CMSC 27130: Honors Discrete Mathematics

Lectures by Alexander Razborov

Notes by Yueheng Zhang

The University of Chicago, Autumn 2018

Notes and references

• The two guest lectures given by Professor Stuart A. Kurtz on algorithms are not covered here since there are complete lecture notes on his website: http://cmsc-27100.cs.uchicago.edu/2018-winter/lectures.php. Topics covered in class are from the second half of lecture 18 to lecture 21 (inclusive).

• When there are detailed definitions or proofs in the textbook, *Discrete Mathematics and Its Applications* (7th Edition) by K. Rosen, that are the same with or very similar to what were presented in class, the section or page number will simply be referred. This book will be referred to as Rosen.

• **Digressions** are materials not required for the grade but beautiful (according to the professor[1]).

• *Exercises* are practice questions or problems related and recommended but not counting towards the grade. Some directions or proofs of them given by Prof. Razborov or Leonardo Nagami Coregliano are included in footnotes.

[1] `confirmed -- A.R.`

Disclaimers: 1. Notes of the first lecture (about ordinary induction and Helly's Theorem) is missing due to an absence. 2. Notes about a digression topic in graph theory about adjacency matrix is not included.

I Ordering

Induction

Ordinary induction

(Rosen 5.1, pp.313, Principle of Mathematical Induction.)

Strong induction

$\forall n((\forall m < n, P(m)) \Rightarrow P(n)) \Rightarrow \forall n P(n)$.

Example.
⌜ An integer p is a prime if $\forall x(x|p \Rightarrow x \in \{1, p\})$.

Theorem. *Fundamental theorem of arithmetic.*
Any positive integer n has a unique representation $n = p_1^{d_1} \cdots p_t^{d_t}$, where $p_1 < p_2 < \cdots < p_t$ are primes and $d_i \geq 1$.

Proof of existence by strong induction on n.

Proof. If n is a prime, we are done. If n is not a prime, n can be written as $n = k \cdot l$, where k and l are positive integers with $k, l < n$. Since k, l each has a prime factorization by inductive hypothesis, n has a prime factorization. □

Note. For the "base" case of the factorization of 1, note that $\bigwedge$(empty set of statement) = TRUE.

Remark. Compare with recursive algorithm:
Procedure $F(n$; non-negative integer)
if $n = 1$
return empty
if n is a prime
return n
if $n = n_1 n_2; 1 < n_1, n_2 < n$
return $F(n_1)F(n_2)$ ⌟

Ordering

Well-ordering

Theorem. *Well-ordering principle*
Every non-empty set of positive integers X has a minimal element.

Let X_n stand for $X \cap \{1, 2, \ldots, n\}$. We prove this by ordinary induction on n.

Proof. Assume $X_n = \emptyset$ or X_n has a minimal element. If $X_{n+1} = \emptyset$, we are done. Otherwise, either $X_n = \emptyset$ or $X_n \neq \emptyset$. If $X_n \neq \emptyset$, the least element in X_n is the least element in X_n. If $X_n = \emptyset, X_{n+1} = \{n + 1\}$ and $n + 1$ is the least element. □

Remark. Strong induction, ordinary induction and well-ordering principle are all equivalent.[1]

[1] OI⇒WOP: proved above, and Rosen 5.2 exercise 41; WOP⇒OI: Rosen 5.1 pp.314; WOP⇒SI: Rosen 5.2 exercise 31; SI ⇒WOP: Rosen 5.2 exercise 41; OI ⇔ SI: Rosen 5.2 exercise 42.

Example.
Minimal counterexample argument. (The existence of a minimal counterexample, a technique often used in proofs.)
$X = \{n \in \mathbb{N} \mid \neg P(n)\}$, then either $X = \emptyset$, or $X \neq \emptyset$ and $\exists n \in X$ that is minimal.
Thus we have $P(1), \cdots, P(n-1), \neg P(n)$.

Linear orderings

Def. Axioms of linear (total) orderings.
S is a set of objects. $R \subseteq S \times S$ is a binary relation $R(a, b)$, $a \leq b$. A linear ordering $(S, \leq)$ satisfies the following axioms.
(a) reflexitivity. $a \leq a$.
(b) antisymmetry. $a \leq b, b \leq a \Rightarrow a = b$.
(c) transitivity. $a \leq b, b \leq c \Rightarrow a \leq c$.
(d) totality. Either $a \leq b$ or $b \leq a$.

A partial-ordered set is a set with a relation that satisfies (a), (b), (c).

Example. The divide relation on integers $m|n$; The subset relation on sets $A \subseteq B$.

Remark. Total order is a special case of partial order.

Theorem. *Extendability theorem.*
For a partial order $(S, \preceq)$, there exists a total order $\leq$ on S, s.t. $a \preceq b \Rightarrow a \leq b$.

Remark. S does not need to be finite, but then we need Axiom of Choice.

Remark. It is easy to construct such an extension, but not easy to count the total number of different extensions.

Reverse induction

Every finite non-empty set of positive integers has a maximal element.

Exercise. Prove the above statement.

Reverse strong induction: $\forall n \leq N((\forall m \in (n, N], P(m)) \Rightarrow P(n)) \Rightarrow \forall n \leq N, P(n)$
Reverse ordinary induction: $(P(N) \wedge (\forall n \leq N(P(n) \Rightarrow P(n-1)))) \Rightarrow \forall n \leq N, P(n)$.

Corresponding to minimal counterexample argument, we have maximal counterexample argument. We can use this to prove the above extendability theorem.

Proof. Suppose $\preceq$ is the largest extension of a given order on S and it is not a total order. Then $\exists a, b \in S$ s.t. $a \not\preceq b, b \not\preceq a$. Define $\preceq' := \preceq \cup \{(c, d) \mid c \preceq a \wedge b \preceq d\}$. Then $\preceq'$ is a larger extension (`Exercise -- A.R.`), which causes a contradiction. □

Transfinite induction

Def. (Rosen 9.6, Definition 4, pp.620) $(S, \preceq)$ is a *well-ordered set* if it is a poset[1] such that $\preceq$ is a total ordering and every nonempty subset of S has a least element.

Theorem. *(Rosen 9.6 Theorem 1, pp.620) The principle of well-ordered induction.*
Suppose that S is a well-ordered set. Then P(x) is true for all $x \in S$, if:
for every $y \in S$, if $P(x)$ is true for all $x \in S$ with $x \prec y$, then $P(y)$ is true.

[1] `In class we mostly discussed linearly well-ordered sets -- A.R.`

Remark. For linearly ordered sets, this is equivalent to the least element principle.

Other sources that might be clarifying: http://mathworld.wolfram.com/TransfiniteInduction.html, http://mathworld.wolfram.com/LimitOrdinal.html.

Induced order

Let $(S, \leq)$ be a linear order, and $T \subseteq S$. Then $(T, \leq_T)$ is an induced ordering. *Isomorphism between two orderings*: $\exists f$, a one-to-one correspondence, s.t. $a \leq b \iff f(a) < f(b)$.

Exercise. Any two infinite subsets of $(\mathbb{N}, \leq)$ are isomorphic.

Ordinal arithmetic

ω - the first infinite ordinal.[1]
addition
$\omega + 1,\ \omega + 2,\ \ldots,\ \omega + n.$

Example. $\omega + 1 = \{\frac{1}{2}, \frac{2}{3}, \ldots, \frac{n-1}{n}, \ldots\} \cup \{1\}$. It is different from ω because it is bounded. (And boundedness is an invariant.)

Note. Addition is associative but not commutative:
$(\alpha + \beta) + \gamma = \alpha + (\beta + \gamma)$;
$1 + \omega = \omega \neq \omega + 1$.
Thus left cancellation holds ($\alpha + \beta = \alpha + \gamma \Rightarrow \beta = \gamma$), but right cancellation does not hold: $1 + \omega = 2 + \omega$.

multiplication
$\omega \cdot 2,\ \omega \cdot 3,\ \ldots,\ \omega \cdot \omega.$

Example.
$\omega \cdot 2 = \omega + \omega = \{1, \frac{1}{2}, \frac{2}{3}, \ldots, 1, 1\frac{1}{2}, 1\frac{2}{3}, \ldots\}$.
$\omega \cdot \omega = \omega^2 = \{k + (1 - \frac{1}{l}) \mid k, l \in \mathbb{N})\}$.
$\omega^\omega := \omega + \omega^2 + \omega^3 + \cdots + \omega^n + \cdots$

Remark. $S \times T$ can be viewed as T copies of S.

Note. Similar to addition, associativity holds but commutativity does not: $2 \cdot \omega = \omega \neq \omega \cdot 2$. Distributivity holds: $\alpha \cdot (\beta + \gamma) = \alpha \cdot \beta + \alpha \cdot \gamma$.

Exercise. Prove that $\omega \cdot 2 \neq \omega$.

Operations on well-ordered sets

addition
$(S, \leq_S), (T, \leq_T)$ are two linearly ordered sets. $S \cap T = \emptyset$.
Define: $x \leq y$ iff $(x, y \in S \wedge x \leq_S y) \vee (x, y \in T \wedge x \leq_T y) \vee (x \in S \wedge y \in T)$.

Theorem. *If $(S, \leq_S), (T, \leq_T)$ are well-ordered, then $(S + T, \leq)$ is well-ordered.*

Proof. For non-empty $X \subseteq S \cup T$,
(a) If $X \cap S \neq \emptyset$, the least element $x_0 \in X \cap S$ is the smallest in X.
(b) If $X \cap S = \emptyset$, then $X = X \cap T$ is non-empty. The least element $x_0 \in X \cap T$ is the smallest in X. □

[1] $\omega = \{0, 1, 2, \ldots\}$ can be identified with the set of all finite ordinals. Reference, http://mathworld.wolfram.com/OrdinalNumber.html

multiplication
$(S, \leq_S), (T, \leq_T)$; $S \times T$.

Def. Lexicographic order.
$(s,t) \leq (s',t')$ iff $s < s' \vee (s = s' \wedge t \leq t')$.

Colex order: the reverse of lexicographic order.

Def. Colex order.
$(s,t) \leq (s',t')$ iff $t < t' \vee (t = t' \wedge s \leq s')$.

Remark. Colex order makes more sense on $\omega \times n : 1, \frac{1}{2}, \frac{2}{3}, \ldots, 1, 1\frac{1}{2}, 1\frac{2}{3}, \ldots, 3, \ldots$. So that we compare the integer part first.

Theorem. *If S, T are well-ordered then colex order $S \times T$ is also well-ordered.*

Proof. Let X be a non-empty subset of $S \times T$. For $t \in T$ define $X_t := \{s \in S | (s,t) \in X\}$; Define a subset Y of T as $\{t | X_t \neq \emptyset\}$($Y$ is the projection of X on T). Y is non-empty because X is non-empty. Let t_0 be the least element in Y, and $X_{t_0} \neq \emptyset$ by definition. Let s_0 be the least element in X_{t_0}, then (s_0, t_0) is the least element in X. □

Remark. By this theorem, we can use double induction on such set $S \times T$. While proving $P(s_0, t_0)$, we can assume that P is true for all (s_0, t) with $t < t_0$ *and* assume that P is true for for all (s, t) with $s < s_0$.

II Number Theory

Recall.
An integer p is a prime if $\forall x(x|p \Longrightarrow x \in \{1, p\})$.

Theorem. *Fundamental Theorem of Arithmetic*
Any positive integer n has a unique representation $n = p_1^{d_1} \cdots p_t^{d_t}$, where $p_1 < p_2 < \cdots < p_t$ are primes and $d_i \geq 1$.

Proof. (Of uniqueness.)

Lemma. (p is a prime and $p|a_1 \cdots a_m$) $\Rightarrow \exists i(p|a_i)$. (Link back to CRT.)

Then we use strong induction on n. Suppose n has two representation $n = p_1^{d_1} \cdots p_t^{d_t} = q_1^{e_1} \cdots q_s^{e_s}$. Then $p_1|n \Rightarrow p_1|q^i$ for some i by lemma. Thus $q_i = p_1$, and $\frac{n}{p_1} = p_1^{d_1 - 1} \cdots p_t^{d_t} = q_1^{e_1} \cdots q_i^{e_i - 1} \cdots q_s^{e_s}$. Since $\frac{n}{p_1}$ has a unique reprensentation by inductive hypothesis, the two prime factorization of n are the same.
Proof of lemma.
It is sufficient to prove that $p|ab \Rightarrow p|a \vee p|b$ then use induction on m. (When $m > 2$, suppose the claim holds for $m = k$, so that $p|a_1 a_2 \cdots a_k \Rightarrow p|a_i, 1 \leq i \leq k$. For $m = k+1$, $p|a_1 a_2 \cdots a_{k+1} \Rightarrow p|(a_1 a_2 \cdots a_k) a_{k+1} \Rightarrow (p|a_1 a_2 \cdots a_k) \vee p|a_{k+1}$. Then by inductive hypothesis ($p|a_1 \vee p|a_2 \vee \cdots \vee p|a_k) \vee (p|a_{k+1})$.

Remark. $p|ab \Rightarrow p|a \vee p|b$ is the cornerstone of this proof.

To prove this, we use another theorem.

Def. a, b are integers, $\gcd(a, b)$ is the largest integer d s.t. $d|a$ and $d|b$.

Theorem. *Bezout's theorem*
For integers a, b, n, $\exists x, y \in \mathbb{Z}$ *s.t.* $ax + by = n \Leftrightarrow \gcd(a, b)|n$. [1]

Apply this theorem to our proof. If $\gcd(p, a) = 1$, $\gcd(p, b) = 1$, then $\exists x, y, u, v \in \mathbb{Z}$ s.t. $px + ay = 1$, $pu + bv = 1$. Thus $p(pxu + xbv + uay) + ab \cdot yv = 1$. Thus $\gcd(p, ab)|1$, and $p \nmid ab$. □

Euclid's Algorithm

Procedure Bezout ($a \geq b > 0$, integers):
    if $b|a$ then output $(x = 0, y = 1)$
    else $a = bq + r,\ b > r > 0$
        (x, y)= Bezout(b, r)
    return $(y, x - qy)$
Analysis:
(a) termination
    $(a, b) \longrightarrow (b, r) \longrightarrow \cdots$ and $(b + r) < (a + b)$, so it termintates.
(b) correctness

Lemma. $\gcd(a, b) = \gcd(b, r)$

Proof. $\forall d(d|a, d|b \equiv d|b, d|r)$, since $a = bq + r$. □

Then we can check the correctness of the return value by reverse induction. $(bx+ry = d)\wedge(a = bq + r) \Rightarrow ay + b(x - qy) = d$.

Remark. In algorithm analysis there is often another step – (c) no failures. It does not apply here since we always output "something" by design.

complexity analysis

Theorem. *Lame's theorem*
The number of steps of procedure Bezout, $t \leq \log_{\frac{3}{2}}(a + b)$.

Proof. This follows from $(b + r) \leq \frac{2}{3}(a + b)$.
Case 1, $b \leq \frac{a}{2}$. Then $\frac{2}{3}(a + b) \geq \frac{2}{3}(2b + b) = 2b \geq b + r$.
Case 2, $b > \frac{a}{2}$. Then $q = 1$ and $a = b + r$. Then $a \leq 2b \Leftrightarrow a \leq \frac{2}{3}(a + b)$. □

Digression

⌜ Uni-variate polynomials, $f(x) = a_0 + a_1x + \cdots + a_nx^n,\ g(x) = b_0 + b_1x + \cdots + b_mx^m$.
$f|g$ if $\exists h$ s.t. $g = f \cdot h$.
If $\gcd(f, g) = h$, then any αh with $\alpha \in \mathbb{Q}^*$ is also gcd of f and g.
Euclidean algorithm
Suppose $n \geq m$ and define f^* as $f - \frac{a_n}{b_m}x^{n-m}g$. Then $\deg(f^*) \leq m - 1$.
$\gcd(f, g) = \gcd(g, f^*)$. ⌟

[1] A proof that might be clarifying: https://en.wikipedia.org/wiki/B%C3%A9zout%27s_identity#Proof

"Simple" commutative algebra

Def. A commutative ring[1] consists of a set, an additive composition, an additive neutral element, a multiplicative composition, a multiplicative identity element ($< A, 0, 1, +, \times >$) and satisfies the follwing axioms.
It is a commutative group with respect to addition:
$\forall a, b, c, \quad (a + b) + c = a + (b + c)$. (associativity)
$\forall a, b, \quad a + b = b + a$. (commutativity)
$\forall a, \quad a + 0 = a$. (identity element)
$\forall a, \quad \exists(-a) \in A$ s.t. $a + (-a) = 0$. (additive inverse)
It is a commutative semi-group with respect to multiplication:
$\forall a, b, c, \quad (a \cdot b) \cdot c = a \cdot (b \cdot c)$. (associativity)
$\forall a, b, \quad a \cdot b = b \cdot a$. (commutativity)
$\forall a, \quad a \cdot 1 = a$. (identity element)
Distributivity of multiplication over addition:
$\forall a, b, c, \quad a \cdot (b + c) = a \cdot b + a \cdot c$.

Note. Commutativity of multiplication holds for a commutative ring, but it does not always hold for a ring. For example it does not hold for the multiplication of matrices.

Note. If there are also multiplicative inverses ($\forall a \neq 0, \exists a^{-1}$ s.t. $a \cdot a^{-1} = 1$), then it is a **field**.

Example. Commutative rings: $\mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$, uni-variate polynomials, multi-variate polynomials, rational functions ...

Def. Let R be a commutative ring. Then a subset $I \subseteq R$ is an ideal iff the following holds:
(a) $0 \in I$
(b) $a, b \in I \Rightarrow a + b \in I$
(c) $(a \in I, r \in R) \Rightarrow a \cdot r \in I$

Example. $R = \mathbb{Z}$, let n be a fixed integer. $n\mathbb{Z} = \{0, n, 2n, \cdots, -n, -2n, \cdots\}$.

Def. The ideal $(a_1, \cdots, a_n)$ generated by $a_1, \cdots, a_n \in R$ is the set of all expressions $\{a_1 r_1 + \cdots + a_n r_n | r_i \in R\}$.

Example. $n\mathbb{Z} = (n)$ is an ideal.

Bezout's Theorem could be used to reduce the cardinality of the generating set.
$\forall a, b, \in \mathbb{Z}$, denote $\gcd(a, b)$ as d, then $(a, b) = (d)$.

Proof. $\exists x, y \in \mathbb{Z}$ s.t. $d = ax + by$, thus $d \in (a, b)$. Thus $(d) \subseteq (a, b)$. While $(a, b) \in (d)$ since $d|a$ and $d|b$. Thus $(a, b) = (d)$. □

Def. An ideal is principal if it can be generated by a single element.

$\mathbb{R}[x, y]$ is not principal.

Remark. By Bezout's Theorem, every ideal over integers is principal.

Def. A ring is Noetherian if every ideal is finitely generated.

Theorem. *Hilbert's Theorem*
If R is a Noetherian ring, then $R[X]$ is a Noetherian ring.

[1] A reference listed on the course website that might be clarifying: S. Lang, *Algebra*, Springer, 8th edition, 2003. (Especially the first two chapters.)

$\mathbb{R}[x_1, \cdots, x_n]$ is Noetherian.

Example.
Gaussian integers, $\mathbb{Z}[i]$ – principal ideals domain.
More generally, the ring of integers in imaginary quadratic fields $Q(\sqrt{-p})$, $p \in \{1, 2, 3, 7, 11, 19, 43, 67, 163\}$.
All ideals in these rings are principal.
Real quadratic field, $Q(\sqrt{p})$. It is much more complicated than imaginary quadratic field. For example, the norm of quadratic integer $\alpha + \beta\sqrt{d}$ is $\alpha^2 - d\beta^2$. There are finitely many integers with a given norm in imaginary quadratic fields (since d is negative), but there are infinitely many in real quadratic fields. This fact accounts for many difficulties.

Modular arithmetic

Def. An equivalence relation $\approx$ on a set A should satisfy:
(a)$\forall a \in A, a \approx a$. (reflexivity)
(b)$\forall a, b \in A, a \approx b \equiv b \approx a$. (symmetry)
(c)$\forall a, b, c \in A, a \approx b \wedge b \approx c \Rightarrow a \approx c$. (transitivity)
The equivalence class of a, $[a]_\approx = \{b |\ a \approx b\}$.

Remark. Compare with axioms of a poset $(A, \preceq)$:
(a)$\forall a \in A, a \preceq a$. (reflexivity)
(b)$\forall a, b \in A, a \preceq b \wedge b \preceq a \Rightarrow a = b$. (antisymmetry)
(c)$\forall a, b, c \in A, a \preceq b \wedge b \preceq c \Rightarrow a \preceq c$. (transitivity)

Theorem. *If $\approx$ is an equivalence relation on S, then $\{[a]_\approx\}_{a \in S}$ makes a partition of S that: (a) covers S; (b) every two equivalence classes either are disjoint or coincide.*

Proof. Take two equivalence classes, $[a]_\approx, [b]_\approx$. If $[a]_\approx \cap [b]_\approx \neq \emptyset$, suppose $c \in [a]_\approx \cap [b]_\approx$. Then $a \approx c$ and $b \approx c$. Then $c \approx b$ by symmetry. Thus $a \approx b$ by transitivity. □

Let R be a commutative ring, and I be an ideal of R. Denote the factor ring as $R \setminus I$.

Def. $a \approx_I b$ iff $b - a \in I$.

Exercise. Prove that $\approx_I$ is an equivalence relation.

Def. $R \setminus I$ is defined as the set of all equivalence classes of $\approx_I$.

Def.
$[a]_I + [b]_I := [a + b]_I$
$[a]_I \cdot [b]_I := [a \cdot b]_I$

Exercise. Check that this definition is the same as Minkowsky's sum and product that is, say, $[a + b]_I = \{x + y | x \in [a],\ y \in [b]\}$.

To check the well-definedness of the multiplicative operation, we need to check that $(a \approx_I a') \wedge (b \approx_I b') \Rightarrow ab \approx_I a'b'$.
Suppose $a' = a + x, x \in I$; $b' = b + y, y \in I$. Then $a'b' = ab + (xb + xy + ay)$ where $(xb + xy + ay) \in I$, thus $ab \approx_I a'b'$.

Example. $R = \mathbb{Z}$. The residue ring $\mod n$ is defined as $\mathbb{Z}_n = \mathbb{Z} \setminus n\mathbb{Z} = \{0, 1, \cdots, n - 1\}$.

Multiplicative group

Def. The multiplicative group $R^* \subseteq R$ is defined as $R^* = \{a \in R \mid \exists x(ax = 1)\}$

Example. $\mathbb{Z}^* = \{1, -1\}$, $\mathbb{Z}[i]^* = \{1, -1, i, -i\}$.
$\mathbb{Z}_n^* = \{a \in \mathbb{Z}_n \mid \exists x(ax = 1 \mod n)\}$.

Theorem. $a \in \mathbb{Z}_n^*$ *iff* $\gcd(a, n) = 1$.

Def. Euler function $\phi(n)$ is the number of $a \in \{1, \ldots, n-1\}$ s.t. $\gcd(a, n) = 1$.

Fact. $\phi(n) = |\mathbb{Z}_n^*|$.

Theorem. *Lagrange's Theorem.*
If G is any finite commutative group of order f, then for any $g \in G$, $g^f = 1$.

Take $G = \mathbb{Z}_n^*$, then $f = \phi(n)$. We have

Theorem. *Euler's theorem.*
If n and a are coprime positive integers (in other words, $a \in \mathbb{Z}_n^*$), then $a^{\phi(n)} \equiv 1 \mod n$.

If p is a prime, $\phi(p) = (p - 1)$.

Theorem. *Fermat's little theorem.*
If p is a prime and $a \in \mathbb{Z}_p^*$, then $a^{p-1} = 1 \mod p$.

Chinese Remainder Theorem

Theorem. *Let $m_1, m_2, \ldots, m_n$ be s.t. $\gcd(m_i, m_j) = 1$ for all $1 \leq i, j \leq n$, $i \neq j$. Let $m = m_1 m_2 \cdots m_n$, then the system of modular equations*

$$\begin{cases} x \equiv a_1 \mod m_1 \\ x \equiv a_2 \mod m_2 \\ \cdots \\ x \equiv a_n \mod m_n \end{cases}$$

has a solution, and this solution is unique mod m.

Lemma. $\gcd(m_{i_1} m_{i_2} \cdots m_{i_n}, m_j) = 1$ $(j \neq i_1, i_2, \ldots, i_n)$ if m_{i_z}, m_j are pairwise coprime for all z.

Proof. If $\gcd(m_{i_1} m_{i_2} \cdots m_{i_n}, m_j) = d > 1$, then there exists a prime $p|d$ s.t. $p|m_{i_1} m_{i_2} \cdots m_{i_n}$. Then $p|m_{i_z}$ (according to this lemma) while $p|m_j$, which contradicts that $\gcd(m_{i_z}, m_j) = 1$. □

Lemma. Assume that $\gcd(m_1, m_2) = 1, m_1|x, m_2|x$. Then $m_1 m_2|X$.

Proof. $\exists a, b \in \mathbb{Z}$ s.t. $am_1 + bm_2 = 1$ since $\gcd(m_1, m_2) = 1$. Then $axm_1 + bxm_2 = x$. And $\exists c, d \in \mathbb{Z}$ s.t. $x = m_2 c = m_1 d$. Thus $acm_1 m_2 + bdm_1 m_2 = x$, and $m_1 m_2|x$. □

Proof. (Of Chinese Remainder Theorem.)
Uniqueness
Suppose x, y are two solutions, then $m_1, m_2, \ldots, m_n|x-y$. Then we can prove that $m_1 m_2 \cdots m_n|x-y$ by induction on n. Namely, $m_1 m_2|x - y$ by the second lemma. Suppose $m_1 m_2 \cdots m_k|x - y$. $\gcd(m_1 m_2 \cdots m_k, m_{k+1}) = 1$ by the first lemma. Then $(m_1 m_2 \cdots m_k) m_{k+1}|x - y$ by the second

lemma. Thus $m_1 m_2 \cdots m_n | x - y$, and the solution is unique mod $m = m_1 m_2 \cdots m_n$.
Existence
We prove this by constructing a solution.
Define $M_k = \frac{m}{m_k}$ for $1 \leq k \leq n$. Then $\gcd(M_k, m_k) = 1$ by the first lemma. For each k, find x_k such that $x_k M_k = 1 \mod m_k$. Then $a_k x_k M_k = a_k \mod m_k$, and $x_k M_k = 0 \mod m_i (1 \leq i \leq n, i \neq k)$. Thus $\sum_{k=1}^{n} a_k x_k M_k$ is a valid solution. □

Remark. Chinese Remainder Theorem implies that there is a bijection between $\mathbb{Z}_m$ and $\mathbb{Z}_{m_1} \times \cdots \times \mathbb{Z}_{m_k}$.[12] Moreover, this bijection is a *ring isomorphism*, that is it respects addition and multiplication.

III Counting

Product rule. If a procedure can be broken into n independent tasks $T_1, T_2, \ldots, T_n$, and they can be performed in $m_1, m_2, \ldots, m_n$ different ways, then there are $m_1 \cdot m_2 \cdot \cdots \cdot m_n$ ways to perform all tasks.
Sum rule. Assume that $X = X_1 \dot{\cup} X_2 \dot{\cup} \cdots \dot{\cup} X_n$, then $|X| = |X_1| + |X_2| + \cdots + |X_n|$.

Example. The number of binary strings of length n, $\{0,1\}^n$ is 2^n.

Example.
$\ulcorner [n] := \{1, 2, \ldots, n\}$, then $|\mathcal{P}(n)| = 2^n$.
Can be proved by showing that there is a one-to-one correspondence between $\mathcal{P}(n)$ and $\{0,1\}^n$, or by induction. Here is a proof by induction.

Proof. $|\mathcal{P}(\emptyset)| = |\{\emptyset\}| = 1 = 2^0$. Suppose $|\mathcal{P}([n-1])| = 2^{n-1}$. $\mathcal{P}([n]) = X \dot{\cup} Y$, where X is the set of subsets without n and Y is the set of subsets with n. Then $\forall A \in Y$ there is a corresponding set $(A - \{n\}) \in X$ and $\forall B \in X$ there is a corresponding set $(B \cup \{n\}) \in Y$. Thus there is a one-to-one correspondence between X and Y, and $|\mathcal{P}([n])| = |X| + |Y| = 2|X| = 2|\mathcal{P}([n-1])| = 2^n$. $\lrcorner$ □

Functions

Def. $f : X \longrightarrow Y$
f is injective iff for any $x \neq x' \in X$, $f(x) \neq f(x')$.
f is surjective(onto) iff $\forall y \in Y, \exists x \in X$ s.t. $f(x) = y$.
f is bijective (one-to-one correspondence) iff it is both injective and surjectve.
$f^{-1} : Y \longrightarrow X$ is an inverse for f iff $f^{-1} \circ f = \text{id}$, $f \circ f^{-1} = \text{id}$.

Theorem. *f is bijective iff it has an inverse.*

Exercise. Prove the above theorem.

Theorem. *Let X, Y be two finite sets.*
An injective function $f : X \longrightarrow Y$ exists iff $|X| \leq |Y|$.
A surjective function $f : X \longrightarrow Y$ exists iff $|X| \geq |Y|$.
A bijective function $f : X \longrightarrow Y$ exists iff $|X| = |Y|$.

Example. To prove $|\{0,1\}^n| = |\mathcal{P}([n])|$, we can show that there is a bijection between $\{0,1\}^n$ and $\mathcal{P}([n])$. For example, $f(a_1 a_2 \cdots a_n) = \{i \mid a_i = 1\}$.

[1] Furthermore, the cardinality of two sides is equal so we only need to prove one directon. (Prof, after class)

[2] Related: There is also a bijection between $(\mathbb{Z}_{m \cdot n})^*$ and $Z_m^* \cdot Z_n^*$ when m, n are coprime. The formal way of finding pairs of correspondence is to fix $xx' = 1 \mod m$, $yy' = 1 \mod n$. Find $a \in \mathbb{Z}_{m \cdot n}$ that corresponds to (x, y) and $b \in \mathbb{Z}_{m \cdot n}$ that corresponds to (x', y') (by Chinese Remainder Theorem). Then $a \cdot b = 1 \mod m \cdot n$. (Prof, after class)

Counting functions

$|X| = m$, $|Y| = n$. $\quad f : X \longrightarrow Y$.
The number of all functions is n^m. (Every element in X has n choices.)
The number of injections is $n(n-1)\cdots(n-m+1) = \frac{n!}{(n-m)!} = P(n,m)$. (Every element in X can be mapped to one unchosen element in Y.)
Generalized product rule
We have m tasks $T_1, \ldots, T_m$ s.t. if we perform them in this order, we have $n_k \geq 0$ choices for T_k, regardless of the history. Then the number of ways to do these tasks is still $n_1 n_2 \cdots n_m$.[1]

Binomial coefficients

Def. X is a finite set. $\binom{X}{m}$ is the family of all m-element subsets of X.
$\binom{n}{m} := |\binom{[n]}{m}|$.[2]

Combinatorial proofs (bijective proofs)[3]

Def. $f : X \longrightarrow Y$ is k-to-1 correspondence if every element $y \in Y$ has exactly k preimages. Then $|X| = k|Y|$.

Theorem. $\binom{n}{m} = \frac{n!}{(n-m)!m!}$

Proof. $X := \{\text{injective functions } f : [m] \longrightarrow [n]\}$, $Y = \binom{[n]}{m}$.
There is a k-to-one correspondence between X and Y, $F(f) = \text{im}(f)$, where $k = m!$ (permutation of the m elements in the image). Thus $\frac{n!}{(n-m)!} = m!\binom{n}{m}$. □

Note. $\binom{n}{0} = 1$. $\quad \binom{n}{m}_{m>n} = 0$

Theorem. *Vandermonde's identity.*
$\binom{m+n}{r} = \sum_{k=0}^{r} \binom{m}{r-k}\binom{n}{k}$.

(To choose r elements in total from $m+n$ elements, we fix two sets with m and n elements respectively, and choose k elements from the m elements then choose $r-k$ elements from the n elements. Sum over all possible choices of k.)

When $n = 1$,

Theorem. *Pascal's identity*
$\binom{m+1}{r} = \binom{m}{r} + \binom{m}{r-1}$

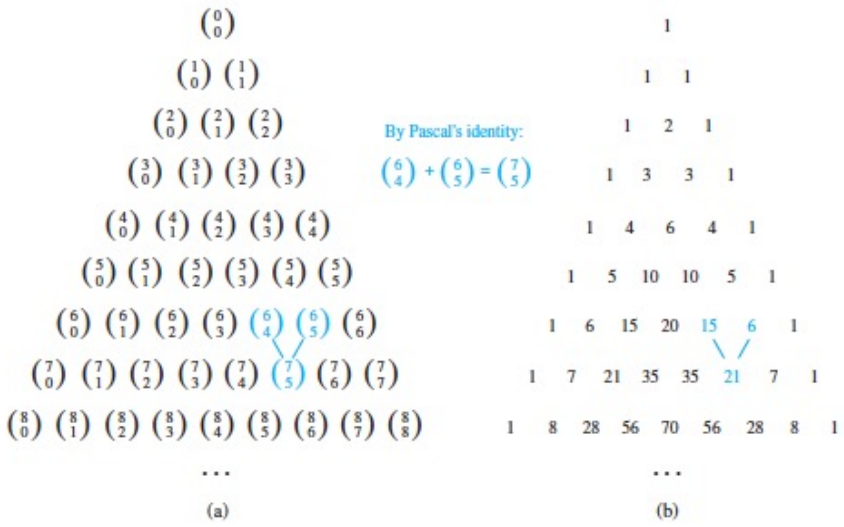
Pascal's Triangle.[4]

[1] Difference from product rule: product rule requires the tasks to be independent, while generalized product rule does not. In each stage the set of choices can be different as long as the number of choices stays the same.

[2] The difference between two notations: one concerns a set and an integer and the other concerns two integers.

[3] "Show two ways of counting the same thing."

[4] Source: Rosen 6.4 Figure 1, pp.419.



Theorem. $\sum_{k=0}^{n} \binom{n}{k} = 2^n$

Proof. We show three ways of proving this theorem.

(1) Induction on Pascal's triangle. Suppose $\sum_{k=0}^{n} \binom{n}{k} = 2^n$. In Pascal's triangle on the next row, every element is summed twice according to Pascal's idnetity. Thus $\sum_{k=0}^{n+1} \binom{n+1}{k} = 2 \cdot 2^n = 2^{n+1}$.

(2) Combinatorial proof

Recall $\binom{n}{k} = |\binom{[n]}{k}|$.

$\sum_{k=0}^{n} \binom{n}{k} = \sum_{k=0}^{n} |\binom{[n]}{k}| = |\binom{[n]}{0} \cup \binom{[n]}{1} \cup \cdots \cup \binom{[n]}{n}| = |\mathcal{P}([n])| = 2^n$.

(3)

Theorem. *Binomial theorem*

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

Remark. This theorem holds in any commutative ring.

Set $x = 1,\ y = 1$, then $2^n = \sum_{k=0}^{n} \binom{n}{k}$. □

Remark. Set $x = 1,\ y = -1$, then $0 = \sum_{k=0}^{n} (-1)^k \binom{n}{k}$.

When m is odd, this is obvious since the terms cancel out.

Exercise. Prove this combinatorially when m is even. Hint: use Pascal's Triangle.

Inclusion-exclusion principle

Theorem. $|A_1 \cup A_2 \cup \cdots \cup A_n| = \sum_{1 \le i \le n} |A_i| - \sum_{1 \le i \le j \le n} |A_i \cap A_j| + \cdots + (-1)^{n-1} |A_1 \cap A_2 \cap \cdots \cap A_n|$

Proof. Fix element x. Let $r = |\{i \mid x \in A_i\}|$
If $r = 0$, x does not appear.
If $r \geq 1$, we need to prove that $1 = \binom{r}{1} - \binom{r}{2} + \cdots + (-1)^{r-1}\binom{r}{r} \Leftrightarrow 0 = \sum_{k=0}^{n}(-1)^k\binom{n}{k}$, which is the above remark. □

Count surjective functions

$f : [m] \longrightarrow [n]$
The number of surjective functions = $n^m -$ the number of non-surjective functions($\exists i \in [n]$ s.t. $i \notin \text{im}(f)$).
Define $A_i = \{f \mid i \notin \text{im}(f)\}$. The number of non-surjective functions $|A_1 \cup A_2 \cup \cdots \cup A_n| = \binom{n}{1}(n-1)^m - \binom{n}{2}(n-2)^m + \cdots + (-1)^{n-2}\binom{n}{n-1}1^m = \sum_{k=1}^{n-1}(-1)^{k-1}\binom{n}{k}(n-k)^m$.

Note. $1 \leq k \leq n-1$ because at most $n-1$ elements in $[n]$ are not mapped to.

The number of surjective functions is $n^m - \sum_{k=1}^{n-1}\binom{n}{k}(n-k)^m = \sum_{k=0}^{n-1}(-1)^k\binom{n}{k}(n-k)^m$.

Stirling number of the second kind

Def. $S(m,n)$ is the number of all equivalence relations on $[m]$ that have precisely n classes.
$S(m,n) = \frac{1}{n!}\sum_{k=0}^{n-1}(-1)^k\binom{n}{k}(n-k)^m$.

Proof. For a surjective function $f : [m] \longrightarrow [n]$, define the corresponding equivalence relation as $x \approx y$ iff $f(x) = f(y)$. This is a $n!$-to-1 correspondence (permutation of the image) between surjective functions and equivalence relations with n classes on $[m]$. □

Theorem. $S(m,n) = S(m-1,n-1) + nS(m-1,n)$

Proof. Consider equivalence relations with n classes on $[m]$. Fix an arbitrary element x. The equivalence relations can be divided into two kinds: (1) $[x]_\approx = \{x\}$ (x is the only element of a class); (2) $[x]_\approx \neq \{x\}$ (x is not the only element of a class). The first kind includes $S(m-1,n-1)$ equivalence relations, while the second kind includes $nS(m-1,n)$ equivalence relations (exclude x first and, for the opposite direction, add it to any existing class). □

Ball Game

m balls, n boxes.
1. Distinguishable balls and boxes
(a) no restrictions: n^m (The number of all functions).
(b) all boxes are non-empty: $n!S(m,n)$ (The number of surjective functions).
(c) boxes with capacities $m_1, m_2, \ldots, m_n$ where $\sum_{i=1}^{n} m_i = m$:

Theorem. *Multinomial coefficient,* $\binom{m}{m_1 m_2 \cdots m_n} = \frac{m!}{m_1!m_2!\cdots m_n!}$.

Proof. There is a $m_1!m_2!\cdots m_n!$-to-one correspondence between all permutations of m elements and the number of ways of distributing balls. (Given a permutation on m elements, we can consider the first m_1 elements to be in the first box, the next m_2 elements to be in the second box, and so on. The order of the elements in one box does not matter.) □

2. Distinguishable balls and indistinguishable boxes
(a) all boxes are non-empty: $S(m, n)$.
(b) no restrictions: $\sum_{j=1}^{n} S(m, j)$.
3. Indistinguishable balls and distinguishable boxes (the number of ordered partitions $m_1 + m_2 + \cdots + m_n$ on m elements.)
a) $m_i \geq 1$: $\binom{m-1}{n-1}$. (Put $n-1$ separating bars into $m-1$ positions.)
b) $m_i \geq 0$:
It's the same as the number of partitions of $m_1 + m_2 + \cdots + m_n$ on $m+n$ elements where $m_i \geq 1$, because there is a one-to-one correspondence between them. (Put one ball into each of the n boxes first.)

Theorem. *The number of ordered partitions* $m_1 + m_2 + \cdots + m_n = m,\ m_i \geq 0$ *is* $\binom{m+n-1}{n-1}$.

4. Indistinguishable balls and indistinguishable boxes
Let $m_1 + m_2 + \cdots + m_n = m,\ m_1 \geq m_2 \geq \cdots \geq m_n \geq 0$.
The number of partitions on m, $P(m)$ is defined as $P^{\infty}(m)$ (number of partitions into as many boxes as we like.)
Young tableau/diagram.
The height is the number of boxes. The width is the number of balls in the box with the most balls (m_1).

Theorem. $P_n(m) = P^n(m)$
The number of partitions with each box having at most n balls is equal to the number of partitions with n boxes.

by symmetry of the diagram.

Pigeonhole principle

The contrapositive form of the criterium for the existence of injective functions.

Theorem. *For two sets* $|X| > |Y|,\ (f : X \longrightarrow Y) \Rightarrow (\exists x_1 \neq x_2, f(x_1) = f(x_2))$.

Theorem. *Generalized Pigeonhole Principle.*
Asssume that n pigeons fly to k holes, then at least one hole has $\lceil \frac{n}{k} \rceil$ pigeons.

Example.
⌜ In a finite poset $(S, \preceq)$, a chain is defined as a set of mutually comparable elements, while an anti-chain is a set of mutually incomparable elements.
The height of $S,\ h(S)$, is the size of the biggest chain. The width of $S,\ w(S)$, is the maximal size of an anti-chain.

Theorem. $|S| \leq h(S)w(S)$.

Proof.

Theorem. *Mirsky's theorem.*
There exists a decomposition of S into h anti-chains, $S = A_1 \dot{\cup} \cdots \dot{\cup} A_h$.

Proof. Define the height of an element, $h(x)$, as the size of the longest chain in which x is the maximal element. Then $h(x) \in \{1, 2, \ldots, h\}$. Let the decomposition be the sets $\{x \mid h(x) = t\},\ 1 \leq t \leq h$. To show that each set is an anti-chain, first assume that $\exists x, y$ with the same height t that are comparable. Assume $x \preceq y$, then $h(y) \geq t + 1$, which contradicts that $h(y) = t$. □

We can prove the above theorem by Mirsky's theorem using Pigeonhole principle. $w(S) \geq \frac{|S|}{h(S)}$. □

Here is a dual theorem that could be used. (More complicated to prove and not proved in class.)

Theorem. *Dilworth's theorem.*
There exists a decomposition $S = C_1 \dot\cup \cdots \dot\cup C_w$ into w chains.

Theorem. *Erdös – Szekeres Theorem*
Assume that $a_1, a_2, \ldots a_{n^2+1}$ are distinct real numbers. Then this sequence contains an either strictly increasing sequence $a_{i_1} < a_{i_2} < \cdots < a_{i_n}$ or strictly decreasing sequence $a_{i_1} > a_{i_2} \ldots > a_{i_n}$.i

Proof. There are two natural orderings on $S = [n^2 + 1]$: (1) By the $\leq$ order of real numbers, $1 < 2 < \ldots < n^2 + 1$. (2) By their corresponding elements in the sequence, $i \preceq j$ iff $a_i \leq a_j$.
Take the Cartesian product of these two orders: we have $i \sqsubseteq j$ iff $i \leq j$ and $a_i \leq a_j$.

Remark. This is a not a linear ordering, unlike lexicographic order.

Apply the theorem we just proved to $(S, \sqsubseteq)$, then $(n^2+1) \leq h(S)w(S)$. Thus $(h \geq n+1) \vee (w \geq n+1)$.
Now by the definition of the ordering, a chain in S is precisely an increasing sequence and an anti-chain a decreasing sequence. (And they are strict since the elements are distinct.) 」 □

Digression

「 *approximation of irrational numbers*
For an irrational number α, for any $q \in \mathbb{Z}$, $\exists n$ s.t. $|\alpha - \frac{n}{q}| \leq \frac{1}{2q}$. (Imagine the real axis partitioned into intervals of length $\frac{1}{q}$, then α falls into one interval, and is of distance at most $\frac{1}{2q}$ from one of the endpoints.)

Theorem. *Dirichlet's approximation theorem.*
For any $n \in \mathbb{Z}$ and any irrational α, $\exists q \leq n$, $p \in \mathbb{Z}$ s.t. $|\alpha - \frac{p}{q}| \leq \frac{1}{nq} \leq \frac{1}{q^2}$.

Proof. (For any $x \in \mathbb{R}$, define $\{x\} \in [0, 1)$ s.t. $(x - \{x\})$ is an integer. $(x \equiv \{x\} \mod 1.)$)
Divide the unit circle into n intervals of $\frac{1}{n}$. Put onto the circle $n+1$ points, $\{\alpha\}\cdot 0, \{\alpha\}\cdot 1, \ldots, \{\alpha\}\cdot n$. Then at least 2 points end up in the same interval by pigeonhole principle. Thus $\exists a, b \in \mathbb{N}$ s.t. $\{\alpha(a - b)\} \leq \frac{1}{n}$; take $a - b$ as q. Then $\exists p \in \mathbb{Z}$ s.t. $|q\alpha - p| \leq \frac{1}{n} \Leftrightarrow |\alpha - \frac{p}{q}| \leq \frac{1}{nq}$. Then $|\alpha - \frac{p}{q}| \leq \frac{1}{q^2}$ since $q \leq n$. □

Def. x is algebraic iff $x^d + a_1 x^{d-1} + \cdots + a_{d-1}x + a_d = 0$ for some $a_i \in \mathbb{Q}$.

Example. $\sqrt{5}$ is algebraic since it is a solution for $x^2 - 5 = 0$.

Exercise. Prove that $\sqrt{5}$ cannot be approximated very well by rational numbers. (For any choice of $p, q \in Z, \exists \epsilon$ s.t. $|\sqrt{5} - \frac{p}{q}| \geq \frac{\epsilon}{q^2}$.)

Theorem. *Thue's Theorem (1909).*
For any algebraic number α, for certain ϵ, $|\alpha - \frac{p}{q}| \geq \frac{\epsilon}{q^{\frac{d}{2}+1+\delta}}$ (with $\delta \to 0$).

Improvement: Roth's Theorem(1955)[1].

[1] https://en.wikipedia.org/wiki/Roth%27s_theorem

Def. Liouville number, $\sum_{k=1}^{\infty} 10^{-k!}$.

It is not algebraic since its difference with a rational number $q = \sum_{k=1}^{n} 10^{-k!}$ can be extremely small. 」

IV Discrete Probability

S is a finite probability space (sample space). (Intuition: the set of all possible outcomes.)

Def. A distribution over S is a function $P : S \longrightarrow \mathbb{R}$ (intuition: $P(x)$ is the chance to see x) satisfying
(a) $\forall x \in S, P(x) \geq 0$.
(b) $\sum_{x \in S} P(x) = 1$.

Def. An event E is a subset of S.
$P(E) = \sum_{x \in E} P(x)$.

Def. Uniform distribution.
$P(x) = \frac{1}{|S|}, \forall x \in S$.

$P(E) = \frac{|E|}{|S|}$.

Example. $S = \mathcal{P}([n])$. Define event E_k as $|A| = k,\ A \in S$. Then $P(E_k) = \frac{\binom{n}{k}}{2^n}$.

Def. Events E and F are mutually disjoint if $E \wedge F = \emptyset$

Note. We will use the logical notation $E \vee F$, $E \wedge F$, $\overline{E}$ to denote the union, intersection, complement of events.

Sum rule.
$P(E_1 \vee E_2 \vee \cdots \vee E_n) = \sum_{i=1}^{n} P(E_i)$ as long as they are mutually disjoint.
Inclusion-exclusion principle.
$|E_1 \vee E_2 \vee \cdots \vee E_n| = \sum_{1 \leq i \leq n} |E_i| - \sum_{1 \leq i \leq j \leq n} |E_i \wedge E_j| + \cdots + (-1)^{n-1} |E_1 \wedge E_2 \wedge \cdots \wedge E_n|$.

Independence

Def. Conditional probability.
$E, F \subseteq S, P(F) > 0$. $P(E|F) = \frac{P(E \wedge F)}{P(F)}$

Def. E is independent from F if $P(E)P(F) = P(E \wedge F)$

$E,\ F$ are independent $\Leftrightarrow P(E|F) = P(E)$.

Note. Notice the symmetry of E and F. $P(E|F) = P(E) \Leftrightarrow P(F|E) = P(F)$.

Example. Two fair cubical dies rolled as i and j. $P(i + j \geq 9) = \frac{5}{18}$. While $P(i + j \geq 9 \,|\, i = 4) = P(j \geq 5 \,|\, i = 4) = P(j \geq 5) = \frac{1}{3}$.

Def. $E_1, \ldots, E_n$ are mutually independent if for any $1 \leq i_1 < \cdots < i_z \leq n, P(E_{i_1} \wedge E_{i_2} \wedge \cdots \wedge E_{i_z}) = P(E_{i_1})P(E_{i_2}) \cdots P(E_{i_z})$.

Theorem. *Let $E_1, \ldots, E_n$ be mutually independent. $I, J \subseteq [n], I \cap J = \emptyset$. Let $F = \bigcup_{i \in I} E_i$ and $G = \bigcup_{j \in J} E_j$ be arbitrary events of this form. Then F and G are independent.*

Exercise. Prove the above theorem.

Theorem. *Formula/Law of complete/total/full probability.*
If $X_1, X_2, \ldots, X_n$ form a partition of the sample space S, then for an event E, $P(E) = \sum_{i=1}^{n} (E \wedge X_i) = \sum_{i=1}^{n} P(E|X_i)P(X_i)$.
$P(E|F) = \sum_{i=1}^{n} P(E|(X_i \wedge F))P(X_i|F)$.

Example. X_i are independent Bernoulli trials of probability $\frac{1}{2}$. Then $P(x_1 + x_2 + x_3 = 1 \mid x_3 + x_4 + x_5 = 1) = P(x_1 + x_2 + x_3 = 1 \mid x_3 = 1 \wedge (x_3 + x_4 + x_5 = 1))P(x_3 = 1 \mid x_3 + x_4 + x_5 = 1) + P(x_1 + x_2 + x_3 = 1 \mid x_3 = 0 \wedge (x_3 + x_4 + x_5 = 1))P(x_3 = 0 \mid x_3 + x_4 + x_5 = 1)$.

Theorem. *Bayesian Law.*
$P(A|B) = P(B|A)\frac{P(A)}{P(B)}$.

Since $P(A|B)P(B) = P(A \wedge B) = P(B|A)P(A)$.

Example. In the above example, we can compute $P(x_3 = 1 | x_3 + x_4 + x_5 = 1)$ by $P(x_3 + x_4 + x_5 = 1 \mid x_3 = 1)\frac{P(x_3=1)}{P(x_3+x_4+x_5=1)}$.

Random variables

Def. A random variable is a function $X : S \longrightarrow \mathbb{R}$.

Remark. It is a special case of push forward distribution $f : S \longrightarrow T$.

$$\begin{array}{ccc} S & \xrightarrow{f} & T \\ \downarrow{\scriptstyle P} & \swarrow_{(f_* P)} & \\ \mathbb{R}_{\geq 0} & & \end{array}$$

Def. $(f_* P)(t) = P(f^{-1}(t)) = \sum_{\{s \in S | f(s) = t\}} P(s)$.

In the case of $X : S \longrightarrow \mathbb{R}$, $P(X = r) = \sum_{\{s | X(s) = r\}} P(s)$.

Example. $E \subseteq S$. The random variable given by the indicator function of E,

$$\mathbb{1}_{E(s)} = \begin{cases} 1, \text{ if } s \in E \\ 0, \text{ otherwise.} \end{cases}$$

Discrete probability distributions

Bernoulli trials
$P(x = 1) = p$, $P(x = 0) = q$, $p + q = 1$.

Denote the number of heads in n trials with probability p of getting head on each trial as $B_{n,p}$.
$P(B_{n,p} = k) = p^k q^{n-k}\binom{n}{k}$.

Poission distribution

The number of Bernoulli trials $n \to \infty$, then the expected number of success $\lambda = pn$.
$p = \frac{\lambda}{n},\ q = 1 - \frac{\lambda}{n}$.
$P(B_{n,p} = k) = \frac{\lambda^k}{n^k}(1 - \frac{\lambda}{n})^{n-k}\frac{n(n-1)\cdots(n-k+1)}{k!}$. $(*)$
$\lim\limits_{n\to\infty}(*) = \frac{\lambda^k}{k!}\lim\limits_{n\to\infty}\frac{n(n-1)\cdots(n-k+1)}{n^k}(1 - \frac{\lambda}{k})^{n-k} = \frac{\lambda^k}{k!}e^{-\lambda}$.

Geometric distribution

Bernoulli trials. Denote the first occurrence of head as X_p.
$P(X_P = k) = pq^{k-1} = p(1-p)^{k-1}$.
Sanity check: $\sum\limits_{k\in\mathbb{N}_+}(1-p)^{k-1} = \frac{1}{1-(1-p)} = \frac{1}{p}$.[1]

Expectations

Def. For $X : S \longrightarrow \mathbb{R},\ E(X) := \sum\limits_{s\in S} P(s)X(s)$.

Theorem. $E(X) = \sum\limits_{r\in R} P(X = r)\cdot r$.

Proof. $\sum\limits_r P(X = r)r = \sum\limits_r \sum\limits_{\{s|X(s)=r\}} P(s)r = \sum\limits_s \sum\limits_{\{r|r=X(s)\}} P(s)r = \sum\limits_s P(s)X(s)$. □

Def. The sum of random variables, $X_1, X_2, \ldots, X_n : S \longrightarrow \mathbb{R}$, defined on the same sample space S.
$(X_1 + \cdots + X_n)(s) := X_1(s) + \cdots + X_n(s)$.

Theorem.
$E(X_1 + X_2 + \cdots + X_n) = E(x_1) + E(X_2) + \cdots + E(x_n)$.
$E(\alpha X) = \alpha E(X)$.

Remark. This does not require that the random variables are independent.

Example. Two ways of calculating $E(B_{n,p})$.
(1) $E(B_{n,p}) = \sum\limits_{k=0}^{n} k\cdot P(B_{n,p} = k) = \sum\limits_{k=0}^{n} k\cdot p^k(1-p)^{n-k}\binom{n}{k}$.
(2) $E(B_{n,p}) = E(B_{1,p}) + E(B_{1,p}) + \cdots + E(B_{1,p}) = pn$.

Conditional expectation

$E(X|F) = \frac{\sum\limits_{s\in F} P(s)X(s)}{P(F)} = \frac{\sum\limits_{s\in F} P(s)X(s)}{\sum\limits_{s\in F} P(s)}$.
If $X = \mathbb{1}_G$ then $E(X|F) = p(G|F)$.

Example.
The event space is $f : [m] \longrightarrow [n]$. We are interested in the expectation of $|\text{im}(f)|$. (f is surjective $\Leftrightarrow |\text{im}(f)| = n$.)
Define X_i as the characteristic function of $i \in \text{im}(f)$ ($X_i = 1 \Leftrightarrow i \in \text{im}(f)$).
$E(|\text{im}(f)|) = E(X_1) + \cdots + E(X_n) = nE(X_1)$ by symmetry.
$E(X_1) = P(X_1 = 1) = 1 - P(1 \notin \text{im}(f))$.
$P(1 \notin \text{im}(f)) = P(f(1) \neq 1 \wedge \cdots \wedge f(m) \neq 1) = P(f(1) \neq 1)^m = (1 - \frac{1}{n})^m$.
$E(|\text{im}(f)|) = n(1 - (1 - \frac{1}{n})^m)$.

[1] There are some useful summations in Rosen 2.4, Table 2, pp.166.

Def. Random variables $X, Y : S \longrightarrow \mathbb{R}$ are independent if $\forall r, s \in \mathbb{R}, P(X = r \wedge Y = s) = P(X = r) \cdot P(Y = s)$.

Theorem. *If X and Y are independent, then any two events of the form $A(X)$[1], $B(Y)$ are independent.*

Sharp concentration inequalities

Theorem. *Markov's inequality.*
Let $X \geq 0$, then $P(X \geq a) \leq \frac{E(X)}{a}$.

Proof. $E(X) = E(X|X \geq a)P(X \geq a) + E(X|x < a)P(x < a)$. The second term is at least 0, the first term is at least $aP(X \geq a)$. Thus $E(X) \geq aP(X \geq a)$. □

Variance

Denote $E(X)$ as c.

Def. Mean deviation of a random variable, $E(|X - c|)$.

Def. Variance of a random variable.
$\text{Var}(X) = E((X - c)^2) = E(X^2) - 2cE(X) + c^2 = E(X^2) - E(X)^2$.

Theorem.
Cauchy-Schwartz.
$E(X^2) \geq E(X)^2$
Jensen's inequality.
$\forall p \geq 1, E(X^p) \geq E(X)^p$. *(Holds for any convex function ϕ: $\phi(E(X)) \leq E(\phi(X))$.)*
Hölder's inequality.
$E(X^p)^{\frac{1}{p}} \geq E(X^q)^{\frac{1}{q}}$ *for any* $p \geq q > 0,\ X \geq 0$. [2]

Theorem. *If random variables X and Y are independent, then Var$(X + Y) = Var(X) +$ Var(Y).*

Proof.
$\text{Var}(X + Y) = E(X^2) + 2E(XY) + E(Y^2) - (E(X)^2 + 2E(X)E(Y) + E(Y)^2) = \text{Var}(X) + \text{Var}(Y) + \text{cov}(X, Y)$,
where $\text{cov}(X, Y) = E(XY) - E(X)E(Y)$.
When X, Y are independent, $\text{cov}(X, Y) = 0$. [3] □

Theorem. *Chebyshev's inequality.*
$P(|X - c| \geq r) \leq \frac{Var(X)}{r^2}$.

Proof. $P(|X - c| \geq r) = P((X - c)^2 \geq r^2) \leq \frac{E((X-c)^2)}{r^2}$ (by Markov's inequality) $= \frac{\text{Var}(X)}{r^2}$. □

Example. Bernoulli trials where p is different for X_i.
Let $X = X_1 + \cdots + X_n,\ X_i \in \{0, 1\}$. Let $P(X_i = 1) = p_i$. Then $\text{Var}(X_i) = p_i - p_i^2$, and $c = E(X) = \sum_{i=1}^{n} E(X_i) = \sum_{i=1}^{n} p_i$ And $\text{Var}(X) = \sum_{i=1}^{n} \text{Var}(X_i) = \sum_{i=1}^{n} p_i(1 - p_i)$.

[1] Any event that depends only on the value of X in an arbitrary way.

[2] This is a special case of Hölder's inequality. For the original form and context, see https://en.wikipedia.org/wiki/H%C3%B6lder%27s_inequality#Probability_measure.

[3] According to Theorem 5 in Rosen 7.4, pp.486.

Let $r = \delta c$. Then $P(|x - c| \geq \delta c) \leq \frac{\sum_{i=1}^{n} p_i(1-p_i)}{\delta^2 c^2} = \frac{\sum_{i=1}^{n} p_i(1-p_i)}{\delta^2 (\sum_{i=1}^{n} p_i)^2}$. $(*)$

If we plug in $p_i = \frac{1}{2}$, then $(*) = \frac{1}{\delta^2 n}$. It starts making sense when $\delta \geq \frac{1}{\sqrt{n}}$, so when, say, c is of order n, r should be of order at most $\sqrt{n}$.

Theorem. *Chernoff's bound.*
If $X = X_1 + X_2 + \cdots + X_n$ where X_i are mutually independent, then $\forall \delta > 0, \exists \epsilon > 0$ s.t. $P((X - c) \geq \delta c) \leq e^{-\epsilon c}$.

Proof. (In the above example.)
For a certain a, $P(X \geq c(1+\delta)) = P(a^X \geq a^{c(1+\delta)}) \leq \frac{E(a^X)}{a^{c(1+\delta)}}$ by Markov's inequality.
$E(a^X) = E(a^{X_1+X_2+\cdots+X_n}) = E(a^{X_1} a^{X_2} \cdots a^{X_n}) = E(a^{X_1})E(a^{X_2}) \cdots E(a^{X_n})$ since they are mutually independent. Also $c = \sum p_i$. Thus $\frac{E(a^X)}{a^{c(1+\delta)}} = \prod_{i=1}^{n} (\frac{E(a^{X_i})}{a^{p_i(1+\delta)}})$. $(*)$
$E(a^{X_i}) = p_i a^1 + (1 - p_i)a^0 = p_i a + (1 - p_i)$.
Fix a as $1 + \delta$, then $E(a^{X_i}) = 1 + \delta p_i \leq e^{\delta p_i}$, since $e^x = 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \cdots$.
Then $(*) \leq \prod_{i=1}^{n} \frac{e^{p_i \delta}}{(1+\delta)^{(1+\delta)p_i}} = \prod_{i=1}^{n} (\frac{e^{\delta}}{(1+\delta)^{1+\delta}})^{p_i} = e^{-\epsilon c}$, where $\epsilon = (1 + \delta)\ln(1 + \delta) - \delta$. □

V Graph Theory

Def. A graph G is a pair (V, E) where V is a finite set of vertices and E is a finite set of edges, and for every edge E, we have one or two vertices associated with it called endpoints.

Def. A loop is an edge with one end point. Multiple edges have a same pair of endpoints. A graph without loops or multiple edges is a simple graph.

For a simple graph, $E \subseteq \binom{V}{2}$.

Def. A directed graph is a graph with an arrow on every edge.

Def. Hypergraphs are graphs in which ≥ 2 vertices are associated with one edge. ($r \geq 2$, $E \subseteq \binom{V}{r}$ is a simple r-graph.)

Some important graphs

K_r, clique with r vertices. (r vertices mutually connected.)
I_r, independent set with r vertices. (With no edges.)
P_r, path with r vertices.
C_r, cycle with r vertices.[1]

Degree sequences

Def. Let G be a graph, $v \in V(G)$. $\deg_G(v)$ is the number of edges incident to v. $\delta(G)$ is the minimal degree of a vertex in the graph. $\Delta(G)$ is the maximal degree.

Deonte $|V(G)|$ as n, and $|E(G)|$ (the number of edges) as m. The degree sequence of G is the sequence of the degrees of its vertices, $d_1 \geq d_2 \geq \cdots \geq d_n$, where $0 \leq d_i \leq n-1$. It is an invariant of a graph.

[1] Path and cycle are defined later.

Theorem. *Handshaking Theorem.*
In a simple graph, $d_1 + d_2 + \cdots + d_n = 2m$.

Proof. Define $A = \{(v, e) \mid v \text{ is an endpoint of } e\}$. Then there are two ways of counting $|A|$. Every vertex has number of its degree of pairs, thus $|A| = d_1 + d_2 + \cdots + d_n$. Every edge has two endpoints thus two pairs; $|A| = 2m$. Thus $d_1 + d_2 + \cdots + d_n = 2m$. □

Cor. $d_1 + d_2 + \cdots + d_n$ is even.

Paths and connectivity

Def. A path in a graph G is a sequence of vertices $< v_0, v_1, \ldots, v_z >$ s.t. $(v_{i-1}, v_i) \in E(G)$.

Empty path with only one vertex is allowed.

Def. A simple path is a path without repeated edges.

Def. A vertex v is *reachable* from u if there is a path from u to v.

Remark. Reachability is an equivalence relation. We use arbitrary (not necessarily simple) paths here, but it also holds for simple paths because if two vertices are reachable by any path, then they are also reachable by a simple path (can be proved by induction).

Def. A connected graph is a graph that has only one connected component. [1]

Metric space
Denote $d_G(u, v)$ as the minimum possible length of a path from u to v.
(a) $d_G(u, u) = 0$.
(b) $d_G(u, v) = d_G(v, u)$.
(c) $d_G(u, w) \leq d_G(u, v) + d_G(v, w)$.

For a simple graph G, denote the number of its vertices as n, its maximal degree as Δ, its diameter (the maximal distance between any pair of vertices in it[2]) as $\text{diam}(G)$.

Theorem. *Moore's bound.*
$n(G) \leq 1 + \Delta(G) \cdot \sum_{i=o}^{diam(G)-1} (\Delta(G) - 1)^i$.

Proof. Fix a vertex v, and arrange the vertices into levels according to their distances to v. Level 0 has one vertex v. Level 1 has at most Δ vertices since the degree of v is at most Δ. For $k \geq 1$, every vertex in level k has at most $(\Delta - 1)$ connections to level $k + 1$, since each one is connected to level $k - 1$ by at least one edge. Thus level 2 has at most $\Delta(\Delta - 1)$ vertices; level 3 has at most $\Delta(\Delta - 1)^2$ vertices; ..., level $k + 1$ has at most $\Delta(\Delta - 1)^k$ vertices. There are at most $\text{diam}(G)$ levels in total, thus the total number of vertices is at most $1 + \Delta \cdot \sum_{i=o}^{\text{diam}(G)-1} (\Delta - 1)^i$. □

Def. A graph that satisfies $n(G) = 1 + \Delta(G) \cdot \sum_{i=o}^{\text{diam}(G)-1} (\Delta(G) - 1)^i$ is a Moore graph.

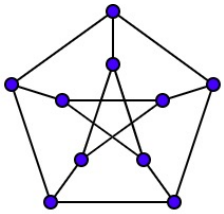
[1] Can refer to Rosen 10.4, pp.682 for definition of connected components.
[2] Can refer to Rosen 10, supplementary exercise 52, pp.741 for the definition of the diameter of a graph and the distance between two vertices.

Def. The girth $g(G)$ of a graph G is the length of the minimal cycle in G.

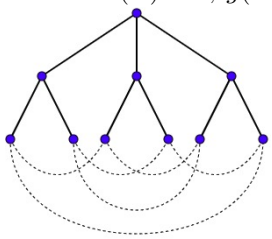
The girth of empty set or a tree is infinite.
An alternative definition of Moore graph by girth,

Def. A regular graph (graph in which the degree of every vertex is the same) G that satisfies $g(G) = 2\text{diam}(G) + 1$ is a Moore graph.

Example. Petersen graph is a Moore graph with $\text{diam}(G) = 2,\ g(G) = 5,\ \Delta(G) = 3$.



Petersen graph.[1]



Petersen graph as a Moore graph.[2]

Adjacency matrices
For a graph G with n vertices, let $V(G) = [n]$ and define a $n \times n$ (0,1) matrix A_G by the edges. $a_{ij} = 1$ iff $(i, j) \in E(G)$.
Example.
⌜ The diagonal of the adjacency matrix of a graph without loops is all 0.

This is the adjacency matrix of a cycle. ⌟

$$\begin{pmatrix} 0 & 1 & 0 & 0 & \cdots & 0 & 1 \\ 1 & 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & 0 & \cdots & 1 & 0 & 1 \\ 1 & 0 & 0 & \cdots & 0 & 1 & 0 \end{pmatrix}$$

Dynamical algorithm – counts the number of paths of length l.
Define $a_{ij}^{(l)} = 1$ to be the number of paths of length l (not necessarily simple) from i to j. Then to count the number of paths of length $l + 1$ from i to j, we can count for every other vertex k the number of paths of length l from i to k, the number of paths of length 1 from k to j, multiply the two results (product rule) and add everything up (sum rule). Thus $a_{ij}^{(l+1)} = \sum_k a_{ik}^{(l)} a_{kj}$.

Digression

⌜ Spectral graph theory.[3] ⌟ (Link back to Notes and references.)

Isomorphism problem

It is still an open problem if there is a poly-time algorithm. Current record (Babai): there is an algorithm of quasi-polynomial time $n^{(\log n)^c}$.
Invariants of a graph
Number of vertices $n(G)$, number of edges $e(G)$, girth $g(G)$, the degree sequence, ...

[1] Source: https://en.wikipedia.org/wiki/Petersen_graph.
[2] Source: https://en.wikipedia.org/wiki/Moore_graph#Bounding_vertices_by_degree_and_diameter.
[3] Notes not included.

Sub-graphs

For two graphs $G = (V, E)$ and $H = (W, F)$,

Def. G is a sub-graph of H iff $V \subseteq W$ and $E \subseteq F$.

Def. G is a spanning sub-graph of H iff $V = W$ and $E \subseteq F$.

Def. G is an induced sub-graph of H iff $E = \{(u, v) \in F \mid u, v \in V\}$.

Extremal graph theory. Use $\text{ex}(n; K_r)$ to denote the maximal number of edges with n vertices without a clique of r vertices.

Theorem. *Mantel's theorem(1907).*

$$ex(n; K_3) = \begin{cases} \frac{n^2}{4} & \text{if } n \text{ is even} \\ \frac{n^2-1}{4} & \text{if } n \text{ is odd.} \end{cases}$$

Proof. For a graph to be without triangles, any two vertices u, v should not have a common neighbor, thus $\deg(u) + \deg(v) \leq n$. Then for at least one of them, suppose u, $\deg(u) \leq \lfloor \frac{n}{2} \rfloor$. Remove u from the graph, we have this recursion

$$\begin{aligned} \text{ex}(n; K_3) \leq & \text{ex}(n-1; K_3) + \lfloor \frac{n}{2} \rfloor \\ \leq & \lfloor \frac{2}{2} \rfloor + \lfloor \frac{3}{2} \rfloor + \cdots + \lfloor \frac{n}{2} \rfloor \\ = & \begin{cases} 2(1 + 2 + \cdots + \frac{n-2}{2}) + \frac{n}{2} = \frac{n^2}{4} & \text{if even,} \\ 2(1 + 2 + \cdots + \frac{n-1}{2}) = \frac{n^2-1}{4} & \text{if odd.} \end{cases} \end{aligned}$$

□

Remark. The case in which the bound is tight is obvious by dividing the vertices into two sets as equal as possible, $\lfloor \frac{n}{2} \rfloor$ and $\lceil \frac{n}{2} \rceil$.

Remark. In extremal problems in 3-graphs, $\text{ex}(n; K_4^{(3)})$ (the maximal number of edges in 3-graphs without tetrahedron) is unsolved. The conjecture is that the maximal density (that is, $\frac{\text{ex}(n; K_4^{(3)})}{\binom{n}{3}}$) is asymptotically $\frac{5}{9}$.

Bipartite graphs

Def. A graph is bi-partite if there exists a partition $V = V_1 \dot{\cup} V_2$, such that every edge has one endpoint in V_1 and another in V_2.

Def. A circuit is a closed path. A cycle is a simple circuit.

A cycle of odd length exists $\Leftrightarrow$ a circuit of odd length exists. Because we can split the circuit at repeated vertices into cycles.

Theorem. *A graph G is bi-partite iff it does not contain odd cycles.*

Proof.
$\Longrightarrow$ If there is a cycle of odd length, as we go along the cycle and put adjacent vertices into

alternating sets, there must be two adjacent vertices in the same set at the end. Hence G is not bi-partite.

$\Longleftarrow$ First we split G into disjoint unions of connected components $G = G_1 \dot\cup G_2 \dot\cup \cdots$ and if each connected components is bi-partite, G is also bi-partite. Thus we only need to prove it for a connected graph.
Fix a vertex u and put it in one of two sets. Put v in the same set as u if the length of a path (u, v) is of even length and in the other set if otherwise. There cannot be both an odd and an even path from u to v because they would form an odd circuit thus there would be an odd cycle by the remark made just before the proof. □

Def. Clique number $\omega(G)$ is the largest r such that K_r is a subgraph of G.

Triangle-free $\equiv \omega(G) \leq 2$.

Def. Chromatic number $\chi(G)$ is the minimum number of colors needed to properly color its vertices. (Properly colored means that if two vertices have the same color, then they are not connected.)

Bi-partite $\equiv \chi(G) \leq 2$.

Remark. No known efficient algorithm exists to check if $\chi(G) \leq 3$ for a graph.

Theorem. $\omega(G) \leq \chi(G)$.

Since all vertices in a clique would need different colors. But this is a pretty bad bound, as

Theorem. *There are triangle-free graphs ($\omega(G_n) = 2$) s.t. $\chi(G_n) \geq n^\epsilon$ for certain ϵ.*

Theorem. $\chi(G) \leq \Delta(G) + 1$.

Since at every vertex at most $\Delta(G)$ colors are used for its neighbors, we can apply a dynamic algorithm.

Def. Independence number, $\alpha(G) := \omega(\overline{G})$, is the largest size of an independent/stable (not connected to each other) set of vertices.

Theorem. $\chi(G) \geq \frac{n(G)}{\alpha(G)}$

Proof. We can divide $V(G)$ into $\chi(G)$ sets according to the coloring, so that vertices in each set have the same color. Then each set is an independent set. Since n vertices are put into $\chi(G)$ sets, at least one set has $\lceil \frac{n}{\chi(G)} \rceil$ vertices by pigeonhole principle. Thus $\alpha \geq \frac{n(G)}{\chi(G)}$. □

Probabilistic method in combinatorics
Ramsey theory. Find the smallest n (call it $R(k, k)$) s.t. for all graphs on n vertices either $\omega(G_n) \geq k$ or $\alpha(G_n) \geq k$.

Theorem. *Erdös(1947).*
There are graphs G_n such that $\omega(G_n) < 2\log_2 n$ and $\alpha(G_n) < 2\log_2 n$.

Proof. Let $k = 2\log_2 n$. $n = 2^{\frac{k}{2}}$. Consider all possible edges in the graph as $\binom{n}{2}$ Bernoulli trials $X_1, X_2, \ldots, X_{\binom{n}{2}}$ of probability $\frac{1}{2}$. By the union bound ($P(A_1 \vee \cdots \vee A_t) \leq \sum_{i=1}^{t} P(A_i)$), the probability that there is at least one k-clique, $P(\omega(G_n) \geq k) \leq \binom{n}{k} P(\{v_1, v_2, \ldots, v_k\} \text{ is a clique}) = \binom{n}{k} 2^{-\binom{k}{2}} < \frac{1}{2}$.
By symmetry $P(\alpha(G_n) \geq k) < \frac{1}{2}$. Thus $P((\omega(G_n) \geq k) \vee (\alpha(G_n) \geq k)) < 1$, and $P((\omega(G_n) < k) \wedge (\alpha(G_n) < k)) > 0$.[1] □

[1] More context of this theorem: Ramsey number can be understood as the minimum value of n for which a counterexample

Trees

Def. A tree is a connected acyclic graph.

Theorem. *Let G be a simple graph. Then the following statements are all equivalent.*
(a) G is a tree.
(b) G is connected and has $n-1$ edges.
(c) G is acyclic and has $n-1$ edges.
(d) For every pair of vertices u, v there exists a unique simple path from u to v.
(e) G is minimally connected. (G is connected and removing an edge disconnects it.)
(f) G is acyclic but adding an edge creates a cycle.[1]

Lemma. Every tree has a vertex of degree 1.

Proof. A tree doesn't have vertex of degree 0 because it is connected. Assume in a tree $\forall v, \deg(v) \geq 2$. Then if we pick an arbitrary vertex and follow a path from it (which connects it to another vertex then another then another), we can go on infinitely. However, the graph is finite, so we are bound to return to a previous vertex. This contradicts that it is a simple acyclic graph. □

Proof. Of (a) $\Longrightarrow$ (b).
We prove by induction on n. We remove the pendant vertex (the vertex of degree 1) and the adjacent edge from the tree. The remaining graph is still a tree because (1) removing a pendant vertex will not disconnect it so it is still connected (2)it is still acyclic. Thus a tree with n vertices has $(n-2)+1 = n-1$ edges by inductive hypothesis. (Base case: $n = 1$.) □

Remark. Every tree is bi-partite because it does not have odd cycles.

Def. A forest is an acyclic graph.

A forest is a disjoint union of trees, $F = T_1 \dot\cup T_2 \dot\cup \cdots \dot\cup T_k$. Suppose the trees have n_i vertices and e_i edges respectively, then the number of edges in the forest $e = e_1 + e_2 + \cdots + e_k = (n_1 - 1) + (n_2 - 1) + \cdots + (n_k - 1) = n - k$.

Def. A spanning tree is a spanning subgraph that happens to be a tree.

Every connected graph has a spanning tree, because a tree is a minimally connected graph and we can always remove edges from a connected graph until it is minimally connected. [2]

Caley's formula[3]

⌜ *Enumerative graph theory.* Count trees on n vertices: (1) count unlabelled trees (up to isomorphism); (2) count labelled trees with distinguishable vertices. In the second case, $E \subseteq \binom{V}{2}$ and there are $2^{\binom{n}{2}}$ possible graphs.
Denote the number of labelled trees with n vertices as N_n.

---

does not exist. This theorem above states that for $n = 2^{\frac{k}{2}}$ a counterexample exists; thus it proves that $2^{\frac{k}{2}}$ is a lower bound for $R(k,k)$. For the definition of Ramsey Number, see Rosen 6.2, Example 13, pp.404. For a more detailed proof of the lower bound of $R(k,k) \geq 2^{\frac{k}{2}}$ including the bounding of $\binom{n}{k}2^{-\binom{k}{2}}$ (which is not included in class), see Rosen 7.2, Theorem 4, pp.465.

[1] Corresponding references in Rosen: (a)⇔(b)&(c), 11.1, Exercise 15, pp.756 (one direction, 11.1, Theorem 2, pp.752); (a)⇔(d), 11.1, Theorem 1, pp.746; (a)⇔(e), 11.1, Exercise 14, pp.756.

[2] A corresponding theorem, Rosen 11.4, Theorem 1, pp.786.

[3] This is a special lecture before Thanksgiving break.

$N_2 = 1.$

$N_3 = 3.$ 

$N_4 = 16 = 4 + \binom{4}{2} \times 2.$ 

Theorem. *The number of labelled trees on n vertices is equal to n^{n-2}.*

First we make the graph directed. [1]

Exercise. Prove that a directed tree is a rooted tree iff the in-degree of any vertex ≤ 1. [2]

The theorem we are trying to prove is equivalent to stating that the number of labelled rooted trees on n vertices is n^{n-1} because every vertex can be made the root and there are n choices. We then prove this statement by a stronger statement on oriented forests.

Theorem. *The number of oriented forests with n vertices and e edges is $N_e = \binom{n-1}{e} n^e$.*

Proof. We prove this by induction on e.
When $e = 0$, there is only one forest with no edges and $N_0 = 1$. When $e = 1$, we pick a vertex as tail first and then pick another vertex as head (defined as the vertex that the arrow is pointing to), thus $N_1 = n(n-1)$. Suppose $N_e = \binom{n-1}{e} n^e$. We prove the case of N_{e+1} by double counting.
We define that there is a connection between an oriented forest F_1 with e edges and an oriented forest F_2 with $e+1$ edges iff F_1 is included in F_2 (F_2 can be obtained by adding a directed edge on F_1). And we count the number of such connections from both sides. If we remove any directed edge from an oriented forest with $e+1$ edges, then the remaining graph is still an oriented forest because the in-degree of any vertex can only decrease. Thus there are $e+1$ ways to obtain an oriented forest with e edges from that forest. If we want to add an edge to an oriented forest with e edges, any vertex can be the tail because there is no restriction on out-degree. However, only the roots of other trees can be the head, because the in-degree of any vertex should be ≤ 1 and there cannot be a cycle. The number of trees in that forest is $n - e$, since $e = n - k$ for a forest. Thus there are $n(n-e-1)$ ways to obtain an oriented forest with $e+1$ edges from that forest. Thus the number of connections is $(e+1)N_{e+1} = n(n-e-1)N_e$.

$$
\begin{aligned}
N_{e+1} =& \frac{n(n-e-1)}{e+1} N_e \\
=& \frac{n(n-e-1)}{e+1} \binom{n-1}{e} n^e \quad \text{by inductive hypothesis} \\
=& \frac{n-e-1}{e+1} \frac{(n-1)!}{e!(n-e-1)!} n^{e+1} \\
=& \frac{(n-1)!}{(e+1)!(n-e-2)!} n^{e+1} \\
=& \binom{n-1}{e+1} n^{e+1}.
\end{aligned}
$$

□

If the forest is one rooted tree, then $e = n - 1$ and there are n^{n-1} such forests. This proves the original statement. ⌟

[1] Related definitions: directed graph, Rosen 11.1, Definition 2, pp.747; in-degree and out-degree, Rosen 10.2, Definition 5, pp.654.

[2] On $\Longleftarrow$ direction: After picking a root (by showing that there can be only one vertex with in-degree 0), we can show by induction that the original alignment is the same as the rooted tree with that vertex as root. (Prof, after class)

Euler circuits

Def. In a simple graph G, an Euler circuit is a circuit that traverses every edge exactly once. (It is a cycle by our definition.)

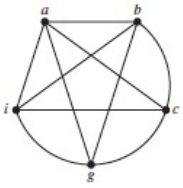
Theorem. *A graph G has an Euler circuit iff G is connected up to the presence of isolated vertices and the degrees of all vertices are even.*

Proof. (Of sufficiency.)
(We consider the connected part and ignore isolated vertices.) We use strong induction on the number of edges. For a graph G in which all vertices are even, there is at least one circuit thus one cycle in it, because otherwise it would be a tree and there would be a vertex of degree 1. Take the longest cycle C (a circuit without repetition) and remove it from the graph. If $G \setminus C = \emptyset$, we are done. Otherwise, take a non-trivial connected component $V \subseteq (G \setminus C)$. V exists because there is at least one edge in $G \setminus C$ since G is connected. And V has an Euler circuit by assumption. (The degrees of vertices are still even.)
We claim that V has a shared vertex with C. Indeed, there has to be one edge leading out of V since G is connected, but it cannot connect to the rest of $G \setminus C$ by definition of a connected component. Thus it has to connect to C. □

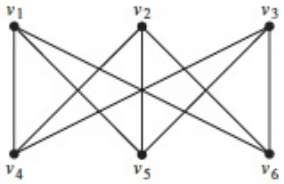
Planar graphs

Def. A planar graph is a graph in which no edges cross over.

non-planar graphs



K_5.[1]



$K_{3,3}$.[2]

Preserve non-planarity
Define graphs obtained from a previous graph by two operations:
(1) super graphs (add vertices and edges).
(2) sub-division (add a vertex on an edge).[3]

Theorem. *Kuratowski's Theorem.*
A graph is non-planar iff it can be obtained from either K_5 or $K_{3,3}$ by these two operations.

Remark. Stating that a graph is thus obtained from K_5 is equivalent to stating that the graph has 5 vertices with vertex disjoint paths among them.

Euler's formula

Four Color Theorem. For a planar graph G, $\chi(G) \leq 4$. This is proved by Appel and Haken in 1976 using exhaustive search by computer. There are no known proofs that do no involve computers. We

[1] Source: Rosen 10.7, Figure 13, pp.724.
[2] Source: Rosen 10.7, Figure 6, pp.719.
[3] For a more detailed definition, see the definition of elementary subdivision and graph homeomorphic in Rosen 10.7, pp. 724.

will prove $\chi(G) \leq 6$ using Euler's formula.
For a connected planar graph G, denote the number of edges as e, the number of vertices as v, the number of regions as r (enclosed by edges, and the outer space is also a region).

Theorem. *Euler's Formula.* $r = e - v + 2$.

We expand our discussion to not necessarily connected graphs, and denote the number of connected components as c.

Theorem. *Euler's characteristic of the plane,* $c - v + e - r$*, is equal to -1.*

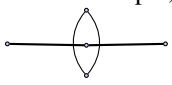
Remark. c, v, e, r are alternating terms by dimension.

Proof. We prove this by induction on the number of edges. Base case: $e = 0,\ r = 1,\ c = v$. To obtain a graph with $e + 1$ edges, we add an edge to a graph with e edges which satisfies the formula by assumption. There are two cases.
Case 1. The added edge is a bridge edge between two originally disconnected parts. Then v is the same, e increases by 1, c decreases by 1. r does not change, because if adding an edge creates a new region, then there has to be another path connecting the two endpoints to bound that region and this contradicts that the two parts are disconnected. Thus $c - v + e - r$ remains the same.
Case 2. The added edge is not a bridge but an edge within a connected component. Then v is the same, e increases by 1, c is the same. r increases by 1 because there has to be a path between the two endpoints since the part is connected, and the added edge forms a region with that path. Thus $c - v + e - r$ remains the same. □

When $c = 1,\ r = e - v + 2$.[1]

Dual graph

Def. The dual graph G^* of a planar graph G is obtained by putting a point in every region in G and connecting all pairs of points in adjacent regions.

Remark. A dual graph could be a multigraph. For example, if G has a vertex of degree 2, there are two edges in G^* surrounding that vertex. 

Remark. G^* depends not only on G but also on the choice of the planar presentation of G.

Remark. If we apply this construction twice, we recover the original graph.

For the dual graph, $e(G^*) = e(G), v(G^*) = r(G), r(G^*) = v$. Thus Euler's formula still holds (sanity check).

For the dual graph of a simple graph, the minimal degree $\delta(G^*) \geq 3$, because every region in G is surrounded by at least three edges.

We apply hand-shaking Theorem to the dual graph of a simple connected planar graph G, $\sum_{v^* \in G^*} d(v^*) = 2e(G^*) \geq 3v(G^*)$. Thus $2e(G) \geq 3r(G) = 3e(G) - 3v(G) + 6$ by Euler's formula on G. Thus for $e \leq 3v - 6$ for G.

Cor. Every planar graph has a vertex of degree ≤ 5.

[1] Another proof listed on the course page which uses dual graph instead of induction and is super interesting: https://www.youtube.com/watch?v=-9OUyo8NFZg&feature=youtu.be.

Proof. If all vertice has degree ≥ 6, then $\sum_{v \in G} d(v) = 2e \geq 6v$. Then $e \geq 3v$, which contradicts our result above. □

We use this corollary to prove $\chi(G) \geq 6$ for a planar graph G.

Proof. We prove this by induction on the number of vertices. Find the vertex v with degree ≤ 5 and remove it from G. The rest can be colored in 6 colors by assumption. Since v has at most 5 neighbors, it can be colored within the 6 colors. □

Cor. K_5 is not planar.

Proof. K_5 has 10 edges and 5 vertices, which contradicts $e \leq 3v - 6$. □

Cor. $K_{3,3}$ is not planar.

Proof. We can improve $\delta(G^*) \geq 3$ to $\delta(G^*) \geq g(G)$, where $g(G)$ is the girth of G, because every region in G is surrounded by at least $g(G)$ edges. Then for a planar graph G, $2e \geq g(e - v + 2)$ by hand-shaking Theorem on G^* and Euler's formula on G. Thus $e \leq \frac{g(v-2)}{g-2}$. When $g = 3$ it is our original result. When $g = 4$, the bound becomes $e \leq 2(v - 2)$.
$g(K_{3,3}) = 4$, while it has 9 edges and 6 vertices, which contradicts the bound. □