Discrete Math 37100-1 Lecture Transcriptions

Originally created by Morgan Sonderegger in 2007.
Edited and expanded by Lars Bergstrom in 2008.

January 14, 2009

1 Preliminaries

Definitions of sets. $A = \{1, 2, 3, 3\}$ is the set with members 1, 2, and 3. $|A| = 3$, read as the *cardinality* or *size* of A is 3. $\forall$ is the "universal quantifier." Notation:

$$(\forall a \in A)(\exists! b \in B)(b = f(a))$$

means "for all a in A there exists a unique b in B such that $b = f(a)$." The number of functions $A \to B$ is $|B|^{|A|} = |B^A|$, define

$$B^A = \{f \mid f : A \to B\}$$

$\neg$ stands for negation. The *Cartesian product* is

$$A \times B = \{(a, b) \mid a \in A,\, b \in B\},$$

and $|A \times B| = |A||B|$. A *relation* is a subset $R \subseteq A \times B$, "relation between A and B." A relation on A is $R \subseteq A \times A$.

Example: $A = \mathbb{R}$, $a < b$ with $a, b \in A$, have relation $<$,

$$R = \{(a, b) \mid a < b\},$$

write aRb.

R is a *transitive* relation if

$$(\forall\, a, b, c)(\text{if } aRb \text{ and } bRc \text{ then } aRc).$$

A *reflexive* relation has $(\forall a)(aRa)$, a *symmetric* relation has $(\forall a, b)(\text{if } aRb \text{ then } bRa)$. R is an *equivalence relation* if it is reflexive, symmetric, and transitive. A partition of a set A is

$$(T_1, \ldots, T_m) : T_i \subseteq A,\, A = T_1 \uplus T_2 \cup \ldots . \uplus T_M,\, T_i \neq \varnothing,$$

where $\uplus$ means disjoint union, applies only if $T_1 \cap T_2 = \varnothing$.

Every partition of A defines a unique equivalence relation on A. In fact, this is a 1-to-1 correspondence (*bijection*), DO. (✎="do it, but do not hand it in.")

✎: $\frac{a}{b} = \frac{c}{d}$ if $ad = bc$, show this is an equivalence relation on $\mathbb{Z} \times (\mathbb{Z} \backslash \{0\})$.

In an equivalence class (partition), a function of two items from the same class should also be in the same class in the result set of that function.

2 Number Theory

a divides b, written $a|b$, if $(\exists\, k)(ak = b)$. For example, $7|21$. $a|1 \iff a = \pm 1$, $1|a$ always, $a|0$ always (take $k = 0$). Note that $0|0$ by this definition. $0|a \iff a = 0$, $(\forall a)(a|a)$, $(\forall a)(a| - a)$. Also, $(a - b)|(a^2 - b^2)$.

Divisibility is

- Reflexive: $a|a$.
- Anti-symmetric ($a|b$ and $b|a$) $\implies a = \pm b$.
- Transitive (✎)

Definition: a is *congruent* to b modulo m, written $a \equiv b \mod m$, if $m|(a - b)$.

(Also called "calendar arithmetic", in relation to mod 7 congruence) Even integers are congruent mod 2, odd integers are congruent mod 2, so congruence mod 2 is an equivalence relation.

✎: $(\forall\, m)$(mod m congruence is an equivalence relation)

✎: If $a \equiv x \mod m$ and $b \equiv y \mod m$, then $a + b \equiv x + y \mod m$, $a \cdot b \equiv x \cdot y$ (all mod m).

Definition: Modulo m *residue classes* are the equivalence classes of the mod m congruence relations.

Theorem 2.1 *There are exactly m of them, and we can do arithmetic on the residue classes.*

Multiplication table modulo 5:

	0	1	2	3	4
0	0	0	0	0	0
1	0	1	2	3	4
2	0	2	4	1	3
3	0	3	1	4	2
4	0	4	3	2	1

Definition: $Div(a) = \{b \mid b|a\}$ is the set of divisors of a.
$Div^+(a) = \{b > 0 \mid b|a\}$ is the set of positive divisors of a.
$Div(a, b) = Div(a) \cap Div(b)$.

Definition: The *greatest common divisor* of a and b is the max element of $Div(a, b)$. Note that $gcd(0, 0)$ is defined to be zero, by point 2 of the definition below.

Theorem 2.2 (*) $(\forall a, b)(\exists d)(Div(a, b) = Div(d))$ *and is unique up to sign.*

Definition: If this holds, then $d = gcd(a, b)$.

Note that the theorem allows negative numbers, but the gcd does not. $(\forall a, b)(gcd(a, b) = max(Div(a) \cap Div(b))$ except when $a = b = 0$. By definition, d is a gcd of a and b if $Div(a, b) = Div(d)$. Equivalently, d must satisfy the following conditions:

1. $d|a$ and $d|b$
2. $(\forall e)$(if $e|a$ and $e|b$ then $e|d$)

Definition: $a \cdot \mathbb{Z} = \{a \cdot x \mid x \in \mathbb{Z}\}$

For example, $3 \cdot \mathbb{Z} = \{0, \pm 3, \pm 6, \pm 9, \ldots\}$.

Definition: Let $A \subseteq \mathbb{Z}$. A is a subgroup if $A \neq 0$ and A is closed under subtraction (i.e. $(\forall a, b \in A)(a - b \in A)$).

Theorem 2.3 (Division Theorem)

$$(\forall a, b \neq 0)(\exists!\, q, r)(a = bq + r \text{ and } 0 \leqslant r < |b|)$$

Example: $a = 100, b = 7$. Solve $100 = 7 \cdot q + r$, where q is quotient and r is remainder. Get $q = 14, r = 2$.

A *module* is a set that is closed under subtraction. Example: $(\forall d)(d\mathbb{Z}$ is a module$)$

Theorem 2.4 *if* $A \subseteq \mathbb{Z}$ *is a module then* $(\exists d)(A = d\mathbb{Z})$

Proof:

1. $0 \in A$ since $A \neq \varnothing \implies \exists a \in A$ such that $a - a = 0$.
2. $-a \in A$ since $0 \in A$, $0 - a = -a$
3. $a, b \in A \implies a + b \in A$ since $-b \in A$, $a - -b \in A$
4. $a \in A \implies a\mathbb{Z} \subseteq A$ (all multiples of a belong to A) NTS: $(\forall n \in \mathbb{Z})(na \in A)$ Simple induction on n

Theorem 2.5 *Let d be the smallest positive number in A. Then $A = d \cdot \mathbb{Z}$.*

Proof:

1. $A \subseteq d \cdot \mathbb{Z}$: Need $(\forall a \in A)(a \in d \cdot \mathbb{Z})$. i.e. $d|a$. So, let $a = dq + r, 0 \leqslant r < d$ (note that this is positive because of the initial claim). $r = a - dq$, $a \in A$ and $d \in A$ and $d \cdot q \in A$, so r cannot be positive, so $r = 0 \implies a = dq \implies d|a$.
2. $A \supseteq d \cdot \mathbb{Z}$: Immediate from $d \in A \implies \{d, d+d, d+d+d, \ldots\}$ and $-d \in A \implies \{-d, -d-d, -d-d-d, \ldots\}$.

Definition: c is a *linear combination* of a and b if $(\exists x, y)(c = ax + by)$

Example: $6 = 18 \cdot \underbrace{2}_{x} + 30 \cdot \underbrace{-1}_{y}$

Theorem 2.6 $(\forall a, b)(\exists x, y)(ax + by$ is a gcd of a and $b)$

Notation: $\forall A, B \subseteq \mathbb{Z}$

1. $A + B = \{a + b \,|\, a \in A, b \in B\}$
2. $A - B = \{a - b \,|\, a \in A, b \in B\}$
3. $A \backslash B = \{a \in A \,|\, a \notin B\}$

So, all linear combinations of a and b are $a \cdot \mathbb{Z} + b \cdot \mathbb{Z}$. Observation: $a \cdot \mathbb{Z} + b \cdot \mathbb{Z}$ is a module. i.e. the difference of two linear combinations of a and b, $(ax + by) - (au + bv) = a(x - u) + b(y - v)$ so $(\exists d)(a\mathbb{Z} + b\mathbb{Z} = d\mathbb{Z})$ so d is a linear combination of a, b because $d \in d\mathbb{Z} = a\mathbb{Z} + b\mathbb{Z}$.

Claim: d is the gcd of a and b.

Proof:

1. $d|a$ because $a = a \cdot 1 + b \cdot 0$
2. $d|b$ similarly
3. let $e|a$ and $e|b$. Claim: $e|d$. $d \in a\mathbb{Z} + b\mathbb{Z} \implies (\exists x, y)(d = ax + by)$. So, $d|a \implies e|ax$ and $d|b \implies e|by$ together imply $a|ax + by = d$

✎**:** Prove if both d and d' satisfy the following then $d = \pm d'$:

1. $d|a$ and $d|b$

2. $(\forall e)$(if $e|a$ and $e|b$ then $e|d$)

Definition: A *prime* is a positive integer $p \geqslant 2$ where $Div^+(p) = \{1, p\}$

Definition: r has the *prime property* if $(\forall a, b)$(if $r|ab$ then $r|a$ or $r|b$ and $r \neq \pm 1$).

Example: $6|3 \cdot 4$ so six does not have the prime property.

Note: 0 has the prime property. Also: if $a \geqslant 2$ and a is not prime, then a does not have the prime property.

Theorem 2.7 *if $p \geqslant 2$ is a prime, then it has the prime property.*

✎**:** The uniqueness of prime factorization (the fundamental theorem of arithmetic) is an immediate consequence.

Proof: Lemma: $gcd(ak, bk) = k \cdot gcd(a, b)$. Let $d = gcd(a, b) = ax + by$. Need a $kd = gcd(ak, bk)$. Know that $kd|ak$ and $kd|bk$, so $d|a$ and $d|b$. If $e|ak$ and $e|bk$ then $e|dk$ because $d = ax + by$, $dk = ak \cdot x + bk \cdot y$ since $e|$ both right terms.

Supposing $p \geqslant 2$, p prime, $p|a \cdot b$, we need $p|a$ or $p|b$. WLOG,[1] assume $p \nmid a$, and prove $p|b$. Then $gcd(a \cdot b, p \cdot b) = b \cdot gcd(a, p)$ by lemma. But that implies $gcd(a, p) = 1$ because $Div^+(p) = \{1, p\}$. Since $p|gcd(ab, pb)$, $p|b$.

✎**:** Learn Euclid's Algorithm.

Proposition 2.8 *$a|b$ and $b|a \iff a = \pm b$.*

Proof: $\Leftarrow \surd$. So for $\Rightarrow$:

$$a|b : \exists k, b = ak$$
$$b|a : \exists l, ab = bl$$

$a = bl = akl$.
$a - akl = 0$, $a = 0 \implies b = ak = a - 0 = 0\surd$,
$a(1 - kl) = 0$, $1 = kl \implies k = \pm 1, b = \pm a\surd$.

As a consequence, the gcd is unique up to sign.

Proof: Suppose d and d' are both gcd's of a and b. Then

1. $d|a$, $d|b$.
2. d is a multiple of all common divisors, including d': $d'|d$. Analogously, $d|d' \implies d = \pm d'$.

Definition: a and b are *relatively prime* if $gcd(a, b) = 1$.

Definition: $a \equiv b \mod m$ means $m|a - b$

Prove that mod is reflexive by $a \equiv a \mod m$ via $(\forall x)(x|0)$. Prove symmetric by $a \equiv b \mod m \implies b \equiv a$ *mod* m via $(\forall x)(x|c \implies x| - c)$. Transitive by $a \equiv b \mod m$ and $b \equiv c \mod m \implies a \equiv c \mod m$ via $m|x$ *and* $m|y \implies m|x + y$ with $x = a - b$ and $y = b - c$.

Theorem 2.9 *If $a = b \mod m$ then $\underbrace{gcd(a, m)}_{d} = \underbrace{gcd(b, m)}_{d'}$.*

[1]WLOG="without loss of generality". Used in a proof when a simplifying assumption is made such that both (a) the proof using the assumption is significantly shorter than the full proof (b) completing the proof without the assumption is straightforward. In the current proof, we know that $p|ab$ and are trying to prove that $p|a$ or $p|b$. In the full proof, we would consider three cases: (1) $p|a$ and $p|b$ (2) $p \nmid a$ (3) $p \nmid b$. In case 1 the claim is trivially true, and if we can prove case 2, the proof of case 3 will be identical. Thus WLOG, we need only consider case 2. (MS)

Proof: $m|a-b$.

$$\mathrm{Div}(d) = \mathrm{Div}(a,m) \stackrel{?}{=} \mathrm{Div}(b,m) = \mathrm{Div}(d'),$$

so need to prove:

$$(\forall x)(x \in \mathrm{Div}(a,m) \stackrel{?}{\iff} x \in \mathrm{Div}(b,m))$$
$$(x|a \text{ and } x|m) \stackrel{?}{\iff} (x|b \text{ and } x|m)$$

$\Rightarrow$, need to prove: $x|b$ and $x|m$. Assume $x|a$ and $x|m$, need to prove $x|b$. Assume $a \equiv b\ (m)$, $x|a$, $x|m$ D.C. $x|b$.

Proof: $m|a-b\ \exists\, y : a-b = my$, then $b = a - my$, put in x's, $\implies x|a-my\surd$. $\Leftarrow$ done the same way.

A residue class mod $m = \{x | x \equiv k\}$, number of residue classes mod m is m. They are equivalence classes.

Corollary 2.10 *If L is a residue class mod m and $(\exists\, x \in L)(gcd(x,m) = 1)$ then $(\forall\, x \in L)(gcd(x,m) = 1)$*

Proof: Thm 2.9

Definition: L is a *reduced* residue class mod m if L is a residue class mod m and its members are relatively prime to m. Denote via $[a]_m$ for the residue class $a \mod m$. The number of reduced residue classes mod m is called $\phi(m)$, called *Euler's phi function.*

So $\phi(m)$ is the # of integers k in the interval $a \leqslant k \leqslant m$ such that $gcd(k,m) = 1$. Have $\phi(1) = 1$, $\phi(2) = 1$, $\phi(3) = 2$, $\phi(4) = 2$, $\phi(5) = 4$, $\phi(6) = 2$, etc.

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

✎: $gcd(a, p^2) \neq 1 \iff p|a$

get that $\phi(p^2) = p^2 - p$. $\phi(p^3) = p^3 - p^2$, in general have

$$\phi(p^k) = p^k - p^{k-1} = p^k(1 - \frac{1}{p})$$

✎: If $\gcd(a,b) = 1$ then $\phi(ab) = \phi(a)\phi(b)$, called "$\phi$ is *multiplicative.*" (Not totally multiplicative, just if gcd has this property.)

✎: $\sum\limits_{d|m} \phi(d) = m$ (but notes slightly more difficult than usual ✎'s)

$$\sum_{d|6} \phi(d) = \phi(1) + \phi(2) + \phi(3) + \phi(6) = 6, \quad \sum_{d|7} \phi(d) = \phi(1) + \phi(7) = 7,$$

where $d|6$ means summation over the positive divisors, etc.

Corollary 2.11 *If $n = p_1^{k_1} \cdots p_2^{k_s}$ and the p_i are distinct primes, then*

$$\phi(n) = \prod_{i=1}^{s} \phi(p_i^{k_i}) = n \prod_{p|n} (1 - \frac{1}{p}),$$

p prime

Example: $\phi(90) = \phi(2 \cdot 9 \cdot 5) = \phi(2) \cdot \phi(9) \cdot \phi(5) = 1 \cdot 6 \cdot 4 = 24 = 0 - (1 - \frac{1}{2})(1 - \frac{1}{3})(1 - \frac{1}{5})$, and $2(1 - \frac{1}{2})9(1 - \frac{1}{3})5(1 - \frac{1}{5})$.

Theorem 2.12 $\sum_p \frac{1}{p} = \infty$

✎**:** Prove that $\inf_n \frac{\phi(n)}{n} = 0$. Note: $lim_{n->\inf} \frac{\phi(p)}{p} = 1$ because $\frac{p-1}{p} = 1 - \frac{1}{p} = 1$

Claim: $x^2 \equiv x \bmod 2$.

Proof: $2 \overset{?}{|} x^2 - x = x(x-1) \implies$ one of them even.

Claim: $x^3 \equiv x \bmod 3$

Proof: $3|x^3 - x = x(x^2-1) = x(x-1)(x+1) = (x-1)x(x+1)$

Claim: $x^5 \equiv x \bmod 5$

Proof: $5|x^5 - x = x(x^4-1) = x(x^2+1)(x^2-1) = (x-1)x(x+1)(x^2+1)$, instead if x^2+1 we would wish $x^2 - 4 = (x-2)(x+2)$, but now $x^2 + 1 \equiv x^2 - 4 \bmod 5$.

✎**:** Prove in a similar manner: $x^7 \equiv x\ (7)$, $x^{11} \equiv x\ (11)$.

Theorem 2.13 (Fermat's Little Theorem) *$x^p \equiv x$ mod p, p prime.*[2]

(Whenever p written without comment, assume is prime.) Call theorem stated this way (1). An equivalent statement, (2), is

$$(\forall x)(\forall p \text{ prime})(\text{if } \gcd(x,p) = 1 \text{ then } x^{p-1} \equiv 1 \bmod p)$$

Proof:

(2) $\implies$ (1): If $\gcd(x,p) = 1$, then (2) $\implies$ $x^{p-1} \equiv 1\ (p)$, $x^p \equiv x\ (p)$. If $\gcd(x,p) \neq 1$, i.e. $p|x$, then $e \equiv 0\ (p)$, $x^p \equiv 0\ (p)$.
(1) $\implies$ (2): We know $x^p \equiv x\ (p) \implies$ divide both sides by x: $x^{p-1} \equiv 1\ (p)$, because we are assuming $\gcd(x,p) = 1$. $x^p \equiv x\ (p)$, $p|x^p - x = x(x^{p-1} - 1)$, $p|x \implies p|x^{p-1} - 1$, by the prime product property.

✎**:** If $ax \equiv ay \mod m$ and $\gcd(a,m) = 1$ then $x = y \mod m$.

✎**:** If $ax \equiv ay \mod am$ then $x \equiv y \mod m$.

Theorem 2.14 (Euler-Fermat) *If* $\gcd(x,m) = 1$ *then* $x^{\phi(m)} \equiv 1 \mod m$.

(Note: $\lim_{p\to\infty} \frac{\phi(p)}{p} = 1$)

Proof: Let $a_1, \ldots, a_{\phi(m)}$ be a set of representatives of all reduced residue classes.

Claim: $xa_1, \ldots, xa_{\phi(m)}$ is again a set of representatives of the reduced residue classes.

Proof: (1) $(\forall i)(\gcd(xa_i, m) = 1)$, proof by $\gcd(x,m) = 1$ and $\gcd(a_i, m) = 1$.
(2) $i \neq j \implies xa_i \neq xa_j \mod m$. Contrapositive is: $xa_i \equiv xa_j \mod m \implies$ (by ✎ex.) $a_i \equiv a_j \mod m$ $\implies i = j$.

$\implies$

$$\prod_{i=1}^{\phi(m)} a_i \equiv \prod_{i=1}^{\phi(m)} (xa_i) \equiv x^{\phi(m)} \underbrace{\prod_{i=1}^{m} a_i}_{A} \mod m$$

get $x^{\phi(m)} A \equiv A \mod m$, $\gcd(A,m) = 1 \implies$ by ✎ex., $x^{\phi(m)} \equiv 1 \mod m$

[2] "So for whatever reason, on one sunny afternoon Little Fermat decided to look at the following..."

A sequence $a_0, a_1, \ldots$ is *periodic* with period t if $(\forall n)(a_{n+t} = a_n)$. t is a period, *the* period is the smallest positive period. Equivalent definition: t is a period if $(\forall k, l)$, if $k \equiv l \pmod t$ then $a_k = a_l$.

✎: *The* period is the gcd of all periods.

✎: Prove that if a/b a fraction, $0 < a < b$, $\gcd(b, 10) = 1$, $\implies$ a/b is a periodic decimal.

$a^0 = 1, a, a^2, \ldots \mod m$, assume $\gcd(a, m) = 1$. $\phi(m)$ is *a* period of this sequence. $a^{\phi(m)} \equiv 1 \mod m$ (Euler-Fermat).

If $k, l \geqslant 0$, $k \equiv l \mod \phi(m)$ then $a^k \equiv a^l \mod m$.

If p prime, $k, p \geqslant 0$, $k \equiv l \mod p-1$, then $a^k \equiv a^l \mod p$. In general, the period divides $\phi(m)$.

The period of the sequence $\{a^k \mod m\}$ is called the *order* of $a \mod m$. (Assume $\gcd(a, m) = 1$.)

In other words, the order of $a \mod m$, $\mathrm{ord}_m(a)$, is the smallest $k > 0$ such that $a^k \equiv 1 \mod m$. [Euler-Fermat tells us $\mathrm{ord}_m(a) | \phi(m)$].

Ex: $\mathrm{ord}_5(2) = 4$, $\mathrm{ord}_7(2) = 3$.

Definition: a is a *primitive root* mod p if $\mathrm{ord}_p(a) = p - 1$.

Theorem 2.15 *For any prime p, $\exists$ a primitive root mod p*

Ex: 2 is primitive root mod 5, 3 is primitive root mod 7. This theorem is non-trivial, can find online, etc.

$10 \equiv 3 \mod 7 \implies 10$ primitive root mod 7.

1/7=0.142857, periodic. Let $A = 142,857$, then $7A = 000,000$. Puzzle: 142,857 is the only 6-digit number A such that $A, 2A, \ldots, 6A$ all have the same digits.

✎: $\frac{1}{p}$ is in decimal periodic; period is $\mathrm{ord}_p(10)$

✎: 10 is a primitive root mod 17. (Note means 1/17=0.BBB.., where B has 16 digits.

✎:
1. Definition of gcd of any number of integers.
2. Prove gcd exists, is repr. as a linear combination.
3. $\gcd(a, b, c) = \gcd(a, \gcd(b, c))$.

2.1 Linear congruences

Claim: $ax \equiv b \mod m$ is solvable $\iff$ $\gcd(a, m) | b$.

Proof:
1. Necessity, i.e. $(\exists x)(\ldots) \implies \gcd \ldots$. Obs: If $a \equiv b \mod m$ and $r | m$ then $a \equiv b \mod r$. Pf: transitivity of divisibilities, ✎.

   Proof: $d = \gcd(a, m)$, $ax \equiv b\ (m) \implies ax \equiv b\ (d)$, $0 \equiv b\ (d)$, so $d | b$ $\surd$.
2. Sufficiency: $\gcd \ldots \implies (\exists x)(\ldots)$. $d := \gcd(a.m)$, assumption $d \,|\, b$. $\exists\, x_0, y_0$, $d = ax_0 + my_0$, $ax_0 \equiv d \mod m$. $a\frac{b}{d}x_0 \equiv \frac{b}{d}d = b \pmod m$, with $\frac{b}{d}x_0 = x$. $\surd$

Case $b = 1$: $ax = 1 \pmod m$, x the multiplicative inverse of a mod m ($a^{-1} \mod m$). It exists $\iff$ $\gcd(a, m) = 1$.

Simultaneous congruences:

$$x \equiv 1(8) => x \equiv 1(2) \tag{1}$$
$$x \equiv 5(7) \tag{2}$$
$$x \equiv 4(6) => x \equiv 4(2) \tag{3}$$

Not solvable, since (1) and (3) contradict each other. In general,

$$x \equiv a(m) \tag{4}$$
$$x \equiv b(n) \tag{5}$$

contradict each other if $a \neq b \mod \gcd(m, n)$.

Corollary 2.16 *If the system (4, 5) is solvable then* $a \equiv b$ *mod* $\gcd(m,n)$ *so this is a necessary condition of solvability.*

✎**:** It is also sufficient.

Corollary 2.17 *If* $\gcd(m,n) = 1$ *then (4, 5) is always solvable.*

Theorem 2.18 (Chinese Remainder Theorem) *Consider the system*

$$x \equiv a_1(m_1) \tag{6}$$

$$\vdots$$

$$x \equiv a_k(m_k) \tag{7}$$

If *the* m_i *are pairwise relatively prime, then a solution exists, and solution is unique modulo* $N := m_1 \cdots m_k$.

Example: System

$$x \equiv 2(5)$$
$$x \equiv 1(6)$$
$$x \equiv 3(7)$$

by CRT $\exists x$ satisfying these. Take 42, 35, 30, $x = 42A + 35B + 30C$, mod 5 gives $42a \equiv 2$, mod 6 gives $35B \equiv 1$, mod 7 gives $30C \equiv 3$. A exists because $\gcd(42,5) = 1$, etc. Literature for the CRT is Wikipedia, very good description of theorem and proof.

Theorem 2.19 *For system (6–7), if* $\exists x$ *it is unique modulo* $lcm(m_1, \ldots, m_k) =: L$*, where lcm stands for "least common multiple"*

Proof: Suppose

$$y \equiv a_1(m_1)$$

$$\vdots$$

$$y \equiv a_k(m_k)$$

Need to prove: $x \equiv y \mod L$, i.e. $L | x - y$.

✎**:** Define lcm in full analogy with definition of gcd, prove $\exists$.

Proof: $L|b \iff (\forall i)(m_i|b)$ by definition of lcm. $\surd$

Theorem 2.20 (Euclid) $\exists$ *infinitely many primes.*

Proof: (Euclid) Assume by contradiction that $p_1, \ldots, p_k$ are all the primes ($p_1 = 2$). Let $N = p_1 p_2 \cdots p_k$. Then $N + 1 \geqslant 2 \implies \exists$ prime $p | N+1 \implies (\exists i)(p = p_i$, $N \equiv -1 \bmod p$, $N \equiv 0 \mod p_i$, so $1 \equiv 0 \mod p$, contradiction.

Example: Find $x : x^2 \equiv 1(187)$ but $x \not\equiv \pm 1(187)$. Via CRT:

$$x \equiv 1(17)$$
$$x \equiv -1(11)$$

Solution in the form: $x = A * 17 + B * 11$.

$$\begin{aligned} B * 11 &\equiv 1(17) & B &\equiv -3(17) \\ A * 17 &\equiv -1(11) & 6A &\equiv -1(11) \\ & & 12A &\equiv -2(11) \\ & & A &\equiv -2(11) \end{aligned}$$

So, $x = -2 * 17 + -3 * 11$ and $x = -67$. Check: $(67)^2 \equiv 1(187)$. That's $-67 \equiv 1(17)$ and $-67 \equiv -1(11)$ so it's good!

3 Counting

An n-set is a set of n elements, $[n] = \{1, \ldots, n\}$. The $\#k$-subset of an n-set is $\binom{n}{k} = \frac{n!}{k!(n-k)!}$.

"n choose k" In poker you get five cards, so $\binom{52}{5} = \frac{52*51*50*49*48}{5!}$. The bottom divides the sequences into equivalence classes based on the "same cards". Remember to make life easy when you can: $\binom{n}{3} = \frac{n!}{3!(n-3)!} = \frac{n(n-1)(n-2)}{3!}$.

A *permutation* of a set A is an $A \to A$ bijection. The # of permutations of an n-set is $n!$. Will be taking $0^0 = 1$.

☠ [3]: $\lim\limits_{x,y\to 0} x^y =$ mostly 1.

Pascal's triangle, Pascal's identity is

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

Combinatorial proof: $\binom{n}{k}$ is $\#(k+1)$ subsets containing special element, $\binom{n}{k+1}$ is $\#(k+1)$-subsets avoiding special elements... and get it from there, then gives binomial theorem.

READ: binomial theorem: $(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k$.

3.1 Asymptotic notation

Definition: For a sequence $\{a_i\}$, $\lim\limits_{n\to\infty} a_n = A$ means $(\forall \epsilon > 0)(\exists N)(\forall n > N)(|a_n - A| < \epsilon)$. Interpret as "for all sufficiently large n, a_n is within a threshold distance of A." For an interval of size ϵ, as N gets large the difference between a_n and A gets smaller.

Definition: $a_n \sim b_n$ are *asymptotically equal* if $\lim\limits_{n\to\infty} \dfrac{a_n}{b_n} = 1$

False version is just the negation of all quantifiers OR no limit exists.

Examples:

1. $a_n = 3n^2 + 5n + 100$ and $b_n = 3n^2$ are asymptotically equal. This is because $\frac{3n^2+5n+100}{3n^2} = 1 + \frac{5n}{3n^2} + \frac{100}{3n^2} \approx 1$.
2. Stirling's formula:(memorize) $n! \sim (\frac{n}{e})^n \sqrt{2\pi n}$
3. $\pi(x) = \#$ primes $\leqslant x$. So $\pi(4) = 2$, $\pi(10) = 4$, $\pi(100) = 25$, etc. One of the biggest theorems in math:

   Theorem 3.1 (Prime Number Theorem)

   $$\pi(x) \sim \frac{x}{\ln x}$$

   Proved in 1896 by Jacque Hadamard and Pierre de la Vallée Poussin.[4]

When is $a_n \sim b_n$? Let

$$c_n = \begin{cases} \frac{a_n}{b_n} : b_n \neq 0 \\ * : a_n \neq 0, b_n = 0 \\ 1 : a_n = b_n = 0 \end{cases}$$

Say $a_n \sim b_n$ if $\lim c_n = 1$. Under this definition, $\sim$ is reflexive (proved), symmetric (proved), and transitive (✎).

Definition: $f(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_0$ is a *polynomial of degree* n if $a_n \neq 0$.

[3]=Challenge

[4]"Hadamard was French, and that means you put letters at the beginning and end which aren't pronounced, to confuse the enemy."

Note that $f(x) \sim a_n x^n$. Also, $\lim\limits_{x\to 0} \dfrac{\ln(1+x)}{x} = 1$.

✎: $\sqrt{1+\frac{1}{n}} - 1 \sim ?$

Example: $\binom{n}{3} = \frac{n(n-1)(n-2)}{3!} \sim \frac{n^3}{6}$

✎: $a_n \sim b_n > 1 \implies \ln a_n \sim \ln b_n$? (Answer is "almost".. find condition.)

Pascal's triangle, define *floor* $\lfloor\ \rfloor$ and *ceiling* $\lceil\ \rceil$, from Pascal's triangle $\binom{n}{\lfloor \frac{n}{2} \rfloor} = \binom{n}{\lceil \frac{n}{2} \rceil}$. Have

$$(1+1)^n = \binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{n} = 2^n,$$

$\implies \binom{n}{k} < 2^n\ \forall k$. So $2^n > \binom{n}{\lfloor \frac{n}{2} \rfloor} > \frac{2^n}{n+1}$, $\frac{2^n}{n} < \binom{n}{\lfloor \frac{n}{2} \rfloor} < 2^n$. This all works because it relies on the rule that the biggest must be bigger than the average.

✎: $\binom{n}{\lfloor \frac{n}{2} \rfloor} \sim c\frac{s^n}{\sqrt{n}}$, $c =$? Use Stirling's formula..

Claim: O_n is odd subsets, E_n even subsets, then $|O_n| = |E_n|$, if $n \geqslant 1$.

Proof: $0^n = (1-1)^n = \binom{n}{0} = [\binom{n}{0} + \binom{n}{2} + \ldots] - [\binom{n}{1} + \binom{n}{3} + \ldots]$. Combinatorial proof, use a bijection: $[n] = \{1, \ldots, n\}$, for $A \subseteq [n]$, odd $\implies \bar{A} = [n]\backslash A$. For n odd, take one element out, etc..

$$\begin{aligned} O_{2k} &= O_{2k-1} + E_{2k-1} \\ E_{2k} &= E_{2k-1} + O_{2k-1} \end{aligned}$$

Make a function f which toggles whether the element is in your subset of not:

$$f : \begin{cases} f(A) = A\backslash\{n\} : n \in A \\ f(A) = A \cup \{n\} : n \notin A \end{cases}$$

f is a bijection between even and odd sets (for $n \geqslant 0$).

Note $\displaystyle\sum_{k=0}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{2k} = 2^{n=1}$

☠: Consider $\displaystyle\sum_{k=0}^{\lfloor \frac{n}{4} \rfloor} \binom{n}{4k} =$? For what n is it 2^{n-2}?

☠: Show that $\displaystyle\left| \sum_{k=0}^{\lfloor \frac{n}{3} \rfloor} \binom{n}{3k} - \frac{2^n}{3} \right| < 1$.

Have $\binom{x}{3} = \frac{x(x-1)(x-2)}{6}$, even for complex numbers. Define $\binom{n}{k} = 0$ for $k > n$, then

$$(1+z)^n = 1 + \binom{n}{1}z + \binom{n}{2}z^2 + \cdots + \binom{n}{n}z^n + \binom{n}{n+1}z^{n+1} = \sum_{k=0}^{\infty} \binom{n}{k} z^k$$

and *Newton's Binomial Theorem* is

$$(1+z)^x = \sum_{k=0}^{\infty} \binom{x}{k} z^k$$

For all complex numbers x, assuming $|z| < 1$. Have

$$\frac{1}{1-z} = 1 + z + z^2 + \cdots = (1-z)^{-1} = \sum_{k=0}^{\infty} \underbrace{\binom{-1}{k}}_{(-1)^k} (-z)^k = \sum_{k=0}^{\infty} z^k,$$

where

$$\binom{-1}{k} = \frac{(-1)(-2)\cdots(-k)}{k!} = (-1)^k.$$

This is how many way to pick k x's, l y's, and m z's. Also, don't forget that $\frac{1}{1+t} = 1 - t + t^2 - t^3 + \cdots$. HW: show

$$\frac{1}{\sqrt{1-z}} = (1-z)^{-\frac{1}{2}} = \sum_{k=0}^{\infty} \binom{-\frac{1}{2}}{k} (-z)^k$$

3.2 Generating functions

Power series are the generating functions of the sequence a_n:

$$\begin{aligned}
f(x) &= \sum_{n=0}^{\infty} a_n x^n \\
f(x) + g(x) &= \sum_{n=0}^{\infty} (a_n + b_n) x^n \\
f(x) * g(x) &= \sum_{n=0}^{\infty} c_n x^n \qquad \left(\text{where } c_n = \sum_{k=0}^{n} a_k b_{n-k}\right) \\
f'(x) &= \sum_{n=1}^{\infty} a_n n x^{n-1}
\end{aligned}$$

Look at fib-gen, where the coefficients are the Fibonacci numbers. So, $f(x) = \sum_{n=0}^{\infty} F_n x^n$. Reduce, pulling out factors and simplifying:

$$\begin{aligned}
f(x) = F_0 + F_1 x + \sum_{n=2}^{\infty} (F_{n-1} + F_{n-2}) x^n \\
f(x) = x + \sum_{n=2}^{\infty} F_{n-1} x^n + \sum_{n=2}^{\infty} F_{n-2} x^n \\
f(x) = x + x * f(x) + x^2 * f(x) \\
f(x) = \frac{x}{1 - x - x^2}
\end{aligned}$$

Theorem 3.2 (Trinomial)

$$(x+y+z)^n = \sum_{k,l,m \geqslant 0, k+l+m-n} \binom{n}{k,l,m} x^k y^l z^m,$$

where $\binom{n}{k,l,m} = \frac{n!}{k!l!m!}$.

For $\binom{n}{k,l,m}$, think of $n!$ total ways to distribute the cards, divided by the ways that $k!, l!, m!$ could have been distributed. Note that $k + l + m = n$.

Theorem 3.3 (Multinomial)

$$(x_1 + \cdots + x_k)^n = \sum_{t_1,\dots,t_k \geqslant 0,\, t_1+\cdots+t_k=n} \left(\frac{n!}{\prod_{i=1}^{k} t_i!} x_1^{t_1} \cdots x_k^{t_k} \right)$$

Claim the number of terms in the k-nomial theorem is $\binom{n+k-1}{k-1}$. Lots of reasoning here on why this would be so: looking for number of solutions to the equation $x_1 + \cdots + x_k = n$, for $x_i \geqslant 0$, $x_i \in \mathbb{Z}$. Easier question is same for $y_1 + \cdots + y_k = n$, $y_i \geqslant 1$, by looking at putting $k-1$ dividers in n places, get $\binom{n-1}{k-1}$. Now, let $y_i := x_i + 1$, $y_i \geqslant 1$, $\sum y_i = n + k$, get $\binom{n+k-1}{k-1}$ ✓

Definition: $a_n \sim b_n$ if $\lim_{n\to\infty} \dfrac{a_n}{b_n} = 1$, replace $\frac{0}{0}$ by 1.

Definition: $a_n = o(b_n)$, if $\lim_{n\to\infty} \dfrac{a_n}{b_n} = 0$, abd $\frac{0}{0} := 0$.

In this notation, $a_n = o(1)$ means $\lim_{n\to\infty} a_n = 0$.
Obs: $a_n \sim b_n \iff a_n = b_n(1 + o(1))$, meaning $\exists c_n$, $a_n = b_n(1 + c_n)$, where $c_n = o(1)$.

Note that $a_n = o(c_n)$, $b_n = o(c_n) \not\Rightarrow a_n b_n = o(c_n)$, i.e. $a_n = b_n\sqrt{n}$, $c_n = n$.

Also, $a_n = o(b_n), a_n = o(c_n) \not\Rightarrow a_n = o(b_n + c_n)$, but want this to be true, it is under condition $b_n, c_n > 0$ (or both negative).

✎**:** Is this statement T or F?

Definition: $a_n = O(b_n)$ if $(\exists C)$(for all sufficiently large n), i.e. $(\exists n_0)(\forall n \geqslant n_0)$. This is equivalent to $|a_n| \leqslant C|b_n|$ for some C.) Say "the order of magnitude of a_n is $\leqslant$ the order of magnitude of b_n."

✎**:** $\frac{100n^2-7}{5n+8} = O(n)$

If $a_n = o(b_n)$, then $a_n = O(b_n)$

Definition: $a_n = \Omega(b_n)$ if $b_n = O(a_n)$.

Definition: $a_n = \Theta(b_n)$ if $a_n = O(b_n)$ and $a_n = \Omega(b_n)$, i.e. $(\exists C, c > 0)(\exists n_0)(\forall n \geqslant n_0)(c|b_n| \leqslant |a_n| \leqslant C|b_n|)$, say "$a_n$ and b_n have the same order of magnitude." We have to have "for n sufficiently large" here because $a_n = 0, b_n \neq 0$ could happen at a finite number of places or "within constant factors of each other."

Definition: a_n *polynomially bounded* if $(\exists C)(a_n = O(n^C)$.

Definition: a_n grows exponentially if $(\exists c > 0)(a_n = \Omega(e^{n^c}))$

Theorem 3.4 *If a_n is polynomially bounded and b_n grows exponentially then $a_n = o(b_n)$.*

In fact, $\frac{b_n}{a_n}$ grows exponentially (assuming $a_n \neq 0$).

Theorem 3.5 $\ln x = o(x)$*, i.e.* $\lim_{x\to\infty} \dfrac{\ln x}{x} = 0$.

Proof: L'Hôpital's rule, get

$$\frac{(\ln x)'}{x'} = \frac{1/x}{1} = \frac{1}{x} \to 0$$

In fact $\forall c > 0$, $\ln x = o(x^c)$, do by re-defining x, same kind of proof w/ LHR.

Now, how to prove that $(\ln x)^{100} = o(x)$? Say

$$\left(\frac{\ln x}{\sqrt[100]{x}}\right)^{100} \implies \frac{(\ln x)^{100}}{x} \to 0$$

.

Theorem 3.6 $(\forall c)(c > 0)(\ln x = o(x^c))$

3.3 Polynomial vs. exponential growth

Theorem 3.7 *$\forall C$, $C > 0$, $n^C = o(e^{n^c})$, ex: $n^{1000} = o(e^{\sqrt[100]{n}})$*

Proof: $\ln x =^{100} \sqrt{n}$, $(\ln x)^{100000} = o(x)$ ✓

✎**:** $\forall c, d > 0$, $(1+c)^{n^d}$ grows exponentially. Meaning: $\exists f, g > 0$ such that $e^{n^f} < (1+c)^{n^d} < e^{n^g}$ for all sufficiently large n

Ω notation read "a_n grows at least exponentially." But O notation read a_n "is" exponential, even though behaves more like an inequality.

✎**:** Θ is an equivalence relation on sequences.

Theorem 3.8 *If $L = \lim_{n\to\infty} \frac{a_n}{b_n}$ exists, then*

1. *If $L = 1$ then $a_n \sim b_n$*
2. *If $L = 0$ then $a_n = o(b_n)$*
3. *If $L = \pm\infty$ then $b_n = o(a_n)$*
4. *If $L \neq 0, \neq \pm\infty$, then $a_n = \Theta(b_n)$*

Proof: ✎

Theorem 3.9 *Suppose $a_n \geqslant 1$. a_n is polynomially bounded $\iff$ $\ln a_n = O(\ln n)$.*

Proof: ✎

$\ln n = \Theta(\log_2 n)$ because $\frac{\log_2 n}{\ln n} = \frac{1}{\ln 2}$. If $\pi(x)$ = number of primes $\leqslant x$, then $\pi(x) = o(x)$, get $x^{.99} = o(\frac{x}{\ln x})$, because PNT gives $\pi(x) \sim \frac{x}{\ln x} > \frac{x}{x^{.01}}$

Theorem 3.10 *If $\lim_{n\to\infty} \frac{a_n}{b_n}$ exists $:= L$, have $\binom{2n}{n} = o(4^n)$, because $\binom{2n}{n} = \Theta(\frac{4^n}{\sqrt{n}})$ (using Stirling's formula)*

Proof: ✎

Theorem 3.11 *If $a_n = \Theta(b_n)$ and $a_n, b_n \to \infty$, then $\ln a_n \sim \ln b_n$*

Suggests writing o with an "ear" to distinguish from O. o and O are notation from Landau $\sim$ 1900, Ω, Θ from Don Knuth. Notation not used here is $\omega(b_n) = a_n$ if $b_n = o(a_n)$.

4 Finite Probability Spaces

Sample space = set of all possible outcomes of the experiment. Each outcome: elementary event. Usually call Ω the sample space, A an event, $A \subseteq \Omega$.

Definition:
1. Non-empty finite set Ω, the *sample space*.
2. A probability distribution P over Ω: $P : \Omega \to \mathbb{R}$. such that
   (a) $(\forall x \in \Omega)(P(x) > 0)$
   (b) $\sum_{x\in\Omega} P(x) = 1$

Elements of Ω are "elementary events", then (Ω, P) is a *finite probability space.*

If $(\forall x \in \Omega)(P(x) = \frac{1}{|\Omega|})$ then the space is *uniform distribution*.

An *event* is $A \subseteq \Omega$, $P(A) = \sum\limits_{x \in A} P(x)$. In particular, $P(\varnothing) = 0$, $P(\Omega) = 1$.

✎: If $A_1, \ldots, A_k \subseteq \Omega$, then $P(A_1 \cup \ldots \cup A_k) \leqslant \sum\limits_{i=1}^{k} P(A_i)$, *union bound.*

Equality holds $\iff$ the A_i are pairwise disjoint, i.e. they are *mutually exclusive.*

✎: $P(A \cup B) + P(A \cap B) = P(A) + P(B)$ (modular equation)

4.1 Conditional probability

$A, B \subseteq \Omega$, $B \neq \varnothing$, then

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

is the *probability of A conditional on B*.

Definition: A, $B \subseteq \Omega$ are *independent* if $P(A \cap B) = P(A)P(B)$.

Definition: The *trivial events* are $\varnothing, \Omega$.

✎: If A is trivial then $(\forall B)(A, B$ are independent$)$.

Consequence: if $B \neq \varnothing$ then A, B are independent $\iff$ $P(A) = P(A \mid B)$.

Theorem 4.1 (Complete probability) *For a partition $\Omega = B_1 \uplus \cdots \uplus B_k$, $B_i \neq \varnothing$, $\uplus$ is "disjoint union",*

$$P(A) = \sum P(A|B_i)P(B_i)$$

Proof: Have

$$P(A \mid B_i)P(B_i) = \frac{P(A \cap B_i)}{P(B_i)} P(B_i) = P(A \cap B_i),$$

$\Omega = B_1 \uplus \cdots \uplus B_k$, $A = (A \cap B_i) \uplus \cdots \uplus (A \cap B_k)$.

Proof of causes: Say we know $P(S \mid B) = 90\%$, $P(S) = 5\%$, $P(B) = 2\%$.
Q: What is $P(B \mid S)$?

$$P(B \mid S) = \frac{P(B \cap S)}{P(S)} = \frac{P(B)P(S \mid B)}{P(S)} = \frac{0.02 \cdot 0.9}{0.05} = \frac{2}{5} \cdot 0.9 = 0.36,$$

so 36%. Note that this used $P(B \cap S) = P(B) * P(S \mid B)$.

Definition: A, B *positively correlated* if $P(A \cap B) > P(A)P(B)$, *negatively correlated* if $P(A \cap B) < P(A)P(B)$.

Example: Roll a die, A event it's prime, B event it's odd. Then $P(A) = \frac{1}{2}$, $P(B) = \frac{1}{2}$, $P(A \cap B) = \frac{1}{3} \implies$ positively correlated.

✎: For what n are the following events independent: $A : 2 \mid x$, $B : 3 \mid x$. Yes if $6 \mid n$. Pick a number x from $\{1, \ldots, n\}$. For $n = 8$, $P(A) = \frac{1}{2}$, $P(B) = \frac{1}{4}$, $P(A \cap B) = \frac{1}{8} = P(A)P(B)$.

If P is uniform, then

$$P(A) = \sum_{x \in A} \underbrace{P(x)}_{\frac{1}{|\Omega|}} = \frac{|A|}{|\Omega|},$$

i.e. "# of good cases"/"# of all cases".

Experiment: n coin flips, get an outcome such as HTTHTTT, $|\Omega| = 2^n$.

Deal 5 cards from standard deck of 52 cards, a "poker hand", then $|\Omega| = \binom{52}{5}$.

For events $A, B, C \subseteq \Omega$, $P(A \cap B \cap C) = P(A)P(B)P(C)$ *plus* pairwise independent. Without this last bit, can have $A = B$ non-trivial and $C = \varnothing$, holds but not pairwise independent.

✎: If A, B, C independent, then

- A, $B \cup C$ also independent,
- A, $B \cap C$ also independent,
- A, $B \backslash C$ also independent.

Means $A, B, \bar{C}$ independent (where $\bar{C} = \Omega \backslash C$).

Definition: $A_1, \ldots, A_k \subseteq \Omega$ are *independent* if $\forall I \subseteq [k]$,

$$P(\bigcap_{i \in I} A_i) = \prod_{i \in I} P(A_i),$$

2^k conditions. Turns out 2^k conditions actually $2^k - k - 1$. (Need only for $|I| \geqslant 2$, it is automatically satisfied for $|I| \leqslant 1$.)

If $|I| = 1$, singleton, $I = \{i\}$, $\bigcap_{j \in I} A_j = A_i$.

If $I = \varnothing$, $\prod_{i \in \varnothing}$ anything $= 1$. $\bigcap_{i \in \varnothing} A_i = \Omega$.

✎: Experiment: n coin flips. Space: uniform. A_i="ith coin comes up heads" $\implies$ $A_1, \ldots, A_n$ are independent, $P(A_i) = \frac{1}{2}$.

4.2 Random variables

Function $X : \Omega \to \mathbb{R}$. The *expected value* of X is

$$E(X) = \sum_{x \in \omega} X(x)P(x),$$

the weighted average of outcomes. Over a *uniform* space,

$$E(X) = \sum X(x)\frac{1}{|\Omega|} = \frac{\sum X(x)}{|\Omega|},$$

the simple average.

✎: $\min X \leqslant E(X) \leqslant \max(X)$

✎: If $X, Y : \Omega \to \mathbb{R}$, then $E(X + Y) = E(X) + E(Y)$.

Have $E(cX) = cE(X)$ for $c \in \mathbb{R}$, so

Theorem 4.2 (Linearity of expectation) *For $a_i \in \mathbb{R}$, $X_i : \Omega \to \mathbb{R}$,*

$$E(\underbrace{\sum_{i=1}^{k} a_i X_i}_{\textit{linear comb.}}) = \sum_{i=1}^{k} a_i E(X_i).$$

or if $X = c_1 Y_1 + c_2 Y_2 + \cdots$ then $E(X) = c_1 E(Y_1) + c_2 E(Y_2) + \cdots$.

Theorem 4.3

$$E(x) = \sum_{r \in R} r P(X = r),$$

but r really $\in$ range(X), because if not the probability is 0.

Why? "$X = r$" is an event, namely $\{x \in \Omega \,|\, X(x) = r\} = X^{-1}(r)$. Anyhow, proof is ✎.

Definition: The *indicator variable* of event A is

$$\vartheta_A(x) = \begin{cases} 1 \text{ if } x \in A \\ 0 \text{ if } x \notin A \end{cases}$$

If $n = |\Omega|$, the # events $= 2^n$, #(0,1)-random variables (indicator variables) $= 2^n$. Every random variable that takes values 0, 1 is the indicator variable of an event: $A = Y^{-1}(1)$, $Y = \vartheta_A$,

$$E(\vartheta_A) = 1 \cdot P(\vartheta_A) = 1) + 0 \cdot P(\vartheta_A = 0) = P(A)$$

This is very important: $E(\vartheta_A) = P(A)$, i.e. the event "$\vartheta_A = 1$" <u>is</u> A.

For X : # heads in n coin flips,

$$E(X) = \sum_{r=0}^{n} r P(X = r) = \sum_{r=0n} \frac{r\binom{n}{r}}{2^n} = \frac{n}{2},$$

the last step by intuition about coin flips. (Notation: $(X = r)$ means $\{a | X(a) = r\}$). Can prove this intuition by knowing $r\binom{n}{r} = n\binom{n-1}{r-1}$ (✎), sum above

$$= n\frac{1}{2^n} \sum_{r=1}^{n} \binom{n-1}{r-1} = n\frac{2^{n-1}}{2^n} = \frac{n}{2}.$$

For Y_i the indicator of event ith coin is H, $X = \sum Y_i$,

$$E(X) = \sum E(Y_i) = \sum_{i=0}^{n} P(Y_i) = \frac{n}{2},$$

so indicator functions nicer.

5 Graph Theory

A *graph* is a set of vertices and edges, for the moment unordered pairs of vertices, called an *undirected graph*. Relation on V is adjacency: $v, w \in V$ are *adjacent* if $\{v, w\} \in E$. The *degree* of vertex x is # of vertices adjacent to x. G is *regular* of degree k if every vertex has degree k. For $k = 1$ it's pairs of points; for $k = 2$ it's a disjoint union of cycles, and for $k = 3$ it's already an infinite set of graphs (trivalent).

Can do some work, convince yourself that:

Theorem 5.1 *If G is regular of degree 3, then $|V|$ is even.*

Proof:

$$\sum_{x \in V} \deg(x) = 2m,$$

where m will always stand for $|E|$.

Call the fact that $\sum_{x \in V} \deg(x) = 2m$ the "handshake theorem." Call K_n the complete graph on n vertices, $m = \binom{n}{2}$. $\overline{K}_n$ the empty graph, $m = 0$. For every graph, $0 \leqslant m \leqslant \binom{n}{2}$.

The *complement* of $G = (V, E)$ is $\overline{G} = (V, \overline{E})$, where $\{x, y\} \in \overline{E} \iff x, y \in V$, $x \neq y$, and $\{x, y\} \notin E$.

Bipartite graph: vertices can be colored red and blue such that adjacent vertices never have the same color. Ex: 6 vertices in a hexagon, put in 3 diagonals intersecting at center. A bipartite graph cannot contain a cycle of length 3, i.e. $K_3 = C_3$. Cycle C_n is bipartite $\iff$ n is even. So, generalization:

Theorem 5.2 *G is bipartite $\iff$ G contains no odd cycles.*

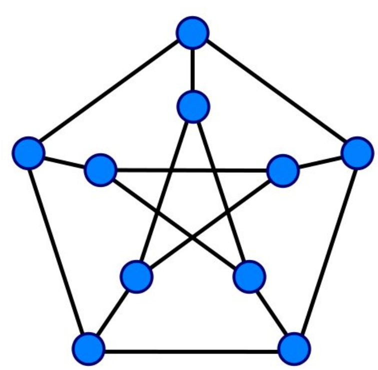
We've done "only if" step. *Walk* of length n in a graph: $v_0 - v_1 - \ldots - v_n$ such that $\{v_{i-1}, v_i\} \in E$, $i = 1, \ldots, n$. A *path* in E is a walk without repeated vertices. Write the number of walks of length k as $\partial_n K_n$ (in the complete graph). If G is regular of degree d, the # walks of length k is nd^k. The number of paths of length k is $n(n-1)\cdots(n-k)/2$.

The *complete bipartite graph $K_{k,l}$* looks like a line of k red next to a line of l blue, edge between every red and blue. $n = k + l$, $m = kl$.

Definition: An *isomorphism* between $G = (V, E)$ and $H = (W, F)$ is a bijection $f : V \to W$ which preserves adjacency: $(\forall v_1, v_2 \in V)$, $v_1 \sim_G v_2 \iff f(v_1) \sim_H f(v_2)$.

Definition: G and H are *isomorphic* if $\exists f : G \to H$ an isomorphism.

It's an open problem whether you can prove non-isomorphism in polynomial time. A graph that is often used as a counterexample is *Petersen's graph*:



Length of shortest cycle is *girth*, *diameter* is $\max_{x,y \in V} \operatorname{dist}(x, y)$, distance (x, y) is length of shortest path from x to y (∞ if $\nexists$ such path). Petersen's graph has girth 5, diameter 2, regular of degree 3.

✎**:** If G has girth 5 and is regular of degree r then $n \geqslant r^2 + 1$.

✎**:** If G has diam=2 and is regular of deg=r then $n \leqslant r^2 + 1$.

Gives a funky graph (get fm someone),

✎**:** This is isomorphic to Petersen's graph.

Definition: $y \in V$ is *accessible* from $x \in V$ if $\exists x \leftrightarrow y$ path.

x acc y, x acc $y \implies y$ acc x, transitive:

✎**:** Prove: if $\exists x \ldots y$ walk then $\exists x \ldots y$ path.

Definition: The equivalence classes of "accessibility" are the *connected components* of G.

Definition: G is *connected* if $\forall x, y$, x acc y, i.e. there is just 1 connected component

Definition: G is a *tree* if G is connected and has no cycles.

Example: P_n, line of n nodes, $m = n - 1$.

Example: star$_n$, one node in middle, $n - 1$ around it in circle, connected to center node. $m = n - 1$.

Proof: By induction on n. Wrong proof: $n - 1$ vertices, just add one more. But:

Lemma 5.3 (1) *Every tree has a vertex of degree 1 ($n \geqslant 2$) (✎)*

I.H.: true for $n - 1$ vertices, D.C. ". Let x be a vertex of degree 1 in tree T with n vertices. Remove it: get graph T', has $n - 1$ vertices, T has no cycles, T' is connected.

Lemma 5.4 (2) *If G is connected, deg$(x) = 1$, then $G \backslash x$ is connected. (✎)*

Say a *legal coloring* is $f : V \to \{\text{colors}\}$ such that $(\forall x, y \in V)(x \sim y \implies f(x) \neq f(y))$. G is *k-colorable* if $\exists$ a legal coloring with $\leqslant k$ colors. The *chromatic number* $\chi(G) := \min\{k \mid G \text{ is } k-\text{colorable}\}$. A graph is bipartite $\iff$ 2-colorable. $\chi(G) = 1 \iff G = \bar{K}_n$, $\chi(K_n) = n$, Do $\chi(G) = n \iff G = K_n$.

Theorem 5.5 (Kuratowski's Theorem) *G is planar $\iff$ G has no K_5 or $K_{3,3}$.*

Definition: A *clique* is a complete subgraph. $\omega(G)$ is the size of the largest clique. $\chi(G) \geqslant \omega(G)$.

Definition: The *independence number* is $\alpha(G)$ and is the size of the largest independent set in a graph. A set of vertices is a subset $A \subseteq G$ such that no two vertices are adjacent. Also, $\alpha(G) = \omega(\bar{G})$.

Definition: A *plane graph* is a plane drawing of a graph without any intersections.

Definition: A *multigraph* is a graph that also allows loops (self-edges) and parallel edges (multiple edges between a pair of vertices).

Note that the handshake lemma remains valid ($2m = \sum deg(n)$).

Definition: *Regions* are connected components of the complement of the plane graph.

Theorem 5.6 (Dual handshake) *number of sides of a region (r) = 2 * number of edges (m)*

The dual plane graph is the set of connected points between regions, going over each of the edges. Note that duals can introduce multigraphs even from a simple graph. Trees have one region. Their dual will be a vertex with $n - 1$ loops (edges).

Theorem 5.7 (Euler's Formula) *For a connected plane graph, $n - m + r = 2$*

Can prove by induction on $n + m$, but need to use the Jordan Curve Theorem, which is too advanced for this class.

Count the number N of trees on n vertices, drawing pictures: $N(2) = 1$, $N(3) = 3$, $N(4) = 16$. Count paths of length k in K_n:

$$\frac{n(n-1)(n-2)\cdots(n-k+1)}{2}$$

Shows $N(5) = 125$. These all suggest one formula:

Theorem 5.8 (Cayley's Formula) *The number of spanning trees of K_n is n^{n-2}.*

Proof: Bijective: Encode every spanning tree by a string of length $n-2$ over an alphabet of size n. "Pr̀ufer code": MN, Wiki.

Another proof: figure here, prescribe: vertex i has degree $d_i \geqslant 1$, and $\sum_{i=1}^{n} d_i = 2n-2$, by the handshake theorem. Then

Theorem 5.9 *Suppose $d_1, \ldots, d_n$ satisfy these conditions, then then the number of trees with these degrees on vertex set $[n] = \{1, \ldots, n\}$ is*

$$\frac{(n-2)!}{\prod_{i=1}^{n}(d_i - 1)!}$$

Proof: Proof: by induction.

Lemma 5.10 *If $d_1, \ldots, d_n$ satisfy the constraints, then $\exists i$, $d_i = 1$.*

Proof: Suppose false: $\forall i$, $d_i \geqslant 2 \implies \sum d_i = (2n-2) \geqslant 2n$, $\Rightarrow\Leftarrow$. Look at vertex n, then

$$N(d_1, \ldots, d_n) = \sum_{i=1, d_i \neq 1}^{n-1} N(d_1, \underbrace{\ldots}_{\substack{\uparrow \\ d_i - 1}}, d_{n-1}) = \frac{(n-3)!}{\prod (d_i - 1)!} \left(\sum_{i=1}^{n-1} (d_i - 1)\right) = \frac{(n-2)(n-3)!}{\prod ()!} = \checkmark$$

Then, proof of Cayley's formula:

$$\#\text{sp. trees of } K_n = \prod_{d_i \geqslant 1, \sum d_i = 2n-2} N(d_1, \ldots, d_n) = \sum_{d_i \geqslant 1, \sum d_i = 2n-2} \frac{(n-2)!}{\prod (d_i - 1)!} = (\underbrace{1 + \cdots + 1}_{n})^{n-2} = n^{n-2},$$

last bit by the multinomial theorem. Note: $\sum (d_i - 1) = \sum d_i - n = (2n-2) - n = n-2$.

Count n digit integers of which (1) all digits are odd, (2) all odd digits occur.

(1): 5^n

(1)+(2): $5^n - 5 \cdot 4^n + \binom{5}{2} \cdot 3^n - \binom{5}{3} 2^n + \binom{5}{4} 1^n$

This is a special case of:

5.1 Inclusion-Exclusion

Universe Ω, subsets $A_1, \ldots, A_k$, given $|\cap_{i \in I} A_i|$ for all $I \subseteq [k]$, want to find $|B|$, where $B = \overline{A_1 \cup \ldots \cup A_k}$.
$|B| = S_0 - S_1 + s_2 - + \ldots$, $p_i := \frac{|S_i|}{|\Omega|}$, uniform dist.
Answer (Inclusion-Exclusion formula):

$$\begin{aligned} S_0 &= |\Omega| \\ S_1 &= |A_1| + \cdots + |A_k| \\ S_2 &= |A_1 \cap A_2| + |A_1 \cap A_2| + \cdots + |A_{k-1} \cap A_k| \\ &\vdots \\ S_j &= \sum_{1 \leqslant i_1 < \cdots < i_j \leqslant k} |A_{i_1} \cap \cdots \cap A_{i_j}| \end{aligned}$$

and number of terms in S_j is $\binom{k}{j}$.

Example: $|\Omega| = 5^n$ strings with digits $1,2,3,4,5$, A_i set of those that miss ith digit, $A_i \cap A_j$. $5^n - 5 * 4^n + \binom{5}{2} * 3^n - \binom{5}{3} * 2^n + \binom{5}{4} * 1^n$.

Proof: For any x (diagram), look at $r(x) = \#\{i \mid x \in A_i\} = r$ and $c(x)$, the contribution of x to $S_0 - S_1 + \cdots$. Need to prove:

$$c(x) = \begin{cases} 1 \text{ if } x \in B, \text{ i.e. } r(x) = 0 \\ 0 \text{ if } x \notin B, \text{ i.e. } r(x) = 1 \end{cases}$$

Now,

$$c(x) = 1 - r + \binom{r}{2} - \binom{r}{3} + \cdots = (1-1)^r = 0^r = \begin{cases} 1 \text{ if } r = 0 \\ 0 \text{ if } r \geqslant 1 \end{cases}$$

More general version (we only have over uniform distribution): over *any* probability distribution:

Theorem 5.11 *If $A_1, \ldots, A_k$ are events and p_i is defined by (*), then* $P(B) = p_0 - p_1 + p_2 - + \ldots = \sum_{I \subseteq [k]} (-1)^{|I|} P(\bigcap_{i \in I} A_i)$.

where (*) is

$$\begin{aligned} p_0 &= P(\Omega) = 1 \\ p_1 &= \sum P(A_i) \\ p_2 &= \sum P(A_i \cap A_j) \\ &\vdots \end{aligned}$$

(general).

✎**:** Adapt previous proof

A, B events, I_A indicator variable:

$$I_A(x) = \begin{cases} 1 \text{ if } x \in A \\ 0 \text{ if } x \notin A \end{cases}$$

✎**:**

$$\begin{aligned} I_{A \cup B} &= I_A I_B \\ I_{\overline{A}} &= 1 - I_A \end{aligned}$$

Now

$$\begin{aligned} (1+x_1)(1+x_2)\cdots(1+x_n) &= \sum_{I \subseteq [n]} \prod_{i \in I} x_i \\ (1-x_1)(1-x_2)\cdots(1-x_n) &= \sum_{I \subseteq [n]} (-1)^{|I|} \prod x_i \end{aligned}$$

Now,

$$B = \overline{A_1 \cup \cdots \cup A_k} = \overline{A_1} \cap \cdots \cap \overline{A_k},$$

$$I_B = \prod I_{\overline{A_i}} = \prod_{i=1}^{k} (1 - I_{A_i}) = \sum_{I \subseteq [k]} (-1)^{|I|} \prod_{i \in I} I_{A_i} = \sum_{I \subseteq [k]} (-1)^{|I|} I_{\cap_{i \in I} A_i}$$

By linearity of expectation,

$$P(B) = E(I_B) = \sum_{I\subseteq[k]} (-1)^{|I|} E(I_{\cap_{i\in I} A_i} = \sum_{I\subseteq[k]} (-1)^{|I|} P(\cap_{i\in I} A_i)$$

Application 1: explicit formula for Euler's ϕ function:

$$n = \prod_{i=1}^{n} p_i^{k_i}, \quad \phi(n) = n\prod_{i=1}^{t}(1-\frac{1}{p_i}).$$

Proof: $\Omega = [n]$, $A_i \subseteq \Omega$ = set of numbers divisible by p_i, $B = \overline{\cup A_i} = \{j : \gcd(j,n) = 1\}$, $\phi(n) = |B|$.
$|A_i| = \frac{n}{p_i}$, $P(A_i) = \frac{1}{p_i}$, $|A_i \cap A_j| = \frac{n}{p_i p_j}$, $P(A_i \cap A_j) = \frac{1}{p_i p_j}$, uniform distribution, $P(B) = \frac{|B|}{n}$.,

$$P(B) = \sum_{|I|\subseteq[t]} (-1)^{|I|} P(\cap_{i\in I} A_i) = \sum (-1)^{|I|} \frac{1}{\prod_{i\in I} p_i} = \sum_{I\subseteq[t]} (-1)^{|I|} \prod_{i\in I} \frac{1}{p_i} = \prod (1 - \frac{1}{p_i}) \checkmark$$

Application: "derangement problem": probability that random permutation is a derangement $\sim \frac{1}{e}$. In MN, read "Hatcheck Lady & Co."

Now, back to random variables.

Definition: X, Y random variables are *independent* if

$$(\forall x, y \in \mathbb{R})(P(X = x, Y = y) = P(X = x)P(Y = y))$$

If $E(XY) > E(X)E(Y)$, X and Y are *positively correlated*, if $E(XY) < E(X)E(Y)$ they're *negatively correlated*, if equal then they're *uncorrelated*.

Note that independence $\implies$ uncorrelated but not the other way around.

Corollary 5.12 *If $P(y = y) \neq 0$, then $P(X = x) = P(x = x \mid Y = y)$.*

✎**:** If X, Y independent, then $E(XY) = E(X)E(Y)$.

✎**:** Events A, B independent $\iff$ I_A, I_B are independent.

Definition: Random variables $X_1, \ldots, X_k$ *independent* (fully independent, mutually independent, collectionwise independent) if

$$(\forall x_1, \ldots, x_k \in \mathbb{R})(P((\forall i)(X_i = x_i)) = \prod_{i=1}^{k} P(X_i = x_i)$$

✎**:** Events $A_1, \ldots, A_k$ are independent $\iff$ $I_{A_1}, \ldots, I_{A_k}$ independent.

Theorem 5.13 *If $X_1, \ldots, X_k$ are independent, then*

$$E(\prod_{i=1}^{k} X_i) = \prod_{i=1}^{k} E(X_i)$$

✎**:** If X, Y, Z, W, T are independent random variables, then $X + Y$, $\cos(Z - W)$, e^T are independent.

✎**:** If $X - 1, \ldots, X_k$ are independent random variables and $[k] = I_1 \uplus \cdots \uplus I_t$ partition and $f_1, \ldots, f_t$ are functions and f_i has $|I_i|$ variables, then $f_1(X_i : i \in I_1)$, $\ldots$, $f_t(X_i : i \in I_t)$ are independent random variables.

The *covariance* of X and Y is

$$\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y)$$

X and Y are *positively correlated* if $\mathrm{Cov}(X,Y) > 0$, *negatively correlated* if $\mathrm{Cov}(X,Y) < 0$, *uncorrelated* if $\mathrm{Cov}(X,Y) = 0$.

The *variance*,or *second moment*, is

$$\begin{aligned}
\mathrm{Var}(X) &= E[(X-m)^2], \quad m = E(X) \\
&= E[X^2 + m^2 - 2Xm] \\
&= E[X^2] + E[m^2] - 2E[mX] \\
&= E[X^2] - m^2 \\
&= E[X^2] - (E[X])^2
\end{aligned}$$

Have

$$\mathrm{Var}(\sum_i X_i) = \sum_i \mathrm{Var}(X_i) + 2\sum_{i<j} \mathrm{Cov}(X_i, X_j).$$

For X_i a random variable, $X = \sum X_i$,

$$E[X] = E[\sum_i X_i] = \sum_i E[X_i].$$

Also,

$$\mathrm{Var}(\sum_i X_i) = E[(\sum X_i)^2] - (E[\sum X_i])^2$$

✎**:** $(x_1 + \cdots + x_n)^2 = \sum_i \sum_j x_i x_j = \sum_i x_i^2 + 2\sum_{i<j} x_i x_j$

Now,

$$\begin{aligned}
\mathrm{Var}(\sum_i X_i) &= E[(\sum X_i)^2] - (E[\sum X_i])^2 \\
&= E[\sum_i X_i^2 + 2\sum_{i<j} X_i X_j] - (\sum_i E(X_i))^2 \\
&= \sum_i E(X_i^2) + 2\sum_{i<j} E(X_i X_j) - (\sum_i E(X_i)^2 + 2\sum_{i<j} E(X_i)E(X_j)) \\
&= \sum_i (E(X_i^2) - E(X_i)^2) + 2\sum_{i<j}(E(X_i, X_j) - E(X_i)E(X_j)) \\
&= \sum_i \mathrm{Var}(X_i) + 2\sum_{i<j} \mathrm{Cov}(X_i, X_j)
\end{aligned}$$

$\mathrm{Var}(\sum X_i) = \sum \mathrm{Var}(X_i)$ and standard deviation (SD) is $\sigma(x) = \sqrt{\mathrm{Var}(X)}$.

$$\mathrm{Var}(X) = E(X^2) - E(X)^2 \geqslant 0 \implies E(X^2) \geqslant E(X)^2$$

The last inequality is the *Cauchy-Schwartz Inequality*, perhaps in a different form than you've seen before. Another representation of Cauchy-Schwartz:

$$\begin{aligned}
(\sum X_i Y_i)^2 &\leqslant (\sum X_i^2)(\sum Y_i^2) \\
E(XY)^2 &\leqslant E(X^2)E(Y^2)
\end{aligned}$$

Setting $a = X_i/\sqrt{E(X_i^2)}$, $b = Y_i/\sqrt{E(Y_i^2)}$,

$$\begin{aligned}
a^2+b^2 \geqslant 2ab \quad &\Longrightarrow \quad E(a^2)+E(b^2) \geqslant 2E(ab) \\
&\Longrightarrow \quad E(\frac{X_i^2}{E(X_i^2)}) + E(\frac{Y_i^2}{E(Y_i^2)}) \geqslant 2E(\frac{X_iY_i}{E(X_i^2)E(Y_i^2)}) \\
&\Longrightarrow \quad 1+1 \geqslant 2\frac{E(X_iY_i)}{\sqrt{E(X_i^2)Y(Y_i^2)}}
\end{aligned}$$

A *Bernoulli trial* means tossing a biased coin: H with prob. p, tails with prob $1-p$. A *k-Bernoulli trial* is a completely independent set of k Bernoulli trials, $X_k = P[n-\text{Bernoulli trial will have k heads}]$.

$$P(X_k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Let Y_i be the outcome of the ith Bernoulli trial (=1 if H, 0 if T). Define $X = \sum Y_i$, get

$$E[X] = E[\sum Y_i] = \sum_i E[Y_i] = \sum_i p = np$$

$$\text{Var}[X] = \text{Var}[\sum Y_i] = \sum_i \text{Var}[Y_i] = \sum_i [E[Y_i^2] - E[Y_i]^2] = n(p-p^2) = np(1-p)$$

The *weak law of large numbers* says that for $\epsilon, p > 0$ fixed,

$$P[|X^n - np| > \epsilon(np)] \to_{n\to\infty} 0$$

To prove, this need the *Markov inequality*: for η a random variable, non-negative,

$$P[\eta > a] \leqslant \frac{E[\eta]}{a}$$

Proof:

$$E[\eta] = \sum_i \mu_i P(\eta = \mu_i) \geqslant \sum_{\mu_i > a} \mu_i P(\eta = \mu_i) > a \sum_{\mu_i > a} P(\eta = \mu_i) = aP[\eta > a]$$

This is the first of the so-called concentration lemmas. Another is Chebyshev's inequality:

$$P[|\eta - m| > a] \leqslant \frac{\text{Var}(\eta)}{a^2}, \quad m = E(\eta)$$

Proof:

$$P[|\eta - m| > a] = P[(n-m)^2 > a^2] \leqslant \frac{E[(\eta-m)^2]}{a^2} = \frac{\text{Var}(\eta)}{a^2}$$

C's inequality proves the WLLN, because

$$\Longrightarrow P[|X^n - np| > \epsilon np] \leqslant \frac{np(1-p)}{\epsilon^2 n^2 p^2} = \frac{(1-p)}{\epsilon^2 pn} \to 0$$

as $n \to \infty$. Now, some problems.

A *random graph* on n vertices: for each (v_i, v_j) toss an unbiased coin to decide whether the edge is in the graph. Then want to show that

$$P[\text{All vertices in the graph has degree close to } \frac{n}{2}] \to_{n\to\infty} 1$$

The expected number of neighbors of a vertex is $(n-1)/2$.

$$\begin{aligned}
P[\forall \sigma \in V, |N(v) - \frac{n-1}{2}| \leqslant \frac{\epsilon(n-1)}{2}] \quad &\to \quad 1 \\
P[\exists v \in V, |N(v) - \frac{n-1}{2}| > \frac{\epsilon(n-1)}{2}] \quad &\to \quad 0
\end{aligned}$$

Y_σ is probability that for vertex v, $|N(v) - \frac{n-1}{2}| > \frac{\epsilon(n-1)}{2}$, want $P[\bigcup_v Y_v] \to 0$. The *union bound* is

$$P[\bigcup_v Y_v] \leqslant \sum_v P[Y_v] \downarrow 0$$

($N(v)$ is the degree of vertex v) Now

$$\begin{aligned} P[Y_\sigma] &= P[|N(v) - \frac{n-1}{2}| \geqslant \frac{\epsilon(n-1)}{2}] \\ &\leqslant \frac{\mathrm{Var}(Y_\sigma)}{(\frac{\epsilon(n-1)}{2})^2} = \frac{n}{4\epsilon^2 \frac{n^2}{4}} = \frac{1}{\epsilon^2 n} = O(\frac{1}{n}) \end{aligned}$$

So here, Chebyshev isn't enough. Need *Chernoff bound*: let η_i be 1 with prob. $\frac{1}{2}$, -1 with prob. $\frac{1}{2}$, $E[\eta_i] = 0$. Then

$$P[\sum_{i=1}^n \eta_i > a] \leqslant \exp(-\frac{a^2}{2n})$$

We can adapt this slightly for our purposes, take η_i =1 w.p $\frac{1}{2}$, = 0 otherwise,

$$P[|\sum_{i=1}^n \eta_i - \frac{n}{2}| > \frac{\epsilon n}{2}] \leqslant \exp(-(\frac{\epsilon^2 n^2}{2n})) = \exp(-O(n))$$

5.2 Random graph

With high probability,[5] when the last vertex is reached by an edge, the graph is connected. There's a whole theory of random graphs. Whole subject stems from one original paper: Erdös & Rényi 1960, "Evolution of random graphs." Choosing m edges at random, $|\Omega| = \binom{\binom{n}{2}}{m}$. Note that if know one edge there, any other edge less likely to occur, so they're negatively correlated. In another model, edges are thrown in independently with probability p, then $m = \binom{n}{2}p = E(\#\,\text{edges})$. The first model is referred to as $G_{n,m}$ model, second is $G_{n,p}$ model, much more studied.

Most frequently studied in an introduction to random graphs is $G_{n,\frac{1}{2}}$.

Definition: The *diameter* of G is $\mathrm{diam}(G) = \max_{x,y\in V} \mathrm{dist}(x,y)$. The *distance* between $x, y \in V$ is $\mathrm{dist}(x,y) = \min \mathrm{length}(x-y \text{ path})$.

For example, $\mathrm{diam}(K_n) = 1$, longest path is $K_n = n-1$.

Theorem 5.14 *Almost all graphs have diameter 2, meaning if $p_n = P(diam(G_{n,\frac{1}{2}}) = 2)$, then* $\lim_{n\to\infty} p_n = 1$

In fact, diam $\neq 2$ is exponentially unlikely: $1 - p_n < C^n$ for some constant $0 < C < 1$, find $< 0.76^n$ ✓.

✎**:** $\forall p > 0$ constant, $P(\mathrm{diam}(G_{n,p}) = 2) \to 1$.

Let

$$\begin{aligned} g_n &:= P(\underbrace{\mathrm{diam}(G_{n,\frac{1}{2}}) \geqslant 3}_{A_n}) \\ r_n &:= P(\underbrace{(\exists x \neq y \in V)((\nexists z)(x \sim z \sim y))}_{B_n}) \end{aligned}$$

$A_n \implies B_n$ because if diam $\geqslant 3$ then $\exists x, y$ such that $\mathrm{dist}(x,y) \geqslant 3$.

[5] Also written "w.h.p." Means in some limit, the probability of an event A occurring is one, which is different from event A *always* occurring. When a coin is flipped n times, the probability a head comes up at least once is small but finite. As $n \to \infty$, $P(\text{at least one H}) = 1$, even though the infinite sequence TTTT... could occur.

Claim: $r_n \to 0$ at an exponential rate.

Proof: For $x \neq y \in V$, $A(x,y) =$ "x, y have no common neighbor"$= \bigcap_{z \neq x,y}$ "z is not a common neighbor", which are $n-2$ independent events. To see this, fix x, y, z, $P(x \sim z, y \sim z) = \frac{1}{4}$, z is a common neighbor, so $P(z$ is not a common neighbor of $x, y) = \frac{3}{4}$, and

$$P(A(x,y)) = (\frac{3}{4})^{n-2}$$

$$P(\underbrace{(\exists x \neq y)}_{\binom{n}{2} choices} A(x,y)) = P(\bigcup_{x \neq y}^{\binom{n}{2} A(x,y)} \underbrace{<}_{unionbd} \binom{n}{2}(\frac{3}{4})^{n-2} < (\frac{3}{4}+\epsilon)^n$$

for $n \geqslant n_0$, $\forall \epsilon > 0 \exists n_0$., where last $<$ is œ.

5.3 Digraphs

Directed graphs. A *digraph* is a relation on V, $E \subseteq V \times V$. We can use graphs as digraphs by replacing an edge with $\rightleftarrows$.

A *directed walk* is $v_0 \to v_1 \to \ldots \to v_k$ of length k, a *directed path* has no repeated vertices, a cycle has $v_0 = v_k$. y is *accessible* from x if $\exists x \to \cdots y$ path (means directed path). Say "y is accessible from x", $k = 0$.

✎: Accessibility is a transitive relation.

Definition: x, y are *mutually accessible* if x is accessible from y and vice versa.

✎: This is an equivalence relation.

Definition: The *strong components* are equivalence classes.

✎: The strong components form a poset under accessibility.

Have sources and sinks, nice figures here.

Definition: *Weakly connected* means "connected" if we ignore orientation.[6]

The *adjacency matrix* of $G = (V,E)$ is a $(0,1) matrix$, $V = [n]$, $A = (a_{ij})$,

$$a_{ij} = \begin{cases} 1 & \text{if } i \to j \\ 0 & \text{otherwise} \end{cases}$$

The transpose of this is $B = A^T$, $b_{ij} = a_{ji}$. G^{reverse} corresponds to adjacent matrix A^T.

Definition: A symmetric matrix is $A = A^T$ ($\iff$ undirected, loops permitted).

A DAG is a "directed acyclic graph" (no cycles) .

✎: Prove G is a DAG $\iff$ has a topological sort.

Now, A is $k \times l$, B is $l \times n$, $C = AB = (c_{ij})$,

$$c_{ij} = \sum_{t=1}^{l} a_{it} b_{tj}$$

For A an adjacency matrix of $G = (V,E)$, $A^2 = (b_{ij})$, $n \times n$, $b_{ij} = \sum_{t=1}^{n} a_{it} a_{tj}$, so $b_{ij} = \#$ 2-step $i \to j$ walks.

[6] "Connected for pedestrians, not for automobiles. Or bikes. I like riding my bike the wrong way.. at least I know who is hitting me."

✎: $A^k = (c_{ij})$, $c_{ij} = \#$ of k-step walks $i \to \cdots \to j$.

A *discrete stochastic process* is a set of states B and transitions between states. A *finite Markov chain* is a finite set of states and fixed transition probabilities, $V = [n]$, $p_{ij} = P(X_{t+1} = j \mid X_t = i)$, X_t is location of particle at time t. Let $T = (p_{ij})$ be the transition matrix of a finite Markov chain, then $T^2 = (g_{ij})$, $g_{ij} = \sum_{l=1}^{n} p_{il}p_{lj}$.

✎: $p_{il}p_{lj} = P(X_{t+1} = l \text{ and } X_{t+2} = j \mid X_t = i)$

So, $= P(X_{t+2} = j \mid X_t = i)$, 2-step transition probabilities

✎: $T^k = (p_{ij}^{(k)})$, $p_{ij}^{(k)} = P(X_{t+k} = j \mid X_t = i)$, k-step transition probabilities

If $T^k \approx$ uniform $= \frac{1}{n}J$, J is matrix with all 1's.

Observation: Every row of T is $\geqslant 0$ and sums to 1: $\sum_{j=1}^{n} p_{ij} = 1$.

Definition: T is a *stochastic matrix* if $t_{ij} \geqslant 0$ and $\sum_{i=1}^{n} t_{ij} = 1$.

Have a digraph associated with T: $a_{ij} = 1 \iff p_{ij} > 0$. One interesting question is whether it's strongly connected.

5.4 Matrix theory and applications to digraphs and finite Markov chains

Entries of A^k count k-step walks $i \to \cdots \to j$. T is a *stochastic matrix* if $p_{ij} \geqslant 0$, $\sum_{j=1}^{n} p_{ij} = 1$, the row sums.

The *adjacency matrix* has entries

$$a_{ij} = \begin{cases} 1 & i \to j \\ 0 & \text{otherwise} \end{cases}$$

The *transition matrix* $T = (p_{ij})$,

$$p_{ij} = P(X_{t+1} = j \mid X_t = i)$$

The entries of $T_k = (p_{ij}^{(k)})$ are the k-step transition probabilities.

Definition: For A an $n \times n$ matrix with real or complex entries, $\lambda \in \mathbb{R}$ or $\mathbb{C}$, $\boldsymbol{x} \in \mathbb{R}^n$ or $\mathbb{C}^n$, $\boldsymbol{x} = [x_1 \cdots x_n]^T$. $\boldsymbol{x}$ is a *right eigenvector* of A to *eigenvalue* λ if $\boldsymbol{x} \neq 0 = [0 \cdots 0]^T$ and $A\boldsymbol{x} = \lambda\boldsymbol{x}$. For $\boldsymbol{y} = [y_1 \cdots y_n]^T$, $\boldsymbol{y}^T = \lambda\boldsymbol{y}$, $\boldsymbol{y}^T$ is a *left eigenvector.* λ is a *right eigenvalue* of A if $\exists$ corresponding right eigenvectors.

Theorem 5.15 *Right $\iff$ left eigenvalues.*

F is the "field of scalars", $F = \mathbb{R}$ or $\mathbb{C}$.

Definition: The vectors $\boldsymbol{x}_1 = [x_{11} \cdots x_{1n}]^T, \ldots, \boldsymbol{x}_k = [x_{k1} \cdots x_{kn}]^T$ are *linearly independent* if only their trivial linear combination is zero, where

$$\lambda_1\boldsymbol{x}_1 + \cdots + \lambda_k\boldsymbol{x}_k = 0$$

is a linear combination, $x_{ij} \in F$, $\lambda_i \in F$, $\lambda_1 = \cdots = \lambda_k = 0$ is the trivial linear combination.

Definition: $\operatorname{rank}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k)$ = maximum number of linearly independent vectors among the $\boldsymbol{x}_i$.

Comments: If $(\exists i)(\boldsymbol{x}_i = \mathbf{0})$ then $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ are *not* linearly independent,

$$0\boldsymbol{x}_1 + \cdots + 1\boldsymbol{x}_i + \cdots_0 \boldsymbol{x}_n = 0.$$

If $(\exists i \neq j)(\boldsymbol{x}_i = \boldsymbol{x}_j)$, then

$$\underbrace{\cdots}_{0} + 1\boldsymbol{x}_i + \underbrace{\cdots}_{0} + (-1)\boldsymbol{x}_j + \underbrace{\cdots}_{0} = \mathbf{0}$$

✎**:** In $\mathbb{R}^n$ find $n+1$ vectors such that every $\boldsymbol{n}$ of them are linearly independent

For A a $k \times l$ matrix over F, the *column rank* of A is $\operatorname{rank}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_l)$ and the *row rank* of A is $\operatorname{rank}(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_k)$, where $\boldsymbol{a}_i$ is the ith column of A, r_i is the ith row of A. By definition, rowrank(A)=colrank(A). But:

> **Miracle # 1 (of linear algebra)**: If $S \subseteq F^n$, then every maximal (=nothing can be added to preserve linear independence) linearly independent subset of S is maximum (=largest).
> **Miracle # 2**: colrank=rowrank, i.e. colrank(A)=colrank(A^T).

Definition: For S a set of vectors, span(S)=set of all linear combinations of S.

Obs: $\forall S \subseteq F^n$, $0 \in \operatorname{Span}(S)$ even if $S = \varnothing$.

Definition: $U \subseteq F^n$ is a *subspace* if $(0 \in U)$ and U is closed under linear combinations.

✎**:** Span(S) is always a subspace.

Definition: If $U \subseteq F^n$, $\dim(U) = \operatorname{rank}(U)$.

If $\dim U = d$ then $\exists d$ linearly independent vectors in U and no more.
For $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_d$,

Claim: $\operatorname{Span}(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_d) = U$

Proof: Let $\boldsymbol{x} \in U$. NTS: $\boldsymbol{x} \in \operatorname{Span}(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_d)$, i.e., $\exists \lambda_1, \ldots, \lambda_d \in F$ such that $x = \lambda_1 \boldsymbol{b}_1 + \cdots + \lambda_d \boldsymbol{b}_d$.

We know that $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_d, \boldsymbol{x}$ are linearly dependent, i.e. $\exists \alpha_1, \ldots, \alpha_d, \alpha_{d+1}$, not all zero, such that $\sum_{i=1}^{d} \alpha_i \boldsymbol{b}_i + \alpha_{d+1}\boldsymbol{x} = 0$.

Claim: $\alpha_{d+1} \neq 0$ because $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_d$ are linearly independent, so $\sum(-\frac{\alpha_i}{\alpha_{d+1}})\boldsymbol{b}_i = \boldsymbol{x}$✓

Definition: A *basis* of a set of vectors S is a linearly independent set of vectors in S which spans S, i.e. $S \subseteq$ Span(those vectors)

Example: Column-basis of a matrix A.

Theorem 5.16 *Every maximal linearly independent subset of S is a basis of S*

Proof: ✎

Theorem 5.17 $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k \in U$ *(subspace) is a basis of U if and only if every $\boldsymbol{x} \in U$ can be written as a unique linear combination of $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k$.*

Obs: If $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k$ are linearly independent and $\sum \alpha_i \boldsymbol{a}_i = \sum \beta_i \boldsymbol{a}_i \implies (\forall i)(\alpha_i = \beta_i)$

Proof: $\sum(\alpha_i - \beta_i)\boldsymbol{a}_i = \mathbf{0}$

Obs: Conversely, if $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k$ are linearly dependent, then every vector in $\operatorname{Span}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k)$ can be written as a linear combination in more than one way.

Proof: $0 = 0\boldsymbol{a}_1 + \cdots + 0\boldsymbol{a}_k = \lambda_1\boldsymbol{a}_1 + \cdots + \lambda_i\boldsymbol{a}_k$ with not all $\lambda_i = 0$. Suppose now $\boldsymbol{x} = \sum \alpha_i\boldsymbol{a}_i = \sum(\alpha_i + \lambda_i)\boldsymbol{a}_i$.

The standard basis of F^n is $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$ (defined as usual, I'm not writing out). This is a basis because

$$\sum \alpha_i\boldsymbol{e}_i = [\alpha_1 \cdots \alpha_n]^T \in F^n,$$

and this decomposition is unique.

Corollary 5.18 $\dim F^n = n$

✎: If $\boldsymbol{a}_1, \dots, \boldsymbol{a}_k \in S$ are linearly independent, then this can be extended to a basis.

Definition: $\boldsymbol{a}_1, \dots, \boldsymbol{a}_m$ *generate* U if $U = \mathrm{Span}(\boldsymbol{a}_1, \dots, \boldsymbol{a}_n)$.

✎: If $\boldsymbol{a}_1, \dots, \boldsymbol{a}_n$ generate U, then $\exists$ a subset of them that is a basis.

Matrix notation: say $A\boldsymbol{x} = \boldsymbol{b}$ for A a $k \times l$ matrix, $\boldsymbol{x} \in F^l$, $\boldsymbol{b} \in F^k$, denotes a system of linear equations

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1l}x_l &= b_1 \\ &\vdots \\ a_{k1}x_1 + \cdots, + a_{kl}x_l &= b_k \end{aligned}$$

$A = [\boldsymbol{a}_1, \dots, \boldsymbol{a}_l]$, $\boldsymbol{a}_j$ the jth column of A, $A\boldsymbol{x} = \boldsymbol{x}_1\boldsymbol{a}_1 + \cdots + \boldsymbol{x}_l\boldsymbol{a}_l$ (✎). So in the system of linear equations $A\boldsymbol{x} = \boldsymbol{b}$, we are looking to express $\boldsymbol{b}$ as a linear combination of the columns of A.

$x_1\boldsymbol{a}_1 + \cdots + \boldsymbol{x}_l\boldsymbol{a}_l = \boldsymbol{b}$, $\boldsymbol{x}_i$ unknown,

Corollary 5.19 *$A\boldsymbol{x} = \boldsymbol{b}$ solvable $\iff b \in \mathit{Span}(\boldsymbol{a}_1, \dots, \boldsymbol{a}_k) \iff \mathit{rank}(A) = \mathit{rank}(A \,|\, b)$*

The method to solve a system of linear equations = method to find rank: this is *Gaussian elimination*, READ anywhere, not going to do here.

A *homogeneous* system of linear equations is $A\boldsymbol{x} = \boldsymbol{0}$. $\boldsymbol{x} = \boldsymbol{0}$ is always a solution (trivial solution). A nontrivial solution exists $\iff \boldsymbol{a}_1, \dots, \boldsymbol{a}_l$ are linearly dependent, $\sum x_i\boldsymbol{a}_i = \boldsymbol{0}$.

Corollary 5.20 *For a $k \times l$ matrix A, the following are equivalent:*

1. *$A\boldsymbol{x} = \boldsymbol{0}$ has no nontrivial solutions.*
2. *The columns of A are linearly independent.*
3. *$\mathit{rank}(A) = l$*
4. *The rows of A span F^l*
5. *A has a* left inverse, *i.e. $\exists B$, $l \times k$ such that $BA = I_l$, the $l \times l$ identity matrix*

✎: Review the proof of the equivalence of (1)-(4)

✎: Show (5)

✎: $\mathrm{rank}(A \cdot B) \leqslant \min\{\mathrm{rank}(A), \mathrm{rank}(B)\}$

✎: Find A, B with rank¿0 such that $A \times B = 0$ (and $A, B \neq 0$).

✎: Find $A \neq 0$ such that $A^2 = 0$

☠ If $F = \mathbb{R}$, then $\mathrm{rank}(A^TA) = \mathrm{rank}(A)$.

Theorem 5.21 *For an $n \times n$ matrix A over F, the following are equivalent:*

1. *$A\boldsymbol{x} = 0$ has no nontrivial solution.*
2. *$\forall \boldsymbol{b} \in F^n$ $(\exists \boldsymbol{x})(A\boldsymbol{x} = b)$*
3. *$(\forall \boldsymbol{b} \in F^n)(\exists! \boldsymbol{x})(A\boldsymbol{x} = \boldsymbol{b})$*
4. *Columns of A are linearly independent.*
5. *Rows "*
6. *Columns span F^n*
7. *Rows span F^n.*
8. *A has a left inverse*
9. *A has a right inverse*
10. *A has a 2-sided inverse*
11. $\det(A) \neq 0$

Name derives from fact that it "determines" whether or not the set of linear equations has a nontrivial solution.

Definition: A is *nonsingular* if $\det(A) \neq 0$

✎**:** Equivalence of all but the last property (det) in the previous theorem.

Definition:

$$det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc$$

Note: If A $n \times n$ then $\det A$ is a sum of $n!$ terms, half +, half -. The *eigenvalue equation* is $A\boldsymbol{x} = \lambda \boldsymbol{x}$, $\boldsymbol{x} \neq 0$, $A\boldsymbol{x} = \lambda I \boldsymbol{x}$, $I = I_n$. So $\lambda$is an eigenvalue $\iff$ $\exists \boldsymbol{x} \neq \mathbf{0}$ such that $(\lambda I - A)\boldsymbol{x} = \mathbf{0}$ $\iff$ $\lambda I - A$ singular.

Theorem 5.22 *λ is an eigenvalue* $\iff$ $\det(\lambda I - A) = 0$.

Fact: $det(A) = det(A^T)$ Example: A as above, eventually get

$$det(\lambda I - A) = \lambda^2 - (a+d)\lambda + (ad - bc),$$

a quadratic equation in λ, say $f_A(t) = det(tI - A)$, this is a polynomial of degree n, the *characteristic polynomial* of A.

Corollary 5.23 *λ is an eigenvalue of A $\iff$ $f_A(\lambda) = 0$, i.e. λ is a root of the characteristic polynomial.*

Corollary 5.24 *Left $\iff$ right eigenvalues the same, because $f_A(t) = f_{A^T}(t)$.*

✎**:** A is stochastic $\iff$ $a_{ij} \geqslant 0$ and 1 is an eigenvalue with right eigenvector $[1 \cdots 1]^T$.

A *group* $(G, \cdot)$, $(G, +)$ has $(\forall a, b \in G)$, $\cdot : G \times G \to G$, $(a, b) \mapsto ab$

1. $(\exists! c \in G)$("$ab = c$")("$a + b = c$")
2. associative: $(ab)c = a(bc)$, $(a + b) + c = a + (b + c)$.
3. identity: $(\exists e)(\forall a)(ae = ea = a)$, e is the *identity*
4. inverse: $(\forall a)(\exists b)(ab = ba = e)$, $b = a^{-1}$, $(\forall a)(\exists b)(a + b = b + a = 0)$, $b = (-a)$.
5. commutativity: $ab = ba$, $a + b = b + a$, if true this is an *abelian group*

Some groups are $(\mathbb{Z}, +)$, $(\mathbb{R}, +)$, $\mathbb{R}^\times = \mathbb{R} \backslash \{0\}$, $(\mathbb{R}^\times, \cdot)$. For $\mathbb{Z}_n$=residue classes mod n, $(\mathbb{Z}_n, +)$ is a group, $(\mathbb{Z}_n, \cdot)$ isn't, define $\mathbb{Z}_n$ = reduced residue classes mod n = residue classes that are relatively prime to n, $|\mathbb{Z}_n^+| = \phi(n)$,

✎**:** $(\mathbb{Z}_n^\times, \cdot)$ is a group.

$GL_n(\mathbb{R})$ is the group of $n \times n$ nonsingular real matrices, $\det \neq 0$, means $\exists$ inverse, full rank. Can have time for $GL_n(\mathbb{F})$, where $\mathbb{F}$ is any field. The identity element is I, the $n \times n$ identity matrix.

✎**:** Give simplest proof that if A, B nonsingular, then AB is nonsingular.

The *symmetric group of degree* n is all permutations of $[n]$, S_n, where a permutation is a bijection

$$f : [n] \to [n] \quad \text{(bijection)}, a \mapsto a^f$$

$\forall a$, $a^{id} = a$, $|S_n| = n!$.

Example:

$$f : \begin{pmatrix} 1\,2\,3\,4\,5 \\ 4\,2\,5\,3\,1 \end{pmatrix} \quad \Longrightarrow \quad f^{-1} : \begin{pmatrix} 4\,2\,5\,3\,1 \\ 1\,2\,3\,4\,5 \end{pmatrix}$$
$$f : \begin{pmatrix} 3\,4\,1\,5\,2 \\ 5\,3\,4\,1\,2 \end{pmatrix} \quad \Longrightarrow \quad f^{-1} : \begin{pmatrix} 1\,2\,3\,4\,5 \\ 5\,2\,4\,1\,3 \end{pmatrix}$$

Composition of permutations:

$$f : \begin{pmatrix} 1\,2\,3\,4\,5 \\ 4\,2\,5\,3\,1 \end{pmatrix}, \quad g : \begin{pmatrix} 1\,2\,3\,4\,5 \\ 4\,3\,5\,1\,2 \end{pmatrix} = \begin{pmatrix} 4\,2\,5\,3\,1 \\ 1\,3\,2\,5\,4 \end{pmatrix} \Longrightarrow fg : \begin{pmatrix} 1\,2\,3\,4\,5 \\ 1\,3\,2\,5\,4 \end{pmatrix}$$

f has 3 inversions, g has 5, fg has 2 (check... not sure I got it right). Let $\mathrm{Inv}(g)$ be # inversions of g, and

$$\mathrm{Inv}(fg) = \mathrm{Inv}(f) + \mathrm{Inv}(g) \mod 2$$

$$t = \begin{pmatrix} 1\,2\,3 \cdots n \\ 2\,1\,3 \cdots n \end{pmatrix}$$

$\mathrm{Inv}(t) = 1$. So for $(\forall f)(\mathrm{Inv}(ft) \equiv \mathrm{Inv}(f) + 1 \mod 2)$, $|S_1| = 1$.

Definition: f is an even permutation if $\mathrm{Inv}(f) \equiv 0 \mod 2$, an *odd permutation* if $\mathrm{Inv}(f) \equiv 0 \mod 2$.

Corollary 5.25 *# even permutations = # odd permutations (assuming $n \geqslant 2$)*

This is because $f \mapsto f \cdot t$ is a bijection between even and odd permutations. A *transposition* switches two elements,

$$t_{ij} = \begin{pmatrix} 1\,2 \cdots i \cdots j \cdots n \\ 1\,2 \cdots j \cdots i \cdots n \end{pmatrix}$$

$\mathrm{Inv}(t_{ij}) = 2(j - i) - 1 \equiv 1 \pmod 2$, so all transpositions are odd.

✎**:** Transpositions generate S_n.[7]

Theorem 5.26 *A permutation f is even $\iff$ f is the product of an even number of transpositions.*

Cycle notation: pictures of $1 \to 4 \to 3 \to 5$, $2 \to$ self, f is a 4-cycle (don't count identity), g is $1 \to 4 \to 1$, $2 \to 3 \to 5$

Definition: A k-cycle is $i_1 \mapsto c_2 \mapsto \cdots \mapsto i_k \mapsto i_1$, everything else fixed.

Notation: $(i_1\, i_2\, \ldots\, i_l)$, so $f = (1435) = (4351)$, $g = (14)(235) = (235)(14)$, this is *cycle notation*

Theorem 5.27 *Every permutation is a product of disjoint cycles, unique up to the order of the factors.*

[7] Says written on an open-shelf math library in Germany: "Dear patrons: please remember that transpositions generate S_n."

Transpositions (ab) are odd, $(123) = (12)(13)$ is even, $(1234) = (12)(13)(14)$, etc., gives

Theorem 5.28 *A k-cycle is even $\iff$ k is odd.*

Sign of permutation is $\text{sgn}(f) = (-1)^{\text{Inv}(f)} = 1$ if f is even, -1 if f is odd.

✎**:** $\text{sgn}(fg) = \text{sgn}(f) \cdot \text{sgn}(g)$

Definition: For A an $n \times n$ matrix,

$$\det(A) = \sum_{f \in S_n} \frac{\text{sgn}(f) \cdot \prod a_{i,if}}{n!}$$

(in definition, the product is the expansion term.)

Theorem 5.29 *Let $A = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n]$, elementary operation is $\boldsymbol{a}_i \mapsto \lambda \boldsymbol{a}_j$, $j \neq i$, λ a scalar,* $\det(A') = \det(A)$.

(There's a whole example here using λ..)

Proof:

$$\det[\boldsymbol{a}_1, \cdots \boldsymbol{a}_i - \lambda \boldsymbol{a}_j \cdots \boldsymbol{a}_j \cdots \boldsymbol{a}_n] = \det A + \det[\boldsymbol{a}, \cdots \underbrace{[-\lambda \boldsymbol{a}_j]}_{i} \boldsymbol{a}_j \cdots \boldsymbol{a}_n] = (-\lambda) \det[\boldsymbol{a}_1, \cdots, \boldsymbol{a}_j, \boldsymbol{a}_j, \ldots, \boldsymbol{a}_n] = 0$$

$$\begin{aligned} A &= [\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n] \quad (\boldsymbol{a}_1 = \boldsymbol{b} + \boldsymbol{c}) \\ B &= [\boldsymbol{b}, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n] \\ C &= [\boldsymbol{c}, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n] \\ D &= [\lambda \boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n] \end{aligned}$$

$\det(D) = \lambda \det A$. Warning: $A \neq B + C$, $D \neq \lambda A$.
If $\boldsymbol{a}_1 = 0$ then $\det A = 0$, if $\exists i \neq j$ such that $\boldsymbol{a}_i = \boldsymbol{a}_j$ then $\det A = 0$

Theorem 5.30 *If two columns of A are equal then* $\det A = 0$

Proof: We can match up the expansion terms into pairs that cancel.

Corollary 5.31 $\det A$ *doesn't change if we subtract any linear combination of columns other than $\boldsymbol{a}_i$ from $\boldsymbol{a}_i$.*

Corollary 5.32 *If rank$A < n$ then* $\det A = 0$

Proof: $\text{rank} A < n \iff$ columns linearly dependent $\implies (\exists i)(\boldsymbol{a}_i \in \text{Span}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_{i-1}, \boldsymbol{a}_{i+1}, \ldots, \boldsymbol{a}_n))$, subtract $\implies$ get 0 column $\implies \det = 0$

This means Gaussian elimination "works."[8] Can look up what Gaussian elimination is online. Another important fact:

Theorem 5.33 *If we switch columns $A \to A'$,* $\det A' = -\det A$.

More generally, if we apply $f \in S_n$ to the columns of A, $A \mapsto A^f$, $\det(A^f) = \text{sgn}(f) \det A$.

✎**:** Elementary operations don't change the <u>rank</u> of A.

[8] "The goal is to tame the determinant, this horrible expression, by making as many zeros as possible."

Corollary 5.34 $\det A = 0 \iff$ *rank*$(A) < n$.

Theorem 5.35 (Fundamental Theorem of Algebra) *If $f(x)$ is a polynomial over $\mathbb{C}$ and $\deg(f) \geqslant 1$ then $(\exists \alpha \in \mathbb{C})(f(\alpha) = 0)$, $\therefore$ if f has degree n then $f(x) = a_n(x - \alpha_1)\cdots(x - \alpha_n)$*

Also (new theorem), if $f(x) = a_0 + a_1x + \cdots + a_nx^n$, $\deg f = n$ if $a_n \neq 0$ then $f(x) = (x - \alpha)g(x)$, g a polynomial, i.e. $x - \alpha \mid f(x)$. (✎)

$\deg(0) = \infty$, where 0 is seen as a polynomial (def of polynomial is that $a_0 \neq 0$.) Also,

1. $\deg(fg) = \deg(f) + \deg(g)$
2. $\deg(f + g) \leqslant \max\{\deg(f), \deg(g)\}$
3. if $\deg(f) \neq \deg(g)$, then $=$ same.

For $f(x) = x^n - 1 = \prod_{i=0}^{n-1}(x - \omega_i)$, where $\omega_0, \omega_1, \ldots, \omega_n$ the nth roots of unity,

$$\omega_j = \cos(\frac{2\pi j}{n}) + i\sin(\frac{2\pi j}{n})$$

The *order* of ω_j is the smallest $k \geqslant 1$ such that $\omega_j^k = 1$, e.g. the order of ω_1 is n.

Definition: ω_j is a *primitive* nth root of unity if its order is n.

✎: Prove: ω_j is a *primitive* nth root of unity $\iff \gcd(j, n) = 1$.

Corollary 5.36 *# primitive nth roots of unity is $\phi(n)$.*

✎: Suppose ω is an nth root of unity, $\omega^n = 1$, then if $k =$ order of ω then $k|n \implies \omega$ is a positive kth root of unity.

Conversely, if $k|n$, then every kth root of unity is also an nth root of unity:

$$z^k = 1 \implies z^n = (z^k)^{\frac{n}{k}} = 1^{\frac{n}{k}} = 1$$

Let $U_n = \{$ set of primitive nth roots of unity$\}$, $V_n = \{$all nth roots of unity$\}$.

$$V_n = \uplus_{d|n} U_d, \quad n = |V_n| = \sum_{d|n} \underbrace{|U_d|}_{\phi(d)}$$

$n = \sum_{d|n} \phi(d)$, $x^n - 1 = \prod_{d|n} \Phi_d(x)$. $x^n - 1 = \prod \omega(x - \omega)$, where ω is an nth root of unity,

$$\Phi_n(x) = \prod_{\omega}(x - \omega)$$

ω the same, $\deg(\Phi_n) = \phi(n)$, the *nth cyclotomic polynomial*

$$\begin{aligned}
\Phi_1(x) &= x - 1 \\
\Phi_2(x) &= x + 1 \\
\Phi_3(x) &= (x + \frac{1}{2} + i\frac{\sqrt{3}}{2})(x + \frac{1}{2} + i\frac{\sqrt{3}}{2}) = x^2 + x + 1 \\
\Phi_4(x) &= (x+i)(x-i) = x^2 + 1 \\
\Phi_5(x) &= \frac{x^5 - 1}{x - 1} = x^4 + x^3 + x^2 + x + 1 \\
&\implies (x-1)\Phi_5(x) = x^5 - 1 \\
x^6 - 1 &= \prod_i = 1^6 \Phi_i(x) = x^2 - x + 1 \\
\Phi_6(x) &= x^2 - x + 1 \\
\Phi_7(x) &= \frac{x^7 - 1}{x - 1} = x^6 + \cdots + x + 1 \\
\Phi_8(x) &= \frac{x^8 - 1}{x^4 - 1} = x^4 + 1
\end{aligned}$$

(There's some algebra in there didn't write down.) Erdös found that the coefficients here get very large.

✎**:** All cyclotomic polynomials have integer coefficients.

$n \times n$ matrix A, if $\boldsymbol{x}$ is a vector $\boldsymbol{x} \neq \mathbf{0}$ and $\exists\lambda$ scalar such that $A\boldsymbol{x} = \lambda\boldsymbol{x}$ then we call $\boldsymbol{x}$ an *eigenvector* to *eigenvalue* λ.

λ is an *eigenvalue* if $\exists \boldsymbol{x} \neq \mathbf{0}$ such that $A\boldsymbol{x} = \lambda\boldsymbol{x}$.

$$Ax = \lambda x = \lambda I x \implies \lambda I x - Ax = 0 \implies (\lambda I - a)\boldsymbol{x} = 0$$

λ an eigenvalue $\iff \exists x \neq 0$: $(\lambda I - A)\boldsymbol{x} = 0 \iff \lambda I - A$ is singular $\iff \det(\lambda - A) = 0$.

An $n \times n$ matrix $f_A(t) = \det(tI - A) =$ polynomial of degree n. Sketches the matrix out, get

$$\det = t^n - \underbrace{(\sum a_{ii})}_{\text{trace}(A)} t^{n-1} \pm \cdots + (-1)^n \det A$$

Corollary 5.37 *λ is an eigenvalue of $A \iff f_A(\lambda) = 0$, λ is a root of the characteristic polynomial.*

Something coordinate related:

$$\begin{pmatrix} \boldsymbol{i}' \\ \boldsymbol{j}' \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \boldsymbol{i} \\ \boldsymbol{j} \end{pmatrix}$$

Say $R_\theta =$ this matrix, the rotation matrix. Then $R_{\alpha+\beta} = R_\alpha + R_\beta$, shows

$$\begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} \cos\beta & -\sin\beta \\ \sin\beta & \cos\beta \end{pmatrix} = \begin{pmatrix} \cos(\alpha+\beta) & -\sin(\alpha+\beta) \\ \sin(\alpha+\beta) & \cos(\alpha+\beta) \end{pmatrix}$$

✎**:** $\boldsymbol{x} = \binom{x_1}{x_2}$ and $\boldsymbol{x}' = R_\theta \boldsymbol{x}$ then θ is the angle between $\boldsymbol{x}$ and $\boldsymbol{x}'$.

So

$$f_{R_\alpha}(t) = \begin{vmatrix} t - \cos\alpha & \sin\alpha \\ -\sin\alpha & t - \cos\alpha \end{vmatrix} = (t - \cos\alpha)^2 + (\sin\alpha)^2 = t^2 - 2\cos\alpha t + 1$$

eventually get $\lambda_{1,2} = \cos\alpha \pm i \sin\alpha$.

Recall that a digraph is strongly connected if $h =$ period $=$ gcd of lengths of all closed walks.

✎: Period(x) = gcd of all closed walks starting at x. If G is strongly connected $\implies (\forall x \in V)(\text{period}(x) \text{ is the same})$

✎: Period is multiple of $k \iff$ graph can be divided into k clusters around a circle such that all edges go from one cluster to the next.

Digraph associated with an $n \times n$ matrix: $i \to j \iff a_{ij} \neq 0$.

Q: For a stochastic matrix A, when does A^n converge?

Assume G is the digraph associated with A, G strongly connected, such A is called *irreducible.*

Theorem 5.38 *A^n converges $\iff$ G is aperiodic.*

means period=1 corollary to Frobenius-Perron theorem.

Stationary distribution. Say $\boldsymbol{x}_{t+1} = \boldsymbol{x}_t T$,

$$\boldsymbol{x}_t = \boldsymbol{x}_0 T^t,$$

evolution of the Markov Chain. The *stationary distribution* is $\boldsymbol{x}$ such that $\boldsymbol{x}T = \boldsymbol{x}$, the left eigenvector to eigenvalue. (note that $\boldsymbol{v}^T = [1 \cdots 1]$ is a right eigenvector). If have a strongly-connected Markov chain, then $\exists$ a unique stationary distribution.

Theorem 5.39

1. *For all finite Markov chains, $\exists$ a stationary distribution.*
2. *If the corresponding graph is strongly connected (=Markov chain irreducible), then the stationary distribution is unique.*

A *regular graph of degree d* has $(\forall x)(\deg(x) = d)$. For A the adjacency matrix,

$$a_{ij} = \begin{cases} 1 \text{ if } i \sim j \\ 0 \text{ otherwise} \end{cases}$$

The transition matrix is then $T = \frac{1}{d}A = (p_{ij})$, $T^t = (p_{ij}^{(t)})$. The largest eigenvalue is $\lambda_1 = d$, then $\lambda_i := \max_{2 \leqslant i \leqslant n} |\lambda_i| \leqslant d$.

Theorem 5.40 $|p_{ij}^{(t)} - \frac{1}{n}| \leqslant (\frac{\lambda}{d})^t$

($n = |V|$ as always.) So the convergence rate is governed by the *eigenvalue gap*, and this is a basic principle.

For A, B $n \times n$ matrices,

Theorem 5.41 $\det(AB) = \det(A)\det(B)$

this can be looked up in "the resources."

✎: $\det(A^{-1}) = \frac{1}{\det(A)}$

Definition: A, B are *similar*, $A \sim B$, if $\exists$ S, S^{-1} such that $B = S^{-1}AS$.

This is an equivalence relation (✎). If $A \sim B$ then $\det(A) = \det(B)$ (✎) (Hint: use $\det(AB)$ formula).

✎: $f_A(x) = \det(xI - A)$ the characteristic polynomial, If $A \sim B$ then $f_A(x) = f_B(x)$.

$$f_D(x) = \det \begin{pmatrix} x - \lambda_1 & \cdots & 0 \\ & \ddots & \\ 0 & & x - \lambda_n \end{pmatrix} = \prod (x - \lambda_i)$$

Example: $A = \left(\begin{smallmatrix}1&1\\0&1\end{smallmatrix}\right)$ is not diagonalizable. Proof by contradiction: assume $\exists\ S$, S^{-1}, then

$$f_A(x) = \det(xI - A) = (x-1)^2$$

So if A is diagonalizable, then $A \sim I$, $S^{-1}AS = I$, $A = SIS^{-1} = I$, $\rightarrow\leftarrow$.

✎**:** (*) Prove: if all roots of f_A are distinct, then A is diagonalizable.

Definition: An *eigenbasis* for A is a basis of F^n consisting of eigenvectors of A, i.e. n linearly independent eigenvectors.

Theorem 5.42 *A is diagonalizable $\iff$ $\exists$ an eigenbasis.*

Proof: A is diagonalizable $\iff$ $\exists\ S$, S^{-1}:

$$S^{-1}AS = D = \begin{pmatrix}\lambda_1 & & 0\\ & \ddots & \\ 0 & & \lambda_n\end{pmatrix}$$

$\iff$ $AS = SD = [\boldsymbol{s}_1 \dots \boldsymbol{s}_n]D = [\lambda_1\boldsymbol{s}_1, \dots, \lambda_n\boldsymbol{s}_n]$ $\iff$ all s_i are eigenvalues $\iff$ eigenbasis. ($S = [\boldsymbol{s}_1, \dots, \boldsymbol{s}_n]$, columns linearly independent.)

The *standard inner product* on $\mathbb{R}^n$ is

$$\boldsymbol{x} \cdot \boldsymbol{y} := \boldsymbol{x}^T\boldsymbol{y} = \sum_{i=1}^{n} x_i y_i,$$

with the dot product defined as usual. Define the *norm* (length) of $\boldsymbol{x}$ to be $||\boldsymbol{x}|| = \sqrt{\boldsymbol{x}^T\boldsymbol{x}} = \sqrt{\sum x_i^2}$.

✎**:** Cauchy-Schwarz: $|\boldsymbol{x}^T\boldsymbol{y}| \leqslant ||\boldsymbol{x}|| \cdot ||\boldsymbol{y}||$. Prove this based on $\mathrm{Var}(X) \geqslant 0$, $E(X^2) \geqslant E(X)^2$

We say $\boldsymbol{x}$ and $\boldsymbol{y}$ are *orthogonal* if $\boldsymbol{x}^T\boldsymbol{y} = 0$, and a set of vectors is orthogonal if they are pairwise orthogonal.

✎**:** If $\boldsymbol{v}_1, \dots, \boldsymbol{v}_k$ are nonzero, orthogonal vectors, then they are linearly independent.

A basis $\boldsymbol{v}_1, \dots, \boldsymbol{v}_n$ is *orthonormal* if it is orthogonal and $||\boldsymbol{v}_i|| = 1$.

✎**:** Any orthonormal set of vectors can be completed to an orthonormal basis.

An $n \times n$ real matrix A is *orthogonal* if $A^TA = I$. (From now on, assume every matrix is $n \times n$ and real.) For $A = [\boldsymbol{a}_1, \dots, \boldsymbol{a}_n]$,

$$\boldsymbol{a}_i^T\boldsymbol{a}_j = \begin{cases}1 \text{ if } i = j\\ 0 \text{ if } i \neq j\end{cases} \iff \boldsymbol{a}_1, \dots, \boldsymbol{a}_n \text{ ONB}$$

Now if $A^TA = I$ then $\exists A^{-1} = A^T$ then $AA^T = I \implies A^T$ orthogonal $\implies$ rows of A are ONB.

Theorem 5.43 *If A is orthogonal then $(A\boldsymbol{x})^T(A\boldsymbol{y}) = \boldsymbol{x}^T\boldsymbol{y}$.*

(Orthogonal matrices correspond to "congruences" of $\mathbb{R}^n$.)

Proof: ✎**:**

$$\begin{aligned}(AB)^T &= B^TA^T\\ (Ax)^T &= x^TA^T\\ (Ax)^T(Ay) = x^TA^TAy = x^TIy &= x^Ty\end{aligned}$$

Define the *spectral norm* of A as

$$||A|| = \max_{\boldsymbol{x}\in\mathbb{R}^n, \boldsymbol{x}\neq\boldsymbol{0}} \frac{||A\boldsymbol{x}||}{||\boldsymbol{x}||}$$

✎**:** $\exists$ max

Note that if λ is an eigenvalue then $||A|| \geqslant |\lambda|$.

Proof:

$$A\boldsymbol{x} = \lambda\boldsymbol{x},\ ||A\boldsymbol{x}|| = ||\lambda\boldsymbol{x}|| = |\lambda|||x||$$

last equality is ✎, then

$$||A|| \geqslant \frac{||Ax||}{||x||} = |\lambda|\checkmark$$

✎**:** (*) This is true if $\lambda \in \mathbb{C}$.

Then states spectral theorem.