CMSC 27100 - Discrete Mathematics Notes

1. Propositional Logic

Definition. Proposition: a true or false statement.

Definition. Implication:

A	B	$A \Rightarrow B$
T	T	T
T	F	F
F	T	T
F	F	T

Definitions. The converse of $p \Rightarrow q$ is $q \Rightarrow p$. Note that these are not equivalent statements. The contrapositive of $p \Rightarrow q$ is $\neg q \Rightarrow \neg p$. These are equivalent statements.

Definition: De Morgan's Laws.

$$\neg(p \vee q) = \neg p \wedge \neg q$$
$$\neg(p \wedge q) = \neg p \vee \neg q$$

Definition: Negating Quanitifers. Suppose $Q(x) = (\forall x, R(x))$. Then:

$$\neg Q(x) = (\exists x : \neg R(x))$$

If, instead, $Q(x) = (\exists x, R(x))$, then:

$$\neg Q(x) = (\forall x, \neg R(x))$$

2. Set Theory

Definition. A set is an unordered collection of objects.

Definition. Two sets are equal if and only if they have the same elements:

$$A = B \iff (\forall x, x \in A \iff x \in B)$$

Definition. The cardinality of a set, $|A|$, is the number of elements in the set.

Definition. The power set of A, $\mathscr{P}(A)$, is the set of all its subsets.

Insight

A set of n elements has 2^n subsets, as each item can either be included or excluded (two options for all n items).

. . . .

Definition. The union of A and B $=$ $A \cup B$:

$$\{x | x \in A \lor x \in B\}$$

The intersection of A and B $=$ $A \cap B$:

$$\{x | x \in A \land x \in B\}$$

Definition. The cartesian product $A \times B$ contains all ordered pairs (x, y) where $x \in A, y \in B$.

Insight

Note that in general, $A \times B \neq B \times A$, however we do have that $|A \times B| = |B \times A| = |A| \cdot |B|$.

. . . .

Theorem. If $A \times B \subseteq A \times C$, and $A \neq \varnothing$, then $B \subseteq C$.

Proof

We will show that for any element $b \in B$, we also have that $b \in C$. Take an arbitrary $a \in A$, and consider $(a, b) \in A \times B$. Because $A \times B \subseteq A \times C$, we have that $(a, b) \in A \times C$. Thus, by the definition of the cartesian product, $b \in C$. □

Definition. For two sets A and B, $A \setminus B$ is the set of all elements in A that are not in B.

Theorem. If $(A \setminus B) \cup (B \setminus A) = A \cup B$, then $A \cap B = \varnothing$.

Proof

We will prove by contrapositive. If $A \cap B$ is not null, then $\exists x \in A \cap B$. Any x in $A \cap B$ is also in $A \cup B$. We know that x cannot be in $A \setminus B$ as it is in B. Further, x cannot be in $B \setminus A$ as it is in A. Thus, if $A \cap B \neq \varnothing$, then $(A \setminus B) \cup (B \setminus A) \neq A \cup B$. □

Lemma. For all sets A and B:

$$\begin{gathered} A = (A \setminus B) \cup (A \cap B) \\ \Rightarrow |A| = |A \setminus B| + |A \cap B| \end{gathered}$$

Theorem. For all finite sets A and B, $|A \cup B| = |A| + |B| - |A \cap B|$.

Proof

Notice that $|A \cup B| = |A \setminus B| + |B \setminus A| + |A \cap B|$. These three sets "partition" $A \cup B$, as they are disjoint and constitute it. We can rewrite this as:

$$|A \cup B| = |A \setminus B| + |B \setminus A| + |A \cap B| + |A \cap B| - |A \cap B|$$

Now, using out lemma to rewrite $|A \cap B|$, we get:

$$\begin{aligned} |A \cup B| &= |A \setminus B| + |B \setminus A| + |A| - |A \setminus B| + |B| - |B \setminus A| - |A \cap B| \\ &= |A| + |B| - |A \cap B| \end{aligned}$$

□

---

3. Functions

Definition. A relation R with domain A and codomain B is a subset of $A \times B$.

Definition. A relation $R \subseteq A \times B$ is total if $\forall a \in A, \exists b \in B$ such that $(a,b) \in R$. In other words, every point in the domain has a corresponding point in the codomain.

Definition. A relation R is single-valued if $\forall a \in A, \forall b_1, b_2 \in B$, we have that:

$$(a,b_1) \in R \wedge (a,b_2) \in R \Rightarrow b_1 = b_2$$

I.e., each input has only 1 output.

Definition. A function $f : A \to B$ is a total, single-valued relation with domain A and codomain B.

Definition. A function $f : A \to B$ is injective (one-to-one) if $\forall x, y \in A$, $f(x) = f(y) \Rightarrow x = y$.

Insight

This is akin to applying single-valuedness to B; no two outputs can be the same.

. . . .

Definition. A function $f : A \to B$ is surjective (onto) if $\forall b \in B, \exists a \in A$ such that $f(a) = b$.

Insight

This is the reciprocal notion of totalness applied to B; every point in B must map back to A.

---

4. Number Theory

Definition. “a divides b” or $a|b \iff \exists d : a \cdot d = b$.

Theorem. $\forall a, b_1, b_2,\ a|b_1 \wedge a|b_2 \Rightarrow a|(b_1 + b_2)$

Proof

We know $\exists d_1, d_2 : a \cdot d_1 = b_1$ and $a \cdot d_2 = b_2$. Notice

$$\begin{aligned} b_1 + b_2 &= ad_1 + ad_2 \\ &= a(d_1 + d_2) \end{aligned}$$

If $(b_1 + b_2) = a(d_1 + d_2)$, then $a|(b_1 + b_2)$. □

Definition. A relation with respect to sets A, B is a function that maps $A \times B \to$ {true, false}. $R(a,b)$ is true or false depending on if (a,b) is in the relation.

Definition. A relation is transitive iff:

$$\forall a,b,c,\ R(a,b) \land R(b,c) \Rightarrow R(a,c)$$

Theorem. Divisibility is transitive:

$$\forall a,b,c,\ a|b \land b|c \Rightarrow a|c$$

Proof

We know $\exists d_1 : a \cdot d_1 = b$ and $\exists d_2 : b \cdot d_2 = c$. By plugging in for b, we get:

$$\begin{aligned} b \cdot d_2 &= c \\ (a \cdot d_1) \cdot d_2 &= c \end{aligned}$$

□

Definition. The set of divisors of n is denoted $\text{Div}(n) = \{a : a|n\}$.

Theorem.

$$a|b \Rightarrow \text{Div}(a) \subseteq \text{Div}(b)$$

Proof

It suffices to show that $\forall a' \in \text{Div}(a), a'|b$. We know that $\exists d : a \cdot d_1 = b$, and we also see that $a' \in \text{Div}(a) \Rightarrow \exists d_2 : a' \cdot d_2 = a$. Substituting for a:

$$\begin{aligned} a \cdot d_1 &= b \\ (a' \cdot d_2) \cdot d_1 &= b \end{aligned}$$

Thus, all divisors of a also divide b if $a|b$. □

Definition. If m,n are $\mathbb{Z}^+$, then $\text{GCD}(m,n)$ is the largest element of $\text{Div}(m) \cap \text{Div}(n)$

Theorem: The Division Theorem. $\forall n,d > 0$, $\exists! q,r$ such that $n = d \cdot q + r$ with $r \in [0,d)$. [a]

[a] Note $\exists!$ denotes unique existence.

Insight

This is just long division. We have our dividend n, our divisor d, our quotient q, and remainder r. Notably, this decomposition is only unique when $0 \leq r < d$.

....

Proof: Nonconstructive Existence of the Division Theorem

Consider the set $A = \{n - d \cdot q | q \in \mathbb{Z}\}$, where d is a positive integer. This is the set of all possible remainders, where $A = \mathbb{Z}$. Because $d \neq 0$, A must have a non-negative number. By the well-ordering principle, A must have a minimum non-negative number r.

Now we prove that $r < d$. FTSOC, assume $r > d$. Then $r - d > 0$, but this contradicts r being the minimum of A. □

....

Proof: Uniqueness of the Division Theorem

FTSOC, suppose that for a fixed but arbitrary $n,d > 0$, $\exists q_1, q_2, r_1, r_2$ such that:

$$\begin{aligned} n &= q_1 \cdot d + r_1 \\ n &= q_2 \cdot d + r_2 \end{aligned}$$

Then, by subtracting both sides:

$$\begin{aligned} 0 &= q_1 \cdot r_1 - q_2 \cdot d - r_2 \\ &= d(q_1 - q_2) + r_1 - r_2 \ (*) \end{aligned}$$

We know that $d|0$, but does $d|(*)$? Yes, because:

$$d|(a+b) \wedge d|a \Rightarrow d|b$$

Thus, we have:

$$d|(d(q_1 - q_2) + r_2 - r_2) \wedge d|(d(q_1 - q_2)) \Rightarrow d|(r_1 - r_2)$$

If, however, $d|(r_1 - r_2)$—which we call $(**)$—and $r_1, r_2 < d$, it must be that $r_1 - r_2 = 0 \Rightarrow r_1 = r_2$. If a larger number divides a smaller one, the smaller one must be 0.

So now we must also have that $d(q_1 - q_2) = 0$, as $(*)$ has become $0 = d(q_1 - q_2) + 0$. But because $d > 0$, it must be that $q_1 - q_2 = 0 \Rightarrow q_1 = q_2$. □

Definition. $\forall n, d > 0$, $n = q \cdot d + r$ where $0 \geq r < d$, and this is unique. We now define:

$$n \textbf{ div } d = q$$
$$n \textbf{ mod } d = r$$

Theorem: Bezout's Identity. Suppose $d = \text{GCD}(a, b)$. There exist integers x, y such that $d = ax + by$.

Proof

First, note that if $a = b = 0$, the GCD is not well defined, so one of $a, b \neq 0$. Consider $S = \{ax + by > 0 : x, y \in \mathbb{Z}\}$. This is a nonempty positive set, so by the well-ordering principle, it has a minimum element $d = as + bt$. We claim $d = \text{GCD}(a, b)$.

First, we show $d|a$ and $d|b$. When we divide a by d, we get $a = dq + r$. We claim that $r \in S \cup \{0\}$, as:

$$\begin{aligned} r &= a - dq \\ &= a - (as + bt)q \\ &= a(1 - qs) - b(qt) \end{aligned}$$

Notice that we have written r in the form $ax + by$, meaning it must be that $r \in S$ or $r = 0$. However, d is the smallest positive integer in S, and $r < d$, so we must have that $r = 0$.

Now we prove d is the greatest of the common divisors. Take $c \in \text{Div}(a) \cap \text{Div}(b)$. We know $\exists u, v : a = cu \land b = cv$. From here,

$$\begin{aligned} d &= as + bt \\ d &= cus + cvt \\ d &= c(us + vt) \\ &\Rightarrow c|d \\ &\Rightarrow d > c \text{ or } c = 0 \end{aligned}$$

Thus, $d \geq c$, which proves that d is the greatest, as no divisor can be greater than d. □

Theorem: Euclid's Algorithm. If $d = \text{GCD}(a,b)$, with $b \neq 0$ and $r = a \mod b$, then $d = \text{GCD}(b,r)$.

Proof

We show $\text{Div}(a) \cap \text{Div}(b) = \text{Div}(b) \cap \text{Div}(r)$. Let $q = a$ **div** b, so $a = q \cdot b + r$. Now let $z \in \text{Div}(a) \cap \text{Div}(b)$. Because $z \in \text{Div}(a)$, we know $z|a$. Further, because $z \in \text{Div}(b)$, we know $z|b \Rightarrow z|(q \cdot b)$. From here:

$$z|a \wedge z|(q \cdot b) \Rightarrow z|(a - qb) = r$$

If we take a new $z \in \text{Div}(b) \cap \text{Div}(r)$, then:

$$z|(q \cdot b + r) = a \Rightarrow z \in \text{Div}(a)$$

Because we have defined $\text{GCD}(a,b) = \max(\text{Div}(a) \cap \text{Div}(b))$, and we now know $\text{Div}(a) \cap \text{Div}(b) = \text{Div}(b) \cap \text{Div}(r)$, it must be that $\text{GCD}(a,b) = \text{GCD}(b,r)$. In algorithmic form, where $a > b$:

```
function gcd(a, b):
  if b = 0:
    return a
  else:
    gcd(b, a mod b)
```

□

Definition. If $a,b,m \in \mathbb{Z}$, with $m > 0$, we say $a \cong b \mod m$ or $a \cong_m b$ if $m|(a-b)$, or equivalently, a **mod** $m = b$ **mod** m.

Theorem. Mod is an equivalence relation.

Proof

Reflexivity:

$\forall a, a \cong_m a$, as $m|(a-a) = 0$.

Symmetry:

If $a \cong_m b$, then $m|(a-b) \Rightarrow m|(b-a) \Rightarrow b \cong_m a$. This is simply multiplication by -1.

Transitivity:

If $a \cong_m b$ and $b \cong_m c$, we have $m|(a-b)$ and $m|(b-c)$. Thus, $m|((a-b)+(b-c)) \Rightarrow m|(a-c)$, meaning $a \cong_m c$.

Definition. $\forall m > 0$ and $a \in \mathbb{Z}$, define:

$$[a]_m = \{b | a \cong b \mod m\}$$

Or, $[a]_m$ is the set of all integers with the same remainder a when divided by m. We define the following operations on these equivalence classes:

$$[a]_m + [b]_m = [a+b]_m$$
$$[a]_m \cdot [b]_m = [a \cdot b]_m$$

There are proofs of the associativity of addition and working with the integers mod m in the notes (3C). They don't seem particularly important, though, so I will omit them for now.

Definition. Two integers are relatively prime (or coprime) if their greatest common divisor is 1.

Definition. An integer a has a multiplicative inverse **mod** m if $\exists b : a \cdot b \cong_m 1$. Equivalently, if $m|(a \cdot b - 1)$.

Theorem. If integers $m > 0$ and a are relatively prime, then a has a multiplicative inverse **mod** m.

Proof

Recall Bezout's identity: $d = \text{GCD}(m, a) = n \cdot m + b \cdot a$. From this we know $\exists n, b$ such that $n \cdot m + b \cdot a = 1$, as m and a are relatively prime. Using our equivalence classes, we can show:

$$\begin{aligned} [1]_m &= [n \cdot m + b \cdot a]_m \\ &= [n]_m \cdot [m]_m + [b]_m \cdot [a]_m \\ [1]_m &= [b \cdot a]_m \end{aligned}$$

because $[m]_m = [0]_m$. Thus, a has a multiplicative inverse **mod** m of b. □

. . . .

Corollary. If p is prime and a,b are integers such that $p|(a \cdot b)$, then $p|a$ or $p|b$.

Proof

Without loss of generality, say that $p \nmid a$. Thus, $\text{GCD}(a,p) = 1$. This means that $\exists x,y : x \cdot a + y \cdot p = 1$ (Bezout's Identity), and multiplying by b yields:

$$xab + ypb = b$$

Notice that $p|(xab)$ because $p|(ab)$, and that $p|(ypb)$. This means that $p|b$, as it divides both of the terms that sum to b. □

Theorem. If a,m are relatively prime, then the multiplicative inverse of a **mod** m is unique.

Proof

Suppose, for the sake of contradiction, that b,c are different multiplicative inverses of a **mod** m. Some arithmetic gets us:

$$\begin{aligned} b &\cong_m b \cdot 1 \\ &\cong_m b \cdot (c \cdot a) \\ &\cong_m c \cdot (b \cdot a) \end{aligned}$$

Because b is a multiplicative inverse of a, we now have:

$$\begin{aligned} b &\cong_m c \cdot 1 \\ b &\cong_m c \end{aligned}$$

□

Notation. We use a^{-1} to denote the multiplicative inverse of a **mod** m, when it exists.

Theorem: Chinese Remainder Theorem. Suppose $m_1, m_2, \ldots, m_k$ are pairwise relatively prime, and that $c_1, c_2, \ldots, c_k$ are integers. Then there exists a solution x to the system:

$$\{x \cong_{m_i} c_i\}$$

Proof: Existence, Constructive

Let $n = m_1 \cdot m_2 \cdot \ldots \cdot m_k$. For each $i \in \{1, 2, \ldots, k\}$, we define $n_i = \frac{n}{m_i}$. Notice that n_i is relatively prime to m_i, as each m_i is relatively prime to all other m_i, which are contained in n_i. We now let $\{a_i\}$ be the set of all multiplicative inverses with respect to $\{n_i\}$:

$$a_i \cdot n_i \cong_{m_i} 1$$

We know these exist because relative primes have multiplicative inverses **mod** m, as proven 2 theorems ago.

We now define $x_i = a_i \cdot n_i$. Note that we now have two cases: x_i **mod** $m_i = 1$, which we call the match case, and x_i **mod** $m_j = 0$, which we call the mismatch case. These cases exist as when the i's match, a_i and n_i are multiplicative inverses mod m_i. When they do not match, $m_j | n_i$.

Finally, we claim the solution to the system is:

$$x = c_1 x_1 + c_2 x_2 + \cdots + c_k x_k$$

To see this, we look at the ith equation:

$$x \cong_{m_i} c_i$$

And we take the ith term of x:

$$x = c_i x_i = c_i \cdot (a_i \cdot n_i)$$

From here we see that in the match case:

$$c_i \cdot a_i n_i \cong_{m_i} c_i$$

Because, tautologically, $a_i n_i$ **mod** $n_i = 1$. In the mismatch case, however, we get:

$$c_j x_j \cong_{m_i} 0$$

Thus, for any i in the system, our x only has 1 match—and thus produces one value—which is equal to c_i **mod** m_i which definitionally solves the $x \cong_{m_i} c_i$. □

Induction. The goal of induction: to prove a universal statement (note that here, $1 = \text{True}$):

$$\forall n > 0, \phi(n) = 1$$

There are two main steps:

1. Base case: $\phi(1) = 1$
2. Inductive step:

$$\forall n > 0, \phi(n) = 1 \Rightarrow \phi(n+1) = 1$$

Theorem. The sum of the first n numbers can be writen as:

$$\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$$

Proof

Our base case is that $\phi(1) = 1$. We know this is true as $1 = \frac{1(2)}{2}$. We now take the inductive step. Suppose $\phi(n)$ is true, meaning:

$$\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$$

We now take $\phi(n+1)$:

$$\begin{aligned} \sum_{i=1}^{n+1} i &= \left(\sum_{i=1}^{n} i\right) + n + 1 \\ &= \frac{n(n+1)}{2} + n + 1 \\ &= (n+1)(\frac{n}{2} + 1) \\ &= \frac{(n+1)(n+2)}{2} \end{aligned}$$

□

Defintion: Strong Induction.

$$(\forall n > 0, \forall k : 0 \leq k < n, (\phi(k) = 1) \Rightarrow (\phi(n) = 1)) \Rightarrow (\forall n > 0, \phi(n) = 1)$$

I.e., if all k up to n are true, then n is true, and so all n are true.

Proof

By contraspositive. Suppose RHS is false, i.e., $\exists n$ such that $\phi(n) = 0$. This implies the set of counter-examples $(\{n | \phi(n) = 1\})$ is nonempty. By the well-ordering principle, there exists a minimum counterexample. Everything smaller than this counterexample must be true, and we call this counterexample n_0. So $\forall k < n_0, \phi(k) = 1$. But this contradicts the LHS, because this implies that $\phi(n_0) \neq 1$. □

Theorem: Existence of a Prime Divisor. Every integer > 1 has a prime divisor.

Proof

Our base case is that $2|2$. Our inductive assumption is that $\forall 2 \leq j < k$, there exists a prime p such that $p|(j)$. We now prove this holds for k. If k is prime, we are done. Otherwise, $k = a \cdot b$. We know that a and b fall into the set of all j, thus they must have a prime divisor. □

Corollary. There are infinitely many primes.

Proof

Consider any finite list of primes $p_1, p_2, \ldots, p_k$. We will prove that some prime is missing. Let $p = p_1 \cdot p_2 \cdot \ldots \cdot p_k$, and let $n = p + 1$. We know that n must have a prime divisor q, which must be in our list of every prime. As a result we know $q|p$. But q also divides n, which means that q must divide 1.

$$q|p \wedge q|n \Rightarrow q|(n-p) \Rightarrow q|1$$

This is impossible, and it means some prime must be missing from out list. □

Lemma: Helpful Lemma. If p is prime and so are $p_1, p_2, \ldots p_k$ (not necessarily distinct), if $p|(p_k \cdot p_{k-1} \cdot \ldots \cdot p_1)$, then $p = p_i$ for some i.

Proof

Base case: if we have 1 prime p_1, and $p|p_1$, then $p = p_1$. Our inductive hypothesis will assume Helpful Lemma true for all k, and prove it for $k+1$.

We know $p|p_{k+1} \cdot p_l \cdot \ldots \cdot p_2 \cdot p_1$, and that for a prime p, if $p|(a \cdot b)$ then $p|a$ or $p|b$. Suppose that $a = p_1 \cdot p_2 \cdot \ldots \cdot p_k$, and $b = p_{k+1}$. Because p divides the product of primes through p_{k+1}, it must be that either $p|a$ or $p|b$. If $p|a$, then $p = p_i$ for some $i \in \{1, 2, \ldots, k\}$. If $p|b$, then $p = p_{k+1}$.

Lemma: Fundamental Theorem of Arithmetic. Every integer $n > 1$ can be written uniquely as a product of primes.

Proof: Uniqueness of the FTA

Suppose that prime factorizations aren't unique. Let n be the least natural number without a unique prime factorization:

$$n = p_1 \cdot p_2 \cdot \ldots \cdot p_k = p'_1 \cdot p'_2 \cdot \ldots \cdot p'_k$$

From this we have that $p_1 | n$. By the Helpful Lemma, we must have that $p_i = p'_j$ for some i and j. But if we divide n by p_1, we get a smaller number written in two different prime factorizations. We claimed, however, that we already had the smallest, and thus every integer > 1 must have a unique prime factorization. □

Definition. Define the factorial function ($n!$) as the product of all positive integers $\leq n$.

Theorem: Wilson's Theorem. For any prime p:

$$(p-1)! \cong_p -1$$

Or equivalently,

$$p | ((p-1)! - 1)$$

Proof

Consider $x \in \{1, 2, \ldots, p-1\}$. Note that all of these numbers are relatively prime to p, which means that each x has a unique multiplicative inverse **mod** p.

We claim that only 1 and $p-1$ are self-inverse **mod** p. Being self inverse means that $x^2 \cong_p 1$, which implies that $x^2 - 1 \cong_p 0$. From here we see that $p | (x-1)(x+1)$. As a result, we have that $x \cong_p \pm 1$, from which we get $1 \cong_p 1$ and $(p-1) \cong_p -1$. These are the only two solutions to the equation, as the degree of x^2 is two.

Now we consider the "middle terms" $\{2, 3, \ldots, p-2\}$. These terms can be "married" into pairs with their multiplicative inverse. This is because multiplicative inverses must be less than p, and that's the set we have.

We're left with just 1 and $p-1$ after $\{2, 3, \ldots, p-2\}$ all go to 1. We see that $1 \cdot (p-1) = p-1$, and $(p-1) \cong_p -1$. □

Lemma: Helpful Lemma 2. Let p be prime and a be such that $\text{GCD}(a,p) = 1$. Then, as sets:

$$\begin{aligned}
&\{a \cdot 1 \mod p,\\
&\ a \cdot 2 \mod p,\\
&\ \dots,\\
&\ a \cdot (p-1) \mod p\} =\\
&\ \{1, 2, \dots, p-1\}
\end{aligned}$$

More compactly we can say that

$$\{a \cdot 1, a \cdot 2, \dots, a \cdot (p-1)\} = \{1, 2, \dots, p-1\}$$

where multiplication is done **mod** p.

Proof

First, each element in LHS is of the form $a \cdot j \not\cong_p 0$, because a and j are relatively prime to p. As a result, each element is in $\{1, 2, \dots, p-1\}$.

Now we prove the elements of the LHS are distinct. Becuase $\text{GCD}(a,p) = 1$, there exists a multiplicative inverse a^{-1} for a. For the sake of contradiction, assume that $(a \cdot j) \cong_p (a \cdot j')$, where $j \neq j'$ (i.e., where j equals the a's in the LHS). We can multiply both sides by a^{-1} to get $j \cong_p j'$, which is a contradiction. Thus, each $(a \cdot j)$ **mod** p in the LHS is distinct and the sets are equal. □

Theorem: Fermat's Little Theorem. For all primes p and integers a such that $\text{GCD}(a,p) = 1$:

$$a^{p-1} \cong_p 1$$

Proof

Consider $n \cong_p (a \cdot 1)(a \cdot 2) \dots (a \cdot (p-1))$. By Helpful Lemma 2, this is a reordering of $(p-1)!$, as we know that the sets $\{a \cdot 1, a \cdot 2, \dots, a \cdot (p-1)\}$ and $\{1, 2, \dots, p-1\}$ are equivalent. By Wilson's Theorem, we know that $(p-1)! \cong_p -1$, which means that $n \cong_p -1$.

Alternatively, we can factor our the $p-1$ occurences of a:

$$\begin{aligned}
n &\cong_p a^{p-1} \cdot (p-1)!\\
n &\cong_p a^{p-1} \cdot -1
\end{aligned}$$

Finally, equating both expressions for n yields:

$$\begin{aligned} -1 &\cong_p a^{p-1} \cdot -1 \\ \Rightarrow a^{p-1} &\cong_p 1 \end{aligned}$$

□

---

5. Combinatorics

5.1. Permutations

Definition. A permutation of a set is an ordered arrangement of its elements.

For example, there are 6 permutations of the set {1, 2, 3}. In general, there are $n!$ ways to arrange n elements. We can also use k-permutations, which are ordered subsets of size k.

Definition. Let $P(n,r)$ be the number of r-permutations of an n-element set.

Theorem.

$$P(n,r) = n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot (n-(r-1))$$

Proof

We have n ways to choose the first item, $n-1$ ways to choose the second, and so on, until we have selected r items, at which point we will just have multiplied $n-(r-1)$.

. . . .

Corollary. If n is positive and $r \in [0,n]$, then

$$P(n,r) = \frac{n!}{(n-r)!}$$

Proof

The number of ways to order an n element set is $n!$. An alternative way of calculating this is first choosing any r elements (of which there are $P(n,r)$ permutations of), and then choosing the remaining $n-r$ elements (of which there are $(n-r)!$ permutations of). Thus,

$$\begin{aligned} n! &= P(n,r)\cdot(n-r)! \\ \Rightarrow P(n,r) &= \frac{n!}{(n-r)!} \end{aligned}$$

Example. How many permutations of "ABCDEFGH" contain "ABC" consecutively?

To answer this problem, we view the letters as a set of: { ABC, D, E, F, G, H }, treating "ABC" as one letter. Thus, any permutation of this set works, giving us 6! permutations.

5.2. Combinations

Definition. Let $C(n,r)$ be the number of subsets of size r of an n-element set.

Theorem. If $n > 0$ and $r \in [0,n]$, then:

$$C(n,r) = \frac{n!}{r!(n-r)!}$$

Insight

Notice that for each combination of size r, there are $r!$ permutations of it, meaning we can derive our formula for combinations by dividing our permutation formula by $r!$.

Proof

$$\begin{aligned} P(n,r) &= C(n,r)\cdot P(r,r) \\ \Rightarrow C(n,r) &= \frac{P(n,r)}{P(r,r)} = \frac{n!}{r!(n-r)!} \end{aligned}$$

Example. A poker hand consists of 5 cards. How many distinct poker hands exist?

$C(52,5)$.

Example. A full house is a poker hand with a three of a kind and a two of a kind. How many distinct full house hands are there?

First, we calculate the ways to permute each "kind", which is equal to $P(13,2)$. We notice that for the three of a kind, we can pick any three cards from the four, and the same for the two of a kind, giving us:

$$P(13,2) \cdot C(4,3) \cdot C(4,2)$$

....

Example. A flush is a poker hand where all cards are of the same suit. How many flushes exist?

Notice this amounts to chosing our suit, followed by any 5 of its 13 cards:

$$C(4,1) \cdot C(13,5)$$

5.3. Combinatorial Proofs

Definition. A combinatorial proof is a proof based on a counting argument, for example proving equality between two numbers by providing two ways of counting the same objects.

Theorem. For $n > 0$, $r \in [0,n]$:

$$C(n,r) = C(n,n-r)$$

Proof

In order to count the number of size-r subsets of n, we can either enumerate them directly (as with $C(n,r)$), or we can enumerate their complements, i.e., "throw out" $n-r$ elements in $C(n,n-r)$ ways, leaving us simply with r elements.

5.4. The Binomial Theorem

Theorem. For $n \in \mathbb{N}$, $n > 0$:

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

Where $\binom{n}{i} = C(n,i)$.

. . . .

Example. What is the coefficient of $x^{11}y^5$ in $(x+y)^{16}$?

Using the binomial theorem, the coefficient is $C(16,11)$, or equivalently, $C(16,5)$.

. . . .

Example. What is the coefficient of x^7y^4 in $(x+2y)^{11}$?

$2^4 \cdot \binom{11}{7}$

5.5. Binomial Identities

Theorem: Binomial Identities. For $n > 0$:

$$\sum_{i=0}^{n} \binom{n}{i} = (1+1)^n = 2^n$$

$$\sum_{i=0}^{n} (-1)^i \binom{n}{i} = (1-1)^n = 0$$

. . . .

Theorem: Pascal's Identity.

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

Insight

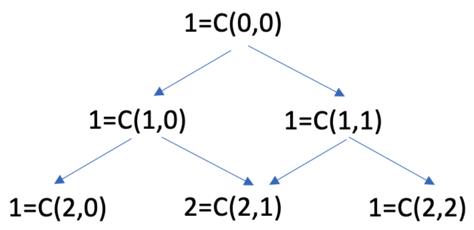
We have $n+1$ rocks, and we want to count subsets of size k. One way is to fix a rock. We can either include or exclude this rock, but either way, we always choose from n rocks.

Proof

Fix $s \in S$ and count subsets K of size k. If $s \in K$, we know that $k-1$ elements of K come from the set $S-\{s\}$. Notice that $|S-\{s\}=n$, and we choose $k-1$, which gets us $\binom{n}{k-1}$. However, if $s \notin K$, all k elements will come from the set $S-\{s\}$, which gives us $\binom{n}{k}$. Because these possibilities are disjoint, we simply add them to get our final result.

Insight

We can visualize Pascal's Identity with Pascal's Triangle:



Theorem: Vandermonde's Identity. For $0 < r \leq n$:

$$\binom{m+n}{r} = \sum_{k=1}^{r} \binom{m}{r-k}\binom{n}{k}$$

Proof

We can "break down" this formula as first taking all r of the subset from m and none from n. Then it takes $r-1$ from m, and 1 from n, so on and so forth. We multiply "inside" each of these pairs because the terms are dependent, but we add because the case of taking all r from m is independent of taking $r-1$ from m.

Corollary.

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

Theorem. Let $0 < r \leq n$.

$$\binom{n+1}{r+1} = \sum_{j=r}^{n} \binom{j}{r}$$

Proof

Let's say we are picking $r+1$ elements of $S = \{0, 1, \dots, m\}$, of size $n+1$. The smallest max element of a subset of size $r+1$ is r, and there is only one subset of S that has r as its max. In general, the number of subsets of S of size $r+1$ that have a largest element $j \geq r$ is $C(j, r)$, as we fix the largest element j in the subset, and have a remaining r to choose to get our total size $r+1$.

5.6. Generalized Permutations and Combinations

Example. How many ways can we select five bills from cash drawers that have an infinite amount of 5, 10, 20, 50, and 100 dollar bills?

We can think of this problem in terms of "moves" that we must make, which work towards two goals: (a) taking 5 bills, and (b) making sure to visit each bill's "drawer". We must perform (a) 5 times—as we must take 5 bills—and we must perform (b) 5 times, as we must "move to the next drawer" 5 times (1 to 5, 5 to 10, and so on). In total we thus have 10 operations, and we want to select every possible subset in which we could take our 5 bills:

$$\binom{10}{5}$$

Theorem. For a set with n distinct elements and replacement, there are

$$\binom{n+r-1}{r}$$

r-combinations.

Example. How many natural number solutions are there to:

$$x + y + z = 11$$

Once again, we have to “move variables” 2 times, and increment our counter 11 times, giving us

$$\binom{13}{11}$$

total solutions.

Example. How many distinct permutations of the word “success” exist?

Proof

The word “success” has 7 letters, so if each were distinct, we would have 7! total permutations. However, there are 2 duplicate letters, meaning we must divide by the number of ways to arrange those two:

$$= \frac{7!}{3! \cdot 2!}$$

This concept has its own notation, defined below.

Definition. The number of distinguishable permutation is given by multinomial coefficients (i.e., $(x_1 + x_2 + \ldots + x_k)^n$):

$$\frac{n!}{n_1! \cdot n_2! \cdot \ldots \cdot n_k!} = \binom{n}{n_1, n_2, \ldots, n_k}$$

5.7. The Pigeonhole Principle

Theorem: The Pigeonhole Principle. Let $A_1, A_2, \ldots, A_m$ be disjoint sets so that $|\bigcup_i A_i| > m$. Then there exists an A_i so that $|A_i| \geq 2$.

Proof: By Contrapositive

If all $|A_i| \leq 1$, then

$$\left|\bigcup_i A_i\right| = \sum_i |A_i| \leq \sum_i 1 = m$$

Insight

The pigeonhole principle really just says that with n pigeons and at most $n-1$ holes, there must be at least 1 hole with 2 pigeons.

....

Corollary. Suppose A, B are finite sets where $|A| > |B|$ and $f : A \to B$. Then f cannot be one-to-one.

This is simply because there are too few "holes" in B to hold all the "pigeons" in A, meaning there must be some element in B with at least 2 corresponding elements in A (thus making the function not one-to-one). □

Example. After a meeting, n people shake others hands. Prove that at least 2 people shook the same number of hands.

Let's enumerate everyone in the set $\{0,1,2,\ldots,n-1\}$. Say person i shakes i hands. This actually can't be! Someone can either shake $n-1$ hands (everybody but themself), or 0 hands; it's impossible for both to happen as someone shaking 0 hands implies that the maximum becomes $n-2$ hands! So this means that we can either have the handshakes be in the integers $\{1,2,\ldots,n-1\}$ or in $\{0,1,\ldots,n-2\}$. Either way, there are $n-1$ choices of the number of hands to shake, and n total people. By the pigeonhole principle, at least two people must shake the same number of hands.

Theorem: Generalized Pigeonhole Principle. Let $A_1, A_2, \ldots, A_m$ be disjoint sets such that $|\bigcup_i A_i| = n$. Then $\exists A_i$ such that $|A_i| \geq \lceil \frac{n}{k} \rceil$.

Proof: By Contrapositive.

Suppose for all i that $|A_i| < \lceil \frac{n}{k} \rceil$. This implies that $|A_i| < \frac{n}{k}$; because $|A_i|$ is always an integer, if it is smaller than $\lceil \frac{n}{k} \rceil$, it is also smaller than $\frac{n}{k}$. For example, $(|A_i| = 3) < (\frac{n}{k} = 3.1) < (\lceil \frac{n}{k} \rceil = 4)$. This implies that

$$\left|\bigcup_i A_i\right| = \sum_i A_i < \sum_i \frac{n}{k} = k(\frac{n}{k}) = n$$

$$\left|\bigcup_i A_i\right| \neq n$$

□

. . . .

Example. A centipede needs 100 same-color socks from a random assortment of red, green, and blue socks. How many must he draw to guarantee he receives 100 same-color socks?

Our answer will be the least n such that $\lceil \frac{n}{3} \rceil = 100$. This implies that $\frac{n}{3} > 99$, so $n > 297$. Thus, our centipede needs to draw 298 socks by the generalized pigeonhole principle.

As a check, this implies that at least one set in $|\bigcup_i A_i| = 298$ must be:

$$A_i \geq \text{ceil}\left(\frac{298}{3}\right) > 99 \Rightarrow A_i = 100$$

6. Probability Theory

Definition. A probability space is a finite set $\Omega \neq \varnothing$ and a function $\Pr : \Omega \to \mathbb{R}$. $\forall \omega \in \Omega$, we have that

$$\Pr[\omega] \geq 0$$

and

$$\sum_{\omega \in \Omega} \Pr[\omega] = 1$$

We call Ω the sample space and Pr the probability distribution.

Definition. An event is a subset of Ω: $A \subseteq \Omega$.
The atomic events are singleton sets such that

$$\Pr[\{\omega\}] = \Pr[\omega]$$

We denote the probability of an event A as

$$\Pr[A] = \sum_{\omega \in A} \Pr[\omega]$$

Definition. The uniform distribution over Ω sets $\Pr[\omega] = \frac{1}{|\Omega|}$ for all ω.

Definition. Events A, B are disjoint if $A \cap B = \varnothing$.

Notice that if $A_1, A_2, ..., A_k$ are disjoint subsets of Ω, then

$$\Pr[A_1 \cup A_2 \cup ... \cup A_k] = \Pr[A_1] + \Pr[A_2] + ... + \Pr[A_k]$$

. . . .

Theorem. For any A and B,

$$\Pr[A] + \Pr[B] = \Pr[A \cup B] - \Pr[A \cap B]$$

which follows from simple inclusion-exclusion.

Lemma.

$$\Pr\left[\bigcup_{i=1}^{k} A_i\right] \leq \sum_{i=1}^{k} \Pr[A_i]$$

Insight

If the sets of all disjoint, these two values are equal, otherwise the right side overcounts the atomic events and thus probability.

6.1. Conditional Probability

Definition. If A, B are events, the conditional probability of A relative to B is

$$\Pr[A|B] = \frac{\Pr[A \cap B]}{\Pr[B]}$$

Sorry that was a lot of definitions. But now we're setup.

. . . .

Example. When rolling 3 dice, what is the probability that the first die is 5, given that the sum is 9?

We let the event A be the first die being 5, and B be the sum of the three dice being 9.

We look to compute

$$\Pr[A|B] = \frac{\Pr[A\cap B]}{\Pr[B]}$$

For $A\cap B$, there are three dice rolls that sum to 9 when the first value is 5: $\{5,2,2\},\{5,1,3\},\{5,3,1\}$. For B, there are 25 rolls that sum to 9 out of 6^3 possible. Thus, our answer is:

$$\Pr[A|B] = \frac{\frac{3}{6^3}}{\frac{25}{6^3}} = \frac{3}{25}$$

6.2. Independence

Definition. Events A and B are independent if

$$\Pr[A\cap B] = \Pr[A]\cdot\Pr[B]$$

Definition. Events A and B are positively correlated if

$$\Pr[A\cap B] > \Pr[A]\cdot\Pr[B]$$

and they are negatively correlated if

$$\Pr[A\cap B] < \Pr[A]\cdot\Pr[B]$$

Example. Consider a dice roll. Let A be the event that the roll is even, and B the event that the roll is prime. Are these events correlated?

Our even rolls are: $\{2,4,6\}$, and our prime rolls are: $\{2,3,5\}$. Thus,

$$\Pr[A\cap B] = \Pr[\{2\}] = \frac{1}{6} < \Pr[A]\cdot\Pr[B] = \frac{1}{2}\cdot\frac{1}{2} = \frac{1}{4}$$

So these events are negatively correlated.

Definition. Events A_i are pairwise independent if $\forall i,j$, A_i and A_j are independent.

Definition. Events $A_1, A_2, \ldots, A_k$ are mutually independent if for all subsets of the events, i.e., $\forall I \subseteq \{1, 2, \ldots, k\}$,

$$\Pr\left[\bigcap_{i \in I} A_i\right] = \prod_{i \in I} \Pr[A_i]$$

Note that mutual independence is stronger than pairwise independence.

6.3. Random Variables

Definition. A random variable is a function $f : \Omega \to \mathbb{R}$ where Ω is the sample space.

Definition. If X is a random variable, we define

$$\Pr[X = r] = \Pr[\{\omega \in \Omega | X(\omega) = r\}]$$

i.e., the probabilty that $X = r$ is the probability of all atomic events ω that X maps to r.

Example. Consider the probability space corresponding with flipping a fair coin 3 times, where the random variable X counts the number of heads. What is $\Pr[X = 3]$?

The event corresponding with three heads happens once out of eight, so $\Pr[X = 3] = \frac{1}{8}$.

6.4. Bernoulli Trials

Definition. A Bernoulli Trial is a random variable B whose codomain is $\{0, 1\}$. We define the event $\{\omega \in \Omega | B(\omega) = 1\}$ as a success, and $B(\omega) = 0$ as a fail.

“The point of Bernoulli trials is to repeat them.” When repeated, we count the number of successes k in n trials, which itself is *also* a random variable!

Example. Consider flipping a biased coin where $\Pr[H] = p$, and $\Pr[T] = 1 - p$. From this, define the random variable X where $X(\{H\}) = 1$ and $X(\{T\}) = 0$. After flipping n coins, define

$$Y = \sum_i X_i$$

For $k \leq n$, what is $\Pr[Y = k]$?

Notice that we have $\binom{n}{k}$ ways to get k heads out of n. We multiply this value by the probability of getting head k times and tails $n - k$ times:

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

This is the probability distribution for a Bernoulli trial where we have k successes out of n trials!

6.5. Expectation

Definition. Let X be a random variable in Ω. We define the expected value of X as

$$\mathrm{E}[X] = \sum_{\omega \in \Omega} X(\omega) \cdot \Pr[\omega]$$

Theorem: Linearity of Expectation. Let $(\Omega, \Pr)$ be a probability space and $X_1, X_2, \ldots, X_n$ be random variables over Ω, with $X = \sum_i X_i$. Then:

$$\mathrm{E}[X] = \mathrm{E}[X_1 + X_2 + \ldots + X_n] = \sum_i \mathrm{E}[X_i]$$

In other words, the expectation of the sum is the sum of the expectations.

Proof

$$
\begin{aligned}
\mathrm{E}[X] &= \sum_{\omega\in\Omega} (X_1(\omega)+X_2(\omega)+\ldots+X_n(\omega))\cdot\Pr[\omega] \\
&= \sum_{\omega\in\Omega}\sum_{i} X_i(\omega)\cdot\Pr[\omega] \\
&= \sum_{i}\sum_{\omega\in\Omega} X_i(\omega)\cdot\Pr[\omega] \\
&= \sum_{i} \mathrm{E}[X_i]
\end{aligned}
$$

. . . .

Example. During a dinner party, n men check in their hats. The hat man turns out to be extremely lazy and gives the men their hats back randomly at the end. What is the expected number of men who get their own hat back?

Let us define a random variable R as the number of men who get their hat back. We know that

$$\mathrm{E}[R] = \sum_{k=0}^{n} 0\cdot\Pr[R=k]$$

This is difficult to calculate, as the order in which we distribute the hats matters (e.g., giving the first man his hat back changes the rest of the calculation).

Instead, we can introduce an “indicator variable” X_i, which equals 1 if the ith man gets his hat back, otherwise it equals 0. We know that

$$\mathrm{E}[X_i] = \frac{1}{n}$$

because the hat man gives the hats back uniformly and randomly. From this, we get:

$$\mathrm{E}[R] = \sum_{i=1}^{n}\mathrm{E}[X_i] = \sum_{i=1}^{n}\frac{1}{n} = 1$$

Theorem: Multiplicativity of Expectation If X and Y are independent random variables, $\mathrm{E}[XY] = \mathrm{E}[X]\cdot\mathrm{E}[Y]$.

Proof

We know that

$$\mathrm{E}[XY] = \sum_{x,y\in\Omega} X(x)Y(y)\Pr[X=x\cap Y=y]$$

And if X and Y are independent random variables over Ω, then

$$\Pr[X = x \cap Y = y] = \Pr[X = x] \cdot \Pr[Y = y]$$

Plugging this into our expected value equation:

$$\begin{aligned}
\mathrm{E}[XY] &= \sum_{x,y\in\Omega} X(x)Y(y)\Pr[X = x]\Pr[Y = y] \\
\mathrm{E}[XY] &= \sum_{x,y\in\Omega} X(x)\Pr[X = x] \cdot Y(y)\Pr[Y = y] \\
\mathrm{E}[XY] &= \left(\sum_{x\in\Omega} X(x)\Pr[X = x]\right)\left(\sum_{y\in\Omega} Y(y)\Pr[Y = y]\right) \\
\mathrm{E}[XY] &= \mathrm{E}[X] \cdot \mathrm{E}[Y]
\end{aligned}$$

6.6. Markov's Inequality

Theorem: Markov's Inequality. If X is a non-negative random variable, then $\forall a > 0$:

$$\Pr[X \geq a] \leq \frac{\mathrm{E}[x]}{a}$$

Alternatively, if we set $a = k \cdot \mathrm{E}[X]$, then:

$$\Pr\left[X \geq k \cdot \mathrm{E}[X]\right] \leq \frac{1}{k}$$

Insight

The second way of interpreting Markov's inequality is hopefully more intuitive; the probability of getting a value k times larger than the expected value is simply $\frac{1}{k}$.

Proof

We know that

$$\mathrm{E}[X] \geq \sum_{\omega\in\Omega, X(\omega)\geq a} X(\omega)\Pr[\omega]$$

This is because by setting a lower bound for X, we can only make the expected value larger (as X is non-negative). From here, we know that

$$\mathrm{E}[X] \geq \sum_{\omega\in\Omega, X(\omega)\geq a} a \cdot \Pr[\omega]$$

because again, we make the summation smaller by multiplying by only the minimum possible output of X. This implies that

$$\begin{aligned}
\mathrm{E}[X] &\geq a \sum_{\omega\in\Omega, X(\omega)\geq a} \Pr[\omega] \\
\mathrm{E}[X] &\geq a \Pr[X \geq a] \\
\frac{\mathrm{E}[X]}{a} &\geq \Pr[X \geq a]
\end{aligned}$$

6.7. Variance

Definition. Let X be a random variable in the probability space $(\Omega, \Pr)$. We define

$$\begin{aligned}
\mathrm{Var}(X) &= \mathrm{E}[(X - \mathrm{E}[X])^2] \\
&= \sum_{\omega\in\Omega} (X(\omega) - \mathrm{E}[X])^2 \cdot \Pr[\omega]
\end{aligned}$$

Definition. We define standard deviation, $\sigma(X)$ as $\sqrt{\mathrm{Var}(X)}$.

Theorem.

$$\mathrm{Var}(X) = \mathrm{E}[X^2] - \mathrm{E}[x]^2$$

Proof

$$\begin{aligned}
\mathrm{Var}(X) &= \sum_{\omega\in\Omega} (X(\omega) - \mathrm{E}[X])^2 \cdot \Pr[\omega] \\
&= \sum_{\omega\in\Omega} \left((X(\omega)^2 - 2X(\omega)\mathrm{E}[X] + \mathrm{E}[X]^2) \cdot \Pr[\omega]\right) \\
&= \sum_{\omega\in\Omega} \left(X(\omega)^2 \Pr[\omega]\right) - 2\mathrm{E}[X] \cdot \sum_{\omega\in\Omega} X(\omega)\Pr[\omega] + \mathrm{E}[X]^2 \cdot \sum_{\omega\in\Omega} \Pr[\omega] \\
&= \mathrm{E}[X^2] - 2\mathrm{E}[X]^2 + \mathrm{E}[X]^2 \\
&= \mathrm{E}[X^2] - \mathrm{E}[X]^2
\end{aligned}$$

Theorem: Independent Variables have Linear Variances. If X and Y are independent,

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$$

Proof

$$\begin{aligned}\mathrm{Var}(X+Y) &= \mathrm{E}[(X+Y)^2]-\mathrm{E}[X+Y]^2\\ &= \mathrm{E}[X^2]+2\mathrm{E}[XY]+\mathrm{E}[Y^2]-\mathrm{E}[X]^2-2\mathrm{E}[X]\mathrm{E}[Y]-\mathrm{E}[Y]^2\\ &= \mathrm{E}[X^2]-\mathrm{E}[X]^2+\mathrm{E}[Y^2]-\mathrm{E}[Y]^2\\ &= \mathrm{Var}(X)+\mathrm{Var}(Y)\end{aligned}$$

Example. Let the random variable X be the sum of two dice rolls. What is $\mathrm{Var}(X)$?

Let X_i be the sum of one dice roll. We know that $\mathrm{E}[X]=2\mathrm{E}[X_i]=7$. From, this, we can compute:

$$\begin{aligned}\mathrm{Var}(X_i) &= \mathrm{E}[X_i^2]-\mathrm{E}[X_i]^2\\ &= \sum_i \frac{i^2}{6}-\left(\frac{7}{2}\right)^2\\ &= \frac{35}{12}\end{aligned}$$

Which implies that

$$\mathrm{Var}(X)=2\cdot\frac{35}{12}=\frac{35}{6}$$

6.8. Chebyshev's Inequality

Definition: Chevbyshev's Inequality. Let X be a random variable. For all $a>0$:

$$\Pr[|X-\mathrm{E}[X]|\geq a]\leq\frac{\mathrm{Var}(X)}{a^2}$$

Proof

This is simply Markov's inequality where $Y=(X-\mathrm{E}[X])^2$. We see that $\mathrm{E}[Y]=\mathrm{Var}(X)$, and apply Markov's inequality with $a=a^2$:

$$\Pr[Y\geq a^2]\leq\frac{\mathrm{E}[Y]}{a^2}$$

From here we can plug in $Y=(X-\mathrm{E}[X])^2$, and notice that if $(X-\mathrm{E}[X])^2\geq a^2$, then $|X-\mathrm{E}[X]|\geq a$, getting us:

$$\Pr[|X-\mathrm{E}[X]|\geq a]\leq\frac{\mathrm{Var}(X)}{a^2}$$

6.9. Law of Total Probability

Theorem: Law of Total Probability. Given disjoint events $H_1, H_2, ..., H_m$ that partition Ω and some other event A:

$$\Pr[A] = \sum_{i=1}^{m} \Pr[A|H_i] \cdot \Pr[H_i]$$

Proof

We can write A as a union of $\{A \cap H_i\}$ for each i, since all H_i are disjoint and partition the sample space. This implies that

$$\begin{aligned} \Pr[A] &= \sum_{i=1}^{m} \Pr[A \cap H_i] \\ &= \sum_{i=1}^{m} \Pr[A|H_i] \cdot \Pr[H_i] \end{aligned}$$

7. Asymptotics

First we must discuss how we should measure the complexity of an algorithm. To do so, we generally make two choices. The first is that we care more about worst-case behavior, although sometimes average-case behavior could be a better measure. The second is that we're generally interested in how many steps an algorithm takes as a function of its input length, i.e., as the length of the input grows, how does the running time scale.

Definition. Let $f, g : \mathbb{N} \to \mathbb{R}$. We say

$$f(x) = O(g(x))$$

if there exist c, k such that

$$\forall x \geq k, |f(x)| \leq c \cdot |g(x)|$$

Insight

This means that once x is large enough (past a certain k), value of f scale within a constant of g.

Example. $x^2 + 2x + 1 = O(x^2)$.

We know that $x^2+2x+1 \le x^2+3x^2+x^2$, but the RHS is simple $4x^2$, which is within a constant of x^2.

Definition. We say that

$$f(x) = \Omega(g(x))$$

if there exists constants c, k such that

$$\forall x \ge k, |f(x)| \ge c \cdot |g(x)|$$

Definition. We say that $f(x) = \Theta(g(x))$ if f is both $O(g)$ and $\Omega(g)$.

This is only true for functions that are constants of each other. For example, $f(n) = 2n$ is $O(n)$ and $\Omega(n)$, while $f(n) = \log n$ is $O(n)$ but not $\Omega(n)$.

---

8. Graphs

Definition. A graph $G = (V, E)$ is a set of vertices V and edges E. Each edge is (v_i, v_j) for $v_i, v_j \in V$.

Definition. Two vertices are *adjacent* if $(v_1, v_2) \in E$.

Graphs can be directed or undirected. In a directed graph, $(v_1, v_2) \ne (v_2, v_2)$.

. . . .

Definition. The degree of a vertex v is the number of edges with v as an endpoint. A digraph has an “indegree” and an “outdegree”.

Note that we count self loops (v_i, v_i) twice when counting degree.

Lemma: Handhsaking Lemma.

$$\sum_{v \in V} \text{degree}(v) = 2|E|$$

Insight

Every edge comes from one vertex and goes to another, so the total degree of the graph is simply two times the number of edges.

Definition. A *walk* of length k from vertex a to b is defined by

$$a = v_1, v_2, \ldots, v_k = b$$

A *closed* walk has the same start and end vertex.

Definition. A *path* is a walk where no vertex is used more than once, and a *cycle* is a closed path.

Definition. A complete graph K_n is an undirected graph where all pairs of vertices are connected by an edge (of which there are $\binom{n}{2}$ total).

8.1. Graph Isomorphism

Definition. An isomorphism between G_1, G_2 is a bijection $f : V_1 \to V_2$ so that vertices a, b are adjacent iff $f(a)$ and $f(b)$ are adjacent.

Graph isomorphism is what we consider a *hard* problem; there are no efficient (polynomial time) algorithms to solve it. We can solve it inefficiently by enumerating all possible bijections from G_1 to G_2, but there are $n!$ total, which is far from polynomial time. László Babai (UChicago prof!) has the best known solution, which solves the problem in $O(2^{(\log n)^c})$ time.

---

9. Complexity Theory

Complexity theory is the study of the hardness of problems. There are two major categories we're concerned with:

- P = problems solvable by polynomial-time algorithms
- NP = nondeterministic polynomial-time solvable; given a solution, we can check it in polynomial time.

We know that P is contained in NP, but whether NP is contained in P is the biggest open problem in theoretical computer science.

Some problems are "NP-Complete". Solving one of these problems would let us solve *any* NP-hard problem. We don't think factoring is NP-complete (which is good for our encryption...).

And those are the notes! I wish we got to more on graph theory an asymptotics but overall good class. There are probably lots of typos and I definitely didn't format this LaTeX doc very well, but I'm definitely going to improve on the latter going forward.

My Definition. This is my definition

My Theorem. This is my theorem

Example X. Here is an example.

Problem X. This is the problem. Math is involved!

$$1 + 1 = 2$$