be used. So we could make f by saying $f(1) = f(2) = f(3) = 1$, where we don't use 2 at all in B . But we can't say $f(1) = f(2) = 2$ and then call it a day because we neglected the 3 in A . You also can't have inputs that are not in A or outputs not in B . So $f(4)$ in this case doesn't exist, since 4 isn't in A . And we can't have $f(1) = 3$, since 3 isn't in B .

Definition 5.28 (Range). The set of outputs of a function f that are actually used

Caveat 5.29. Do NOT conflate codomain and range!

With that $A = \{1, 2, 3\}$ and $B = \{1, 2\}$ for a function f example, we could do $f(1) = f(2) = f(3) = 2$. The codomain is still $\{1, 2\}$. But the range is $\{2\}$, since only 2 actually shows up as an output.

Definition 5.30 (Surjective function). A function is surjective when its codomain equals its range. Another way of saying this is, every value in the codomain is actually used in the range.

If a function $f : A \rightarrow B$ is surjective, then we can say $|A| \geq |B|$. Here's how you see this intuitively. In a surjective function, we have to use up each output at least once. If there are less possible inputs than there are possible outputs, there is no way we can use up each output at least once. Also, don't think this holds the other way around: just because $|A| \geq |B|$, that doesn't mean f is surjective. This implication only works one way.

Definition 5.31 (Injective function). A function is injective when each input maps to a different output. Another way of saying this is, no two inputs will lead to the same exact output.

If a function $f : A \rightarrow B$ is injective, then we can say $|A| \leq |B|$. Another intuitive explanation. If there are more possible inputs than there are possible outputs, some output will have to be used more than once. But an injective function can't have an output being used more than once, as that would contradict its definition. Similar to a surjective function, don't think this holds the other way around: just because $|A| \leq |B|$, that doesn't mean f is injective. This implication only works one way.

In fact, for an injective function f , the elements in the range of f form a partial permutation of the codomain of f . See if you can intuitively understand why this is true.

§5.5 Bijections

Definition 5.32 (Bijective function). A bijective function is both injective and surjective. This is also known as a one-to-one correspondence. Another way of thinking of this is each output in the codomain is actually used and used by exactly one input. Essentially, each output is always associated with exactly one input.

Generally, if the course asks you to prove a function is bijective, the formal way of going about it is showing it is injective and surjective.

However, bijective doesn't merely appear in the context of functions. A bijection can refer to anything where there is a one-to-one correspondence or association. Consider these two problems:

- Find the number of nonnegative integer solutions to $x_1 + x_2 + x_3 = 7$.
- Find the number of ways to distribute 7 indistinguishable marbles into 3 distinguishable urns.

Remember stars and bars? Knowledge and application of stars and bars should show both have the exact same methodology. But let's think about the actual things being counted.

Suppose we consider a tuple for the first problem like $(4, 2, 1)$. But this is associated with a way to distribute the marbles in the second problem: 4 marbles in the first urn, 2 marbles in the second urn, and 1 marble in the third urn. So each tuple can be turned into a marble distribution.

It is also clear this works the other way around: we can turn a marble distribution into a tuple. This establishes a bijection between the tuples and distributions, since we now know each tuple is associated with a unique distribution and vice versa.

Caveat 5.33. To establish a bijection between two types of things, say for the sake of example, donuts and chairs. You have to show each donut corresponds to a unique chair (so if two donuts correspond to the same chair, it's game over for the bijection). Then, you have to show each chair corresponds to a unique donut. If you don't do both of these steps, you don't have a bijection.

§6 Probability

Probability, in Layman's terms, is just a measure of how likely something is to occur. "There is a 50% chance it will rain today" is one example of probability at hand. It's a simple concept but with a lot to learn. Some of the counting stuff will come in handy here.

§6.1 Probability Space and Events

Definition 6.1 (Probability Space). A probability space basically has 2 essential components: a set of outcomes and a probability function.

An outcome is basically a result out of the random procedure. However, you want to make sure the way you define your outcomes, actually covers every possible outcome, and doesn't cover an outcome more than once.

For example, if you're rolling a 6 sided dice and checking the number, you don't want to list rolling 1, 2, or 3 as the only possible outcomes because you could get 4, 5, or 6 as well. You also don't want to list "rolling a 1" more than once.

Definition 6.2 (Probability function). A probability function for a probability space is a function that basically tells us how likely each outcome is.

A probability space is usually defined as (Ω, Pr) , where Ω corresponds to the set of all possible outcomes, while Pr is the probability function.

Definition 6.3 (Event). Some subset of the possible outcomes in a probability space that fall under a certain category or condition. For example, if I'm rolling a dice once, the outcomes are I get 1, 2, 3, 4, 5, or 6. So me rolling a 1 is an example of an event. Me rolling an even number is also an example of an event. Events will make use of the notation associated with sets like cardinality or intersection.

Caveat 6.4. You ALWAYS, ALWAYS have to define the probability space at the start of any probability problem if they don't define it for you. They'll take points off if you don't. Let's just do an example to get us all on the same page.

Suppose I have a magic urn. I draw a red ball from it with probability $\frac{1}{2}$, a blue ball with probability $\frac{1}{3}$, and a green ball with probability $\frac{1}{6}$. When defining the probability space, just be sure to say what's the action being taken that will result in an outcome (in this case, drawing the ball from the urn) and also what we are paying attention to as the outcome (in this case, the ball's color).

So here, I would say: Let (Ω, \Pr) be the probability space associated with me drawing a ball from the urn and checking its color.

That's all I need to say. I don't need to give the layout of the probability function, unless they specifically say for me to do it.

As for events and the probability function, you define events in the way you define a new variable. Like, we can say "Let A be the event that the ball I draw is red." Usually, you'll be tasked with finding the probability that some event occurs, but you should still define other events to get to that point when necessary. Also, we know the probability that A occurs, so we can say $\Pr[A] = \frac{1}{2}$ (essentially saying A has a $\frac{1}{2}$ probability of occurring).

Also, $\Pr[\Omega] = 1$ (the whole space has probability 1). Also, any event that happens in any probability space will have a \Pr between 0 and 1, inclusive.

§6.2 Uniformity

So Ω is basically a set with $|\Omega|$ outcomes. So we say Ω is uniform when each of the $|\Omega|$ outcomes have an equal probability of occurring.

An easy example of a uniform probability space is a fair 6 sided dice, where we check what number we roll (a fair dice is one where every number has an equal chance of showing up).

However, there are other instances of a uniform probability space.

Advice 6.5 — Keep your eyes peeled for the words “uniformly at random.” If there are multiple random actions being done, like rolling a dice multiple times or multiple people all doing random actions, keep your eyes peeled for “independently from the other [stuff].”

So suppose it is only 1 action, like a single dice roll or a single draw from a basket, and you are told the result is uniform and random. Then, you can say because of the “uniform and random,” you can conclude Ω is a uniform space.

But now suppose it is multiple actions. You can only say Ω is uniform if each action is independent of all the others and each action individually is a uniform and random choice.

A nifty formula of a uniform space Ω is that for any event A in Ω , we have

$$\Pr[A] = \frac{|A|}{|\Omega|}.$$

If you identify Ω is uniform, you should always consider this formula if you are tasked with finding a probability. This formula should make sense intuitively because in a uniform space, every outcome has a $\frac{1}{|\Omega|}$ chance of occurring, since each outcome is equally likely to occur. Then, there are $|A|$ outcomes that fit under the event A .

Caveat 6.6. Do NOT ever use this formula if you can't justify that Ω is uniform. You also need to have a sentence or two saying why you can conclude Ω is uniform (generally from the words in the problem statement like “uniform and random” as mentioned earlier).

Also, you'll usually need to find the actual value of $|\Omega|$ when this formula is applicable. When it is one action being done, usually the number of outcomes is easy. But when it is multiple independent outcomes, where each action is made uniformly and randomly, we use our good old friend, the MR. (You still have to explicitly cite it even at this stage in the course)

Suppose we're rolling a 6-sided die and then flipping a coin, where the die and coin are fair (so both are uniform and random) and are independent of each other. And suppose Ω is the probability space associated with rolling the die and flipping the coin and noting the combination of the two results.

So there are 6 possibilities for what the dice can give you and 2 possibilities for what the coin can give you. Steps are independent, so by the MR, we have $6 \cdot 2 = 12$ possible outcomes. So $|\Omega| = 12$.

§6.3 Bernoulli Trials

Definition 6.7 (Bernoulli trial). A Bernoulli trial is basically is a trial, with only two outcomes: “success” and “failure.” Also, the probability of success must always be the same any time the trial is performed.

Here’s an example of a Bernoulli trial: suppose I’m playing baseball, and for some reason, I know when I’m batting, I always have a $\frac{3}{10}$ chance of hitting the ball.

That can be described as a Bernoulli trial, where success is when I hit the ball. I have a $\frac{3}{10}$ chance of success and a $\frac{7}{10}$ chance of failure.

On the flipside, I could make the objective for “success” the opposite. Meaning, I could make a Bernoulli trial where “success” is when I miss the ball. Then, I have a $\frac{7}{10}$ chance of success and a $\frac{3}{10}$ chance of failure.

Flipping a coin and seeing whether it is heads, is an example of a Bernoulli trial.

Rolling a 6-sided dice and checking the number is NOT a Bernoulli trial because there are 6 outcomes, not 2. However, rolling a 6-sided dice and checking whether the number I roll is 3 or not, is a Bernoulli trial because there are 2 outcomes.

§6.4 Probability Properties

Remember some of the techniques from the counting section? The intuition from those will come in handy here.

For an event A , define \bar{A} to be the complement of A . The complement of an event A , is basically the event that A doesn’t occur at all. Then, we have

$$\Pr[A] = 1 - \Pr[\bar{A}].$$

The tip-offs to try this formula are the same as complementary counting: do it when you think the opposite is easier to find. Also, do it when you are tasked with finding the probability that “at least one of [this] satisfies [this].”

Also, Principle of Inclusion/Exclusion returns. This holds for any events A and B in the same probability space:

$$\Pr[A \cup B] = \Pr[A] + \Pr[B] - \Pr[A \cap B].$$

And like with the formula for PIE, a similar results comes out for 3 events:

$$\Pr[A \cup B \cup C] = \Pr[A] + \Pr[B] + \Pr[C] - \Pr[A \cap B] - \Pr[B \cap C] - \Pr[C \cap A] + \Pr[A \cap B \cap C]$$

A result of the Principle of Inclusion/Exclusion with 2 events is if A and B are disjoint:

$$\Pr[A \cup B] = \Pr[A] + \Pr[B]$$

because $\Pr[A \cap B] = 0$ when A and B are disjoint (no overlap).

Do NOT use the above equation if A and B overlap.

This is essentially like the addition rule for in counting, where the cases are disjoint and exhaustive (recall that if $z \in A \cup B$, then it must be the case that $z \in A$ or $z \in B$, which shows exhaustivity).

You can also extend this property. If you have events A_1, A_2, \dots, A_n where any two of them are disjoint (or if each one is independent of all the others), you can say

$$\Pr[A_1 \cup A_2 \cup \dots \cup A_n] = \Pr[A_1] + \Pr[A_2] + \dots + \Pr[A_n]$$

Do NOT use the above equation if any two of the events overlap. But this is like breaking into multiple disjoint cases.

So basically, we have tools for dealing with non-overlapping cases and tools for dealing with overlapping cases (as well as complementary counting), just in a probability context.

§6.5 Independence

Let A and B two events in the same probability space. We say A and B are independent if and only if:

$$\Pr[A \cap B] = \Pr[A] \cdot \Pr[B].$$

The notation for A and B being independent is $A \perp B$. You may be tasked on a homework to prove two events A and B are independent. The way they want you to do that is by using the above formula. But most of the time, you will be given problems where actions are independent, you figure out $\Pr[A]$ and $\Pr[B]$, and then you apply the formula to find $\Pr[A \cap B]$.

Example 6.8

Suppose we're rolling a 6-sided die and then flipping a coin, where the die and coin are fair (so both are uniform and random) and are independent of each other. What is the probability that the number on the die is a 6 and the coin flips heads?

Same drill, define Ω as the probability space associated with rolling the die and flipping the coin and noting the combination of the two results. Let A be the event that we roll a 6 on the die, and B be the event the coin flips heads. We know $\Pr[A] = \frac{1}{6}$ because it is a 6-sided fair die (so each side is equally likely). Also, $\Pr[B] = \frac{1}{2}$. Our goal is to find $\Pr[A \cap B]$ (the event that both A and B occur). But we know the die and coin are independent of each other, as given. So because of independence, we can use the formula:

$$\Pr[A \cap B] = \Pr[A] \cdot \Pr[B] = \frac{1}{6} \cdot \frac{1}{2} = \frac{1}{12}.$$

Caveat 6.9. Always explicitly cite independence when using that formula. Also, they usually expect some justification on why you can conclude the actions are independent, even if it is merely restating stuff in the problem statement.

Definition 6.10 (Pairwise independence). A collection of events A_1, A_2, \dots, A_n are pairwise independent if any two of them are independent.

Definition 6.11 (Mutual independence). A collection of events A_1, A_2, \dots, A_n are mutually independent if each event in the collection is not effected by any combination of the other events in the collection at all.

Note that if the collection of events is mutually independent, then they are pairwise independent events, too. But the other way around isn't always true. You're not going to be using pairwise independence as much as mutual independence.

Advice 6.12 — If you're ever given something like “a dice is rolled 5 times, **where each roll is independent of other rolls,**” you can use mutual independence. Make sure to restate the words in the problem that allow you to conclude mutual independence.

The most important property of mutually independent events A_1, A_2, \dots, A_n is this:

$$\Pr[A_1 \cap A_2 \cap \dots \cap A_n] = \Pr[A_1] \cdot \Pr[A_2] \cdot \dots \cdot \Pr[A_n],$$

basically an extension of the original independence equation. Very useful formula when you have multiple choices going on and the condition counts on you to monitor the combination of choices.

Caveat 6.13. Like with independence with two events, you have to mention the events are mutually independent when using the formula above.

Example 6.14

A fair 6-sided die is rolled 3 times, where each roll is independent of all the others. What is the probability the first roll is a 4, the second roll is even, and the third roll is NOT 6?

Let Ω be the probability space associated with rolling the die 3 times, and checking the 3 rolls.

So let A be the event the first roll is a 4, B be the event the second roll is even, and C be the event the third roll is NOT 6. The event we want the probability of, expressed as the events we defined already, is $A \cap B \cap C$. Since A , B , and C each focus on different rolls, and since we know each roll is independent of all the others, we can say A , B , and C are mutually independent. So because of mutual independence, we know

$$\Pr[A \cap B \cap C] = \Pr[A] \cdot \Pr[B] \cdot \Pr[C].$$

Obviously, we know $\Pr[A] = \frac{1}{6}$. For finding $\Pr[B]$, well there are 6 possible rolls, each equally likely. And only 3 of them result in even numbers. So $\Pr[B] = \frac{3}{6} = \frac{1}{2}$.

As for $\Pr[C]$, we can try the complement, $\Pr[C] = 1 - \Pr[\overline{C}]$. The event \overline{C} basically means the event we roll a 6, so $\Pr[\overline{C}] = \frac{1}{6}$. Hence, $\Pr[C] = \frac{5}{6}$.

Putting everything together,

$$\Pr[A \cap B \cap C] = \Pr[A] \cdot \Pr[B] \cdot \Pr[C] = \frac{1}{6} \cdot \frac{1}{2} \cdot \frac{5}{6} = \frac{5}{72}.$$

Advice 6.15 — Also, as you can see with the examples, your target may be some event filled with “and”s and “or”s when try writing the conditions down to fulfill the event in plain English. It helps to define simpler events and then express your target event with the simpler events and stuff like \cap , \cup , and the complement. This makes it easier to see what tools you should use for the job.

§6.6 Conditional Probability

In a probability space, $\Pr[A \mid B]$ is notation for conditional probability, where we want the probability that A happens **given** that B happened.

The “given” concept may be tricky, but think of it this way. Thus far, when working with probability, our scope is Ω . That is, when we want to compute the probability some event A happens, we essentially scan through Ω , find the stuff that fits the bill for A , and use that to make our probability. So Ω is the area we search basically.

Now, let's suppose we're working on $A \mid B$, or A conditioned on B . Instead of our scope being Ω , our scope is B . So we scan through B , find the stuff that fits the ball for A , and use that for our conditional probability.

The formula for conditional probability is

$$\Pr[A \mid B] = \frac{\Pr[A \cap B]}{\Pr[B]}.$$

The explanation earlier should make this formula easier to memorize. We scan through B , so that's why the denominator is $\Pr[B]$. Then, the stuff we want is the stuff that fits under A as we search through the stuff under B . In other words, the stuff we want is $A \cap B$, so that's why the numerator is $\Pr[A \cap B]$.

§6.7 Chain Rule

We can rearrange the conditional probability formula as follows:

$$\Pr[A \cap B] = \Pr[B] \cdot \Pr[A \mid B].$$

Same equation, we just multiplied both sides by $\Pr[B]$.

This is the chain rule. We can extend by letting $B = X \cap Y$, where X and Y are some events and plugging it in:

$$\Pr[A \cap X \cap Y] = \Pr[X \cap Y] \cdot \Pr[A \mid (X \cap Y)].$$

But from above, we know $\Pr[X \cap Y] = \Pr[Y] \cdot \Pr[X \mid Y]$, so

$$\Pr[A \cap X \cap Y] = \Pr[Y] \cdot \Pr[X \mid Y] \cdot \Pr[A \mid (X \cap Y)].$$

I'll replace X with B and Y with C to make it easier to see.

$$\Pr[A \cap B \cap C] = \Pr[C] \cdot \Pr[B \mid C] \cdot \Pr[A \mid (B \cap C)].$$

I'll just tell you, the chain rule with 4 events is:

$$\Pr[A \cap B \cap C \cap D] = \Pr[D] \cdot \Pr[C \mid D] \cdot \Pr[B \mid (C \cap D)] \cdot \Pr[A \mid (B \cap C \cap D)].$$

Can you see where this is going? We basically start with needing D to be fulfilled. Then, we work our way through the events, and use the fact that all the previous events have been fulfilled, so act as “givens” for each step of the way.

Anyway, the chain rule is useful in scenarios where you have two events A and B , where A and B aren't independent, but as long as A is done, it doesn't effect the conditions that make it easy to calculate B after. I'll explain what I mean in an example.

Example 6.16

I have 10 balls, numbered with the integers 1, 2, \dots , 10. I take two balls without replacement, uniformly, randomly, and independently. What is the probability the first ball has an odd number and the second ball has an even number?

So first things first, let Ω be the probability space associated with drawing two balls and checking their numbers. Let A be the event the first ball has an odd number, and B be the event the second ball has an even number. The event we want is $A \cap B$.

So there are 5 balls with an even number and 5 balls with an odd number. So $\Pr[A] = \frac{5}{10} = \frac{1}{2}$. But by the chain rule, we know

$$\Pr[A \cap B] = \Pr[A] \cdot \Pr[B \mid A].$$

So all that's left is finding $\Pr[B \mid A]$. Let's think about B conditioned on A , and what that really means. So we know A has happened. That means we drew an odd ball first. No matter which odd ball it is, we have 4 odd balls and 5 even balls left for the second choice. Aha, so $\Pr[B \mid A]$ is easy to calculate, being $\Pr[B \mid A] = \frac{5}{9}$. This is because regardless of how A was performed, the conditions always allow the same chance that B is performed successfully after. So

$$\Pr[A \cap B] = \Pr[A] \cdot \Pr[B \mid A] = \frac{1}{2} \cdot \frac{5}{9} = \frac{5}{18}.$$

Now, if it weren't always the case after fulfilling A , that we were left with the same distribution of even vs odd balls, the chain rule wouldn't be as nice.

§6.8 Random Variables

A random variable is basically a quantity whose value is based on the results of the random events in a probability space.

Suppose I flip a coin 5 times and I make a random variable H denoting the number of heads I flip of those 5 times. One time, I could flip 4 heads. Then, $H = 4$. Another time, I could make another 5 flips and say, 0 of them end up heads. Then, $H = 0$.

Basically, a random variable is like a function that takes an outcome from a probability space and returns whatever quantity the random variable is defined for. In this case, it takes in a sequence of coin flips and returns the number of heads.

You can also use the probability function with a random variable to find the probability a random variable will equal a specific value. Like $\Pr[H = 4]$ denotes the probability that H will equal 4.

§6.9 Expectation

The notation $E[X]$, where X is a random variable, denotes the expected value of X . Think of expected value as a weighted average. Basically, the weighted average takes into account each possible value of X and how likely each possible value of X is to occur to make its “expected value.”

Here's an intuitive example. Suppose X is a random variable such that either X is 1 or 2 and $\Pr[X = 1] = \frac{1}{2023}$ and $\Pr[X = 2] = \frac{2022}{2023}$. Clearly, $X = 2$ is so much more likely than $X = 1$. So you'd expect X to usually come up as 2, meaning $E[X]$ should be closer to 2 than 1. Normally though, there is more than 2 values of X and the probabilities don't make an intuitive explanation easy to comprehend. However, here is the formula for expected value:

$$E[X] = \sum x \cdot \Pr[X = x],$$

where x ranges across every possible value X can equal. The formula should be intuitive: we take every possible value of X and weight it based on how likely that value is to occur. Values of X that are unlikely to occur have a smaller probability associated with them and thus, a smaller weight to the expected value. But values of X that are likely to occur have a bigger probability associated with them and thus, a bigger contribution to the expected value.

§6.10 Linearity of Expectation and Indicators

Linearity of expectation goes as follows. Let X_1, X_2, \dots, X_n be random variables and let c_1, c_2, \dots, c_n be real constants (NOT random variables). Then, we have

$$E[c_1X_1 + c_2X_2 + \dots + c_nX_n] = c_1E[X_1] + c_2E[X_2] + \dots + c_nE[X_n].$$

The nifty part is, X_1, X_2, \dots, X_n don't even have to be independent in any capacity for the formula to work. Essentially, we can break the expected value over addition and pull constants out of the expected value. Simple applications of this formula would be $E[cX] = cE[X]$, where c is a constant and X is a random variable. But the form you're most often going to be using is

$$E[X_1 + X_2 + \dots + X_n] = E[X_1] + E[X_2] + \dots + E[X_n].$$

Remember Bernoulli trials? These will be extremely helpful here. You'll see.

Example 6.17

Suppose we're rolling a fair 6-sided dice 2023 times, where each roll is independent of all the others. Find the expected number of times a 1 shows up.

So let X be the random variable that returns the number of times a 1 shows up in the 2023 rolls. Imagine trying to find the probability each possible value of X shows up!

But here's the trick: instead of thinking about the 2023 rolls as a whole, maybe consider each roll's expected contribution to X .

First, we're going to need to revisit Bernoulli trials. So let's suppose we have some event E that succeeds with probability p . Then, E fails with probability $1 - p$ (the complement).

Definition 6.18 (Indicator). For an event E , the indicator I of the event is a random variable. I will return 1 if E succeeds and 0 if E fails.

Using this definition of an indicator, we can think about a dice roll's contribution to X . It should contribute 1 to X should a 1 be rolled (success of the event) and 0 to X if a 1 isn't rolled (failure of the event).

So let's define the 2023 events $A_1, A_2, \dots, A_{2023}$, where A_k is the event that a 1 is rolled in the k th roll. Let I_k be the indicator variable of the event A_k .

So I_k returns 1 if A_k succeeds, and 0 if A_k fails. Clearly, X is the number of A_k that succeeds, by definition of X . So $X = I_1 + I_2 + \dots + I_{2023}$.

So we want

$$E[X] = [I_1 + I_2 + \dots + I_{2023}].$$

Here's where LoE (linearity of expectation) comes in:

$$E[X] = E[I_1 + I_2 + \dots + I_{2023}] = E[I_1] + E[I_2] + \dots + E[I_{2023}].$$

Now, any indicator variable I has nice properties with expected value. Suppose let's say $\Pr[I = 1] = p$. Remember that I is only ever 0 or 1. Using the expected value formula,

$$E[I] = 0 \cdot \Pr[I = 0] + 1 \cdot \Pr[I = 1] = 1 \cdot p.$$

If A is the event associated with I , then $p = \Pr[A]$. So $E[I] = \Pr[A]$. This is a nifty property.

So using this, we can change the expected indicators into probabilities:

$$E[X] = E[I_1] + E[I_2] + \cdots + E[I_{2023}] = \Pr[A_1] + \Pr[A_2] + \cdots + \Pr[A_{2023}].$$

The rest is easy. We know a 1 is rolled with probability $\frac{1}{6}$ on any roll, since the dice is fair and the rolls are independent. So we have

$$E[X] = 2023 \cdot \frac{1}{6} = \frac{2023}{6}.$$

So that was a lot to take in but here's the breakdown, so you can apply this to other problems:

- Strongly consider this solution path when you have a large number of things and want to see the expected number that “succeeds” with some condition.
- Define separate events that each thing succeeds.
- Make indicator variables for each of those events.
- Suppose $E[X]$ is your target, where X is a random variable. Express X as the sum of the indicators (the contribution of each individual thing).
- Apply LoE to get terms of the form $E[\text{indicator}]$ in a sum.
- Use the property $E[I] = \Pr[A]$ if I is the indicator of some event A .
- Find the probability of each of those events, add them up, and you're done.

Caveat 6.19. ALWAYS cite LoE when using it.

§6.11 Independent Random Variables

Two random variables X and Y in the same probability space are considered independent if and only if

$$\Pr[(X = x) \cap (Y = y)] = \Pr[X = x] \cdot \Pr[Y = y]$$

for any value of x that X can equal and for any value of y that Y can equal. Another way of saying this condition for random variables being independent is, the events $X = x$ and $Y = y$ are always independent, no matter what x and y are.

Generally, how they want you to prove random variables are independent is to show

$$\Pr[(X = x) \cap (Y = y)] = \Pr[X = x] \cdot \Pr[Y = y]$$

for an arbitrary combination of real numbers x and y . In other words, to show independence of random variables, we need to show no matter how we select x and y , that above equation holds.

Example 6.20

Let X and Y be independent random variables. Prove that $2X$ and $3Y$ are also independent random variables.

This should be obviously intuitively: why should scaling X or Y effect independence at all? But that's not how they want you to prove it. Instead we need to show for an arbitrary combination of real numbers a and b , we have

$$\Pr[(2X = a) \cap (3Y = b)] = \Pr[2X = a] \cdot \Pr[3Y = b].$$

Let's work on the left hand side. The event $2X = a$ holds only when $X = \frac{a}{2}$. Similarly, the event $3Y = b$ holds only when $Y = \frac{b}{3}$. So we can rewrite the event:

$$\Pr[(2X = a) \cap (3Y = b)] = \Pr[(X = \frac{a}{2}) \cap (Y = \frac{b}{3})].$$

This is better to work with. Why? Because we know X and Y are independent and can use the property about it:

$$\Pr[(X = \frac{a}{2}) \cap (Y = \frac{b}{3})] = \Pr[X = \frac{a}{2}] \cdot \Pr[Y = \frac{b}{3}].$$

But we can convert those events back: $\Pr[X = \frac{a}{2}] = \Pr[2X = a]$ and $\Pr[Y = \frac{b}{3}] = \Pr[3Y = b]$. We can combine all our work into a string of equalities:

$$\Pr[(2X = a) \cap (3Y = b)] = \Pr[(X = \frac{a}{2}) \cap (Y = \frac{b}{3})] = \Pr[X = \frac{a}{2}] \cdot \Pr[Y = \frac{b}{3}] = \Pr[2X = a] \cdot \Pr[3Y = b].$$

This is exactly what we wanted to show in the first place. So we are done.

§6.12 Variance

The variance of a random variable X is essentially a measure of how spread out the values are from the expected value of X . The formula is

$$\text{Var}[X] = E[X^2] - E[X]^2.$$

Unfortunately, variance problems are usually messy because dealing with $E[X^2]$ is challenging.

Advice 6.21 — Your best bets for dealing with $E[X^2]$ are computing each possible value of X and its associated probability, or using indicator variables. Note the algebraic identity

$$X^2 = (I_1 + I_2 + \cdots + I_k)^2 = I_1^2 + I_2^2 + \cdots + I_k^2 + 2(I_1I_2 + I_1I_3 + \cdots + I_{k-1}I_k),$$

where in the second term, you have the pairwise products of the k indicator variables.

Also, a property that might show up in a HW or test is that for a constant c and random variable X ,

$$\text{Var}[cX] = c^2\text{Var}[X].$$

Unfortunately, you can't break up variance like you could with expectation.

§6.13 Binomial Distribution

Remember Bernoulli trials? What if we did multiple of them? That's basically how we arrive at a binomial distribution, so long as the probability of success for each trial remains the same each time and each trial is independent of all the others. Also, the Binomial Theorem will come into play here.

Suppose we perform n trials where each trial has a probability p of success and each trial was independent of the others. Let B be the random variable associated with the number of trials that succeed of the n . Now, suppose we wanted the probability exactly k of the trials succeeded, which would be $\Pr[B = k]$.

The formula is

$$\Pr[B = k] = \binom{n}{k} p^k (1-p)^{n-k}.$$

Looks ugly, but we can think of our probability/counting techniques thus far to make it easier to memorize. So we have n total trials. We want k to succeed, so the complement means that $n - k$ of them should fail. We also have to decide where we want the k succeeding trials to occur among n possible trials. There are $\binom{n}{k}$ ways to choose k positions for the succeeding trials. The failing trials occupy everything else.

Now, we want the k succeeding trials to actually be succeeding. A trial succeeds with probability p , so that's where the p^k comes in.

Then, we want the $n - k$ failing trials to actually be failing. A trial fails with probability $1 - p$, so that's where the $(1 - p)^{n-k}$ comes in.

Some properties that you should memorize for a random variable B specifically associated with a binomial distribution (doing a bunch of Bernoulli trials):

- $E[B] = np$
- $\text{Var}[B] = np(1 - p)$

§7 Graphs

A graph in this course is a mathematical structure used to model objects and any connections or relations between the objects.

Graphs are fundamental to networks. On social media like Facebook, you may have friendships and your friends may have friendships of their own. That can be modeled with a graph. Look at a map on your phone and see the intersections and roads. That can also be modeled with a graph. Interactions between objects is such a widespread occurrence that graphs can appear in places you wouldn't expect at first.

§7.1 Basics and Definitions

§7.1.1 Introductory

A graph has two key components: vertices and edges.

Think of a vertex as a dot. Then, edges connect pairs of dots.

For now, we will work with simple graphs. This means there is never more than one edge between any pair of vertices. This also means the endpoints of each edge cannot be the same; we cannot have an edge that starts and ends at the same vertex.

The graphs will also be undirected, meaning the edges are all two-way. Later, we will work with graphs that have one-way edges.

So to construct a graph, first, we establish how many vertices we have. Then, we take every possible pair of two vertices and decide whether or not we want to draw an edge between them. After all that is done, boom, we have a graph. It is possible for a graph to have no edges at all.

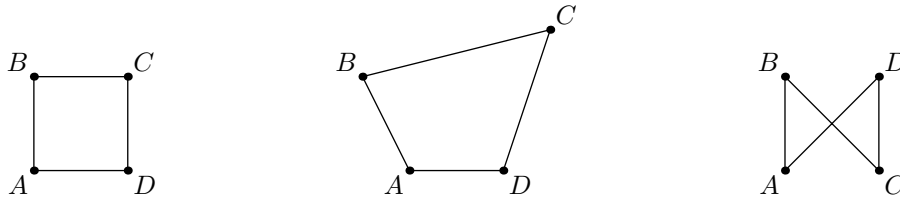
Definition 7.1 (Neighbors). Vertex a is neighbors with vertex b if a and b are connected via an edge.

Definition 7.2 (Degree). The degree of a vertex a is the number of neighbors that a has. For a simple undirected graph that we are working with now, the degree of a vertex is equal to the number of edges emanating from it. One way of denoting the degree of a vertex a is $\deg(a)$.

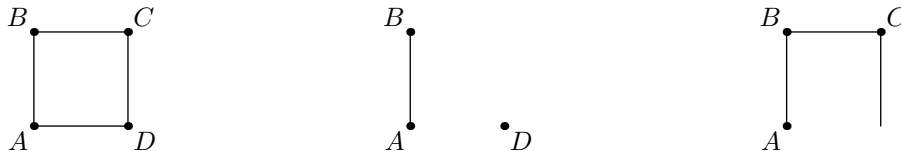
Advice 7.3 — If you are stuck on a graph problem, try proof by contradiction or proof by contrapositive. What makes graph theory problems challenging is how unrestricted graphs can be. Think about it: with just 7 vertices, you can have all sorts of configurations of edges. And the number of configurations only increase when there are more vertices. By considering proof by contradiction or contrapositive, you often create a heavy restriction on the graph that you can work with.

§7.1.2 Intermediate

Definition 7.4 (Isomorphic graphs). The mathematical definition is hard to understand, so here's the plain English way. Take any graph. Now, move the vertices around, but preserve any edges while moving the vertices around. Do not add or remove any edges either. Boom, the new graph you get is **isomorphic** with the original graph. The three graphs below are all isomorphic with each other.



Definition 7.5 (Subgraph). A subgraph of a graph G is a graph whose vertices is a subset of G 's vertices and whose edges is a subset of G 's edges. Below, the second graph is a subgraph of the first graph. However, just because we take a subset of the vertices and a subset of the edges, that doesn't mean we'll actually get a graph all the time. See the third "subgraph" below that isn't actually a graph because an edge is missing an endpoint.



Definition 7.6 (Induced subgraph). Another definition that is easier to explain in plain English. So take a graph and take a subset of its vertices. Examine every pair of vertices in the subset. If there was an edge between the two vertices in the original graph, you keep it in the new graph. Otherwise, you don't draw an edge. Basically, edges are preserved within the vertices selected for the subset. Boom, the new graph is an induced subgraph of the original. Below, the graph on the right is an induced subgraph of the graph on the left.

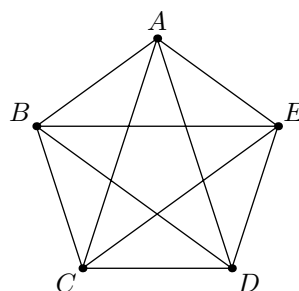


Definition 7.7 (Complement of a graph). Yet, another definition that is easier to explain in plain English. So take a graph and copy and paste the vertices for the new graph, but don't draw edges yet. Consider every pair of vertices for the new graph. If there was an edge between the two vertices in the old graph, don't draw that edge in the new graph. If there was NOT an edge between some vertices in the old graph, draw that edge in the new graph. The two graphs below are complements of each other.

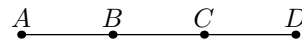


§7.2 Special Graphs

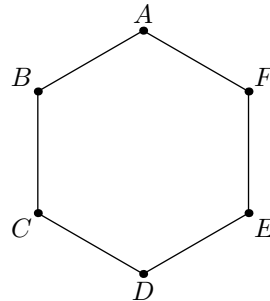
Definition 7.8 (Complete graph). The complete graph with n vertices (denoted as K_n) is the graph where an edge is drawn between every single possible pair of vertices. Below is K_5 .



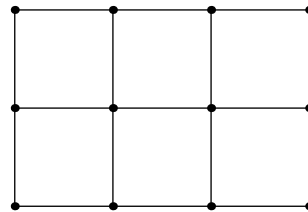
Definition 7.9 (Path graph). The path graph with n vertices (denoted as P_n) has the vertices arranged in a line with edges connecting the vertices along the line. Below is P_4 .



Definition 7.10 (Cycle graph). The cycle graph with n vertices (denotes as C_n) has the vertices arranged in a circle with edges connected the vertices around the circle. Below is C_6 .



Definition 7.11 (Grid graph). The $m \times n$ grid graph has the vertices arranged like a grid with m rows and n columns with edges creating gridlines. Below is the 3×4 grid graph.



§7.3 Handshaking Lemma

Lemma 7.12 (Handshaking Lemma)

For any graph with V as its set of vertices and E as its set of edges,

$$\sum_{a \in V} \deg(a) = 2|E|.$$

In plain English, the left hand side is saying that we take every a in the set V and sum the degrees up. So this lemma is saying the sum of the degrees of all the vertices in the graph is twice the number of edges in the graph.

To prove this, consider how we construct the graph, where we first decide the vertices but not drawing any edges yet. At this point, the total sum degree is 0 because no vertex has edges out of it yet. So the equation in the lemma makes sense at this point with both sides being 0. If we add an edge, which involves two endpoints, those endpoints are having their degree increased by 1. So the degree sum increases by 2, increasing the left hand side by 2. The number of edges increases by 1, increasing the right hand side by 2. So the equation still holds if we add an edge. If we add any more edges, it is easy to see, repeating this logic, that the equation still holds. So this proves the lemma because it will hold no matter how many edges or which edges we add.

Advice 7.13 — Whenever you have a problem where you're given a specific condition about the degree of each vertex (like every vertex is of degree 1 or 3), that is generally a sign to use Handshaking Lemma. It also works well with problems involving the parity of the degree of each vertex (like every vertex is of odd degree).

§7.4 Walks and Paths

Definition 7.14 (Walk). A walk is simply a sequence of vertices, where vertices that are adjacent in the sequence have an edge connecting them.

Definition 7.15 (Path). A path is a walk where a vertex never shows up more than once in the sequence. All paths are walks, but not all walks are paths.

Lemma 7.16

A walk in any graph can always be converted into a path.

If the walk contains no repeated vertices, then it's a path by definition.

But what if there are repeated vertices? Here's the intuitive proof that isn't formal, but is meant to show the general idea. Suppose the walk contains the vertex a more than once. So the walk will start somewhere, visit a at some point, do some stuff in between, and then revisit a , and finally continue to finish. Do we really need to do that stuff in between? Why not just continue to the finish right when we get to a instead of goofing off with the stuff in between? We can just delete that "stuff in between" and remove the repeated use of a . If there are still repeated vertices, we can reuse this logic to remove unnecessary parts to the walk until we no longer have repeated vertices. Then, it becomes a path.

Lemma 7.17 (Maximal Path Technique)

There is always a path of maximal length in any graph.

This relies on something called the Well-Ordering Principle, which says in any finite set of real numbers, there is a largest element. You don't have to prove this, just take their word for it.

Imagine if we take a graph, find all the possible paths, and made a set with the lengths of each possible path. This is a finite set. Why? Intuitively, a path can never use the same edge twice (that would easily lead to repeated vertices), and there are only a finite number of edges in the graph. So there is only a finite number of sets of edges and a finite number of orders those edges can be traversed in. Not all sets of edges and orders lead to an actual path, but that means the number of paths in the graph is less than or equal to a finite number.

So the set with the lengths of each possible path is a finite set. Using the Well-Ordering Principle, this set has a largest element. So there is a path of maximal length in the graph. This path may not be unique; there could be more than one such path. The point though is, there is a path of maximal length.

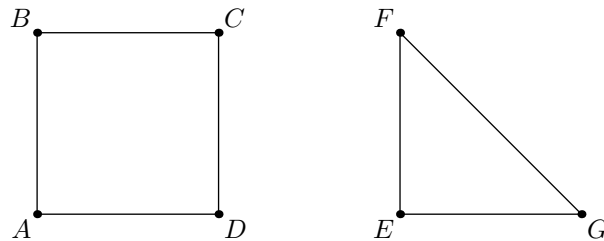
Advice 7.18 — Don't forget this technique. It can be very useful in the homework problems. We'll see later on how it can be powerful when we get to trees and cycles.

§7.5 Connected Components

Definition 7.19 (Connected vertices). Two vertices a and b are connected in a graph if there exists a walk (and by extension, a path, since a walk can be converted to a path) from a to b .

Definition 7.20 (Connected component). A set of vertices such that any two vertices in the set are connected, and we cannot add another vertex to that set so that any two vertices in the set are connected.

It is easier to understand this definition intuitively and with an example. Consider the graph below with vertices A, B, C, D, E, F , and G .



Consider the set of vertices $\{A, B, C\}$. It is obvious any two vertices in the set are connected. However, it is not a connected component as it violates “we cannot add another vertex to that set so that any two vertices in the set are connected.” If we add D to the set, then any two vertices in the set are connected. But the set of vertices $\{A, B, C, D\}$ cannot be expanded any further to have any two vertices in the set being connected. So we know $\{A, B, C, D\}$ is a connected component. So we can think of connected components as these isolated “islands” in the graph that include all the vertices in the island. It is also easy to tell by the logic that $\{E, F, G\}$ forms a connected component but something like $\{E, F\}$ would not.

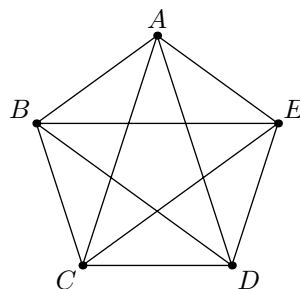
Note that a lone vertex without any edges coming out of it is a connected component on its own.

Definition 7.21 (Connected graph). A connected graph is a graph where any two vertices are connected. The graph is one connected component.

Definition 7.22 (Disconnected graph). The opposite of a connected graph: the graph has more than one connected component. Some two vertices in the graph are not connected.

§7.6 Cycles

Definition 7.23 (Cycle). A cycle in a graph is a walk of at least length 3 where the first and last vertices are the same, but other than that, no two vertices are the same. Consider K_5 below:



So $A - B - E - A$ and $B - D - E - C - B$ would form a cycle. However, $A - B - A$ and A are walks that aren't cycles because their length is not at least 3. $B - D - E - A - D - B$ is not a cycle because a vertex, other than the first and last, repeats (namely D).

Advice 7.24 — Examining cycles can be useful in problems that involve connectivity. A useful property of a cycle in a graph is that you can delete any one edge from the cycle and the graph will still be connected. Why? Because if some path between two vertices needed the deleted edge, you can go the other way around on the cycle as a detour. However, if you delete more than one edge from a cycle, then connectivity may be compromised.

§7.7 Trees, Forests, and Leaves

Definition 7.25 (Tree). A graph that is connected and acyclic (doesn't have any cycles)

Here are useful properties of trees to keep in mind:

- A tree is connected and acyclic (like the above definition)
- For any two vertices in a tree, there is a unique path between the two vertices

- It is minimally connected (if you remove any edge from a tree, the graph is no longer connected)
- If a tree has n vertices, it has $n - 1$ edges
- It is maximally acyclic (If you add a new edge between any two vertices, a cycle WILL form)

The neat thing about these properties is, if you find any one of those bullet points is true for some graph, you can conclude that graph is a tree and then you unlock all the other properties on the list. Also, if you're given a graph is a tree, you immediately have access to all 5 of those properties.

Advice 7.26 — This is where proof by contradiction can be very helpful, especially when the goal is to prove the graph is disconnected or has a cycle. Because by making a contradiction statement, you sometimes force the graph to be a tree by the connected and acyclic property. Then, you get to use multiple restrictive properties of trees that most graphs don't have.

Definition 7.27 (Forest). A graph where each connected component is a tree.

Definition 7.28 (Leaf). A vertex in a graph (tree or non-tree) that is of degree 1.

Lemma 7.29

A tree with at least one edge always has at least two leaves.

Remember that maximal path technique? This will come in handy. So we know a path of maximal length exists in a tree. Let's call this path $v_1 - v_2 - \dots - v_{k-1} - v_k$, where $v_1, v_2, \dots, v_{k-1}, v_k$ are vertices. Since there is an edge in the graph, that means there is a path using at least 2 vertices. So $k \geq 2$, since this path is of maximal length.

Now, we will show v_1 is a leaf through proof by contradiction. So assume for the sake of contradiction that v_1 is not a leaf, meaning $\deg(v_1) \neq 1$. Since v_1 is neighbors with v_2 , v_1 would have to have at least 1 other neighbor for $\deg(v_1) \neq 1$. Suppose this other neighbor is w . If a maximal path were actually maximal, then we should never be able to extend the path to get a longer one, as that would mean the "maximal path" isn't maximal.

- If w is not already in the path, then $w - v_1 - v_2 - \dots - v_k$ forms a path longer than our alleged maximal path. But this is a contradiction because this means the "maximal path" wasn't actually maximal.
- So then w has to be a vertex already in the path. But drawing an edge from v_1 to any vertex already in the path would either mean two edges between the same two vertices (not allowed), or it would form a cycle. But a tree is acyclic, so the presence of a cycle creates a contradiction.

All cases lead to contradiction, so we can conclude v_1 is a leaf. Similarly, v_k is a leaf, so we have the two leaves.

Advice 7.30 — Note that the maximal path technique can be used in non-tree graphs. As you can see from the example, there were heavy restrictions posed on the endpoints of the paths. The endpoints can't have neighbors outside the path, or the path can be extended. It can be helpful to prove a vertex with a degree property exists, or that a cycle with a specific property exists.

Advice 7.31 — When working with trees though, maximal path is a good idea, especially concerning leaves.

§7.8 Cut Edge

Definition 7.32 (Cut edge). An edge that, when removed from a graph G , results in an increase in the number of connected components of G

Lemma 7.33

An edge is a cut edge if and only if it is not part of any cycle.

Here's the intuitive explanation and not formal. Just in a way that's easy to understand.

If an edge is removed from a cycle, that doesn't create another connected component. Why? Because any path that used the deleted edge can just detour by going the other way around the cycle. So no edges part of a cycle can be a cut edge.

But which edges not part of a cycle are cut edges? All of them? Only some of them in certain conditions?

Well, let's consider an arbitrary edge $u - v$ in a graph G that isn't part of a cycle. There cannot be another path from u to v besides $u - v$, as that would imply the existence of a cycle. So the only way from u to v is through that edge. So deleting it would destroy connectivity, creating another connected component. So any arbitrary edge not part of a cycle is a cut edge, as deleting it would destroy connectivity.

This also links back to that property about trees, where it is minimally connected, since trees have no cycles.

§7.9 Spanning Trees

Definition 7.34 (Spanning subgraph). A subgraph that reuses all the vertices from the original graph.

Definition 7.35 (Spanning tree). A spanning subgraph that is a tree. Only works if the original graph is connected.

Definition 7.36 (Spanning forest). A spanning subgraph that is a forest.

Lemma 7.37

All connected graphs have a spanning tree.

Here's an intuitive, informal proof that is easy to understand.

If the original graph G is a tree (meaning it has no cycles), then great. The subgraph itself counts as a spanning tree.

Otherwise, G has cycles but is connected. Remember that edges from cycles are not cut edges, meaning you can delete them without destroying connectivity. So is delete any of the edges in a cycle. If there are no more cycles, great, it is a spanning tree. Otherwise, keep deleting edges to destroy the cycles that are still there after prior deletions. Since G has a finite number of cycles, you will eventually remove them all by continuing this process. Once you no longer have cycles, you stop.

§7.10 Coloring

A coloring of a graph refers to where we have a selection of colors and we color each vertex.

Definition 7.38 (Proper coloring). A coloring of a graph where there any two vertices that are neighbors have different colors.

For the most part, colorings in this course focus on proper colorings. Also, a graph that can be properly colored using k colors is k -colorable.

For a graph G , the minimum number of colors necessary to color G is given by $\chi(G)$. There are cases where it is easy to find $\chi(G)$, but other than that, finding the exact value can be challenging.

Advice 7.39 — Suppose you want to prove $\chi(G) = n$ for some integer n . You need to do 2 things:

1. Show an example of a proper coloring done with n colors
2. Show that it is impossible to properly color the graph with $n - 1$ colors

The second one is important because, just because I find a way to color G with n colors, that doesn't mean it is actually minimal.

Some easy chromatic numbers:

- $\chi(P_n) = 2$ for $n \geq 2$
- $\chi(C_n) = 2$ for even $n \geq 4$
- $\chi(C_n) = 3$ for odd $n \geq 3$
- $\chi(K_n) = n$

First one should be easy to see (alternating in colors). Same with the second one. The third one comes from trying to use 2 colors in an alternating fashion until you reach a conflict trying to color the last vertex, so that requires 3 colors. Fourth one follows from the fact that any two vertices are connecting in K_n , which forces every vertex to be a different color.

Also, any graph with an edge obviously has $\chi(G) \geq 2$.

Lemma 7.40

A graph can be properly colored with 2 colors if and only if it has no odd cycles.

Informal proof.

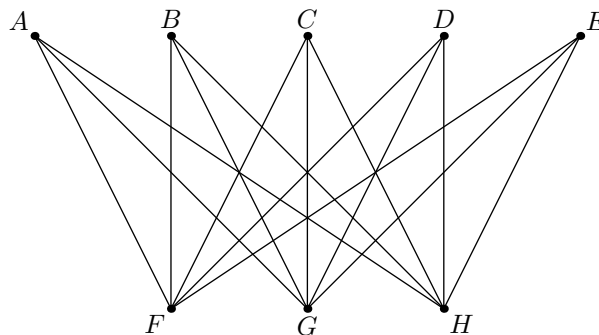
Let G be the graph. An odd cycle means a cycle of odd length. From our chromatic numbers above, an odd cycle requires 3 colors. So if we have an odd cycle, we're using more than 2 colors. This forces G to have no odd cycles.

So consider an arbitrary graph G with no odd cycles. Will it always be 2-colorable?

The intuitive way to think about this is: choose a random vertex v and color it red. Measure the distance from a vertex v to w to be the minimum number of edges necessary to get from v to w . Suppose we color every vertex of an even distance from v red, and every vertex of an odd distance from v blue.

The only problem is if two vertices of even distances from v are adjacent, or vertices of odd distances from v are adjacent. But it turns out either of those two things would require an odd cycle (try playing around with it yourself to convince yourself), which is a no-no, so we're good.

Definition 7.41 (Complete bipartite graph). The complete bipartite graph $K_{m,n}$ has two sets of vertices: V_1 and V_2 . Basically, you make $|V_1| = m$ and $|V_2| = n$. Then, for every vertex v in V_1 , you draw an edge from v to every vertex in V_2 . You do NOT draw any edges between vertices both in V_1 , or edges between vertices both in V_2 . Below is $K_{5,3}$.



§7.11 Cliques and Independent Sets

Definition 7.42 (Clique). A subset S of the vertices of a graph G , where any two vertices in S are neighbors.

Definition 7.43 (Independent Set). Pretty much the opposite of a clique. A subset S of the vertices of a graph G , where any two vertices in S are NOT neighbors.

Remember \overline{G} , the complement of G ? Well, any clique in G becomes an independent set in \overline{G} . Why? Because all the edges become non-edges in transforming G to \overline{G} . So the set of pairwise neighbors (clique) becomes a set of pairwise non-neighbors (independent set).

Advice 7.44 — Sometimes cliques and independent sets show up in coloring problems. Remember, in a clique, all vertices must be of different color because the vertices are pairwise neighbors. In an independent set, you can all the vertices be of the same color, since no two are neighbors.

§7.12 Directed Graphs

Definition 7.45 (Digraph). A graph where there are direction on the edges. An edge between u and v where the direction is from u to v is noted by $u \rightarrow v$. We say the edges in a digraph are directed edges.

Some important changes that digraphs allow that non-directed graphs don't allow:

- You can have self-loops, where an edge exits and enters the same vertex ($u \rightarrow u$)
- You can have two edges between the same two vertices in opposite directions ($u \rightarrow v$ and $v \rightarrow u$ simultaneously)

However, the exact same directed edge cannot be there more than once, as in, you can't have $u \rightarrow v$ twice.

Definition 7.46 (Outdegree). As noted by $\text{out}(u)$, the outdegree of a vertex u is the number of directed edges coming out of u

Definition 7.47 (Indegree). As noted by $\text{in}(u)$, the indegree of a vertex u is the number of directed edges coming in to u

Definition 7.48 (Sink). A vertex with outdegree 0. Meaning all edges connected to it point in to it.

Definition 7.49 (Source). A vertex with indegree 0. Meaning all edges connected to it point out of it.

The degree of u in a directed graph is $\text{out}(u) + \text{in}(u)$.

Definition 7.50 (Directed walk). Basically the same as a walk in a non-directed graph, but the walk has to obey the directions on the edges

Definition 7.51 (Directed path). Basically the same as a path in a non-directed graph, but the path has to obey the directions on the edges

Remember how we can always turn a walk into a path in a non-directed graph? Same idea applies here: if there is unnecessary goofing off in the walk, remove it. Keep doing so until you have a path.

Definition 7.52 (Directed cycle). Basically the same as a cycle in a non-directed graph: a directed walk with no repeating vertices except the first and last ones. Self-loops count as directed cycles. Two edges between the same two vertices in opposite directions (anti-parallel edges) also count as directed cycles.

Advice 7.53 — While adding direction on the edges may seem things got a lot more complicated than they need to be, some problems on digraphs use the same techniques for non-directed graph. Still try proof by contradiction whenever you are stuck. Maximal path still works. Looking at cycles might not be as useful anymore though. Looking at sources and sinks are useful because of their restrictive properties (no edges coming in, and no edges coming out, respectively).

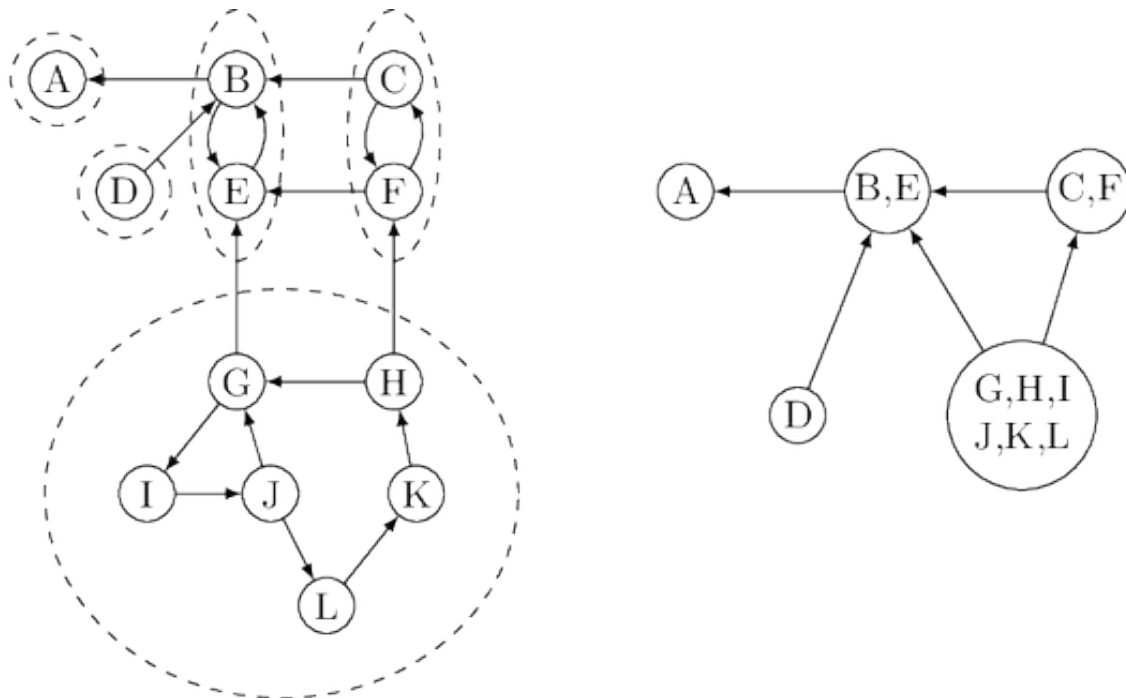
§7.13 Reachability and Strong Connectivity

Definition 7.54 (Reachability). We say v is reachable from u if there exists a directed walk/path from u to v . We note reachability from u to v as $u \rightarrow v$.

Definition 7.55 (Strong connectivity). An extension of reachability is strong connectivity. We say u and v are strongly connected when u and v are both reachable from each other.

Definition 7.56 (Strongly connected component). Similar to a connected component in a non-directed graph. Any two vertices in a strongly connected component are strongly connected, and the component is maximal (meaning you can't add another vertex to the component so that all vertices in the component are pairwise strongly connected).

One easy way of sniffing out the connected components is through a reduced graph:



In a reduced graph, what happens is, vertices in the same strongly connected component get “melded” into a vertex of its own.

Note that a reduced graph will never have directed cycles. Because in a directed cycle, any two vertices in it are strongly connected. So the vertices in the cycle should really be fused together.

Also, we don't have to get the reduced graph on our first try. Say like we noticed G , I , and J were pairwise strongly connected, but we mistakenly thought was maximal. But it makes it easier to realize (G, I, J) (collectively as one vertex), L , K , and H are part of a cycle with less noise in the picture by doing SOME reduction with G , I , and J . Then, we can add H , K , and L to the connected component. We can keep adding vertices we find should be melded together until the resulting graph has no cycles. Then, we know we have found the reduced graph.

§7.14 Directed Acyclic Graphs (DAGs)

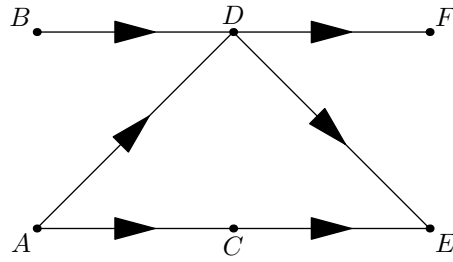
A DAG is a directed graph without any cycles (no self-loops).

Advice 7.57 — You can use the maximal path technique here again to show any DAG must have a source and a sink. In fact, always consider maximal path regarding sources and sink (like leaves in a non-directed acyclic graph, aka trees).

Definition 7.58 (Topological Sort). A topological sort of the vertices of a DAG G goes like this: it is a sequence of the vertices of G such that, whenever $u \rightarrow v$ happens, u will be somewhere before v in the sequence

Topological sorts do not exist in the context of digraphs with cycles (trying to order the vertices in the cycle topologically easily shows this is impossible).

Anyway, a topological sort is best shown with an example.



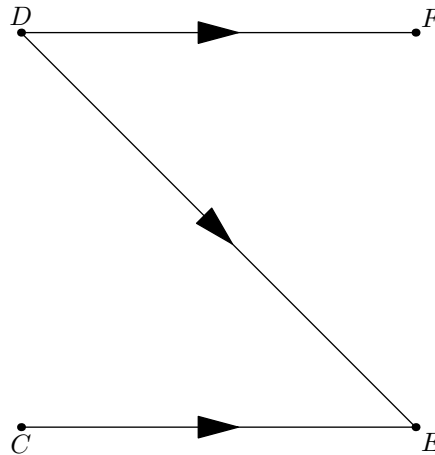
One example of a topological sort that works is A, B, D, C, E, F . You can test it out for yourself. Whenever $u \rightarrow v$ happens, u is before v in the sort. For example, $A \rightarrow D$, which forces A to appear before D .

But perhaps there is a way to actually get a topological sort instead of trial and error?

Well, let's consider the first thing we could add to the sort. We can't just add any vertex first can we? I mean what happens if we add D for example first? We know from the graph that $B \rightarrow D$, which means B needs to be before D in the topological sort. But that's impossible if D is first. So the reason D failed was because it had an arrow pointing in to it.

That lesson means, perhaps we should try something that doesn't have anything pointing in to it first, which would be a source.

So maybe let's let B go first. We might as well delete B and any edges attached to B , as it is no longer necessary. Well, A is another source in the graph now. So let's add that second. Delete A and all its edges. Here's the graph now:



Wait a minute. C and D became sources! Thanks to us removing the sources beforehand, it opened up new sources. There's no issue in adding them to the topological sort now. The sort right now is B, A, C, D . So now C and D are deleted. Then, E and F are sources and can be added. So the sort is B, A, C, D, E, F .

Essentially, the algorithm is always be adding sources to the sort. Then, delete the sources and any edges on them. This opens up new sources. Add those new sources to the sort. Rinse and repeat.

§7.15 Rooted Tree

Take any tree graph T without directions on the edges. Then, choose any vertex and call it the "root" r . Then, we can define some terminology related to the rooted tree:

- The level of a vertex v from the root is the number of edges needed to get from r to v .
- Suppose we have vertices u and v in the tree such that $u - v$. Suppose the vertices are labeled such that vertex u has level ℓ and vertex v has level $\ell + 1$. Then, u is the parent of v , while v is the child of u . Basically, it is like a family tree starting from the root and branching out into children as you get further from the root.
- The height of the rooted tree is the largest possible level in the tree across all vertices.

- Don't cram the day of. Problem solving intuition is something that is built over a period of time and is seen through solving and reflecting on problems.
- Try sitting down and do the practice problems they give you in a timed format and back to back, with write ups being handwritten. That way, you're not as shocked with having to do back to back problems.
- To be more prepared on what will appear, it often helps to think as if you're a TA constructing the test. Look at the possible topics and think "If I were the TA, what would I want to test the students on?"

§8.2 Midterm 1

When I took the course, the semester begun on January 11, while Midterm 1 was February 16. So it takes place a little over a month into the course.

Test statistics (out of 120 points):

- Mean: 74.75
- Standard Deviation: 20.62
- Median: 76
- Max: 113

Tips:

- Expect a combinatorial proof problem.
- The AR and MR rule will show up. Know when to use them and remember their associated buzzwords.
- Know your counting techniques (stars and bars for example) and the different ways you constructively count stuff.
- Expect a problem where you have to prove some set is a subset of another.
- There may be a problem or two at the end that is very hit or miss. For these, just try whatever feels relevant and the solution may emerge.

§8.3 Midterm 2

Midterm 2 was March 30. So it takes place two and a half months into the course.

Test statistics (out of 120 points):

- Mean: 90.71
- Standard Deviation: 18.81
- Median: 95
- Max: 120

Tips:

- Specify uniformity and independence whenever it is necessary.
- Expect multi-part questions.
- Expect a problem involving Pigeonhole Principle.

- Be able to use conditional probability and the chain rule.
- Remember how to prove if two random variables are independent.
- Expect an induction problem. Remember all the things they're looking for in an induction proof.

Note that Val said he made this midterm easier after midterm 1. So this might not be the case when you take midterm 2. But it shouldn't be a shocker that scores are higher in this midterm. People now know what to expect on the midterm and aren't easily losing points over forgotten buzzwords. They also are no longer surprised by the fast pace of the midterm and the time pressure.

§8.4 Final

Test statistics (out of 240 points):

- Mean: 191.07
- Standard Deviation: 30.03
- Median: 195
- Max: 240

Tips:

- Graph theory will be heavy on this exam.
- Luckily, there will be some problems that don't have you rigorously justifying something. These will usually be like "give an example of a graph that satisfies this conditions."
- Know your graph theory vocabulary. That will help with the bullet point above when they make you apply the definition of a vocabulary term.
- Don't forget your probability stuff. Expect a couple questions on probability.
- ALWAYS remember to try proof by contradiction for the proof based graph theory problems if you get stuck.
- Stuff from the first third of the course didn't really show up in the final in Spring 2023, but that could be different for a final in another iteration of the course.
- Again, if you breeze through the first few questions, don't let that be an excuse to tackle the rest of the problems at an excessively slow pace. Time can creep up on you before you know it.
- Good thing to be prepared for a PHP or LoE problem.