CSI 2350 Discrete Mathematics

Dr. Greg Speegle

Fall 2018
Section 1

Binary, Octal and Hex
Overview

» Needs to be added for Spring 2019
Section 2

Propositional Logic
Propositional Logic

- Needs to be added for Spring 2019
Section 3

Knights and Knave
Knights and Knaves

- Needs to be added for Spring 2019
Section 4

CNF, DNF and Proofs
CNF, DNF and Proofs

- Needs to be added for Spring 2019
Section 5

Set Theory
Sets

- Needs to be added for Spring 2019
Section 6

Functions and Sequences
Functions and Sequences

- Needs to be added for Spring 2019
Section 7

Summations and Matrices
Summations

- Add the terms in a sequence
- Uses $\Sigma$
- Examples:
  - $\sum_{j=1}^{n} j = \frac{n(n + 1)}{2}$
  - $\sum_{j=1}^{n} j^2 = \frac{n(n + 1)(2n + 1)}{6}$
  - $\sum_{j=1}^{n} 2j + 3 = 2\sum_{j=1}^{n} j + \sum_{j=1}^{n} 3 = 2\cdot\frac{n(n + 1)}{2} + 3n = n^2 + 4n$
  - $\sum_{j=0}^{n} ar^j = \frac{ar^{n+1} - a}{r - 1}[r \neq 1], (n + 1)a[r = 1]$
  - Let $S = \{1, 2, 4, 8\}$ then $\sum_{x \in S} x = 1 + 2 + 4 + 8 = 15$
Double Summations

- Summation over two variables
- Inner and outer loop
- Examples:
  - $\sum_{i=1}^{4} \sum_{j=1}^{3} i \times j = 1 + 2 + 3 + 2 + 4 + 6 + 3 + 6 + 9 + 4 + 8 + 12 = 60$
  - $\sum_{i=1}^{4} \sum_{j=1}^{3} i \times j = \sum_{i=1}^{4} (i + 2i + 3i) = \sum_{i=1}^{4} 6i = 6 \times \sum_{i=1}^{4} i = 6 \times 10 = 60$
Countable and Uncountable Sets

- Sets $A$ and $B$ have the same cardinality iff there is a 1-1 correspondence (1-1 and onto function) from $A$ to $B$
- Finite sets are countable
- Infinite sets with 1-1 correspondence to integers are countable
- Set of real numbers is uncountable
- Fun fact: Some problems cannot be solved by a computer. Halting problem (does program $P$ halt?).
Matrices

- Express relationships between elements of sets
- Example - Powers of numbers (Matrix M)

\[ M = \begin{bmatrix}
1 & 1 & 1 & 1 \\
2 & 1 & 2 & 4 \\
3 & 1 & 3 & 9 \\
\end{bmatrix} \]

- \( m \) rows and \( n \) cols is \( m \times n \) matrix (\( M \) is 3x4).
- \( m=n \) is square matrix
- Element \( a_{i,j} \) where \( i \) is row number and \( j \) is column number
- \( M_{2,2} = 1, M_{3,3} = 3 \)
Addition

- Requires $M, N$ be same size
- $M + N = L \rightarrow \forall i, j L_{i,j} = M_{i,j} + N_{i,j}$

\[
\begin{bmatrix}
1 & 1 & 1 & 1 \\
2 & 1 & 2 & 4 \\
3 & 1 & 3 & 9
\end{bmatrix}
+ 
\begin{bmatrix}
1 & 2 & 3 & 4 \\
5 & 6 & 7 & 8 \\
9 & 10 & 11 & 12
\end{bmatrix}
= 
\begin{bmatrix}
2 & 3 & 4 & 5 \\
7 & 7 & 9 & 12 \\
12 & 11 & 14 & 21
\end{bmatrix}
\]
Multiplication

- Columns in $M$ must equal rows in $N$
- $M$ is $m \times k$, $N$ is $k \times n$, $L$ is $m \times N$
- $L_{i,j} = \sum_{h=1}^{k} M_{i,h} \times N_{h,j}$

\[
\begin{bmatrix}
1 & 1 & 1 & 1 \\
2 & 1 & 2 & 4 \\
3 & 1 & 3 & 9
\end{bmatrix}
\times
\begin{bmatrix}
1 & 2 \\
5 & 6 \\
9 & 10 \\
11 & 12
\end{bmatrix}
= 
\begin{bmatrix}
26 & 30 \\
69 & 78 \\
134 & 150
\end{bmatrix}
\]

- Example:
  \[L_{2,2} = \sum_{h=1}^{4} M_{2,h} \times N_{h,2} = 2 \times 2 + 1 \times 6 + 2 \times 10 + 4 \times 12 = 78\]
- Example:
  \[L_{1,2} = \sum_{h=1}^{4} M_{1,h} \times N_{h,2} = 1 \times 2 + 1 \times 6 + 1 \times 10 + 1 \times 12 = 30\]
- NOT commutative (e.g., $M \times N \neq N \times M$, see above)
Identity Matrix and Transposition

- **Identity matrix** (usually $I$)
  - Square
  - Diagonal values are 1; all other values are 0
  - $M \times I = M$

- **Transpose of $M = M^t$**
  - $M^t_{i,j} = M_{j,i}$
  - $M$ is symmetric if $M = M^t$ (note: must be square)

\[
M = \begin{bmatrix}
1 & 1 & 1 & 1 \\
2 & 1 & 2 & 4 \\
3 & 1 & 3 & 9 \\
\end{bmatrix} \quad \rightarrow \quad M^t = \begin{bmatrix}
1 & 2 & 3 \\
1 & 1 & 1 \\
1 & 2 & 3 \\
1 & 4 & 9 \\
\end{bmatrix}
\]
Section 8

The Big $\Omega\Theta$ Fraternity
Binary Search

- Input list of elements \([a_1, a_2, \ldots, a_n]\), search key \(x\)
- Output index of element matching \(x\), or 0 if not found
- Pseudocode
  - \(i \leftarrow 1\)
  - \(j \leftarrow n\)
  - while \((i < j)\)
    - \(m \leftarrow \lfloor (i + j)/2 \rfloor\)
    - if \(x > a_m\) then \(i \leftarrow m + 1\)
    - else \(j \leftarrow m\)
  - if \(x = a_i\) then return \(i\)
  - else return 0
Binary Search Runtime

- Assume \( n = 2^k \) elements in list
- After iteration, remaining list half of previous list size
- After \( k \) iterations, list is size 1
- \( \log(n) = k \)
- Worst case running time is \( \log(n) \)
Let $f$ and $g$ be functions. $f(x)$ is $O(g(x))$ if $\exists C, k$ such that $f(x) \leq C * g(x)$ for all $n > k$. Assume $x \geq 0$.

Function $f$ grows slower than function $g$ for $n > k$.

NOTE: Can say ”is” or $\in$ but $=$ is misleading (although used frequently).

NOTE: Existence proof. Find one $C$ and one $k$ (many may exist).

Technique: Start with known ($x > k$). Remember, we can select $k$.

Apply formulas to generate $C * g(x) \geq f(x)$.

Examples:

- $f(x) = 17x + 11 \in O(g(x) = x^2)$. Let $C = 2, k = 17$.
  - $x \geq 17 \rightarrow x^2 \geq 17x$
  - $\rightarrow x^2 + x^2 \geq 17x + 11$
  - $\rightarrow 2 * x^2 \geq 17x + 11$
Big O

- Examples:
  - $f(x) = x \log(x) \in O(g(x) = x^2)$. Let $C = 1$, $k = 1$
    - $x \geq 1 \rightarrow x \geq \log(x) \rightarrow x^2 \geq x \log(x)$
  - $f(x) = x^2 + x + 1 \in O(g(x) = x^2)$ Let $C = 3$, $k = 1$
    - $x \geq 1 \rightarrow x^2 \geq x \rightarrow x^2 + x^2 \geq x^2 + x$
    - $\rightarrow x^2 + x^2 + x^2 \geq x^2 + x + 1 \rightarrow 3 \cdot x^2 \geq x^2 + x + 1$

- To show $f(x)$ is not $O(g(x))$ must show no such $C$ and $k$ can exist.

- Example:
  - $f(x) = x^2$ is not $O(g(x) = x)$.
    - Assume not. Therefore $\exists C, k$ such that $f(x) \leq C \cdot g(x) \forall x > k$
    - Therefore, $x^2 \leq C \cdot x \forall x \geq k$
    - Therefore, $x \leq C \forall x \geq k$, which is a contradiction.
Big $\Omega$

- Let $f$ and $g$ be functions. $f(x)$ is $\Omega(g(x))$ if $\exists C, k$ such that $f(x) \geq C \times g(x)$ for all $n > k$. Assume $x \geq 0$.
- Function $f$ grows faster than function $g$ for $n > k$.
- $f(x)$ in $\Omega(g(x)) \iff g(x) \in O(f(x))$.
- Example:
  - $f(x) = x^2$ is $\Omega(g(x) = x)$ because $g(x) = x$ is $O(f(x) = x^2)$.
  - $f(x) = x^4/2$ is $\Omega(g(x) = x^2)$. Show $g(x) = x^2$ is $O(f(x) = x^4/2)$.
  - Let $C = 1, k = 2$.
  - $x \geq 2 \rightarrow x^2 \geq 2x \rightarrow x^3 \geq 2x^2 \rightarrow x^4 \geq 2x^2 \rightarrow x^4/2 \geq x^2$.
Let \( f \) and \( g \) be functions. if \( f(x) \) is \( \Omega(g(x)) \) and \( f(x) \) is \( O(g(x)) \), then \( f(x) \) is \( \Theta(g(x)) \).

Equivalent: \( f(x) \) is \( O(g(x)) \) and \( g(x) \) is \( O(f(x)) \).

\( f(x) \) is bounded above and below by \( g(x) \).

Example:

\[
f(x) = x^2 + x + 1 \quad \text{is} \quad \Theta(g(x) = x^2)
\]

Show \( f(x) \) is \( O(g(x)) \), previous slide.

Show \( g(x) \) is \( O(f(x)) \). Let \( C = 1, k = 1 \).

\[
x \geq 1 \rightarrow x^2 + x \geq x^2 \rightarrow x^2 + x + 1 \geq x^2.
\]
Intractable Problems

- Show $f(x) = x^2$ is $O(g(x) = 2^x)$ Let $C = 1, k = 4$
- $x \geq 4 \rightarrow x \geq 2 \log(x) \rightarrow x \geq \log(x^2) \rightarrow \log(2^x) \geq \log(x^2) \rightarrow 2^x \geq x^2$
- Show $g(x) = 2^x$ is NOT $O(f(x) = x^2)$. Assume not.
- $\exists C, k$ such that $Cx^2 \geq 2^x, \forall x \geq k$.
- Therefore, $\log(Cx^2) \geq \log(2^x), \forall x \geq k$
- Let $a = \log(C)$. Therefore, $a + 2 \log(x) \geq x, \forall x \geq k$
- Note that $\forall a \exists x$ s.t. $x > a + 2 \log(x)$. Consider $\max(x = 2^a, 16)$. Then $2^a > 3a$, which is true for $a \geq 4$.
- Therefore, problems requiring exponential time are “harder” than quadratic problems (or any polynomial)
- Such problems are called **intractable**
P=NP?

- P - class of problems solvable in polynomial time
- NP - class of problems best solutions require exponential time
- Unknown if NP problems can be solved in polynomial time
- One of grand challenges of mathematics for 21st century (Millennium Problems)
- Example: Satisfiability
  - Predicate $P$ in CNF $(p \lor q \lor r) \land (\neg p \lor s \lor \neg t) \ldots$
  - Can assign truth values to variables such that $P$ can be satisfied?
  - DNF is trivial $(p \land q \land r) \lor (\neg p \land s \land \neg t) \ldots$
Section 9

Induction
Induction

- Proof technique for showing infinite series is true
- Require incremental progress
- Basis – Initial step (usually very easy to show)
- Induction Hypothesis (IH) – Theorem is true for $k$ elements
- NOTE! Have not proved it is true. Just assuming it is.
- Show Theorem is true for $k + 1$ elements
- Therefore, initially true. Basis is now IH for next step. Now have IH for following step, etc.
Induction Example

- Summations “easily” shown by induction (proof simple; algebra can be tricky)
- Prove $\sum_{i=1}^{n} i^3 = (n(n + 1)/2)^2$
- Basis: $P(1). 1^3 = 1 = (1(2)/2)^2$
- IH: $\sum_{i=1}^{k} i^3 = (k(k + 1)/2)^2$
- Let $n = k + 1$.
- $\sum_{i=1}^{k+1} i^3 = \sum_{i=1}^{k} i^3 + (k + 1)^3$
- By IH, $\sum_{i=1}^{k+1} i^3 = (k(k + 1)/2)^2 + (k + 1)^3$
- $= ((k^2 + k)/2)^2 + k^3 + 3k^2 + 3k + 1$
- $= (k^4 + 2k^3 + k^2)/4 + k^3 + 3k^2 + 3k + 1$
- $= (k^4 + 2k^3 + k^2 + 4k^3 + 12k^2 + 12k + 4)/4$
- $= (k^4 + 6k^3 + 13k^2 + 12K + 4)/4$
- Note that $((k + 1)(k + 2)/2)^2 = ((k^2 + 3k + 2)/2)^2 = (k^4 + 3k^3 + 2k^2 + 3k^3 + 9k^2 + 6k + 2k^2 + 6k + 4)/4$
- $= (k^4 + 6k^3 + 13k^2 + 12k + 4)/4$
- So $\sum_{i=1}^{k+1} i^3 = ((k + 1)(k + 2)/2)^2$
Induction Example

- Prove \( x \geq 4 \rightarrow x \geq 2 \log(x) \)
- Basis: P(4). \( 4 \geq 2 \log(4) = 4 \)
- IH: \( x = k \rightarrow x \geq 2 \log(x) \)
- Let \( x = k + 1 \).
- By IH, \( x \geq 2 \log(k) + 1 \)
- \( x \geq \log(k^2) + \log(2) \)
- \( x \geq \log(2k^2) \)
- \( x > \log(k^2 + 2k + 1) \)
- \( x > \log((k + 1)^2) = 2 \log(k + 1) = 2 \log(x) \)
Tag! You’re Out!

Need four volunteers.
Odd Man Pie Fights (from text)
- People stand in yard at distinct distances
- Throw pie at nearest neighbor
- Odd number of participants ensures at least 1 person is not hit

Technique - Induct over $n$ for $P(2n + 1)$

Basis: $n = 1, P(3)$. Let $(a, b)$ be closest pair. Then $a$ hits $b$ and $b$ hits $a$. $c$ hits whoever is closest to $c$, but nobody hits $c$
Odd Man Out Proof (cont’d)

- IH: Assume theorem is true for \( n = k, P(2 \times k + 1) \).
- Let \( n = k + 1 \). Therefore, want to show \( P(2 \times (k + 1) + 1) \) or \( P(2 \times k + 3) \).
- Let \((a, b)\) be the closest pair (all distances unique implies least element exists).
- Therefore, \( a \) hits \( b \) and \( b \) hits \( a \).
- Case I:
  - Someone else throws a pie at \( a \) or \( b \)
  - Therefore, at least 3 pies thrown at \( a \) and \( b \), leaving at most \( 2k \) pies for \( 2k + 1 \) people.
- Case II:
  - Nobody throws a pie at \( a \) or \( b \)
  - Now \( 2 \times (k + 1) \) people remain with pies.
  - By IH, at least one is not hit.
Induction Example

- Number of elements in $2^S$ (from text)
- Show $|2^S| = 2^{|S|}$
- Basis: $\emptyset$. The only subset of $\emptyset$ is itself. $|2^S| = 1 = 2^0$
- IH: $|S| = k \rightarrow |2^S| = 2^k$
- Let $|T| = k + 1$. $T = S \cup \{a\}$ such that $S = T - \{a\}$.
- Let $X \subseteq S$. Therefore, $X \subseteq T$ and $X \cup \{a\} \subseteq T$.
- By IH. there are $2^k$ such subsets of $S$, and $2 \times 2^k = 2^{k+1}$ subsets of $T$. 
Section 10

Complete Induction & Recursion
Strong (Complete) Induction

- IH is not just that $P(k)$ is true, but $P(1) \land P(2) \land P(3) \ldots P(k)$ is true
- Equivalent to Induction, but sometimes easier to use
- Example: Prove that every positive integer $n$ can written as a sum of distinct powers of 2
  - Basis: $n = 1$. $2^0 = 1 = n$
  - IH: $\forall n \leq k$, $n$ can be written as a sum of distinct powers of 2
  - Let $n = k + 1$
  - Case I:
    - Let $n$ be odd.
    - Since $n$ is odd, $k$ must be even.
    - By IH, let $S$ be representation of the sum for $k$
    - Therefore, $2^0$ cannot be in $S$
    - Therefore $n = (\text{representation for } k) + 2^0$. 
Proof Continued

- Case II
  - Let $n$ be even.
  - Therefore, $n/2$ is an integer such that $n/2 \leq k$.
  - By IH, let $S$ be representation of the sum for $n/2$
  - Multiplying $S$ by 2 increases each exponent by 1.
  - Therefore, $S$ with each exponent increased by 1 is $n$
Function Recursion

- Compute values of a function based on previous values in the function
- Specify the value at 0 (or first $k$ values)
- Provide rule for $f(n)$ based on lower values of $n$
- Factorial: $f(0) = 1. \forall n, n > 0 \rightarrow f(n) = n \cdot f(n - 1)$.
- Fibonacci:
  $f(0) = 0, f(1) = 1. \forall n, n > 1 \rightarrow f(n) = f(n - 1) + f(n - 2)$
- Paradox:
  $f(0) = 0.5. \forall n, n > 1 \rightarrow f(n) = (1 - f(n - 1))/2 + f(n - 1)$
- Min: $f(1) = a_1. \forall n, n > 1 \rightarrow f(n) = \min(a_n, f(n - 1))$
Set Recursion

- Compute elements in a set based on previous elements in the set
- Provide rule for initial elements. Provide rule for adding new elements.
- Initial - \(1 \in S\). Step - \(s, t \in S \rightarrow s + t \in S\) Prove \(S = \mathbb{Z}^+\) (Note! \(s\) and \(t\) do not have to be unique!):
  - Clearly, \(S \subseteq \mathbb{Z}^+\)
  - Basis: \(1 \in S\) by definition of \(S\)
  - IH: \(k \in S\).
  - Since \(k\) and \(1\) are in \(S\), by rule, \(k + 1 \in S\)
- Transitive Closure: Let \(V = \{v_0 \ldots v_{n-1}\}\) and \(E \subset V \times V\). Initial - \((v_i, v_j) \in E \rightarrow (v_i, v_j) \in S\). Step - \((v_i, v_j), (v_j, v_k) \in S \rightarrow (v_i, v_k) \in S\)
- DAG on the board
Recursive Algorithm

- Compute results using previously computed results
- Provide rule for terminal case. Provide rule for recursive call.
- Example: factorial
  - fact(n)
    - if $n \leq 0$ return 1
    - else return $n \times \text{fact}(n - 1)$
- Example: gcd
  - gcd(a,b)
    - if a=0 return b
    - return gcd(b%a,a)
Recursive Algorithm

- Example: fibonacci (Problem?)
  - fib(a)
  - if a=0 return 0
  - if a=1 return 1
  - return fib(a-1)+fib(a-2)

- Example: Binary Search (Problem?)
  - Let $A = [a_1, a_2, \ldots, a_n]$
  - find($A$, $x$, $i$, $j$)
  - if $i \geq j$, then
    - if $a_i = x$ return $i$
    - else return 0
  - $m \leftarrow \lfloor (i + j)/2 \rfloor$
  - if $x > a_m$, then return find($A$, $x$, $m + 1$, $j$)
  - else return find($A$, $x$, $i$, $m$)
Loop Invariants

- Only part of Chapter 5 we’ll cover
- Proposition true for every iteration of loop
- Example:
  - Loop invariant: \( f = i! \)
  - \( i \leftarrow 1 \)
  - \( f \leftarrow 1 \)
  - while \((i < n)\)
    - \( i \leftarrow i + 1 \)
    - \( f \leftarrow f \times i \)
- At end of loop, loop invariant and \( \neg(i < n) \rightarrow f = n! \)
Section 11

Basic Counting
Overview

- Needs to be added for Spring 2019
Section 12

Permutations and Combinations
Overview

- Permutations are the number of arrangements of a set of items of a specific size
- Combinations are the number of subsets of a specific size
- The Binomial Theorem is a general representation of binomial coefficients
- Playing cards
  - 4 suits - spades, hearts, diamonds, clubs
  - 13 values - A,K,Q,J,10,9,8,7,6,5,4,3,2
  - 52 total cards (+ jokers)
Permutations

- Battle - Card game where highest value wins (special rules for ties)
- Demonstration (Hearts only)
  - How many ways to arrange 6 cards? (Hint: Product rule.)
  - How many ways to arrange 3 of 6 cards?
  - How many ways to arrange k of 6 cards?
  - How many ways to arrange k of n cards?
Permutation Formulae

- All arrangements: \( n(n - 1)(n - 2) \ldots (1) = n! \)
- Arrangements of size \( r \): \( n(n - 1)(n - 2) \ldots n - (r - 1) \)
  - Last term is also \( n - r + 1 \)
  - \( (n - r)! = (n - r)(n - r - 1)(n - r - 2) \ldots (1) \)
  - \( \frac{n!}{(n-r)!} \) yields arrangements of size \( r \)
- Formula: \( P(n, r) = \frac{n!}{(n-r)!} \)
- Note: if \( n == r \) then \( P(n, n) \) or \( P(n) = \frac{n!}{(n-n)!} = \frac{n!}{0!} = n! \)
Combinations

- Oh Heck - Card game with hands and “tricks”
- Demonstration (Hearts only)
  - Order of cards in hand does not matter
  - How many different hands of size 3 can be dealt?
  - Given my hand, how many different hands of size 3 can opponent have?
  - Given my hand, how many different hands of size 3 can opponent have with all cards lower than my highest?
Combination Formulae

- Number of permutations divided by the number that are the same (division rule)

\[ P(n, r) / P(r) = \frac{n!}{(n-r)!} \cdot \frac{r!}{r!(n-r)!} = \frac{n!}{r!(n-r)!} \]

- Number of subsets of set S a given size
  - Size 0 is ∅, only one \( \frac{n!}{0!(n-0)!} = 1 \)
  - Size \( n \) is S, only one \( \frac{n!}{n!(n-n)!} = 1 \)
  - Size 1 subsets, each element is \( |S| \), \( \frac{n!}{1!(n-1)!} = n \)
  - Size \( n - 1 \) subsets, S with each element removed, number is \( |S|, \frac{n!}{(n-1)!(n-(n-1))!} = n \)

- Formula: \( C(n, r) \) or \( \binom{n}{r} = \frac{n!}{r!(n-r)!} \)

- Note: \( \binom{n}{n} = \frac{n!}{n!0!} = 1 \) and \( \binom{n}{0} = \frac{n!}{n!0!} = 1 \)
Binomial Theorem

\( (x + y)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k = \binom{n}{0} x^n + \binom{n}{1} x^{n-1} y + \binom{n}{2} x^{n-2} y^2 + \ldots + \binom{n}{n-1} x y^{n-1} + \binom{n}{n} y^n \)

- Can find coefficient for any term in binomial expansion
  - Given \((2x + 3y)^4\), what is the coefficient for the \(x^2y^2\) term?
  - \( \binom{4}{2}(2x)^2(3y)^2 = 6 \times 4 \times 9 = 216 \)

- Shows \( \sum_{k=0}^{n} \binom{n}{k} = 2^n \) (from text)

- \( 2^n = (1 + 1)^n = \sum_{k=0}^{n} \binom{n}{k} 1^{n-k} 1^k = \sum_{k=0}^{n} \binom{n}{k} \)
Pascal’s Triangle

- Example on board
- Pascal’s Identity: $\binom{n+1}{k} = \binom{n}{k-1} + \binom{n}{k}$
- Algebraic proof:
  - $\binom{n+1}{k} = \frac{(n+1)!}{k!(n+1-k)!}$
  - $\binom{n}{k-1} = \frac{n!}{(k-1)!(n-(k-1))!}$
  - $\frac{k*\binom{n}{k-1}}{k*\binom{n}{k}} = \frac{\binom{n}{k}*(n+1-k)}{k*(n+1-k)!}$
  - $\binom{n}{k-1} + \binom{n}{k} = \frac{k*n!}{k*(n+1-k)!} = \frac{n!(k+n-k+1)}{k!(n+1-k)!} = \frac{(n+1)!}{k!(n+1-k)!}$
Section 13

Probability
Finite Probability

- Let \( S \) be a set of equally likely outcomes – sample space.
- Let \( E \subseteq S \) be a set of desired outcomes – event.
- \( p(E) = \frac{|E|}{|S|} \) – probability of \( E \)

Examples (sample space is deck of cards):
- Probability of drawing a heart \( p(\text{heart}) = \frac{13}{52} = \frac{1}{4} \)
- Probability of drawing a king \( p(\text{king}) = \frac{4}{52} = \frac{1}{13} \)
- Probability of drawing king of hearts \( p(K\text{heart}) = \frac{1}{52} \)
- Probability of drawing two hearts in a row (replacing your card) \( p(2\text{hearts}) = \frac{169}{2704} \)
  - Why? Two draws makes total outcomes is \( 52 \times 52 \)
  - 13 successes in each draw means 169 successful outcomes (product rule)
- Probability of drawing 5 hearts in a row (without replacement) \( p(\text{flush}) = \frac{13 \times 12 \times 11 \times 10 \times 9}{52 \times 51 \times 50 \times 49 \times 48} \)
The complement of $E$ is $\bar{E}$ is $S - E$. $p(\bar{E} = 1 - p(E)$

Probability of not drawing a heart $p(\text{heart} = 1 - 1/4 = 3/4$

Sometimes much easier to calculate the complement

The union of two events is $E_1 \cup E_2$.

$p(E_1 \cup E_2) = p(E_1) + p(E_2) - p(E_1 \cap E_2)$

$p(\text{king or heart}) = p(\text{king}) + p(\text{heart}) - p(K\text{heart}) = 1/13 + 1/4 - 1/52 = 16/52$
Probability Theory

- Let $S$ be a sample space
- Each element in $S$ is assigned a probability ($p(s)$).
- $0 \leq p(s) \leq 1$
- $\sum_{s \in S} p(s) = 1$
- Function from $S$ to set of probabilities is called probability distribution function
- If all probabilities are the same, uniform distribution
  $|S| = n, \forall s \in S, p(s) = 1/n$
- $p(E) = \sum_{s \in E} p(s)$
Conditional Probability

- Probability of event $E$ given that event $F$ has happened
  $p(E|F)$

  $p(E|F) = \frac{p(E \cap F)}{p(F)}$

- Example:
  - Given that 4 hearts in a row have been drawn, what is the probability that a 5th heart will be drawn?
  - $p(F) = p(4\text{hearts}) = \frac{13 \times 12 \times 11 \times 10}{52 \times 51 \times 50 \times 49}$
  - In this case, $p(E \cap F) = p(E) = p(\text{flush}) = \frac{13 \times 12 \times 11 \times 10 \times 9}{52 \times 51 \times 50 \times 49 \times 48}$
  - $p(E|F) = \frac{9}{48}$
Independence

- Two events are independent if one happening has no effect on the other happening
- Examples:
  - I draw a 7 from a deck of cards and you draw a 10 from a different deck.
  - I wear a hat and President Livingstone wears a hat.
  - We have an exam in 2350 and Dr. Donahoo gives an exam in 4321.
- $E$ and $F$ are independent iff $p(E) \times p(F) = p(E \cap F)$
- Different decks:
  - $p(7) = 1/13$, $p(10) = 1/13$, $p(7 \land 10) = 1/169 \approx .0060$
- Same decks:
  - $p(7) = 1/13$, $p(10) = 1/13$, $p(7 \land 10) = 1/13 \times 4/51 = 4/663 \approx .0059$
Bernoulli Trials

- Probability of $k$ successes of $n$ independent trials with success $p$ and failure $q = 1 - p$ is $\binom{n}{k} p^k q^{n-k}$

- Examples (replacing cards back in deck)
  - Probability of drawing 4 hearts out of 5 cards
    - $n = 5$, $k = 4$, $p = 0.25$, $q = 0.75$, $p(4H) = \binom{5}{4} \cdot 0.25^4 \cdot 0.75 \approx 0.0146$
  - Probability of drawing at least 4 hearts out of 5 cards
    - $p(4H) + p(5H) = \binom{5}{4} \cdot 0.25^4 \cdot 0.75 + \binom{5}{5} \cdot 0.25^5 \cdot 0.75^0 = 0.015625$
  - Probability of drawing at least 2 hearts out of 5 cards
    - $1 - (p(0H) + p(1H)) = 1 - \binom{5}{0} \cdot 0.25^0 \cdot 0.75^5 + \binom{5}{1} \cdot 0.25^1 \cdot 0.75^4 \approx 0.367$
Random Variable

- Not random and not a variable
- Function from $S$ to $\mathbb{R}$ (assigns real number to each possible outcome)
- Example: (Replacing cards as before)
  - Draw 3 cards from a deck. The set of all possible outcomes is $S$.
  - Let $X(s)$ be the random variable of the number of times a heart is drawn, where $s \in S$
  - Quick demo
  - $\forall s \in S, 0 \leq X(s) \leq 3$
- Distribution of $X$ on $S$ is the set of pairs $(r, p(X = r))$ where $r \in X(s)$
- From example:
  $(0, 0.421875), (1, 0.421875), (2, 0.140625), (3, 0.015625)$
Section 14

Expected Value and Variance
Expected Value

- Given a random variable $X$, the *expected value* of $X$ is 
  
  $E(X) = \sum_{s \in \mathcal{S}} p(s)X(s)$

- Recall $X(s) = y$ means $y$ is the number of interesting occurrences in event $s$

- Examples:
  - Assume cards 2-10 of hearts. Let $X$ be the value of the card. Expected value of drawing a card:
    
    $\frac{1}{9} \times 2 + \frac{1}{9} \times 3 + \frac{1}{9} \times 4 + \frac{1}{9} \times 5 + \frac{1}{9} \times 6 + \frac{1}{9} \times 7 + \frac{1}{9} \times 8 + \frac{1}{9} \times 9 + \frac{1}{9} \times 10 = 6$
  
  - Assume J,Q,K have value 10 and A has value 11. Expected value of a card:
    
    $E(X) = \frac{1}{13} \times (\sum_{k=2}^{9} k + 11) + \frac{4}{13} \times 10 = \frac{95}{13}$
  
  - Assume 5 cards, consisting of 4 2s and 1 3.
    
    $E(X) = .8 \times 2 + .2 \times 3 = 2.2$
The expected number of success of $n$ Bernoulli trials with success $p$ is $n \times p$

Proof: (from text)

Let $X(s)$ be the number of successes out of $n$ trials.

\[ p(X = k) = \binom{n}{k} p^k q^{n-k} \]

\[ E(X) = \sum_{k=1}^{n} k \times \binom{n}{k} p^k q^{n-k} \quad \text{Note: } k = 0 \text{ adds } 0 \text{ to } E(X) \]

\[ = \sum_{k=1}^{n} n \binom{n-1}{k-1} p^k q^{n-k} \]

\[ = np \times \sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1} q^{n-k} \]

\[ = np \times \sum_{j=0}^{n-1} \binom{n-1}{j} p^j q^{n-1-j} \quad \text{(by shifting index)} \]

\[ = np \times (p + q)^n - 1 \quad \text{(by binomial theorem)} \]

\[ = np \quad \text{(by } p + q = 1) \]
Linearity

- Expected value of sum of random variables is sum of expected values
  \[ E(X_1 + X_2 + \ldots + X_n) = E(X_1) + E(X_2) + \ldots + E(X_n) \]
- \[ E(aX + b) = a \cdot E(X) + b \]
- Examples:
  - Sum of two cards (with replacement) is \( 14 \frac{8}{13} \)
  - Sum of two die rolls is \( 2 \cdot E(X) \) where
  \[ E(X) = \frac{1}{6} \cdot \sum_{k=1}^{6} k = \frac{21}{6} = \frac{7}{2} \]
Complex Example - Expected number of inversions in a polynomial

- A permutation $P$ of integers $1 \ldots n$ is an arrangement of the numbers
- An inversion is where $i < j$ but $j \prec i$ in $P$
- Example: $P = (1, 3, 5, 2, 4)$ the inversions are $(2, 3), (2, 5), (4, 5)$
- Let $I_{i,j}$ be the random variable on the set of all permutations of the first $n$ integers with $I_{i,j} = 1$ if $(i, j)$ is an inversion on the permutation
- For the example, $I_{2,3}(P) = 1, I_{1,4} = 0$
- Let $X$ be the random variable equal to the number of inversions, $X = \Sigma_{1 \leq i < j \leq n} I_{i,j}$
- For example, $X(P) = 3$
- $E(I_{i,j}) = 1 \times p(I_{i,j} = 1) + 0 \times p(I_{i,j} = 0) = 1/2$ (equally likely inversion as not)
- There are $\binom{n}{2}$ ways for 2 numbers to be arranged out of $n$
- $E(X) = \binom{n}{2} E(I_{i,j}) = \frac{n!}{(n-2)! \times 2^2 \times 2} = \frac{n \times (n-1)}{4}$
Average Case Complexity

- $S$ is the possible inputs to the program
- $X : S \to \mathbb{R}$, such that $\forall s \in S, X(s)$ is the number of operations performed
- Let $p(s)$ be the probability of $s$ being the input to the program
- $\sum_{s \in S} p(s)X(s)$ is the expected (or average) number of operations
Average Complexity Linear Search (text)

- Let $p$ be the probability $x \in A$. Assume $x$ is equally likely to be in any other location.
- Counting number of comparisons
- For each element, check to see if at end of array and compare value (2 comparisons per element)
- After loop, one comparison to see if past end of array
- Probability $x$ is at element $k$ is $p/n$
- Probability $x$ is not in list is $q = 1 - p$
- If $x \in A$, then $\sum_{k=1}^{n} \frac{p}{n} (2k + 1) =$
- $\frac{p}{n} \sum_{k=1}^{n} (2k + 1) = \frac{p}{n} * (n + 2 \sum_{k=1}^{n} k)$
- $= \frac{p}{n} * (n + 2 * \frac{n(n+1)}{2}) = \frac{p}{n} * (n + n(n+1)) = p * (1 + n + 1) = p(n+2)$
- If $x \not\in A$, then $(2n + 2)q$
- $E(X) = p(n + 2) + (2n + 2)q$
Variance

- Let $X$ be a random variable on $S$
- Variance on $X$ (denoted $V(X)$) indicates the spread of values in $X(S)$
- $V(X) = \sum_{s \in S} (X(s) - E(X))^2 p(s)$
- Standard deviation $\sigma(X) = \sqrt{V(X)}$
- Example:
  - Blackjack cards: $V(X) = \sum_{s \in S} (X(s) - 7 \frac{4}{13})^2 p(s) =$
  - $(-5 \frac{4}{13})^2 \times 1/13 + (-4 \frac{4}{13})^2 \times 1/13 + \ldots + (2 \frac{9}{13})^2 \times 4/13 + (3 \frac{9}{13})^2 \times 1/13$
  - $\approx 8.5$
  - $\sigma(X) \approx 2.9$
Variance Continued

- $V(X) = E(X^2) - E(X)^2$
- Example:
  - Blackjack cards: $E(X)^2 = (\frac{74}{13})^2 \approx 53.4$
  - $E(X^2) = 1/13 \times (\Sigma_{k=2}^{9} k^2 + 121) + 4/13 \times 100 = \frac{784}{13} \approx 61.9$
  - $\approx 8.5$
  - $\sigma(X) \approx 2.9$
- Let $E(X) = \mu$. Then $V(X) = E((X - \mu)^2)$
  - $= 1/13 \times (\Sigma_{k=2}^{9} (k-7 \frac{4}{13})^2 + (11-7 \frac{4}{13})^2) + 4/13(10-7 \frac{4}{13})^2 = 8.5$
- Let $X$ be a random variable such that $X(t) = 1$ if a Bernoulli trial is successful and $X(t) = 0$ otherwise.
  - Note: Single trial, so $n = 1$ for Bernoulli distribution.
  - $E(X) = p \times 1 + q \times 0 = p$. $E(X^2) = p \times 1^2 + q \times 0^2 = p$.
  - $E(X)^2 = p^2$. $V(X) = p - p^2 = p(1 - p) = pq$. 
Variance Equations

- Bienayme’s Formula
  - If $X$ and $Y$ are independent random variables on $S$, then $V(X + Y) = V(X) + V(Y)$

- Chebyshev’s Inequality
  - If $X$ is a random variable on $S$ with probability function $p$, then $p(|X(s) - E(X)| \geq r) \leq V(X)/(r^2)$

- Example:
  - Probability draw a card 3 or more from the mean of blackjack cards
    - $V(X)/r^2 \approx 8.5/9 \approx 0.94$
    - Actual is 2, 3, 4, $A = 4/13$
Section 15

Recurrence Relations
A *recurrence relation* is an equation that expresses \( a_n \) in terms of one of more of the previous terms.

A sequence is a *solution* to a recurrence relation if its terms satisfy the equations.

**Examples:**

- **Recurrence Relation:** \( a_n = a_{n-1} + 3, \ a_0 = 2 \).
  
  Solution: \([2, 5, 8, \ldots]\)

- **Fibonacci:** \( a_n = a_{n-1} + a_{n-2}, \ a_0 = 1, \ a_1 = 1 \).
  
  Solution: \([0, 1, 1, 2, 3, 5, \ldots]\)

- A *closed form solution* is an equation for each term that does not reference other terms.
Linear Homogeneous Recurrence Relation

Definition

- A *linear homogeneous recurrence relation of degree k with constant coefficients* is a recurrence relation of the form: \( a_n = c_1a_{n-1} + c_2a_{n-2} + \ldots + c_k a_{n-k} \) such that every \( c_i \in \mathbb{R} \) and \( c_k \neq 0 \).
- Linear because each right-hand side term is sum of previous terms
- Homogeneous because no terms occur that are not multiples of previous terms
- Constant coefficients means no \( c_i \) can reference \( n \) (but note that 0 is allowed for all but last coefficient)
- The degree is determined by the number of terms required
Linear Homogeneous Recurrence Relation

▶ Examples

▶ $a_n = \frac{3a_{n-1}}{2}$ is l.h.r.r. of degree 1
▶ Fibonacci is l.h.r.r of degree 2
▶ $a_n = 2 \times a_{n-5}$ is l.h.r.r of degree 5 (4 terms with 0 as coefficient)
▶ $a_n = a_{n-1} + 3$ is not l.h.r.r because 3 is not multiple of previous term
▶ $a_n = 2^n a_{n-1}$ is not l.h.r.r because $2^n$ is not constant coefficient
▶ $a_n = a_{n-1}^2$ is not l.h.r.r because squared term is not linear
Solving L.H.R.R. of degree 2

- Find closed form equation (of the form $a_n = r^n$).
- $r^n = c_1 r^{n-1} + c_2 r^{n-2} + \ldots + c_k r^{n-k}$
- $r^k = c_1 r^{k-1} + c_2 r^{k-2} + \ldots + c_k$ – divide both sides by $r^{n-k}$
- $r^k - c_1 r^{k-1} - c_2 r^{k-2} - \ldots - c_k = 0$ – is the characteristic equation
- Solutions to the characteristic equation are the characteristic roots
- Assuming degree=2 and distinct roots $r_0$ and $r_1$, $a_n = \alpha_1 r_0^n + \alpha_2 r_1^n$
- Use initial terms to solve for $\alpha_1$ and $\alpha_2$
Solving L.H.R.R.

- \( a_n = 5a_{n-1} - 6a_{n-2}, a_0 = 1, a_1 = 0 \)
- \( r^n = 5r^{n-1} - 6r^{n-2} \rightarrow r^2 = 5r - 6 \rightarrow r^2 - 5r + 6 = 0 \)
- Characteristic Roots are 3, 2
- \( a_n = \alpha_1 3^n + \alpha_2 2^n \)
- \( 1 = \alpha_1 + \alpha_2 \rightarrow 1 - \alpha_1 = \alpha_2 \)
- \( 0 = \alpha_1 * 3 + \alpha_2 * 2 \rightarrow 0 = \alpha_1 * 3 + (1 - \alpha_1) * 2 \rightarrow \alpha_1 = -2 \rightarrow \alpha_2 = 3. \)
- \( a_n = -2(3^n) + 3(2^n) \)
- Check: Sequence solution is [1, 0, -6, -30, ...]
- \( a_3 = -2(3^3) + 3(2^3) = -54 + 24 = -30 \)
Solving L.H.R.R.

- Fibonacci: \( a_n = a_{n-1} + a_{n-2}, a_0 = 0, a_1 = 1 \)
- \( r^2 = 1r^1 + 1r^0 \)
- \( r^2 - r^1 - 1 = 0 \)
- Quadratic Equation: \( \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \)
- \( a = 1, b = -1, c = -1, \frac{1+\sqrt{1+4}}{2}, \frac{1-\sqrt{1+4}}{2} \)
- \( a_n = \alpha_1 \left( \frac{1+\sqrt{5}}{2} \right)^n + \alpha_2 \left( \frac{1-\sqrt{5}}{2} \right)^n \)
- Use \( a_0 \) and \( a_1 \) to determine values for alpha
Solving L.H.R.R. continued

\[0 = \alpha_1 \left(\frac{1+\sqrt{5}}{2}\right)^0 + \alpha_2 \left(\frac{1-\sqrt{5}}{2}\right)^0 = \alpha_1 + \alpha_2\]

Therefore, \(-\alpha_1 = \alpha_2\)

\[1 = \alpha_1 \left(\frac{1+\sqrt{5}}{2}\right)^1 + \alpha_2 \left(\frac{1-\sqrt{5}}{2}\right)^1\]

Substituting: \(1 = \alpha_1 \left(\frac{1+\sqrt{5}}{2}\right)^1 - \alpha_1 \left(\frac{1-\sqrt{5}}{2}\right)\)

\[1 = \alpha_1 \left(\frac{1+\sqrt{5}}{2} - \frac{1-\sqrt{5}}{2}\right) = \alpha_1 \frac{1+\sqrt{5}-1+\sqrt{5}}{2} = \alpha_1 \sqrt{5}\]

\[\alpha_1 = \frac{1}{\sqrt{5}}, \alpha_2 = \frac{-1}{\sqrt{5}}\]

\[a_n = \frac{1}{\sqrt{5}} \left(\frac{1+\sqrt{5}}{2}\right)^n + \frac{-1}{\sqrt{5}} \left(\frac{1-\sqrt{5}}{2}\right)^n\]

Test cases:

- \(\text{fib}(0) = 0.\) \(\frac{1}{\sqrt{5}} \left(\frac{1+\sqrt{5}}{2}\right)^0 + \frac{-1}{\sqrt{5}} \left(\frac{1-\sqrt{5}}{2}\right)^0 = \frac{1}{\sqrt{5}} - \frac{1}{\sqrt{5}} = 0\)

- \(\text{fib}(5) = 5.\) \(\frac{1}{\sqrt{5}} \left(\frac{1+\sqrt{5}}{2}\right)^5 + \frac{-1}{\sqrt{5}} \left(\frac{1-\sqrt{5}}{2}\right)^5 \approx 4.96 - 0.04 = 5\)

- \(\text{fib}(18) = 2584.\) \(\frac{1}{\sqrt{5}} \left(\frac{1+\sqrt{5}}{2}\right)^{18} + \frac{-1}{\sqrt{5}} \left(\frac{1-\sqrt{5}}{2}\right)^{18} \approx 2584.00007 - (7 \times 10^{-5} = 2584)\)
Let $r^2 - c_1 r^1 - c_2 r^0$ have only one real root $x$.

Closed form solution for $a_n = \alpha_1 x^n + \alpha_2 n * x^n$

Example: $a_n = 4a_{n-1} - 4a_{n-2}$, $a_0 = 6$, $a_1 = 8$

$r^2 = 4r - 4 \rightarrow r^2 - 4r + 4 = 0$

$\frac{4 \pm \sqrt{16-4*1*4}}{2} = \frac{4}{2} = 2$

$a_n = \alpha_1 2^n + \alpha_2 n * 2^n$

$a_0 = \alpha_1 2^0 + \alpha_2 0 * 2^0 \rightarrow 6 = \alpha_1$

$a_1 = \alpha_1 2^1 + \alpha_2 1 * 2^1 \rightarrow 8 = 6 * 2 + \alpha_2 * 2 \rightarrow -2 = \alpha_2$

$a_n = 6(2^n) - 2n(2^n)$

Check: [6, 8, 8, 0, -32, ...]

$a_4 = 6 * 16 - 8 * 16 = 96 - 128 = -32$
Section 16

Relations
Definitions of Relations

- A *relation* between sets $A$ and $B$ is a subset of $A \times B$
- Typically, a relation defines a connection between elements of the set
- Example 1: $A = \{ \text{students in class} \}$, $B = \{ \text{side of room} \}$, $R(a, b) \leftrightarrow a \text{ sits on } b \text{ side of the room.}$ (on board)
- Example 2: $A = \{ \text{volunteers} \}$, $B = \{ \text{food} \}$, $R(a, b) \rightarrow a \text{ likes } b.$ (on board)
- Functions are relations restricted such that elements from $A$ appear only once (Example 1)
- Graphs can show relations
- Relations can be on one set $A = \{ \text{food} \}$, $B = \{ \text{food} \}$, $R(a, b) \leftrightarrow a \text{ is the same color as } b$ (on board). Usually written as $R(a, a)$
Properties of Relations

- Consider relations on $\mathbb{Z} \times \mathbb{Z}$

  - **Reflexive**: $\forall a \in A, R(a, a)$
    - $R(z_0, z_1) \iff z_0 \leq z_1$ is reflexive
    - $R(z_0, z_1) \iff z_0 < z_1$ is NOT reflexive

  - **Symmetric**: $\forall a \in A, \forall b \in B, R(a, b) \rightarrow R(b, a)$
    - $R(z_0, z_1) \iff z_0 \leq z_1$ is NOT symmetric
    - $R(z_0, z_1) \iff z_0 = z_1$ is symmetric (and reflexive)
    - $R(z_0, z_1) \iff z_0$ and $z_1$ are relatively prime is symmetric (and not reflexive)

  - **Antisymmetric** (poorly named):
    $\forall a \in A, \forall b \in B, R(a, b) \land R(b, a) \rightarrow a = b$
    - $R(z_0, z_1) \iff z_0 \leq z_1$ is antisymmetric
    - $R(z_0, z_1) \iff z_0 = z_1$ is antisymmetric
    - $R(z_0, z_1) \iff z_0$ and $z_1$ are relatively prime is NOT antisymmetric
Properties of Relations

- **Transitive**: \( R(a, b) \wedge R(b, c) \rightarrow R(a, c) \)
  - \( R(z_0, z_1) \leftrightarrow z_0 \leq z_1 \) is transitive
  - \( R(z_0, z_1) \leftrightarrow z_0 = z_1 \) is transitive
  - \( R(z_0, z_1) \leftrightarrow z_0 \) and \( z_1 \) are relatively prime is **NOT** transitive
Sets and Relations

- Relations are sets of ordered pairs. Therefore, all set operations apply.
- $R(z_0, z_1) \leftrightarrow z_0 \leq z_1 - R(z_0, z_1) \leftrightarrow z_0 < z_1 = R(z_0, z_1) \leftrightarrow z_0 = z_1$

- Proof:
  - Let $LEQ = R(z_0, z_1) \leftrightarrow z_0 \leq z_1$, $LT = R(z_0, z_1) \leftrightarrow z_0 < z_1$ and $EQ = R(z_0, z_1) \leftrightarrow z_0 = z_1$.
  - Let $(a, b) \in LEQ - LT$. Therefore, $(a, b) \in LEQ$ and $(a, b) \notin LT$
    - Therefore, $a \leq b$ and $a \geq b$ (not less than).
    - Therefore, $a = b$ and $EQ(a, b)$. Reverse direction is similar.

- Let $A = \{1, 2, 3\}$. Let $R(a, a) = \{(1, 1), (2, 2), (1, 2)\}$ and $S(a, a) = \{(1, 1), (2, 2), (2, 1)\}$.
  - $R \cup S = \{(1, 1), (2, 2), (1, 2), (2, 1)\}$
  - $R \cap S = \{(1, 1), (2, 2)\}$
  - $R \oplus S = \{(1, 2), (2, 1)\}$
N-ary Relations

- Text is awkward with notation
- Extend notion to n-wise cross product.
- Given sets $S_0, S_1, \ldots S_{n-1}$, $R \subseteq S_0 \times S_1 \times \ldots \times S_{n-1}$
- $r \in R$ is a $n$-tuple. Note that the ordering is important.
- Relational databases (Oracle, MySQL, SQLServer, etc.) use tables as relations with attributes representing sets
- Example: Students(Id, Name, Major, Favorite Number)
- Collection of attributes is the schema
- Note: Databases allow duplicate elements – database tables are bags of n-tuples
- Id is primary key uniquely identifies row in table
Basic Relational Algebra

- Let $R(A, B, C) = \{(1, 2, 3), (2, 3, 4)\}$ and $S(C, D, E) = \{(3, 4, 5), (3, 2, 1)\}$
- $\sigma_P R$ (selection) creates new table with same schema as $R$. A row is in $\sigma_P R$ if it is in $R$ and it satisfies predicate $P$
- $\sigma_{A=1} R = \{(1, 2, 3)\}$
- $\Pi_{A,B} R$ (projection) creates new table with columns $A$ and $B$. There is a 1-1 mapping from each row in $R$ to each row in $\Pi_{A,B} R$
- $\Pi_{B,C} R = \{(2, 3), (3, 4)\}$
- $R \bowtie S$ (natural join) creates new table with union of columns in $R$ and in $S$. A row is in $R \bowtie S$ if $\exists r \in R \land s \in S$ such that $r[R \cap S] = s[R \cap S]$.
- $R \bowtie S = \{(1, 2, 3, 4, 5), (1, 2, 3, 2, 1)\}$
- Note: Results of relational algebra operators are relations. Operations can be composed.
- $\Pi_{A,E}(\sigma_{B=D}(R \bowtie S) = \{(1, 1)\}$
Queries

- Example queries using relational algebra
- To be added Spring 2019
Section 17

Relations, Matrices and Digraphs
Matrices and Relations

- Matrix representation of $R(a, a)$ (can be any relation – see text)
- $M_R[i, j] = 1 \iff R(i, J)$. Otherwise, $M_R[i, j] = 0$.
- Matrix $M_R$ is the representation of $R = \{(1, 1), (1, 2), (2, 1), (3, 3)\}$
- Matrix $M_S$ is the representation of $S = \{(2, 2), (2, 3), (3, 2), (3, 3)\}$
- Matrix diagonal all 1's implies relation is reflexive
- $M = M^t$ implies relation is symmetric

\[
M_R = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad M_S = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}
\]
Matrix Operations and Relational Operations

- $A \lor B$ is the join of matrices $A$ and $B$ – logical OR of corresponding elements
- $A \land B$ is the meet of matrices $A$ and $B$ – logical AND of corresponding elements
- $A \odot B$ is the Boolean product of matrices $A$ and $B$ –
  \[ C = A \odot B \rightarrow c_{ij} = (a_{i1} \land b_{1j}) \lor (a_{i2} \land b_{2j}) \lor \ldots \lor (a_{in} \land b_{nj}) \]
  – see below (note similarity to matrix multiplication)
- $M_R \lor M_S = R \cup S$
- $M_S \land M_S = R \cap S$
- Boolean product is composition of relations
- $M_R \odot M_R = M_R \rightarrow R$ is transitive

\[
\begin{bmatrix}
1 & 1 & 0 \\
1 & 0 & 0 \\
0 & 0 & 1
\end{bmatrix}
\odot
\begin{bmatrix}
0 & 0 & 0 \\
0 & 1 & 1 \\
0 & 1 & 1
\end{bmatrix}
=
\begin{bmatrix}
0 & 1 & 1 \\
0 & 0 & 0 \\
0 & 1 & 1
\end{bmatrix}
\]
Let $R$ be a relation on set $A$

$G_R = (A, E)$ where $E \subseteq A \times A$ such that $(a_i, a_j) \in E \leftrightarrow R(a_i, a_j)$

Example on board for $R = \{(1, 1), (1, 2), (2, 3), (3, 1)\}$

Boolean product of $M_R \odot M_R$ contains edges two steps away

\[
\begin{bmatrix}
1 & 1 & 0 \\
0 & 0 & 1 \\
1 & 0 & 0
\end{bmatrix} \odot \begin{bmatrix}
1 & 1 & 0 \\
0 & 0 & 1 \\
1 & 0 & 0
\end{bmatrix} = \begin{bmatrix}
1 & 1 & 1 \\
1 & 0 & 0 \\
1 & 1 & 0
\end{bmatrix}
\]
Paths and Boolean Products

- $M^n_R[i,j] = 1$ (Boolean product of $M_R$ with itself $n$ times) iff $G_R$ contains a path of length $n$ from $a_i$ to $a_j$

- Proof (by induction):
  - WLOG, let $|A| = p$
  - Basis: By definition, $G_R$ contains an edge from $a_i$ to $a_j$ exactly when $M_R[i,j] = 1$
  - Inductive Hypothesis: If $M^k_R[i,j] = 1$, there exists a path of length $k$ from $a_i$ to $a_j$ in $G_R$.
  - Consider $M^{k+1}_R = M^k_R \odot M_R$. $M^{k+1}_R[i,j] = 1$ iff $\exists m, 1 \leq m \leq p$ such that $M^k_R[i,m] = 1$ and $M_R[m,j] = 1$.
  - By the IH, $M^k_R[i,m] = 1$ means there is a path of length $k$ from $a_i$ to $a_m$. Call this path $P$.
  - $M_R[m,j] = 1$ means there is an edge from $a_m$ to $a_j$.
  - The path $P'$ which follows $P$ for $k$ steps, then takes the edge from $a_m$ to $a_j$ is a path from $a_i$ to $a_j$ and is of length $k + 1$

- The completion of the proof (showing if there is a path of length $k$ in $G_R$ then $M^k_R[i,j] = 1$) is in the homework.
Let $|A| = n$.

$\circ_{k=1}^{n} M_{R}^{k}$ is the transitive closure of $R$

Also called the connectivity relation $R^*$.

Given a graph $G = (A, E)$, we can define $R(a_i, a_j) \leftrightarrow (a_i, a_j) \in E$ (e.g., derive relation from graph).

Note: If there is a path from $a_i$ to $a_j$ in $G$, then the shortest path cannot be longer than $|A|$.

$\circ_{k=1}^{n} M_{R^*}^{k}$ is the set of all paths in $G$
An equivalence relation is a relation that is reflexive, symmetric and transitive.

If $R(A, A)$ is an equivalence relation, then $R(a_i, a_j)$ means $a_i \sim a_j$ ($a_i$ and $a_j$ are equivalent).

Let $R(A, A)$ be an equivalence relation. The equivalence class of $a_i \in A$, $[a]_R = \{a_j \mid R(a_i, a_j)\}$.

Note: $a_i \in [a_i]_R$, since $R$ is reflexive.

Note: $\bigcup_{a \in A} [a]_R = A$

Note: $a_j \notin [a_i]_R \rightarrow [a_j]_R \cap [a_i]_R = \emptyset$

Note: $a_j \in [a_i]_R \rightarrow [a_j]_R = [a_i]_R$

The equivalence classes of $R$ form a partition of $A$. 
Partial Orders

- $R(A, A)$ is a *partial order* on $A$ iff it is reflexive, antisymmetric and transitive.
- $(A, R)$ is a *partially ordered set* or a *poset*.
- Consider $R = \{(1, 1), (2, 2), (3, 3), (4, 4), (1, 2), (1, 3), (2, 4)\}$ ($M_R$ is below).
- Arbitrary relation symbol is $\preceq$, so $(A, \preceq)$ is a poset with arbitrary relation.

\[
M_R = \begin{bmatrix}
1 & 1 & 0 & 1 \\
0 & 1 & 0 & 1 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{bmatrix}
\]
Poset Properties

- Not all elements are related – 2 and 3 from previous
- if \( a \preceq b \) holds a and b are *comparable*.
- if all elements in A are comparable, \((S, \preceq)\) is a *total ordering*.
- \( \forall a_i \in A, a_j \not\preceq a_i \rightarrow a_j \) is a *maximal* element (3 and 4 are maximal in example)
- \( \forall a_i \in A, a_i \not\preceq a_j \rightarrow a_j \) is a *minimal* element (1 is minimal in example)
- every poset has at least one minimal and one maximal element (can have more)
Section 18

Graphs
Terminology

- A graph $G = (V, E)$ where $V$ is a set of vertices (or nodes) and $E \subseteq V \times V$ is a set of edges.
- A graph can be *directed*
  - first vertex in an edge is the *source*
  - second vertex is the *destination*
  - connectivity is from source to destination
  - edges represented as arrows pointing at the destination
  - example on board
- A graph can be *undirected*
  - both vertices are incident on edge
  - connectivity is bidirectional
  - edges represented as lines between vertices
  - example on board
- A graph is *simple* if $E$ is a set and there are no self-loops.
Neighborhoods

- Vertex \( v_i \) is adjacent to \( v \) if there is an edge from \( v \) to \( v_i \)
- Vertex \( v_i \) is a **neighbor** of \( v \) if it is adjacent to \( v \)
- Set of neighbors of \( v \) are the **neighborhood** of \( v \), denoted \( N(v) \)
- Directed graph definition: \( N(v) = \{ v_i \in V | (v, v_i) \in E \} \)
- Undirected graph definition: \( N(v) = \{ v_i \in V | (v, v_i) \in E \lor (v_i, v) \in E \} \)
- Neighborhood can apply to \( A \subset V \). \( N(A) = \bigcup_{v \in A} N(v) \).  
- Example on board
Degree of a node

- The degree \( \deg(v) \) of a node in an undirected graph is the number of times an edge connects to the node.

\[
\deg(v) = |\{(v, v_i) | (v, v_i) \in E\}| + |\{(v_i, v) | (v_i, v) \in E\}|
\]

- Example on board

- **Handshake Theorem**: Let \( G = (V, E) \) be an undirected graph. \( 2|E| = \sum_{v \in V} \deg(v) \).

- Proof by induction
  - Basis: Let \( E = \emptyset \). Therefore, \( \forall v \in V, \deg(v) = 0 \), so \( \sum_{v \in V} \deg(v) = 0 = 2 * |E| \)
  - Inductive Hypothesis: Let \( G = (V, E) \) be a graph with \( k \) edges. Therefore, \( 2|E| = \sum_{v \in V} \deg(v) \).
  - Let \( G' = (V', E') \) such that \( V' = V \) and \( E' = E \cup (v_i, v_j) \).
  - Case 1: \( (v_i, v_j) \in E \). Therefore, \( E = E' \) and by the inductive hypothesis, the theorem holds.
  - Case 2: \( (v_i, v_j) \notin E \). Therefore, \( |E'| = |E| + 1 \).
  - Note that, \( \deg(v_i) \) and \( \deg(v_j) \) both increase by one. Therefore, \( \sum_{v \in V} \deg(v) = \sum_{v \in V} \deg(v) + 2 \).
  - Therefore, \( 2|E'| = 2|E| + 2 = \sum_{v \in V} \deg(v) + 2 = \sum_{v \in V} \deg(v) \)
Indegree and outdegree

- The **indegree** ($\text{deg}^-(v)$) of a node $v$ in a directed graph is the number of edges with $v$ as the destination.
- The **outdegree** ($\text{deg}^+(v)$) of a node $v$ in a directed graph is the number of edges with $v$ as the source.
- Example on board
- $\sum_{v \in V} \text{deg}^-(v) = \sum_{v \in V} \text{deg}^+(v) = |E|$
Special Graphs

- Let $G = (V, E)$ be a simple graph. $G$ is a **complete graph** iff $\forall v_i, v_j \in V, v_j \in N(v_i)$.
- Example on board.
- Let $G = (V_1 \cup V_2, E)$ be a simple graph. $G$ is a **bipartite graph** iff $\forall (v_i, v_j) \in E, v_i \in V_1 \rightarrow v_j \in V_2 \land v_i \in V_2 \rightarrow v_j \in V_1$
- Example on board.
- Let $G = (V_1 \cup V_2, E)$ be a simple graph. $G$ is a **complete bipartite graph** iff it is bipartite and every node in $V_1$ is connected to every node in $V_2$. 
Matchings

- A **matching** of a bipartite graph $G = (V_1 \cup V_2, E)$ is a subgraph $G' = (V_1 \cup V_2, E' \subseteq E)$ such that $\forall v \in V_1 \cup V_2, \deg(v) \leq 1$.

- A **maximum matching** is a matching with the largest number of edges.

- A **complete matching** is a matching from $V_1$ to $V_2$ such that all nodes in $V_1$ are incident on an edge ($|E'| = |V_1|$)

- **Hall’s Marriage Theorem**: A bipartite graph $G = (V_1 \cup V_2, E)$ has a complete matching iff $\forall A \in 2^{V_1}, |N(A)| \geq |A|$. 
Section 19

Graphs, Paths and Circuits
Graph Isomorphism

- $G_1 = (V_1, E_1)$ is isomorphic to $G_2 = (V_2, E_2)$ iff there exists a bijective function $f : V_1 \rightarrow V_2$ such that $v_i \in N(v_j)$ in $G_1$ iff $f(v_i) \in N(f(v_j))$ in $G_2$.
- Example on board (isomorphic and not isomorphic)
- Properties which must hold under isomorphism
  - Number of vertices
  - Number of edges
  - Number of vertices with same degree
- $M_G$ is adjacency matrix of $G$
- Rearrange rows and columns of $M_{G_1}$ until $M_{G_1} = M_{G_2}$, then $G_1$ is isomorphic to $G_2$
- Example on board
Paths

- A **path** in a simple graph $G$ is a sequence $P = [x_0, x_1, \ldots, x_{n-1}]$ of vertices such that $x_{i+1} \in N(x_i)$.
- A **circuit** is a path such that $x_0 = x_{n-1}$
- The length of $P$ is $n - 1$.
- Example on board (undirected and directed)
- A graph is **connected** if there is a path (in both directions) between every pair of distinct vertices
- A **connected component** of $G$ is a subgraph $G'$ such that $G'$ is connected and there does not exists a connected subgraph $G''$ of $G$ such that $G'$ is a proper subgraph of $G''$.
- An **articulation point** is a vertex $v \in V$ such that the removal of $v$ would make $G$ no longer connected
- A **bridge** is an edge $(v_i, v_j) \in E$ such that the removal of $(v_i, v_j)$ would make $G$ no longer connected
- Example on board
- Paths are preserved under isomorphism. Therefore, if $G_1$ has a circuit of length $k$ and $G_2$ does not, then $G_1$ and $G_2$ are not isomorphic.
Euler Circuits and Paths

- A multigraph $G$ contains multiple edges from two vertices (i.e., $E$ is not a set)
- A Euler circuit in $G$ is a simple circuit containing every edge in $G$
- A Euler path in $G$ is a simple path containing every edge in $G$
- Example on board
- Conditions for Euler circuits:
  - $G$ must be connected
  - Every vertex must have even degree
- Conditions for Euler path, but NOT Euler circuit
  - $G$ must be connected
  - Exactly two vertices with odd degree
- Excellent proofs in text
A Hamilton circuit in $G$ is a simple circuit containing every vertex in $G$

A Hamilton path in $G$ is a simple path containing every vertex in $G$

Example on board

No known simple criteria for Hamilton circuits or paths (necessary and sufficient)

Sufficient criteria for circuit in simple, undirected graph $G = (V, E)$

- $G$ is connected
- $|V| \geq 3 \land \forall v \in V, \deg(v) \geq n/2$ (Dirac’s Theorem)
- $|V| \geq 3 \land \forall u, v \in V, u \notin N(v) \rightarrow \deg(u) + \deg(v) \geq n$ (Ore’s Theorem)
Section 20

Shortest Path and Trees
Shortest Path

- A weighted graph $G^+(V, E)$ where $E$ is a set of 3-tuples $(v_i, v_j, w)$ such that $w$ is the weight of the edge $(v_i, v_j)$
- Can be directed or undirected
- The matrix representation of $G^+$ uses weights as values (assuming simple graph with all weights positive)
- Example on board using matrix below
- The shortest path has the least sum of weights
- Example on board from 0 to 4

\[
M = \begin{bmatrix}
0 & 10 & 0 & 20 & 0 \\
12 & 12 & 12 & 12 & 12 \\
0 & 0 & 0 & 0 & 0 \\
20 & 12 & 0 & 0 & 10 \\
0 & 12 & 5 & 10 & 0
\end{bmatrix}
\]
Dijkstra’s Algorithm

- Inputs: $G^+, v_i, v_j \in V$
- Output: Shortest path from $v_i$ to $v_j$ in $G^+$
- Keep list of shortest known paths from $v_i$ to all $v \in V$
- Initialize list so that all nodes have unknown path ($P$) with infinite length ($L$)
- Set $P(v_i)$ to $[v_i]$ and $L(v_i) = 0$
- While $v_j$ is unmarked
  - Let $v$ be the vertex with the shortest path so far (choose randomly for ties)
  - Mark $v$
  - For all unmarked $v_k \in N(v)$
  - Let $w$ be from the edge $(v, v_k, w)$
  - If $L(v) + w < L(v_k)$, then set $P(v_k) = P(v).v$ and $L(v_k) = L(v) + w$

- Example on board
A directed acyclic graph (DAG) is a directed graph $G = (V, E)$ such that in the transitive closure of $G$ $G^*_R = (V, E^*), \forall v \in V, v \notin N(v)$

Example on board
$G = (\{0, 1, 2, 3\}, \{(0, 1), (1, 2), (0, 2), (3, 0)\})$

The relation corresponding to a DAG $G$ is not reflexive but it is antisymmetric. It may or may not be transitive.
Trees

- A *tree* is an undirected graph with no simple cycles.
- A *rooted tree* is a DAG such that:
  - The undirected form of the graph is a tree
  - A root is a node with no incoming edges
  - All edges are directed away from the root
- Any edge in tree can be selected as root (different roots yield different trees)
- A *leaf* is a node with no outgoing edges
- The *branching factor* of a tree is the maximum number of children for each node
- If $m$ is the branching factor, than the tree is *$m$-ary*. If $m = 2$, the tree is binary
Tree Properties

- A tree with \( n \) vertices has \( n - 1 \) edges.
- Inductive Proof (from text):
  - Basis: \( n=1 \). One node. No edges.
  - Inductive Hypothesis: Every tree with \( k \) vertices has \( k - 1 \) edges.
  - Let \( T \) be a tree with \( k + 1 \) nodes. Let \( v \) be a leaf in \( T \).
  - Removing \( v \) from \( T \) generates a tree \( T' \) with \( k \) nodes (\( T' \) has no simple circuits).
  - Therefore, \( T' \) has \( k - 1 \) edges.
  - There can be only one edge from any node in \( T' \) to \( v \), otherwise a cycle would exist (can you prove why?)

- The level of a node is the length of the path from the root to the node.
- The height of a tree is the maximum level of any node.
- A tree of height \( h \) is balanced if all leaves are at level \( h \) or \( h - 1 \).
- There are at most \( m^h \) leaves in an m-ary tree of height \( h \).
Section 21

Tree Traversals and Heaps
Alternative Representation for Sparse Graph

- Consider complete binary tree of height \( k \)
- \( 2^k \) nodes. Matrix is \( 2^k \times 2^k \) with \( 2^{2k} \) entries
- Each non-leaf node has 2 neighbors. All leaves have zero neighbors.
- Matrix has \( 2^k \) non-zero values and \( 2^k(2^k - 1) \) zeros.
- Examples:
  - \( k = 3 \). Matrix has 64 entries. Eight are one. 56 are zero.
  - \( k = 10 \) Matrix has 1,048,576 entries. 1,047,552 are zero.
- Adjacency List
  - Array of values for each node (size \( 2^k \)).
  - Root is index location 0
  - Parallel array with list of neighbors – “in order” if applicable
  - Weights become parallel array with list of (neighbor, weight) pair
- Node Object
  - Value with node references (pointers) (size \( 2^k \))
  - References stored “in order” (if applicable)
  - Root reference stored in special location
  - Weights stored with references
Tree Traversals

- Recursive procedure for processing nodes
- Preorder traversal
  - Visit node first
  - Visit children in order
  - Return
- Inorder traversal (more common in binary trees)
  - Visit first child
  - Visit node
  - Visit remaining children in order
  - Return
- Postorder traversal
  - Visit children in order
  - Visit node
  - Return
Depth First Search

- Values can be complex.
- Example: Game of tic-tac-toe (below)
- Spanning tree of graph (tree containing all nodes of graph)
- Algorithm
  - Visit children in order
  - Visit node
  - Return
- Example on board

<table>
<thead>
<tr>
<th></th>
<th>O</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>O</td>
</tr>
<tr>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>
Breadth First Search

- Same applications
- Algorithm
  - Place root in queue
  - While queue not empty
    - Pop node N from queue
    - Visit N
    - Append children of N to queue
- Example on board

```
  O
  X O X
  X
```
Heap

- Tree structure (often binary)
- Parent always greater than or equal to children (max heap)
- Insertion:
  - Add new value to first available leaf
  - if child greater than parent, swap (continue until swap root)
- Example: 10,5,15,8,12,20,2
- Pop: (remove top element)
  - Move greater child into empty slot
  - continue until leaf moved
- Different insertion order can yield different heaps
- Example: 2,5,8,10,12,15,20