Propositional Logic

$P \Rightarrow Q \equiv \neg P \vee Q$ $P \Rightarrow Q \equiv \neg Q \Rightarrow \neg P$

P	Q	$\neg P$ Negation	$P \wedge Q$ Conjunction	$P \vee Q$ Disjunction	$P \Rightarrow Q$ Implication
T	T	F	T	T	T
T	F	F	F	T	F
F	T	T	F	T	T
F	F	T	F	F	T

De Morgan's Laws

- $(P \Rightarrow Q) \equiv (\neg P \vee Q)$
- $\neg(P \vee Q) \equiv (\neg P \wedge \neg Q)$
- $\neg(P \wedge Q) \equiv (\neg P \vee \neg Q)$
- $\neg \forall x\, P(x) \equiv \exists x\, \neg P(x)$
- $\neg \exists x\, P(x) \equiv \forall x\, \neg P(x)$

Domains

- $\mathbb{N}$: all natural #'s (including 0)
  - $(0, 1, 2, \ldots)$
- $\mathbb{Z}$: all integers
  - $(\ldots, -2, -1, 0, 1, 2, \ldots)$
- $\mathbb{Q}$: all rational numbers
  - $x = \frac{a}{b}$, ratio of 2 ints
- $\mathbb{R}$: all real numbers
  - rational & irrational
- $\mathbb{C}$: all complex numbers
  - imaginary numbers

Proof Techniques

- Direct Proof:
  - Prove $P \Rightarrow Q$:
  - Assume P, derive Q.
- Contraposition
  - Prove $P \Rightarrow Q$
  - Assume $\neg Q$, derive $\neg P$
- Contradiction
  - Prove P
  - Assume $\neg P$, find contradiction, so $\neg P$ is wrong, thus P.
- Proof by cases:
  - Prove Q
  - Show that all cases imply Q. Make sure cases cover every possibility.

Induction

Weak: Assume $P(k)$

Strong: Assume $P(0), P(1), \ldots, P(k)$

Prove $P(n)$

1. Base Case: Prove $P(0)$
2. Inductive Hypothesis: Assume $P(k)$
3. Inductive Step: Prove $P(k+1)$ using $P(k)$

$P(0)$, $P(k) \Rightarrow P(k+1)$

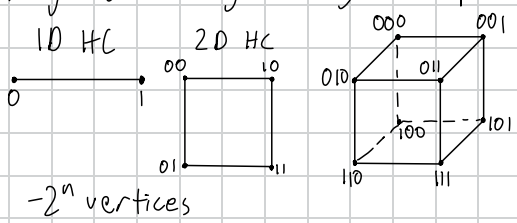
Stable Matching

Proof by Contradiction ☆

- A matching that does not have a rogue couple.
- Rogue couple: A job and candidate that prefer each other over their current matching.
- Optimal Candidate: Highest ranked candidate a job can be matched with in a stable matching.
- Pessimal Candidate: Lowest ranked candidate a job can be matched with in a stable matching.
- P&R Algorithm always terminates and outputs a stable, job optimal, and candidate pessimal matching.
- Job optimal ⇒ Candidate pessimal and vice versa.
- Improvement lemma: If job J makes an offer to candidate C on the kth day, C has a job offer in hand every day after with preference equal to or better than J.
- Terminates in $(n-1)^2+1$ days.
- Jobs proposing ⇒ Job optimal
- Assume stable and prove rogue couple or vice versa

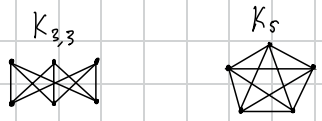
Graphs ☆ Proof by cases or induction

- A walk is any sequence of adjacent edges and vertices.
- A path is a walk without repeated edges or vertices.
- A tour is a walk with the same start and end.
- A cycle is a series of vertices in a closed chain with no repeated edges and the same start and end.
- An Eulerian tour visits every edge exactly once with the same start and end vertex. An undirected graph has an Eulerian tour iff G is even degree, connected, and every vertex is even degree.
- A Hamiltonian tour visits every vertex exactly once with the same start and end vertex. Every hypercube has a Hamiltonian tour.
- The degree of a given vertex is the # of neighbors it has.
- Handshake lemma: The sum of degrees of a graph is $2e$, where e = # of edges.
- Connectivity: There exists a path between any 2 vertices.
- Coloring: A graph is n-colorable if each vertex can have a different color than all of its adjacent vertices.
- Bipartite graphs can be split up into two sets of vertices such that all edges are only between the two sets. A bipartite graph is two colorable and vice versa.
- Complete graph: Each vertex has an edge to every other vertex. Complete graphs have $\frac{v(v-1)}{2}$ edges, where v = # of vertices.
- Hypercubes of n dimensions assign each vertex an n digit bit string, in which two vertices are connected iff their bit strings differ by exactly one position.



2^n vertices

- Trees are minimally connected graphs with no cycles ($v-1$ edges). A graph is a tree iff it is connected and acyclic or connected and $v-1$ edges or connected and removing an edge disconnects the graph or its acyclic and adding an edge creates a cycle.
- Planar graphs can be drawn without edges intersecting each other.
  - $e \leq 3v-6$ for a planar graph
  - $e \leq 2v-4$ for a planar bipartite graph
  - $v+f=e+2$ for all connected planar graphs, f is a face (bounded by edges).
  - Nonplanar if it contains $K_{3,3}$ or K_5.



- Vertex colored in at most 4 colors.
- Complement of G is $\bar{G}$. G and $\bar{G}$ have opposite edges.

Common Definitions

$a \mid b \longrightarrow b = ax$

rational $x = \frac{a}{b}$, where a and b are integers.

Bijection: You can pair every element is set A to a unique element in set B.

Modular arithmetic

- To find inverse of $a \pmod m$, $ax \equiv 1 \pmod m$, where x is the multiplicative inverse.

Exponentiation – Repeated Squaring:

$3^{13} \bmod 7$

$13 = 1101_2 = 8+4+1$

$3^1 \equiv 3 \pmod 7$
$3^2 \equiv 2 \pmod 7$
$3^4 \equiv 4 \pmod 7$
$3^8 \equiv 2 \pmod 7$

$3^{13} \equiv 3^8 \times 3^4 \times 3^1 \equiv 2 \cdot 4 \cdot 3 = 24 \pmod 7$
$3^{13} \equiv 3 \pmod 7$

$x \equiv a \pmod p$ is equal to $x = a + kp$ for some integer k.

$2^{(p-1)(q-1)} \equiv 1 \pmod{pq}$

Euclidean Algorithm: Find GCD

Steps to find GCD(a, b)

1. $a \div b = x$ remainder r.
2. Replace a with b and b with r.
3. Repeat until $r = 0$. The last nonzero r is the GCD.

GCD(48, 18)
$48 \div 18 = 2$ rem 12
$18 \div 12 = 1$ rem 6
$12 \div 6 = 2$ rem 0
GCD(48, 18) = 6

Extended Euclidean Algorithm

- Same as Euclidean Algorithm, but track equation $a = qb + r$.

GCD(48, 18)
$48 = 2 \cdot 18 + 12$
$18 = 1 \cdot 12 + 6$
$12 = 2 \cdot 6 + 0$

$12 = 48 - 2 \cdot 18$
$6 = 18 - 1 \cdot 12$
$6 = 18 - (48 - 2 \cdot 18)$
$6 = 18 - 48 + 2 \cdot 18$

Final result: GCD(48, 18) = 6
$6 = 18 - 1 \cdot (48 - 2 \cdot 18)$

Fermat's Little Theorem

For int a and prime p,
$a^{p-1} \equiv 1 \pmod p$ and
$a^p \equiv a \pmod p$

Fundamental Thm of Arithmetic

Every positive integer $n > 1$ can be expressed uniquely in the form $p_1 p_2 \cdots p_k$, where each p_i is a prime.

Chinese Remainder Theorem

$x \equiv a \pmod l$ $\quad \gcd(l, m) = 1$
$x \equiv b \pmod m$ $\quad \gcd(m, n) = 1$
$x \equiv c \pmod n$ $\quad \gcd(l, n) = 1$

There is exactly one $x \pmod{l \cdot m \cdot n}$ that satisfies the above equations.

Ex: $x \equiv 2 \pmod 3$ $\quad \gcd(3, 4) = 1$ ✓
$x \equiv 1 \pmod 4$ $\quad \gcd(3, 11) = 1$ ✓
$x \equiv 7 \pmod{11}$ $\quad \gcd(4, 11) = 1$ ✓

$l \cdot m \cdot n = 132$ Find inverses

$M_1 = 132/l = 132/3 = 44 \rightarrow 44^{-1} \equiv 2 \pmod 3$
$M_2 = 132/m = 132/4 = 33 \rightarrow 33^{-1} \equiv 1 \pmod 4$
$M_3 = 132/n = 132/11 = 12 \rightarrow 12^{-1} \equiv 1 \pmod{11}$

$x = 2 \cdot 44 \cdot 2 + 1 \cdot 33 \cdot 1 + 7 \cdot 12 \cdot 1 = 293 \bmod 132 \equiv 29 \pmod{132}$

$x = a \cdot mn \cdot (mn)^{-1} + b \cdot ln \cdot (ln)^{-1} + c \cdot lm \cdot (lm)^{-1}$

RSA Public Key Cryptography

- message = m
- The sender sends y, which equals $m^e \pmod N$
- The public key is (N, e). N is pq where p and q are primes and unknown to the sender.
- The private key has primes p and q and also $d = e^{-1} \pmod{(p-1)(q-1)}$
- e and $(p-1)(q-1)$ must be coprime.
- The receiver receives y and raises it to the d power.

$y = m^e \pmod N$
$y^d = m^{ed} \pmod N$
$y^d \pmod N = m$

Example: m = 2 Public key: (15, 3) Private key: d = 3, p = 3, q = 5

Encrypt: $y = m^e \pmod N = 2^3 \pmod{15} = 8 \pmod{15}$

Decrypt: $8^3 \pmod{15} = 512 \pmod{15} \equiv 2 \pmod{15}$

so m = 2.

- Use repeated squaring when encrypting and decrypting with large exponents.

Pigeonhole Principle: Let n and k be positive integers. Place n objects into k boxes. If $n > k$, at least one box must contain multiple objects.

Polynomials

- $P(x) = a_d x^d + a_{d-1} x^{d-1} + \ldots + a_0$
- degree d
- A polynomial of degree d has at most d roots (or infinity if $P(x) = 0$)
- A polynomial of degree d can be uniquely defined by a set of d+1 points.
- Galois Field (GF(p)): Fancy way of saying working in mod p, for a prime p.
- Given $(x_1, y_1), (x_2, y_2), (x_3, y_3)$

$$P(x) = \frac{y_1(x-x_2)(x-x_3)}{(x_1-x_2)(x_1-x_3)} + \frac{y_2(x-x_1)(x-x_3)}{(x_2-x_1)(x_2-x_3)} + \frac{y_3(x-x_1)(x-x_2)}{(x_3-x_1)(x_3-x_2)}$$

Secret Sharing

- You need m+1 packets to recreate a polynomial of degree m
- Packet: $(x, P(x))$

Error Type 1: Erasure

- Packets can get lost.
- Solution: Send N+K+1 packets where K is the maximum number of packets that will get lost.
- All N+K+1 packets must belong to the same degree N polynomial, so you can still reconstruct the polynomial.

Error Type 2: Corruption

- Packets get altered
- Unknown set of errors in the locations $\{e_1, e_2, \ldots, e_K\}$
- Solution: Send N+2K+1 packets for a polynomial of degree N.

Berlekamp Welch Algorithm

- $P(x)$ = message polynomial (unknown)
- $E(x)$ = error polynomial, roots at corrupted x's.
- $Q(x) = P(x)E(x)$
- Received packets (x_i, r_i)
- $Q(x_i) = P(x_i)E(x_i) = r_i E(x_i)$ at each packet (x_i, r_i)
  - If r_i is not corrupted, $P(x_i) = r_i$, so the equation holds.
  - If r_i is corrupted, $E(x_i) = 0$ since the roots of $E(x)$ are at the corrupted locations. Thus, the equation still holds.
  - Create $Q(x_i) = r_i E(x_i)$ for all packets and solve for $Q(x)$ and $E(x)$. Then, $P(x) = \frac{Q(x)}{E(x)}$

Counting

1. Does order matter? Ex: King, Queen vs. Queen, King
2. Is there replacement for the outcome? Can the same outcome happen again?
   Ex: Drawing from deck of cards has no replacement, but flipping a coin has replacement.

	Ordered (Permutations)	Unordered (Combinations)
No replacement	$\frac{n!}{(n-k)!}$	$\binom{n}{k} = \frac{n!}{k!(n-k)!}$
Replacement	n^k	$\binom{n+k-1}{k-1} = \binom{n+k-1}{n}$

n = # of options (2 for a coin: H & T)
k = # of times (How many flips, etc.)

Zeroth rule: If a bijection exists between sets A and B, $|A| = |B|$, they have the same size.
First rule: For a sequence of independent choices, multiply the options for each choice.
Second rule: If every object in B corresponds to m objects in A, $|A| = m \cdot |B|$
Ex: $B = (\{a,b,c\}, \{a,c,d\}, \text{etc.})$ → order doesn't matter
$A = ((a,b,c), (a,c,b), \text{etc.})$ → order matters

Graph Theory

**Except for start and end of cycle

Name	No Repeated Vertices	No Repeated Edges	Start = End
Path	✓	✓	
Cycle	✓**	✓	✓
Hamiltonian Path	✓	✓	
Hamiltonian Cycle	✓**	✓	✓
Walk			
Tour			✓
Eulerian Walk		✓	
Eulerian Tour		✓	✓

Hamiltonian visits every vertex once; Eulerian visits every edge once

Galois Fields

GF(p), where p is prime = work in (mod p).
Property 1: A degree d polynomial has at most d roots.
Property 2: Given d+1 points with distinct coordinates, there is a unique polynomial of degree at most d that passes through all d+1 points.

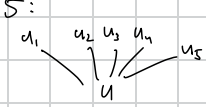
For GF(p), # of points = d − x → # of polynomials = p^{x+1}

Inductive Graph Proofs

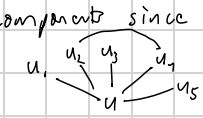
Euler's formula:
For every connected planar graph, v+f=e+2
Proof:
BC: When e=0, and v=f=1, 2=2 ✓
Two cases: G is a tree, then f=1 and e=v−1.
OR
G is not a tree, then find a cycle and remove one of its edges, then e=e−1 and f=f−1. From the inductive hypothesis, the formula is true, so it must be true in the inductive step.

5 color Theorem for Planar Graphs:

Assume vertex u has degree 5. If its 4 or less, it is obviously 5 colorable.
If its 5:



If u_2 and u_4 are in different components (not including u) then we can swap every 2 to 4 and 4 to 2 in u_2's component. This will leave u_2 as color 4 and u can be colored with 2. If u_2 and u_4 are in the same component, then u_1 and u_3 must be in different components since the graph is planar



There is no way to connect u_1 and u_3 without intersecting lines. So since u_1 and u_3 are in different components, we can swap the 3's and 1's in u_1's component. Therefore u_1 is color 3 and u can be colored 1.

Inductive Tree Proof

The statements "G is connected and contains no cycles" and "G is connected and has n−1 edges" are equivalent.

Forward direction: contains no cycles → n−1 edges
BC: n=1: no cycles and 0 edges ✓
IH: Assume for tree n=k, k−1 edges.
IS: For n=k+1, remove leaf node and corresponding edge. Still acyclic so n=k so from IH, k−1 edges. Add back leaf node and edge → n=k+1, k edges.

Backward Direction: n−1 edges → no cycles
Assume G has a cycle with n−1 edges. Removing edge e, means n−2 edges. But trees must have n−1 edges so contradiction.

Stable Matching Proofs

The P&R algorithm always halts.
Proof: At least one candidate eliminated from a job each day. n cands & n jobs → Algorithm must terminate in at most n^2 days.

Improvement Lemma

Induct on day i=k, i=k+1. Offer on day k → Offer on day k+1

P&R algorithm terminates with a matching.

Proof by contradiction: J left unpaired → rejected by n candidates who got better offers, so n jobs + J = n+1 jobs contradicts assumption of n jobs.

The matching produced by P&R is always stable.

J likes C* more than C and so offered C* before C but got rejected. That means C* likes its final job at least as much as J, so not a rogue couple → stable.

P&R is proposer optimal for output T

Assume ¬optimal, then J rejected by C* in favor of job J*, but there exists some stable T, where (J, C*) and (J*, C) exist. We know C* prefers J*. Assume J* proposed to C* on day k and k is the first day any job was rejected by its optimal candidate. Then C* must be equal in preference to J*'s optimal candidate. Therefore, (J*, C*) is a rogue couple in T, which means it can't be stable. Contradiction.

Proposer Optimal ⇒ Receiver Pessimal

Let T be the job optimal output which contains (J, C). Assume S is stable and contains (J, C*) and (J*, C) where J* < J for C. So C prefers J to J* and J prefers C to C* because C is matched with J in the optimal matching T. Thus, (J, C) is a rogue couple in S and so is not stable. So, job optimal ⇒ candidate pessimal.

Graphs

A complete graph K_n with n vertices has $\binom{n}{2}$ edges.

$\sum_{i=1}^{f} s_i = 2e$. The total sum of the size of all faces in a planar graph G is 2e.

Finding the multiplicative inverse of 7 (mod 68)

$7(0) + 68(1) = 68$ Reduce until 1
$7(1) + 68(0) = 7$ −29 is the inverse.
$7(-9) + 68(1) = 5$ $-29 = 39 \bmod 68$.
$7(10) + 68(-1) = 2$
$7(-29) + 68(3) = 1$

In mod p, a polynomial can have a leading coefficient up to p−1, while the other d coefficients can be up to p.

For example, there are $(p-1)p^d$ polynomials of degree d in mod p.

There are $(p-1)p$ polynomials of the form $a(x-r)^2$.

There are $(p-1)\binom{p}{2}$ polynomials of the form $a(x-r_1)(x-r_2)$.

There are $(p-1)(p^2 - \binom{p}{2} - p)$ polynomials with 0 roots and degree d. $(p-1)\binom{p}{2}$ represents all degree 2 polynomials with 2 roots and $(p-1)p$ represents all degree 2 polynomials with 1 root. $(p-1)p^2$ represents all polynomials of degree 2.

CANTSTANFURD can be rearranged in $\frac{12!}{2!2!2!} = x$
Factorial of total letters / Factorial of each repeat

X letters can be arranged X! ways
Divide by additional conditions.
Ex: $\frac{X}{3!}$ so C is before S is before D in any ordering

Countability

Find a bijection between the set and a known countable set.

Countable sets: Finite strings (with finite alphabet)
Naturals
Pairs of countable sets

Uncountable sets can be diagonalized.

Diagonalization:

1	(A)	B	C	D
2	B	(C)	D	A
3	C	D	(A)	B
4	D	A	B	(C)

Assume this set contains all 4 letter strings using ABCD. Take the diagonals and change them so now its BDBD. This string differs from string 1 at position 1, string 2 at position 2, etc.

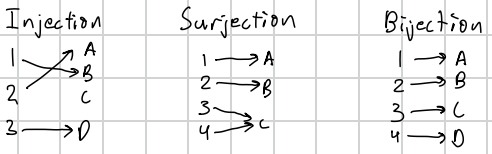
Proof by contradiction: Assume the diagonalized string is the kth string. However, we know it differs from the kth string at position k, so it cannot be the kth string.

Bijections

Injection: f is an injection if it maps distinct inputs to distinct outputs (one to one function)

Surjection: f is a surjection if all items in B are being mapped to.

Bijection: f is a bijection iff f is an injection and surjection



Computability

Given some task, is it possible for a computer to solve it?

Halting Problem: TestHalt(P, x) = true if program P halts on input x, false if P loops on x. Define Turing(P) to loop if TestHalt(P,P) = true, otherwise halt. Call Turing(Turing). If it loops, it means TestHalt(Turing, Turing) returned true, which means Turing(Turing) halted. This is a contradiction and the same applies if the call halts initially. So therefore, our assumption of TestHalt existing and working for all inputs is incorrect.

Discrete Probability

Sample point (ω): The outcome of a single random experiment. (HHHH, HHHT, etc)

Sample Space (Ω): The set of all possible outcomes. (HHHH → TTTT)

Probability Space: The sample space Ω together with P[ω] for each point ω, s.t. all probabilities are between 0 and 1 inclusive, and the sum of all probabilities is 1.

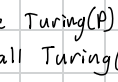
Event (A): Subset of the sample space Ω. $P[A] = \sum_{\omega \in A} P[\omega]$

Complement ($\bar{A}$): Set of all sample points not in event A. $P[A] + P[\bar{A}] = 1$

Uniform Probability Space: A probability space in which each sample point has the same probability.

$$P[A] = \frac{|A|}{|\Omega|} = \frac{\#\text{ outcomes that satisfy A}}{\#\text{ total outcomes}}$$

Conditional Probability



Event Partioning: $A = A_1 \cup A_2 \cup \ldots \cup A_n$ and $A_1, \ldots, A_n$ are mutually exclusive.

Total Probability Rule: $P[B] = \sum_{i=1}^{n} P[B \cap A_i] = \sum_{i=1}^{n} P[B|A_i]P[A_i]$

Since A_i covers all of the sample space, so you can sum the probabilities of B and A_i happening at the same time.

Independence: Events A and B are independent if knowing one happened doesn't change the probability of the other.

$P(A|B) = P(A)$ OR
$P(B|A) = P(B)$ OR
$P(A \cap B) = P(A)P(B)$

Pairwise Independence: Assume events A, B, C. These are pairwise independent if AB, BC, and AC are all independent

Mutual Independence

Assume events $A_1, A_2, \ldots, A_n$. They are mutually independent if every possible group of them is independent. That means every pair, triple, etc. are independent. Pairwise independence doesn't imply mutual independence.

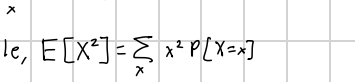
$$P[\cap_{i \in I} A_i] = \prod_{i \in I} P(A_i)$$

The set of events A_i is mutually independent if the probability that they all happen at the same time is equal to the product of their individual probabilities.

Inclusion-Exclusion

When finding $P(A \cup B)$, adding their individual probabilities double counts the shared outcome.

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$



$$P(A \cup B \cup C) = P(A)+P(B)+P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

Mutually Exclusive Events

$P(\cup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)$. If events are mutually exclusive, the probability of their union is the sum of their individual probabilities as there is no double counting.

Union Bound

$P(\cup_{i=1}^{n} A_i) \leq \sum_{i=1}^{n} P(A_i)$. The probability of the union of the events is at most the sum of their individual probabilities. This occurs when the events are mutually exclusive, otherwise the probability of their union will be less.

Random Variable

Define random variable x = # of heads

For outcome w = HHH, x = 3
w = HHT, x = 2, etc.

Expectation - Center of mass of a distribution.

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

Linearity of Expectation: $E[X+Y] = E[X] + E[Y]$
$E[cX] = cE[X], \forall c \in \mathbb{R}$

Law of the Unconscious Statistician (LOTUS):

$$E[f(X)] = \sum_{x} f(x) P(X = x)$$

For example, $E[X^2] = \sum_{x} x^2 P[X = x]$

Variance: Average squared distance of a random variable from its mean. $Var(X+Y) = Var(X) + Var(Y)$ if independent.

$$Var(X) = E[(X - E[X])^2] = E[X^2] - (E[X])^2$$

Covariance: Measures the linear association between two random variables

$$Cov(X,Y) = E[(X - E[X])(Y - E[Y])] = E[XY] - E[X]E[Y]$$

If X and Y are independent, the $Cov(X,Y) = 0$ b/c $E[XY] = E[X]E[Y]$

Correlation: Normalized covariance

$$Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}} = P(X \cap Y) - P(X)P(Y)$$

Properties:

$$Var(cX) = c^2 Var(X)$$
$$Var(X + b) = Var(X)$$
$$Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$$
$$Cov(X,X) = Var(X)$$

Bernoulli Distribution

X is a Bernoulli random variable with parameter p. $X \in \{0, 1\}$. $P(X=1) = p$, $P(X=0) = 1-p$.

Let I_A be the Bernoulli variable for event A w/ parameter P(A).

$$E[I_A] = P(A) = p$$
$$Var(I_A) = p(1-p)$$

Binomial Distribution

Models the number of successes in n independent identical trials, where each trial succeeds with probability p. For example, flipping a coin multiple times. Converges to Poisson as $n \to \infty$

Let $X_i = \begin{cases} 1 & \text{if flip } i \text{ is H} \\ 0 & \text{otherwise} \end{cases}$

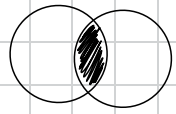
$X_i \sim Bernoulli(p)$ X_i is Bernoulli RV with probability p

Let $X = \sum_{i=1}^{n} X_i$ $P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$

$X \sim Bin(n, p)$ X is Binomial RV with n trials and probability p.

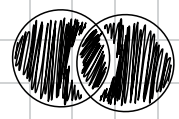
$$E[X] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} p = np \qquad Var(X) = \sum_{i=1}^{n} Var(X_i) = \sum_{i=1}^{n} p(1-p) = np(1-p)$$

Intersection $P(A \cap B) \Rightarrow$



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

Union $P(A \cup B) \Rightarrow$



Σ = summation
Π = product
$\sigma^2 = Var(X)$

Geometric Distribution

Models how many trials are needed until the first success. Let X = # of flips until first H.

$X \sim Geometric(p)$ X is a geometric RV with probability p.

$$P(X = x) = (1-p)^{x-1} p \qquad E[X] = \frac{1}{p} \qquad Var(X) = \frac{1-p}{p^2}$$

Memoryless property: If m trials have already failed, the remaining waiting time is distributed the same as if you are starting over. Essentially, the coin doesn't remember its past flips.

$$P(x > m+n \mid x > m) = P(X > n)$$

Poisson Distribution

Models how many times an event happens in a fixed interval.

λ = average number of events per interval

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!} \qquad E[X] = \lambda \qquad Var(X) = \lambda$$

Distributions

Name	Parameters	Abbv.	PMF ($P(X=x)$)	E[X]	Var(X)
Bernoulli	p = probability of success	$X \sim B(p)$	$P(X=0) = 1-p$, $P(X=1) = p$	$E[X] = p$	$Var(X) = p(1-p)$
Binomial	n = number of trials, p = probability of success	$X \sim Bin(n,p)$	$P(X=x) = \binom{n}{p} p^x (1-p)^{n-x}$	$E[X] = np$	$Var(X) = np(1-p)$
Geometric	p = probability of success	$X \sim Geom(p)$	$P(X=x) = p(1-p)^{x-1}$	$E[X] = \frac{1}{p}$	$Var(X) = \frac{1-p}{p^2}$
Poisson	λ = mean number of events	$X \sim Pois(\lambda)$	$P(X=x) = \frac{\lambda^x e^{-\lambda}}{x!}$	$E[X] = \lambda$	$Var(X) = \lambda$

Markov's Inequality

$$P(X \geq c) \leq \frac{E(X)}{c}$$ for a nonnegative RV, X with a finite mean.

The probability that $X \geq c$ is at most the mean of X over c. The chance that X is at least a is at most the average size of X divided by a. Given a small mean, the probability of getting a large value is small.

Chebyshev's Inequality

$$P(|X - E[X]| \geq c) \leq \frac{Var(X)}{c^2}$$ for both tails

The probability that a RV X is at least c away from its mean is at most $Var(X)/c^2$. The probability a RV is far from its mean is small and it shrinks with 1/(distance)²

$$P(X - E[X] \geq c) \leq P(|X - E[X]| \geq c) \leq \frac{Var(X)}{c^2}$$

Single tail Both tails

Central Limit Theorem

When n is large, the distribution of the sample mean or sample sum of identical distribution RVs looks Normal, regardless of the original distribution.

$X_1, X_2, X_3, \ldots, X_n$: Independent and identically distributed (i.i.d.) RVs

Each X_i has mean, μ, and standard deviation, σ.

Sample Sum: $S_n = \sum_{i=1}^{n} X_i$ $E[S_n] = n\mu$ $SD(S_n) = \sqrt{n}\sigma$

$$S_n \approx N(\underbrace{n\mu}_{\text{mean}}, \underbrace{n\sigma^2}_{\text{variance}}) \text{ for large } n$$

Sample Mean: $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ $E[\bar{X}_n] = \mu$ $SD(\bar{X}_n) = \frac{\sigma}{\sqrt{n}}$

$$\bar{X}_n \approx N(\underbrace{\mu}_{\text{mean}}, \underbrace{\frac{\sigma^2}{n}}_{\text{variance}}) \text{ for large } n$$

Continuous Random Variable

Takes uncountably many values (any real number in an interval)

$P(X=a) = 0$ because a single point has no width. Instead of $P(x=a)$, $P(a \leq x \leq b)$. What is the probability that a RV lies in some interval?

Probability Density Function

The PDF is the actual curve of the distribution. Probability is the area under the curve.

Properties:

① Non-negativity: $f_X(x) \geq 0$ for all x.

② Interval Probability: $P(a \leq X \leq b) = \int_a^b f_X(x)\,dx$

③ Normalization: $\int_{-\infty}^{\infty} f_X(t)\,dt = 1$ Total probability sums to 1

$$E = \int_{-\infty}^{\infty} x \cdot f_X(x)\,dx = \int_0^{\infty} P(X \geq x)\,dx$$

Cumulative Distribution Function

$F_X(a) = P(X \leq a)$. Gives the probability that X is to the left of a.

$$P(a \leq X \leq b) = P(a < X < b)$$

$$F_X(a) = \int_{-\infty}^{a} f_X(x)\,dx \qquad \frac{d}{dx} F_X(x) = f_X(x)$$

Properties:

① Non decreasing: $F_X(x)$ never goes down, only add probability as you go right.

② Limits: $\lim_{x \to -\infty} F_X(x) = 0$ $\lim_{x \to \infty} F_X(x) = 1$

-∞ and ∞ refer to the bounds of the domain of X. So if X is only over c to d, use c and d instead of -∞ and ∞.

Discrete vs. Continuous

	Discrete	Continuous
Expectation	$E[X] = \sum_{a \in A} a \cdot P[X=a]$	$E[X] = \int_{-\infty}^{\infty} a \cdot f_x(a)\, da$
LOTUS	$E[g(X)] = \sum_x g(x) P(X=x)$	$E[g(X)] = \int_{-\infty}^{\infty} g(t) f_x(t)\, dt$
Total Prob	$P[B] = \sum_{i=1}^{n} P[B\|A_i] P[A_i]$	$P[B] = \int_{-\infty}^{\infty} P[B\|A_i] f_x(t)\, dt$
Variance	$Var(X) = E[X^2] - (E[X])^2$	$Var(X) = \int_{-\infty}^{\infty} x^2 f(x)\, dx - \left(\int_{-\infty}^{\infty} x f(x)\, dx\right)^2$

Uniform Distribution

A random variable is uniform on an interval if every value of the RV within the interval is equally likely

$X \sim Uniform(a,b)$, where a,b is the interval

$$PDF: f_x(x) = \begin{cases} \frac{1}{b-a} & x \in [a,b] \\ 0 & \text{otherwise} \end{cases} \qquad CDF: F_x(x) = \begin{cases} 0 & x<a \\ \frac{x-a}{b-a} & a \le x \le b \\ 1 & x>b \end{cases}$$

$$P(c \le x \le d) = \frac{d-c}{b-a} = \frac{\text{interval of probability}}{\text{interval of distribution}}$$

$$E[X] = \frac{a+b}{2} \qquad Var(X) = \frac{(b-a)^2}{12}$$

Exponential Distributions

Models waiting time until an event happens

$X \sim Exp(\lambda) \rightarrow X$ = time until first success, λ = rate at which successes happen

$$PDF: f_x(x) = \begin{cases} \lambda e^{-\lambda x} & x \ge 0 \\ 0 & \text{otherwise} \end{cases} \qquad CDF: F_x(t) = P(x \le t) = 1 - e^{-\lambda t}$$

Probability that X happens by t

$$P(a \le x \le b) = \int_a^b \lambda e^{-\lambda x}\, dx$$

Survival function: Probability that object survives past t

$$S_x(t) = P(X > t) = 1 - F_x(t) = e^{-\lambda t}$$

$$E[X] = \frac{1}{\lambda} \qquad Var(X) = \frac{1}{\lambda^2}$$

Memoryless property: If m time has already past, the additional waiting time does not depend on the past.

$$P(x > m+n \mid x > m) = P(X > n)$$

Normal Distribution

$X \sim N(\mu, \sigma^2)$ means X is normally distributed with mean μ and variance σ^2

$$PDF: f_x(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/(2\sigma^2)}$$

$$P(a \le X \le b) = \int_a^b f_x(x)\, dx$$

$$CDF: F_x(x) = P(X \le x) = \frac{x-\mu}{\sigma} \cdot \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}}\, dt$$

$$E[X] = \mu \qquad Var(X) = \sigma^2$$

$X \sim N(\mu, \sigma^2)$

$Z = \frac{X-\mu}{\sigma}$

$Z \sim N(0,1)$

$P(X \le x) = P\left(Z \le \frac{x-\mu}{\sigma}\right)$

Sum of Independent Normals:

$X \sim N(0,1)$ and $Y \sim N(0,1)$

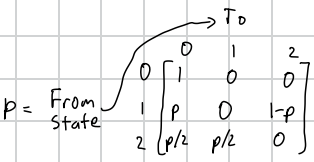
$Z = aX + bY \sim N(0, a^2+b^2)$

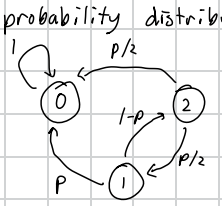
Markov Chains

States - represented by circles

Transition Probabilities going out of any state should add to 1.

$\Pi = \Pi P$, Π is a vector of the probability distribution of X_n. $\Pi = [\pi_0\ \pi_1\ \dots\ \pi_n]$





All rows add to 0.

Represent final states with a self loop of $p=1$.

Solving:

① Write out balance equations and $\sum_{i=0}^{n} \pi_i = 1$

$$[\pi_0\ \pi_1\ \pi_2] = [\pi_0\ \pi_1\ \pi_2] \begin{bmatrix} 1 & 0 & 0 \\ p & 0 & 1-p \\ p/2 & p/2 & 0 \end{bmatrix}$$

$\pi_0 = \pi_0 + \pi_1 p + \pi_2 \frac{p}{2}$ Solve for all π_i in terms of p.

$\pi_1 = \pi_2 \frac{p}{2}$

$\pi_2 = \pi_1 (1-p)$

$\pi_0 + \pi_1 + \pi_2 = 1$

Markov Chains Definitions

Irreducible: Can go from every state i to every other state in a finite # of steps. Also means Π exists.

Period of State i: Find all the possible paths to go from state i back to state i. Record the step count for each. Take the gcd of all the step counts to get the period of State i.

IF A MARKOV CHAIN IS IRREDUCIBLE, ALL STATES HAVE THE SAME PERIOD!!!

Aperiodic

Period of all states is 1.

Fundamental Theorem of Markov Chains

If a Markov Chain is irreducible and aperiodic, then for any starting distribution (i.e. [0.5 0.5] means 50% chance start in state 0 and 50% in state 1), the probability of being in state i at step n converges to the same value π_i from the invariant distribution Π.

Formally, if a Markov Chain is irreducible and aperiodic, then for any initial distribution Π_0, we have that $\Pi_n \rightarrow \Pi$ as $n \rightarrow \infty$, and Π is the unique invariant distribution for the Markov chain.

Markov Chain Applications

Let X = # of steps before reaching state A:

$a(i) = 0$ if $i = A \Rightarrow$ Already at A

$a(i) = 1 + \sum P(i,j)\, a(j) \Rightarrow$ Step from i to j plus future steps

Probability of Reaching A before B:

$a(i) = 1$ if $i = A \Rightarrow$ Already at A

$a(i) = 0$ if $i = B \Rightarrow$ Already at B

$a(i) = \sum P(i,j)\, a(j) \Rightarrow$ Land in state j with $p = P(i,j)$ and future steps $= a(j)$

Euclidean Algorithm for Finding Inverses

Find inverse of $63 \pmod{71}$

$63(0) + 71(1) = 71$	E_1
$63(1) + 71(0) = 63$	E_2
$63(-1) + 71(1) = 8$	$E_3 = E_1 - E_2$
$63(8) + 71(-7) = 7$	$E_4 = E_2 - 7E_3$
$63(-9) + 71(8) = 1$	$E_5 = E_3 - E_4$

So $63(-9) \equiv 1 \pmod{71}$ so inverse is $-9 \equiv 62 \pmod{71}$

Confidence Interval

Using Chebyshev's Inequality, give a 95% confidence interval for p, given that 50 people in your sample of 100 people tested positive for having the flu.

Let X = proportion of ppl who test positive

$$X = \frac{1}{100} \sum_{i=1}^{100} X_i$$

$$Var(X) = Var\left(\frac{1}{100} \sum_{i=1}^{100} X_i\right) = \frac{1}{100^2} \sum_{i=1}^{100} Var(X_i) = \frac{1}{100} Var(X_i)$$

Assume $p = \frac{1}{2}$. $Var(X_i) = p(1-p) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$

$$Var(X) = \frac{1}{100} Var(X_i) = \frac{1}{400}$$

$$P[|X - E[X]| \ge c] \le \frac{Var(x)}{c^2} = \frac{1}{400c^2}$$

We want $\frac{1}{400c^2}$ to be at most $1 - 95\% = 5\% = 0.05 = \frac{1}{20}$

$$\frac{1}{20} = \frac{1}{400c^2} \Rightarrow 400c^2 = 20 \Rightarrow c^2 = \frac{1}{20} \Rightarrow c = \frac{1}{\sqrt{20}}$$

Actual center $= \frac{50}{100} = \frac{1}{2}$

$$CI: \left[\frac{1}{2} - \frac{1}{\sqrt{20}}, \frac{1}{2} + \frac{1}{\sqrt{20}}\right]$$

Halting Problems

① Identify the problem being solved: Does P(x) halt, loop, etc.

② Identify the magic tool (special function)

③ Create an inner function and pass it as an argument to the magic tool.

④ Use some sort of signal in the inner function, usually call the P(x) for the program P and input x, given as parameters.