N1.

Proposition: a statement which is either true or false.

e.g. $\sqrt{3}$ is irrational

not: $2+2$; $x^2+3x=5$

implication: $P \Rightarrow Q$. P implies Q. $\equiv \neg P \vee Q$.
if P then Q.

P	Q	$P \Rightarrow Q$	$\neg P \vee Q$
T	T	T	T
T	F	F	F
F	T	T	T
F	F	T	T

False imply everything

N2. Proof. $P \Rightarrow Q$.

Direct proof. To prove $P \Rightarrow Q$.
Assume P
⋮
therefore Q

e.g. $a, b, c \in \mathbb{Z}$. if $a|b$, and $a|c \Rightarrow a|(b-c)$

Proof By contradiction.
$P \Rightarrow Q$.
Assume $\neg P$
⋮
R
⋮
$\neg R$.
$\neg P \Rightarrow \neg R \wedge R$.

e.g. There are infinite prime number
$p_1, p_2 \cdots, p_k$.
$q: p_1 p_2 \cdots p_k + 1$ not prime. then it has a prime divisor
$r = p_1 p_2 \cdots p_k$.
$p|q$ & $p|r \Rightarrow p | q-r \Rightarrow q \leq 1$ false

N3. Induction

Simple induction

Base case: Prove $P(0)$ True.

Inductive Hypothesis: Arbitary $k \geq 0$. assume $P(k)$ True

Inductive Step: with assumption. show $P(k+1)$ True

If $[P(0) \wedge \forall n \in \mathbb{N}\ P(n) \Rightarrow P(n+1)]$ then
$\forall\ [\forall n \in \mathbb{N}\ P(n)]$

Strong induction.
if $[P(0) \wedge \forall n \in \mathbb{N}\ P(0) \wedge P(1) \cdots \wedge P(n) \Rightarrow P(n+1)]$
then $[\forall n \in \mathbb{N}\ P(n)]$

e.g. Every natural number $n > 1$ can be written as a product of primes.

Quantifiers: $\forall$. $\exists$.
$(\forall n \in \mathbb{N})(n^2+n+41$ is prime$)$ × $n=41$

Quantifiers doesn't commute.
$(\forall x \in \mathbb{Z})(\exists y \in \mathbb{Z})(x<y)$ ✓
$(\exists y \in \mathbb{Z})(\forall x \in \mathbb{Z})(x<y)$ ×

Demorgan's law.
$\neg(P \wedge Q) \equiv \neg P \vee \neg Q$
$\neg(P \vee Q) \equiv \neg P \wedge \neg Q$.
$\neg(\exists x \forall y\ W(x,y))$
$= \forall x \neg(\forall y\ W(x,y)) = \forall x \exists y\ \neg W(x,y)$

Proof By Contraposition
To prove $P \Rightarrow Q$
Assume $\neg Q$.
⋮
Therefore $\neg P$.
Conclusion $\neg Q \Rightarrow \neg P$. $\equiv P \Rightarrow Q$.
e.g. $n \in \mathbb{Z}^+$. $d|n$. if n is odd then d odd

Proof by case.
e.g. there exists irrational numbers x and y such that x^y is rational
a) $\sqrt{2}^{\sqrt{2}}$ is rational or
b) $\sqrt{2}^{\sqrt{2}}$ is irrational
$(\sqrt{2}^{\sqrt{2}})^{\sqrt{2}} \Rightarrow \sqrt{2}^2 = 2$.

Strengthening the hypothesis.
e.g. $a_1 = 1$. $a_{n+1} = 3a_n^2$.
prove $a_n \leq 3^{2^n}$
too weak.
prove $a_n \leq 3^{2^n-1}$

N4 Stable Matching Algorithm.

Job	Candidates:	Candidates	Jobs
1	A > B > C	A	2 > 1 > 3
2	B > A > C	B	1 > 3 > 2
3	A > B > C	C	1 > 2 > 3

	A	B	C
D1	1, 3	2	
D2	1	2, 3	
D3	1, 2	3	
D4	2	3, 4	
D5	2	1	3

5 days. man
Morning: propose woman
Afternoon: may be / never
Evening: cross-off woman say never

Lemma. 4.1 The stable marriage algorithm always halt.

Proof: if everyday not halt. must be a man cross-off one woman from his list.
n men. n women. So at most n^2

Lemma 4.2. Improvement Lemma.
If man M. proposes to woman W on k^{th} day. then on every subsequent day W has someone on a string whom she likes at least as much as M.

well ordering Principle.
if $S \subseteq \mathbb{N}$ and $S \neq 0$. then S has a minimal element.

Lemma 4.3. The TMA pairing is always stable.
suppose not. (M, W^*) M prefer W^* than his final partner W. However, by improvement lemma. W^* must prefer her final partner than M. So no rogue couple.
M proposes to W^* before W.

Lemma 4.4. TMA pairing is man optimal

Lemma 4.5. TMA pairing is woman-pessimal.

Theorem 6.2. m, x positive integers. $\gcd(x, m) = 1$. then x has an unique multiplicative inverse mod m.

$0, x, 2x, \ldots (m-1)x$. all unique. then must be $ax \equiv 1 \bmod m$.

suppose not.

$ax \equiv bx \bmod m \Rightarrow (a-b)x \equiv 0 \bmod m \Rightarrow (a-b)x = km$

$\gcd(x, m) = 1$. $1 \le a-b \le m-1$. so contradiction

Theorem 6.3. Euclid's Algorithm.

Let $x \ge y > 0$. Then $\gcd(x, y) = \gcd(y, x \bmod y)$

$O(\log N) = O(n)$ n is # bits

Division in Modular Arithmetic.

$8x \equiv 9 \bmod 15 \Rightarrow 8^{-1} \equiv 2 \bmod 15$.

$x \equiv 18 \bmod 15 \equiv 3 \bmod 15$.

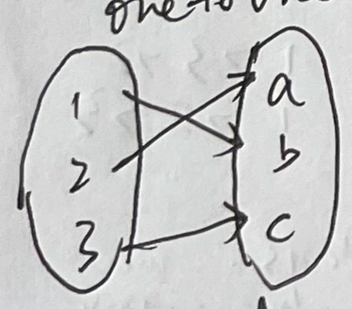
Chinese remainder theorem.

$g(m, n) = 1$. $\exists$ exactly one x mod mn. that $x = a \bmod n$ and $x = b \bmod n$.

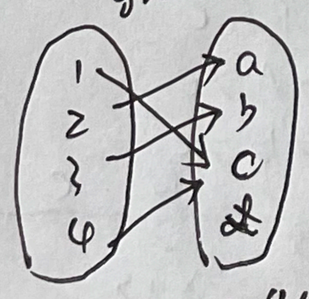
Bijections:

f. function. then $x \in A$. exactly one element $f(x) \in B$

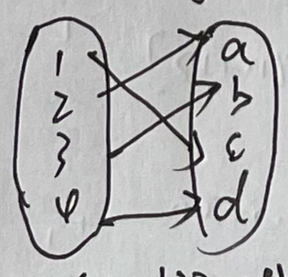
one-to one



onto



both (bijection)



$x \ne y \Rightarrow f(x) \ne f(y)$ $(\forall y \exists x)(f(x) = y)$ f is bijection iff one-to one and onto

---

N7. RSA.

Th. 7.2 Fermat's Little Theorem.

For any prime p and any $a \in \{1, 2 \ldots p-1\}$ (or coprime with p) we have $a^{p-1} \equiv 1 \bmod p$.

$S = \{1, 2, \ldots p-1\}$ $S' = \{a \bmod p, 2a \bmod p \ldots (p-1)a \bmod p\}$

they are the same set. proof by contradiction any two in S' subtract. $\not\equiv 0 \bmod p$.

$1 \times 2 \times \cdots \times (p-1) = (p-1)! \bmod p$. $a \times 2a \cdots \times (p-1)a = (p-1)!\, a^{p-1} \bmod p$

$(p-1)! \equiv a^{p-1}(p-1)! \bmod p$.

since every number $\in \{1, \ldots p-1\}$ is coprime with p. multiplicative inverse exists. so $a^{p-1} \equiv 1 \bmod p$.

RSA. pick p, q. two large primes

$N = p \times q$. e relatively prime to $(p-1)(q-1)$. like 3, 5. small value

$d \equiv e^{-1} \bmod (p-1)(q-1)$

public key: (N, e) private key.

$E(m) = m^e \bmod N$

$D(E(m)) = m^{ed} \bmod N$

$= m \bmod N$

N6. Modular Arithmetic

Theorem. 6.1. $a \equiv c \pmod m$ and $b \equiv d \pmod m$ then $a+b \equiv c+d \pmod m$ and $a \cdot b \equiv c \cdot d \pmod m$.

Exponentiation.

Algorithm mod-exp (x, y, m)
if $y = 0$ then return (1)
else
$z = $ mod_exp (x, y div 2, m)
if y mod 2 = 0 then return ($z \cdot z$ mod m)
else return ($x \cdot z \cdot z$ mod m)

$x^{2a} = (x^a)^2$. $x^{2a+1} = x \cdot (x^a)^2$.

To calculate the inverse. use e-gcd (x, y)

$d = \gcd(x, y) = ax + by$

$1 = \gcd(m, x) = am + bx$.

then b is the multiplicative inverse of x

Extended-gcd (x, y)
if $y = 0$ then return (x, 1, 0)
else
(d, a, b) := extended-gcd (y, x mod y)
return ((d, b, a − (x div y) * b))

$n_1, n_2, \ldots n_k$ coprime to each other. unique integer x between 0 and $N = \prod_{i=1}^{k} n_i$ such that

$x \equiv a_1 \pmod{n_1} \ldots$ $x = a_k \pmod{n_k}$

Moreover. $x = \sum_{i=1}^{k} a_i b_i \bmod N$.

let $f: A \to B$. be a bijection. f has an inverse $f^{-1}: B \to A$ that satisfies $f^{-1}(f(a)) = a$. for all $a \in A$. and f^{-1} is also a bijection.

Chinese Remainder Theorem:

$\gcd(m,n)=1$. exactly one $x \pmod{mn}$ that. $x \equiv a \pmod m$ $x \equiv b \pmod n$

$x \equiv a_1 \pmod{n_1}$ $x \equiv a_2 \pmod{n_2}$ ...

$x = (\sum_{i=1}^{k} a_i b_i) \bmod N$. $N = \prod_{i=1}^{k} n_i$

$b_i = \frac{N}{n_i} (\frac{N}{n_i})_{n_i}^{-1}$ mul inverse $(\bmod n_i)$ of $\frac{N}{n_i}$. $\searrow$ $(1 \bmod n_j)$ and $(0 \bmod n_j)$ $i \neq j$

[Graph Theory]

path simple $v_1, \dots v_n$ distinct.

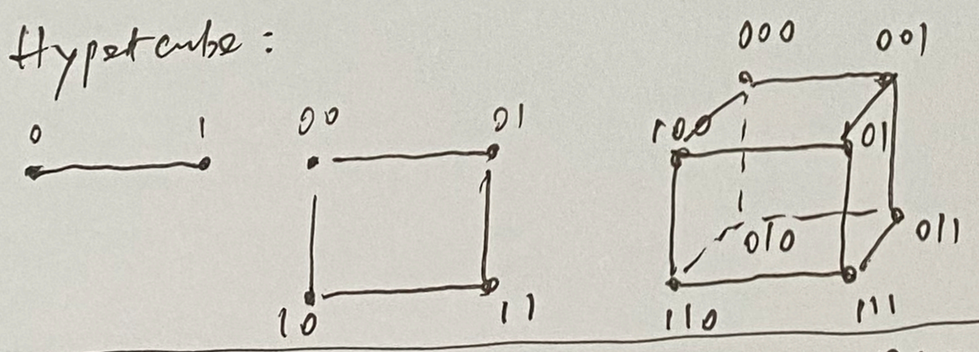
undirected graph $G=(V,E)$ has an Euleriun tour iff G is even degree and connected

Function Euler(G,s):
  $T =$ FindTour(G,s)
  Let $G_1, \dots G_k$ be the connected components when edges in T removed from G. let s_i be the first vertex in T intersects G_i
  output Splice$(T, \text{Euler}(G_1,s_1), \dots \text{Euler}(G_k,s_k))$
End Euler.

Theorem 5.2.
For every connected planar graph. $v+f=e+2$
(planar graph: a graph is planar if it can be drawn on plane without crossings).

Trees: connected and no cycles.
connected and has $n-1$ edges.

Hypercube:



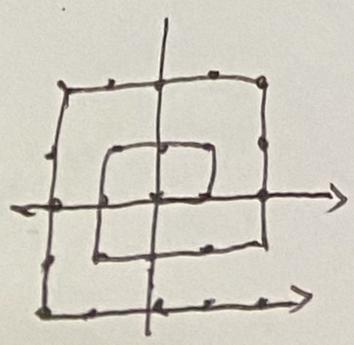
$P(x)$	$E(x)$	$Q(x)$
deg: $n-1$	k	$n+k-1$

$P(x) = \frac{Q(x)}{E(x)}$

$Q(x) = a_{n+k-1}x^{n+k-1} + \dots + a_1 x + a_0$ # coeff $= n+k$

$E(x) = x^k + b_{k-1}x^{k-1} + \dots b_1 x + b_0$ # coeff $= k$.

$Q = \{\frac{x}{y} \mid x,y \in \mathbb{Z}.\ y \neq 0\}$ $Q \to N$



Any set of binary strings / polynomials with natural number coeff (ternary strings) is countable

Cantor's Diagnolization. $N \xrightarrow{x} R$
can construct a new real number digit $d \Rightarrow d+2$.
the new real number s not in original list
nor $f(n) = s$

$f(0) = 0.0216 93 \dots$
$f(1) = 0.1016 29$
$f(2) = 0.9408 27$
$f(3) = 0.5308175$
$\vdots$

Note 8. Polynomials.

property 1: A non-zero polynomial of degree d has at most d roots.

property 2: Given $d+1$ pairs $(x_1,y_1) \dots (x_{d+1}, y_{d+1})$. with all x_i distinct. There is a unique polynomial $p(x)$ of degree at most d. such that $p(x_i) = y_i$ for $1 \le i \le d+1$.

polynomial Interpolation

$(1,1)$ $(2,2)$ $(3,4)$

$\Delta_1(x) = \frac{(x-2)(x-3)}{(1-2)(1-3)} = \frac{1}{2}x^2 - \frac{5}{2}x + 3$

$\Delta_2(x) = \frac{(x-1)(x-3)}{(2-1)(2-3)} = \frac{(x-1)(x-3)}{-1} = -x^2 + 4x - 3$

$\Delta_3(x) = \frac{(x-1)(x-2)}{(3-1)(3-2)} = \frac{1}{2}x^2 - \frac{3}{2}x + 1$

$p(x) = 1\Delta(x) + 2\Delta(x) + 4\Delta(x) = \frac{1}{2}x^2 - \frac{1}{2}x + 1$.

C1: If a is a root of $p(x)$ with deg d. then $p(x) = (x-a)q(x)$.

C2: $p(x)$ of deg d. with distinct roots $a_1 \dots a_d$. can be written as $p(x) = C(x-a_1)\dots(x-a_d)$

GF(m). work with numbers mod a prime m. add. subtract multiply divide.. same as usual

secret sharing

$p(x)$ of degree $k-1$. st. $p(0)=s$. give $p(1)$. $p(2)$ ... $p(n)$ to n officials.
Any k officials know the secret
any 2 of them have no more info than 1 person does.

Note 9. Error Correcting Codes

Erasure errors. send $n+k$ packets with at most k info lost.

General errors: send $n+2k$ to endure k errors.

$E(x) = (x-e_1)(x-e_2)\dots(x-e_k)$

$p(i)E(i) = r_i E_i$ for $1 \le i \le n+2k$.

solve linear equations $Q(x_i) = r(x_i)E(x)$

Note 10. Countability.

$N \to Z^+$. $N \to 2N$. $f(x) = \begin{cases} \frac{x}{2} & x \text{ even} \\ -\frac{(x+1)}{2} & x \text{ odd.} \end{cases}$

☆ countable: there is a bijection between S and N or subset of N

Chinese Remainder Theorem:

$\gcd(m,n)=1$. exactly one $x \pmod{mn}$ that. $x \equiv a \pmod m$ $x \equiv b \pmod n$

$x \equiv a_1 \pmod{n_1}$ $x \equiv a_2 \pmod{n_2}$ ...

$x = (\sum_{i=1}^{k} a_i b_i) \bmod N$. $N = \prod_{i=1}^{k} n_i$

$b_i = \frac{N}{n_i} (\frac{N}{n_i})_{n_i}^{-1}$ mul inverse $(\bmod n_i)$ of $\frac{N}{n_i}$. $\searrow$ $(1 \bmod n_j)$ and $(0 \bmod n_j)$ $i \neq j$

[Graph Theory]

path simple $v_1, \dots v_n$ distinct.

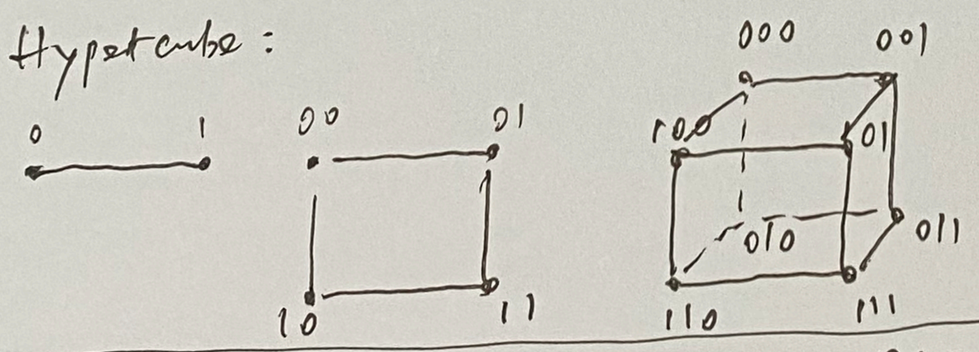
undirected graph $G=(V,E)$ has an Euleriun tour iff G is even degree and connected

Function Euler(G,s):
  $T =$ FindTour(G,s)
  Let $G_1, \dots G_k$ be the connected components when edges in T removed from G. let s_i be the first vertex in T intersects G_i
  output Splice$(T, \text{Euler}(G_1,s_1), \dots \text{Euler}(G_k,s_k))$
End Euler.

Theorem 5.2.
For every connected planar graph. $v+f=e+2$
(planar graph: a graph is planar if it can be drawn on plane without crossings).

Trees: connected and no cycles.
connected and has $n-1$ edges.

Hypercube:



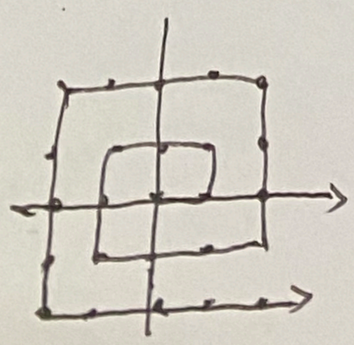
$P(x)$	$E(x)$	$Q(x)$
deg: $n-1$	k	$n+k-1$

$P(x) = \frac{Q(x)}{E(x)}$

$Q(x) = a_{n+k-1}x^{n+k-1} + \dots + a_1 x + a_0$ # coeff $= n+k$

$E(x) = x^k + b_{k-1}x^{k-1} + \dots b_1 x + b_0$ # coeff $= k$.

$Q = \{\frac{x}{y} \mid x,y \in \mathbb{Z}.\ y \neq 0\}$ $Q \to N$



Any set of binary strings / polynomials with natural number coeff (ternary strings) is countable

Cantor's Diagnolization. $N \xrightarrow{x} R$
can construct a new real number digit $d \Rightarrow d+2$.
the new real number s not in original list
nor $f(n) = s$

$f(0) = 0.0216 93 \dots$
$f(1) = 0.1016 29$
$f(2) = 0.9408 27$
$f(3) = 0.5308175$
$\vdots$

Note 8. Polynomials.

property 1: A non-zero polynomial of degree d has at most d roots.

property 2: Given $d+1$ pairs $(x_1,y_1) \dots (x_{d+1}, y_{d+1})$. with all x_i distinct. There is a unique polynomial $p(x)$ of degree at most d. such that $p(x_i) = y_i$ for $1 \le i \le d+1$.

polynomial Interpolation

$(1,1)$ $(2,2)$ $(3,4)$

$\Delta_1(x) = \frac{(x-2)(x-3)}{(1-2)(1-3)} = \frac{1}{2}x^2 - \frac{5}{2}x + 3$

$\Delta_2(x) = \frac{(x-1)(x-3)}{(2-1)(2-3)} = \frac{(x-1)(x-3)}{-1} = -x^2 + 4x - 3$

$\Delta_3(x) = \frac{(x-1)(x-2)}{(3-1)(3-2)} = \frac{1}{2}x^2 - \frac{3}{2}x + 1$

$p(x) = 1\Delta(x) + 2\Delta(x) + 4\Delta(x) = \frac{1}{2}x^2 - \frac{1}{2}x + 1$.

C1: If a is a root of $p(x)$ with deg d. then $p(x) = (x-a)q(x)$.

C2: $p(x)$ of deg d. with distinct roots $a_1 \dots a_d$. can be written as $p(x) = C(x-a_1)\dots(x-a_d)$

GF(m). work with numbers mod a prime m. add. subtract multiply divide.. same as usual

secret sharing

$p(x)$ of degree $k-1$. st. $p(0)=s$. give $p(1)$. $p(2)$ ... $p(n)$ to n officials.
Any k officials know the secret
any 2 of them have no more info than 1 person does.

Note 9. Error Correcting Codes

Erasure errors. send $n+k$ packets with at most k info lost.

General errors: send $n+2k$ to endure k errors.

$E(x) = (x-e_1)(x-e_2)\dots(x-e_k)$

$p(i)E(i) = r_i E_i$ for $1 \le i \le n+2k$.

solve linear equations $Q(x_i) = r(x_i)E(x)$

Note 10. Countability.

$N \to Z^+$. $N \to 2N$. $f(x) = \begin{cases} \frac{x}{2} & x \text{ even} \\ -\frac{(x+1)}{2} & x \text{ odd.} \end{cases}$

☆ countable: there is a bijection between S and N or subset of N

Note 11. Computability.

Halting Problem:

$\text{TestHalt}(P, x) = \begin{cases} Y & \text{if } P \text{ halts on } x \\ N & \text{if } P \text{ loops on } x. \end{cases}$

Turing(P):

if TestHalt(P, P) = "Y". then loop

else halt.

	P_0	P_1	P_2	P_3	P_4
P_0	H	L	H	L	H
P_1	L	L	H	L	L
P_2	H	H	H	H	L
P_3	H	H	H	H	H
P_4	L	H	L	L	L

The opposite of the diagnol sequence is the result of the Turing. so Turing is not in the list.

Note 12 Counting

$C_n^k = \frac{n!}{(n-k)!\,k!}$

Note 13. Discrete Probability

The outcome of a random variable: sample point w

The set of all possible sample points: sample space Ω

$\Omega = \{HHH\ HHT.\ HTH\ HTT.\ THH\ THT\ TTH\ TTT\}$

probability space $\sum_{w \in \Omega} P[w] = 1$.

event A is subset of sample space. Ω.

$A = \{HHH.\ HHT.\ HTH\ THH\}$ At least 2 Hs.

C: flip n times! a coin with r Hs.

$P[C] = \binom{n}{r} p^r (1-p)^{n-r}$

$P[A] = \frac{\#\text{ sample points in } A}{\#\text{ sample points in } \Omega} = \frac{|A|}{|\Omega|}$

- A_k be the event that the load of some processor is at least k (n jobs, n processors)

Find the smallest value k s.t. $P[A_k] \le \frac{1}{2}$

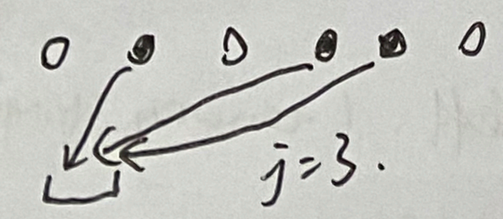
$A_j(1)$ = bin 1 has $\ge j$ balls.

$P[A_j(1)] \le \frac{1}{2n}$. choose j such that.

$A_j = A_j(1) \cup A_j(2) \cup \cdots \cup A_j(n)$

$P[A_j] \le \sum_{i=1}^{n} P(A_j(i)) = n P[A_j(1)] = n \cdot \frac{1}{2n} = \frac{1}{2}$

B_S $S \subseteq \{1, \ldots, n\}$ $|S| = j$ S ends up in bin 1



$j = 3$.

$A_j(1) = \bigcup_{S:|S|=j} B_S$ $\quad P[A_j(1)] \le \sum_{S:|S|=j} P[B_S] = \binom{n}{j}\frac{1}{n^j}$

$\le \frac{n^j}{j!\,n^j} = \frac{1}{j!}$

$P[A_j(1)] \le \frac{1}{j!}$ $\quad \frac{1}{j!} \approx \frac{1}{2n}$ $\quad j \approx \frac{\ln n}{\ln \ln n}$

Mathematics is incomplete.

$S(F)$: "This statement is not provable in F"

F is a formal system. of axioms. rules.

1. case 1. $S(F)$ is provable. Then $S(F)$ should be true but the content of the statement itself says $S(F)$ should not be provable

2. case 2. $S(F)$ is not provable. then incomplete.

Use Halting problem to solve Gödel's Proof

$\exists z$ (z encodes a valid halting execution sequence of P on input x)?

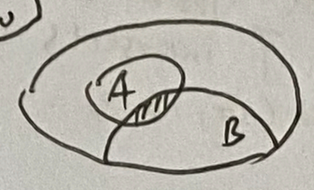
Search(P, x):

for every proof q:

if q is proof of $\exists p, x$. output "Yes"

if q is a proof of $\neg S_{p,x}$. output "No"

Note 14. Conditional probability.



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

$= \frac{P(B|A)\,P(A)}{P(B)}$

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

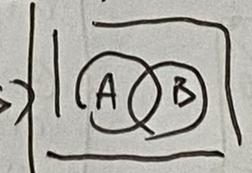
Union of Events.

$P[\cup_{i=1}^{n} A_i] = \sum_{i=1}^{n} P[A_i] - \sum_{i<j} P[A_i \cap A_j] + \sum_{i<j<k} P[A_i \cap A_j \cap A_k]$

$- \cdots + (-1)^{n-1} P[A_1 \cap A_2 \cap \cdots \cap A_k]$

Note 17 Hashing and Loading Balance.

- Table is size n. randomly hash keys to it. how many keys can we hash before the prob of a collision exceeds (say) $\frac{1}{2}$



$Pr(A \cup B) \le Pr(A) + Pr(B)$

C_i denotes event that pair i has collision

$P[\bar{A}] = P[\cup_{i=1}^{\binom{m}{2}} C_i] \le \sum_{i=1}^{\binom{m}{2}} P[C_i] = \binom{m}{2} \times \frac{1}{n} \approx \frac{m^2}{2n}$

so $m = O(\sqrt{n})$ to make $P[\bar{A}] \le \frac{1}{2}$

Note 15. Random Variables. Expectation.

A Random Variable X on a sample ~~point~~ space Ω is a function $X: \Omega \to R$ that assigns to each sample point $w \in \Omega$. a real number $X(w)$

1. ~~set~~ of values it can take.
2. probability with which it takes on the values

$P(X = a) = \sum_{w: X(w) = a} P(w)$

Bernoulli Distribution

$P[X = i] = \begin{cases} p & \text{if } i = 1 \\ 1-p & \text{if } i = 0 \end{cases}$

Binomial Distribution

flip n times a coin

$P[X = i] = \binom{n}{i} p^i (1-p)^{n-i}$

Note 15. Expectation

$\sum_{a \in A} a \times P[X=a] = E[X]$.

$E[X+Y] = E[X] + E[Y]$. linearity of Expectation

Note 16. Variance and covariance.

$Var(X) = E[(X-\mu)^2]$. variance of X

$\sigma(X) = \sqrt{Var(X)}$ standard deviation of X

$Var(cX) = c^2 Var(X)$

Independent random variables X, Y.

$E[X \cdot Y] = E[X]E[Y]$.

$Var[X+Y] = Var[X] + Var[Y] + 2(E[XY] - E[X]E[Y])$.

$Var[X+Y] = Var[X] + Var[Y]$

$cov(X, Y) = E[(X-\mu_X)(Y-\mu_Y)] = E[XY] - E[X]E[Y]$

1. if X, Y independent. $cov[X,Y] = 0$. converse not True.

2. $cov(X,X) = Var(X)$.

$corr(X,Y) = \frac{cov(X,Y)}{\sigma(X)\sigma(Y)}$.

$corr(X,Y) \in [-1, 1]$.

$\tilde{X} = (X-\mu_X)/\sigma(X)$. $\tilde{Y} = (Y-\mu_Y)/\sigma(Y)$

$corr(X,Y) = cov(\tilde{X}, \tilde{Y}) = E[\tilde{X}\tilde{Y}]$

if $corr(X,Y) = \pm 1$. $Y = aX + b$.

$a > 0$. $corr(X,Y) = +1$. $a < 0$. $corr(X,Y) = -1$.

N18 Markov. Chebyshev Inequalities.

Markov. Inequality

$P[X \geq c] \leq \frac{E[X]}{c}$

$E[X] = \sum_{a \in A} a P[X=a] \geq \sum_{a \geq c} a P[X=a]$

$\geq \sum_{a \geq c} c P[X=a] = c \sum_{a \geq c} P[X=a]$

$= c P[X \geq c]$

"It might not provide a very good upper bound".

Chebyshev's Inequality

$P[|X-\mu| \geq c] \leq \frac{Var[X]}{c^2}$.

Proof: Let $Y = (X-\mu)^2$.

$P[|X-\mu| \geq c] = P[Y \geq c^2] \leq \frac{E[Y]}{c^2} = \frac{Var[X]}{c^2}$

Large Number Theorem:

$P[|\frac{1}{n}S_n - \mu| < \varepsilon] \to 1$ as $n \to \infty$

$\leq \frac{Var(\frac{1}{n}S_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \to 0$.

If we observe some Random Variables many times. then the average of observations will converge to the mean (Expectation of the random variable)

2. Possoin Distribution.

Measuring Rare Events.

$P[X=i] = \frac{\lambda^i}{i!} e^{-\lambda}$

Taylor Series: $e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$

Approximation of Binomial Distribution especially $n \to \infty$. p small. ($n \geq 20, p \leq 0.05$)

App: Average 1.5 typo per page

$P[X=5] = \frac{1.5^5}{5!} e^{-1} = \frac{1}{120e} \approx \frac{1}{326}$

$E[X] = \lambda$. $Var[X] = \lambda$

Estimating bias of a coin using Chebyshev's

$E[\hat{p}] = E[\frac{1}{n}S_n] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = p$.

($\hat{p} = \frac{1}{n}S_n$, $S_n = X_1 + \cdots + X_n$. X_i is indicator of coming up heads)

$Var[\hat{p}] = Var(\frac{1}{n}S_n) = \frac{1}{n^2} Var(S_n) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{\sigma^2}{n}$

Then $P[|\hat{p}-p| \geq \varepsilon] \leq \frac{Var(\hat{p})}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2}$

ε is the error. $P[|\hat{p}-p| \geq \varepsilon] = \delta$ is the confidential probability

Then $n \geq \frac{\sigma^2}{\varepsilon^2 \delta}$

So for bias coin estimation

$n \geq \frac{p(1-p)}{\varepsilon^2 \delta}$. And $p = \frac{1}{2}$. maximize

so $n \geq \frac{1}{4\varepsilon^2\delta}$ to achieve error of 0.1 with probability at least 95%. we need 800 people.

N19. Some Distribution.

1. Geometric Distribution

$P[X=i] = (1-p)^{i-1} p$. i: How many times before we get the event (say head).

Theorem 19.1. $E[X] = \sum_{i=1}^{\infty} P[X \geq i]$

Theorem 19.2. $E[X] = \frac{1}{p}$

Or. Because each coin toss is independent

$E[X] = \underbrace{p \cdot 1}_{\text{first toss is H}} + \underbrace{(1-p)(1+E[X])}_{\text{first toss is T. then toss again}}$ $E[X] = \frac{1}{p}$

Application: Coupon Collector problem.

How many boxes do we need to buy to get n cards.

$S_n = X_1 + X_2 + \cdots + X_n$. X_i 步 times to get ith new card.

$E[S_n] = \sum_{i=1}^{n} E[X_i] = \frac{n}{n} + \frac{n}{n-1} + \cdots \frac{n}{1} = n\sum_{i=1}^{n}\frac{1}{i} \approx n(\ln n)$

($E[X_i] = \frac{n}{n-i+1}$, $P[\text{get } i^{th} \text{ new card}] = \frac{n-i+1}{n}$, Geometric distribution for X_i).

N20

Continous Probability:

Uniform Distribution $\frac{\text{Length of } [a,b]}{\text{Length of } [0,1]} = \frac{b-a}{l}$

PDF. Probability Density Function

e.g. $[0,l]$. $E[X] = \int_0^l x \frac{1}{l} dx = \frac{l}{2}$

$Var[X] = \int_0^l x^2 \frac{1}{l} dx - E(X)^2 = \frac{l^2}{12}$

Exponential Distribution.

$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{o.w.} \end{cases}$ $E[X] = \frac{1}{\lambda}$ $Var[X] = \frac{1}{\lambda^2}$

time continuous

Approx. to Geometric distribution

$P[X > t] = \int_t^{\infty} \lambda e^{-\lambda x} dx = e^{-\lambda t}$

Geom:

$P[Y > t] = (1-\lambda\delta)^{t/\delta} \approx e^{-\lambda t}$ let $\lambda = p/\delta$.

Normal Distribution.

$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ $E[X] = \mu$. $Var[X] = \sigma^2$

$P[X \le a] = P[Y \le \frac{a-\mu}{\sigma}]$. Y is std normal distribution

Central Limit Theorem:

X_i independent random variables. with mean μ and variance σ^2

$A_n = \frac{1}{n}\sum_{i=1}^{n} X_i$

$E[A_n] = \mu$. $Var[A_n] = \frac{\sigma^2}{n}$.

$A_n' = \frac{(A_n - \mu)\sqrt{n}}{\sigma} = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}}$

$A_n' \sim N(0,1)$

Gaussian Model

$p(x) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)$

$\hat{\mu}_j = \frac{1}{n_j}\sum_{x \in class(j)} x$ $\mu_j \in R^{784}$ $\Sigma_j \in R^{784\times784}$

$\hat{\Sigma}_j = \frac{1}{n_j}\sum_{x \in class(j)} (x-\hat{\mu}_j)(x-\hat{\mu}_j)^T$

Similarly. μ_{ji} is class j's values mean on pixel i.

And Σ_j is the covariance matrix considering each

argmax $P_j(x)$ we can get the classification result.

Inference.

Baye's rules:

$Pr[y=j|x] = \frac{P[x|y=j]\,Pr[y=j]}{P(x)} = \frac{\pi_j P_j(x)}{\sum_i \pi_i P_i(x)}$

$y=j$. the digit is j (5).

x. given the image x.

π_i. each digit natural distribution say all 0.1 for 10 digits.

$P_j(x) = P_{j1}(x_1)P_{j2}(x_2)\cdots P_{j784}(x_{784})$

$P_{ji}(x) = \begin{cases} P_{ji} & \text{if } x_i = 1 \\ 1-P_{ji} & \text{if } x_i = 0 \end{cases} = P_{ji}^{x_i}(1-P_{ji})^{1-x_i}$

$\hat{P}_{ji} = \frac{n_{ji}}{n_j}$ $n_{ji} \Rightarrow$ # samples of class j with $x_i = 1$, $n_j \Rightarrow$ # samples of class j

$P_j(x) = \prod_{i=1}^{784} P_{ji}(x_i) = \prod_{i=1}^{784} P_{ji}^{x_i}(1-P_{ji})^{1-x_i}$

To classify x. (take the log)

$\arg\max_j \pi_j \prod_{i=1}^{784} P_{ji}^{x_i}(1-P_{ji})^{1-x_i}$

"But the model might understand the data bad"

Lec 25. Power Law Distribution

1). N

2). $N/2$

3) $N/3$

$\vdots$

Incomes: $P[\text{income} > x] = \frac{D}{x^{\alpha-1}}$

$\frac{\#\text{people with income} > x}{N} = \frac{ND}{x^{\alpha-1}}$

Highest income: $\frac{ND}{x^{\alpha-1}} = 1$

ith Highest income: $\frac{ND}{x^{\alpha-1}} = i$ $x = \left(\frac{ND}{i}\right)^{\frac{1}{\alpha-1}} = \frac{I_1}{i^{\frac{1}{\alpha-1}}}$