7/11/14

HW 2 due NOW

Quiz 2 on Monday.

The material for Quiz 2 is done. It will be about the properties of $\vee, \wedge, \neg$, truth tables, and Venn diagrams. (In other words, Boolean algebra).

Now I want to discuss HW 2.

Q: Let S be a set with n elements. How many subsets does S have?

A: 2^n.

To begin, suppose that S and T are finite sets. We will name their elements

$$S := \{s_1, s_2, \ldots, s_m\}$$
$$T := \{t_1, t_2, \ldots, t_n\}.$$

We use the notations

$$|S| = \#S = m$$
$$|T| = \#T = n$$

for the numbers of elements in S and T

Theorem: Then we have

$$\#(S \times T) = \#S \times \#T.$$

Proof: I can arrange the elements of $S \times T$ in a rectangular array, with

- rows indexed by S
- columns indexed by T.

Picture:

$$\begin{array}{c|cccc}
 & t_1 & t_2 & & t_n \\
\hline
s_1 & (s_1, t_1) & (s_2, t_2) & \cdots & (s_1, t_n) \\
s_2 & (s_2, t_1) & (s_1, t_2) & \cdots & (s_2, t_n) \\
 & \vdots & \vdots & \ddots & \vdots \\
s_m & (s_m, t_1) & (s_m, t_2) & \cdots & (s_m, t_n)
\end{array}$$

How many entries does this rectangle have?

Answer: $m \times n$.

(Isn't this the <u>definition</u> of multiplication? Yes it is.)

We conclude that "$S \times T$" is a good notation for the cartesian product.

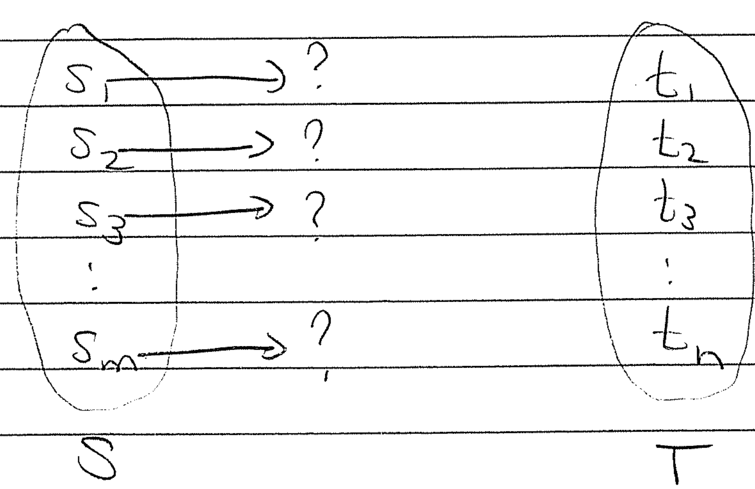
<u>Theorem</u>: If S and T are finite ~~sets~~, then the number of different functions from S to T is

$$\#T^{\#S}.$$

<u>Proof</u>: Recall that a <u>function</u> is a subset $F \subseteq S \times T$ (we think of an element $(s,t) \in F$ as an "arrow" $s \to t$) satisfying <u>one</u> <u>axiom</u>:

- Each element of S has exactly one arrow pointing from it.

Let $S = \{s_1, s_2, \ldots, s_m\}$ and $T = \{t_1, t_2, \ldots, t_n\}$. Then a function $S \to T$ looks like



S $\qquad\qquad\qquad$ T

There are exactly $m = \#S$ arrows, and each arrow has exactly $n = \#T$ choices for its target. We conclude that the total number of choices is

$$\underbrace{n \times n \times n \times \cdots \times n}_{m \text{ times}} = n^m = \#T^{\#S}$$

//

<u>Example</u>: The number of Boolean functions in 2 variables $\left(\{T,F\}^2 \to \{T,F\}\right)$ is

$$\#\{T,F\}^{\#\{T,F\}^2} = 2^{2 \times 2} = 2^4 = 16.$$

We have met quite a few of these 16,

e.g. $\vee, \wedge, \uparrow, \downarrow, \oplus$

The theorem suggests a <u>cute</u> notation. Given sets S and T we let

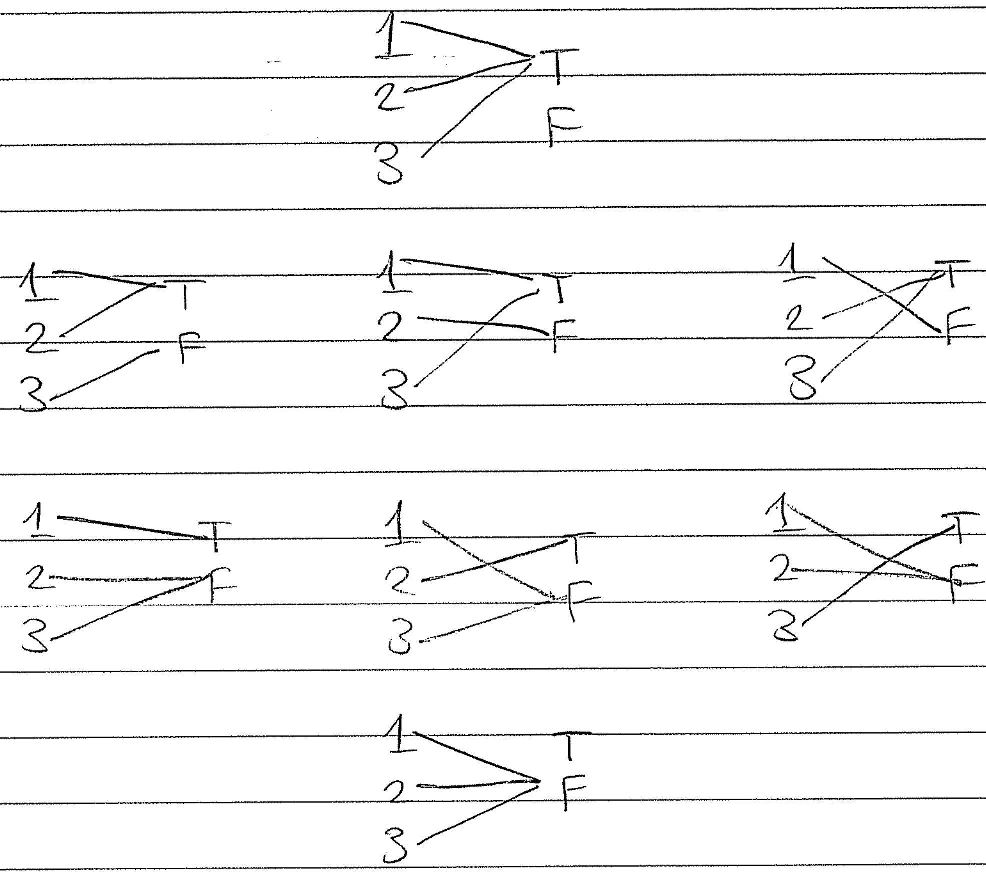
$T^S :=$ The <u>set</u> of functions from S to T.

Note that $T^S \subseteq \wp(S \times T)$. Indeed,

$\wp(S \times T) = \{$ All sets of arrows from S to T $\}$

$T^S = \{$ Just those sets that satisfy the function axiom. $\}$

Example: Here is a picture of $\{T, F\}^{\{1,2,3\}}$.



There are $\#\{T,F\}^{\#\{1,2,3\}} = 2^3 = 8$

as expected.

Now write down <u>all</u> of the subsets of $\{1,2,3\}$:

$$\{1,2,3\}$$

$$\{1,2\} \quad \{1,3\} \quad \{2,3\}$$

$$\{1\} \quad \{2\} \quad \{3\}$$

$$\emptyset$$

There are 8 of them. In fact, we see that there is a "1-to-1 correspondence"

$$\wp(\{1,2,3\}) \longleftrightarrow \{T,F\}^{\{1,2,3\}}$$

$$\text{subsets of } \{1,2,3\} \longleftrightarrow \text{functions from } \{1,2,3\} \text{ to } \{T,F\}$$

and we conclude that

$$\begin{aligned}\#\wp(\{1,2,3\}) &= \#\left(\{T,F\}^{\{1,2,3\}}\right)\\ &= \#\{T,F\}^{\#\{1,2,3\}}\\ &= 2^3 = 8.\end{aligned}$$

Theorem: Let S be a finite set. Then

$$\#\wp(S) = 2^{\#S}$$

Proof: There is a "1-1 correspondence" between $\wp(S)$ and $\{T,F\}^S$. Hence

$$\begin{aligned} \#\wp(S) &= \#(\{T,F\}^S) \\ &= \#\{T,F\}^{\#S} \\ &= 2^{\#S} \end{aligned}$$

//

This theorem leads to a very cute notation. Given any set S, we define

$$\begin{aligned} 2^S &:= \wp(S) \\ &= \text{The set of subsets of } S \end{aligned}$$

The motivation: If S is finite, then

$$\#(2^S) = 2^{\#S}$$

:)

Do you like this?

Let me give you one more way to think about the Boolean algebra of subsets of S (i.e. $\wp(S)$, or 2^S). We might as well take

$$S = \{1, 2, 3, \ldots, n\}.$$

Then each subset $A \subseteq S$ corresponds to a <u>binary</u> <u>string</u> of length n,

$$b_1 b_2 b_3 \cdots b_n$$

where $b_i = \begin{cases} 1 & \text{if } i \in A \\ 0 & \text{if } i \notin A. \end{cases}$

<u>Example</u>: The subset

$$\{2, 3, 5\} \subseteq \{1, 2, 3, 4, 5, 6\}$$

corresponds to binary string

$$011010$$

This is a very efficient language to use:

111

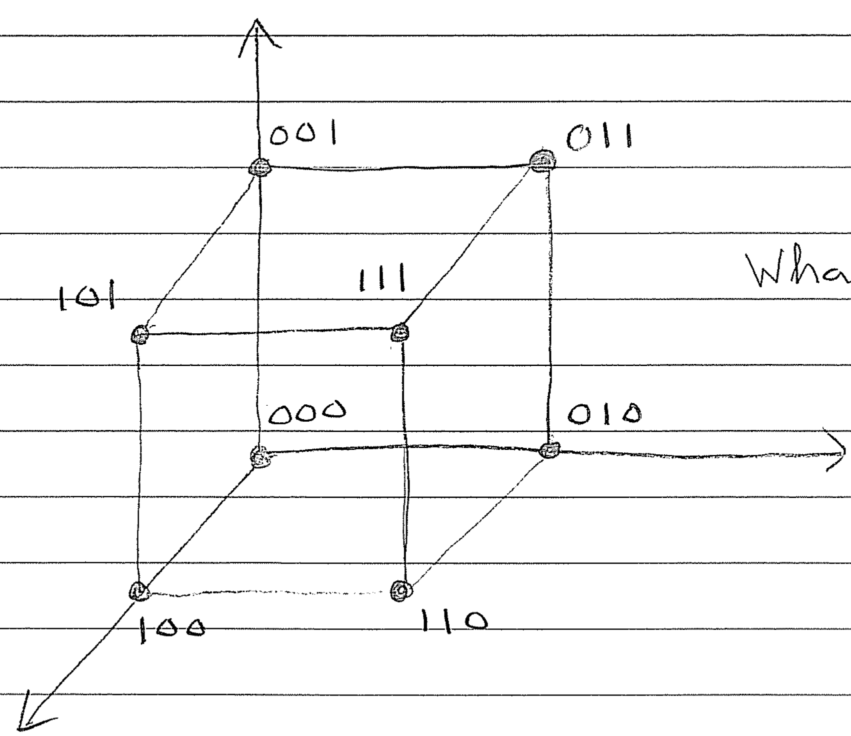
110 101 011

100 010 001

000

But there is more to it. We can think of a binary string as a point in Cartesian space.

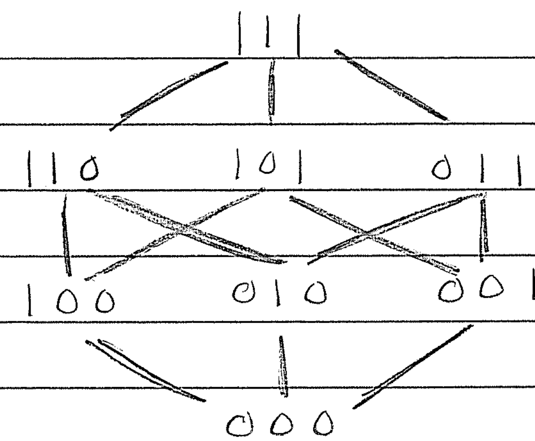
Example:



What is this?

Observation: The Boolean algebra $\wp(\{1,2,3\})$ is just a cube!

So I should draw edges in my diagrams



(Do you see the cube?)

Last issue for today.

Q: Is there a natural way to list / order the subsets of a set S?

A: Sure, I guess, as long as the set S is <u>finite</u>.

Let $\#S = n$ and think of its subsets as binary strings $b_1 b_2 b_3 \cdots b_n$.

(Clever Trick) Then we send the binary string to a number

$$b_1 b_2 b_3 \cdots b_n \longmapsto \sum_{i=1}^{n} b_i \cdot 2^{n-i}$$

Example:

$$\begin{array}{lcll}
000 & \longmapsto & 0\cdot4+0\cdot2+0\cdot1 & = 0 \\
001 & \longmapsto & 0\cdot4+0\cdot2+1\cdot1 & = 1 \\
010 & \longmapsto & 0\cdot4+1\cdot2+0\cdot1 & = 2 \\
011 & \longmapsto & 0\cdot4+1\cdot2+1\cdot1 & = 3 \\
100 & \longmapsto & 1\cdot4+0\cdot2+0\cdot1 & = 4 \\
101 & \longmapsto & 1\cdot4+0\cdot2+1\cdot1 & = 5 \\
110 & \longmapsto & 1\cdot4+1\cdot2+0\cdot1 & = 6 \\
111 & \longmapsto & 1\cdot4+1\cdot2+1\cdot1 & = 7.
\end{array}$$

We have <u>ordered</u> the subsets of $\{1,2,3\}$.

This was cute, but it was not very "natural". There may be other interesting ways to do it.

However,

WARNING: In 1891, Georg Cantor proved that if S is an <u>infinite</u> set then it is <u>impossible</u> to list the subsets of S.

<u>Proof</u> (Cantor's Diagonal Argument):

For convenience, let $S := \mathbb{N} = \{0, 1, 2, 3, 4, \ldots\}$. We can think of its subsets as infinite binary strings

$$\text{e.g.} \quad \begin{array}{cccccccccccc} & 1 & 2 & \cdot & 4 & \cdot & \cdot & 7 & \cdot & \cdot & 10 & 11 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 \end{array} \cdots \quad \text{etc.}$$

Suppose that we <u>could</u> put all of these binary strings in a vertical list, say

$$\begin{array}{ll} \text{1st} & 0\,1\,1\,0\,1\,0\,0\,1\,0\,0\,1\,1 \cdots \\ \text{2nd} & 1\,0\,1\,0\,0\,1\,0\,1\,0\,0\,0\,1 \cdots \\ \text{3rd} & 1\,1\,1\,0\,1\,1\,0\,0\,1\,0\,0 \cdots \\ \text{4th} & 1\,0\,0\,0\,0\,0\,0\,1\,0\,0\,1\,1 \cdots \\ \text{etc.} & \vdots \qquad \ddots \end{array}$$

Suppose that the list eventually contains <u>every</u> binary string.

Now define a new string

$$b_1 b_2 b_3 b_4 b_5 \cdots$$

that disagrees with the ith string in the ith position. ~~In~~ our example, we have

$$1\ 1\ 0\ 1 \cdots \text{ etc.}$$

OOPS. This "diagonal" string is a new string that was NOT in our list.

Conclusion: No listing of the subsets of $\mathbb{N}$ can ever be complete!

Does that surprise you?

Moral of the story. You should never write

$$\wp(S) = \{X_1, X_2, X_3, \ldots\}.$$

It will make you seem unsophisticated.

7/14/14

No HW 3 yet.
Quiz 2 Now (20 minutes)

I think we've done enough logic for the moment (it needs to sink in), so this week we'll discuss

The Binomial Theorem

Let S be a set with n elements, say

$$S := \{x_1, x_2, x_3, \ldots, x_n\}.$$

To each subset $X \subseteq S$ we associate the binary string of length n,

$$b_1 b_2 b_3 \cdots b_n \quad \text{where } b_i = \begin{cases} 1 & \text{if } x_i \in X \\ 0 & \text{if } x_i \notin X \end{cases}$$

This gives a "1-1 correspondence" between subsets of S and binary strings of length n. Recall the notation

$$2^S = \wp(S) = \text{The set of subsets of } S.$$

$\{0,1\}^n$ = The set of "ordered n-tuples" of 0's and 1's

which we can think of as binary strings by erasing the commas and parentheses.

Then we have

$$2^S \longleftrightarrow \{0,1\}^n$$

Picture:

	$\{x_1,x_2,x_3\}$			111	
$\{x_1,x_2\}$	$\{x_1,x_3\}$	$\{x_2,x_3\}$	110	101	011
$\{x_1\}$	$\{x_2\}$	$\{x_3\}$	100	010	001
	$\emptyset$			000	

Recall that the number of subsets is

$$2^n.$$

Problem for Today:

Let S be a set with n elements and consider an integer $0 \leq k \leq n$. How many subsets with k elements does S have?

Equivalently, how many binary strings of length n have exactly k 1's (and hence $n-k$ 0's)?

To solve this, we need a "preliminary fact" (which we call a "lemma").

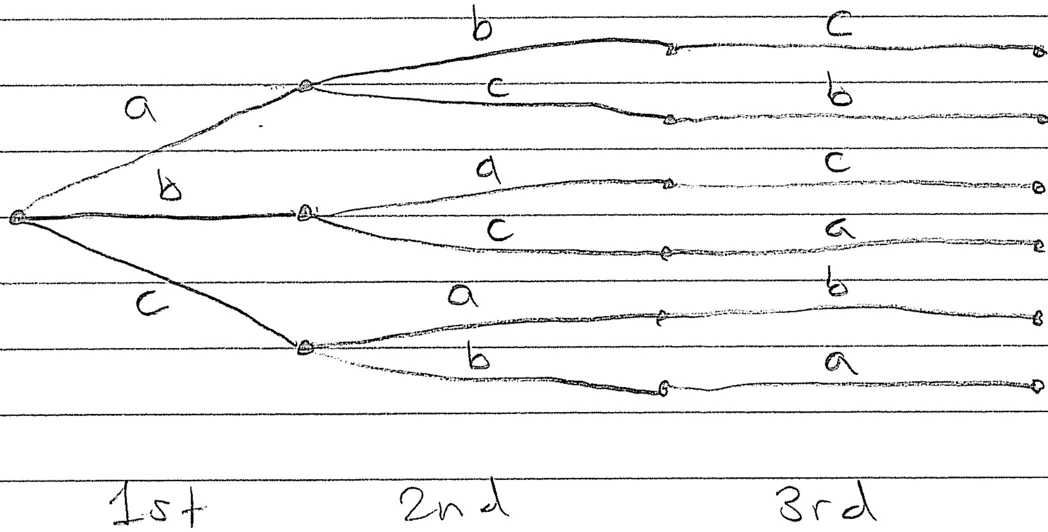
Q: Given n different symbols, in how many ways can I write them in a line?

Example: Use the symbols a, b, c. The possibilities are

abc, acb,
bac, bca,
cab, cba

There are 6 possibilities.

We could arrange them in a tree like this:



So we really want to count the branches of this tree. The total number of branches is

$$3 \times 2 \times 1 = 6.$$

In general, given a positive integer n, we define

$$n! := n(n-1)(n-2) \cdots 3 \cdot 2 \cdot 1$$

We call this "n factorial".

Lemma: Given n different symbols, the number of ways to place them on a line (i.e. put them in order) is

$$n!$$

Proof: There are n ways to choose the first / leftmost symbol. Then there are $n-1$ ways to choose the next symbol. Continuing in this way, the total number of choices is

$$\underbrace{n}_{1st} \times \underbrace{n-1}_{2nd} \times \underbrace{n-2}_{3rd} \ldots \underbrace{2}_{} \times \underbrace{1}_{n^{th}} = n!$$

//

These numbers grow fast:

n	1	2	3	4	5	6	7
$n!$	1	2	6	24	120	720	5040

James Stirling (1692–1770) gave a charming and surprising formula for their rate of growth.

He proved that

$$n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$$

"Stirling's approximation"

Now we want to count the ways to put the symbols

$$0, 0, 0, 1, 1, 1, 1$$

in order. Problem: The symbols are not all different. What can we do?

Let $N :=$ The number of ways to put $0, 0, 0, 1, 1, 1, 1$ in order. Right now N is unknown to us. We need some kind of equation.

(Trick) Let's temporarily consider the labeled symbols

$$0_1, 0_2, 0_3, 1_1, 1_2, 1_3, 1_4.$$

We know there are

$$7! = 5040$$

ways to put these symbols in order.
But this number is <u>bigger</u> than N.
For example, the orderings

$$0_3\, 1_4\, 0_1\, 1_1\, 1_3\, 0_2\, 1_2 \quad \text{and} \quad 0_1\, 1_3\, 0_3\, 1_1\, 1_2\, 0_2\, 1_4$$

correspond to the <u>same</u> ordering of unlabeled symbols. We have

$$5040 > N$$

but how <u>much</u> bigger?

The answer is very nice. Given an ordering of $0, 0, 0, 1, 1, 1, 1$, for example

$$0101101,$$

there are $3!$ to put labels on the 0's and $4!$ ways to put labels on the 1's.

$$0101101 \longrightarrow \left\{\begin{array}{c} 0_3\,1_4\,0_1\,1_1\,1_3\,0_2\,1_2 \\ 0_1\,1_3\,0_3\,1_1\,1_2\,0_2\,1_4 \\ \text{etc.} \\ \vdots \end{array}\right.$$

$$3! \times 4! = 6 \times 24 = 144$$

To each <u>unlabeled</u> ordering, there correspond $3! \times 4! = 144$ <u>labeled</u> orderings.

Finally, this gives us an <u>equation</u> for N:

$$7! = N \times 3! \times 4!$$

$$\Longrightarrow \quad N = \frac{7!}{3!\,4!} = \frac{5040}{144} = 35.$$

Conclusion: There are 35 ways to order the symbols 0, 0, 0, 1, 1, 1, 1. (And we didn't even have to find them all!)

<u>Theorem</u>: Let S be a set with n elements and let k be an integer such that $0 \leq k \leq n$. Then the number of subsets of S with k-elements is

$$\frac{n!}{k!(n-k)!}$$

<u>Proof</u>: This is the same as counting orderings of the symbols

$$\underbrace{1, 1, \dots, 1}_{k \text{ of these}}, \underbrace{0, 0, \dots, 0}_{n-k \text{ of these}}\ ,$$

i.e., binary strings of length n containing k 1's. Let N be the number of orderings. We want an equation for N.

So instead we will count the orderings of the n <u>different</u> symbols

$$1_1, 1_2, \ldots, 1_k, 0_1, 0_2, \ldots, 0_{n-k}.$$

On one hand, there are $n!$ ways to order these symbols.

On the other hand, these orderings break up into groups of equal size: For each ordering of the un<u>labeled</u> symbols, there are $k! \times (n-k)!$ ways to label it.

We conclude that

$$\underset{\substack{\text{order the}\\\text{labeled}\\\text{symbols}}}{n!} = \underset{\substack{\text{order the}\\\text{unlabeled}\\\text{symbols}}}{N} \times \underbrace{k! \times (n-k)!}_{\text{label them.}}$$

Hence

$$N = \frac{n!}{k!(n-k)!}$$

//

We have a shorthand notation for this:

$$\binom{n}{k} := \frac{n!}{k!(n-k)!}$$

We call this number "n choose k" because it is the number of ways to choose k things from a set of n.

Observe that

$$\sum_{k=0}^{n} \binom{n}{k} = 2^n. \quad (\text{Why?})$$

Thinking Problem:

$$\sum_{k=0}^{n} k \cdot \binom{n}{k} = ?$$

7/15/14

HW 3 due Friday
Quiz 3 on Monday

Last time we proved:

If S is a set with n elements and k is an integer such that $0 \le k \le n$, then the number of subsets of S with k elements is

$$\binom{n}{k} := \frac{n!}{k!(n-k)!}$$

We say "n choose k"

Example: I have 10 (different) books and I want to give you 3. In how many ways can I do this?

$$\binom{10}{3} = \frac{10!}{3!\,7!}$$

$$= \frac{10 \cdot 9 \cdot 8 \cdot \cancel{7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}}{3 \cdot 2 \cdot 1 \cdot \cancel{7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}}$$

$$= \frac{10 \cdot \cancel{9}^{3} \cdot \cancel{8}^{4}}{\cancel{3} \cdot \cancel{2} \cdot 1} = 10 \cdot 3 \cdot 4 = 120$$

To prove the formula we used the correspondence

subsets of S with k elements $\longleftrightarrow$ binary strings of length n with k 1's.

Let N be the number of such strings. Then

$$n! = N \times \underbrace{k! \times (n-k)!}_{\text{label them.}}$$

$n!$: strings with labeled 0's & 1's

N: strings with <u>unlabeled</u> 0's & 1's

Solve the equation:

$$\Longrightarrow \quad N = \frac{n!}{k!\,(n-k)!}$$

Today we will give a more algebraic interpretation of these numbers.

Today's Problem: Compute the expansion of the polynomial $(1+x)^n$.

Examples:

$$(1+x)^0 = 1$$
$$(1+x)^1 = 1+x$$
$$\begin{aligned}(1+x)^2 &= (1+x)(1+x) \\ &= 1+2x+x^2\end{aligned}$$
$$\begin{aligned}(1+x)^3 &= (1+x)(1+x)^2 \\ &= (1+x)(1+2x+x^2) \\ &= 1+3x+3x^2+x^3.\end{aligned}$$
$$\vdots$$

Is there a formula for $(1+x)^n$?

Yes. The formula is

$$\begin{aligned}(1+x)^n &= \binom{n}{0} + \binom{n}{1}x + \binom{n}{2}x^2 + \cdots + \binom{n}{n}x^n \\ &= \sum_{k=0}^{n} \binom{n}{k} x^k\end{aligned}$$

In fact, we will show something more general.

☆ The Binomial Theorem: For all real or complex numbers x, y we have

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

$$= \binom{n}{0} y^n + \binom{n}{1} x y^{n-1} + \binom{n}{2} x^2 y^{n-2} + \cdots + \binom{n}{n} x^n$$

<u>Proof</u>: I will temporarily pretend that x and y don't commute ($xy \neq yx$) and see what happens.

$$(x+y)^0 = 1$$
$$(x+y)^1 = x+y$$
$$(x+y)^2 = (x+y)(x+y)$$

$$= xx + \left\{ \begin{matrix} xy \\ +yx \end{matrix} \right\} + yy.$$

$$(x+y)^3 = (x+y)(x+y)(x+y)$$

$$= xxx + \left\{ \begin{matrix} yxx \\ +xyx \\ +xxy \end{matrix} \right\} + \left\{ \begin{matrix} yyx \\ +yxy \\ +xyy \end{matrix} \right\} + yyy$$

Then if we let $xy = yx$ we get

$$(x+y)^3 = \underbrace{xxx}_{\downarrow} + \underbrace{\begin{pmatrix} yxx \\ + xyx \\ + xxy \end{pmatrix}}_{\downarrow} + \underbrace{\begin{pmatrix} yyx \\ + yxy \\ + xyy \end{pmatrix}}_{\downarrow} + \underbrace{yyy}_{\downarrow}$$

$$= x^3 + 3x^2y + 3xy^2 + y^3$$

In other words, $(x+y)^n$ is the sum of all the words of length n using the letters x and y.

If we assume that $xy = yx$ and collect all the x's to the left then every term has the form $x^k y^{n-k}$ for some $0 \leq k \leq n$.

The coefficient of $x^k y^{n-k}$ in $(x+y)^n$ equals the number of words of length n with k x's and $n-k$ y's, and we know this equals $\binom{n}{k}$.

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

//

Observations:

1. Note that

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!} = \frac{n!}{(n-k)!\,k!} = \binom{n}{n-k}.$$

$k = n-(n-k)$

The number of words with k x's and $(n-k)$ y's equals the number of words with $(n-k)$ x's and k y's. (Just switch $x \leftrightarrow y$.).
So we could also write

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

2. Substitute $y=1$ to get

$$(1+x)^n = \sum_{k=0}^{n} \binom{n}{k} x^k = \sum_{k=0}^{n} \binom{n}{k} x^{n-k}$$

$$= \sum_{k=0}^{n} \binom{n}{n-k} x^{n-k} = \sum_{k=0}^{n} \binom{n}{n-k} x^k.$$

Take your pick.

3. Evaluate at $x=1$ to get

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

$$2^n = \binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \cdots + \binom{n}{n}.$$

We can also see this by counting subsets of a set with n elements.

4. Evaluate at $x=-1$ to get

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

$$0 = \binom{n}{0} - \binom{n}{1} + \binom{n}{2} - \binom{n}{3} + \cdots + (-1)^n \binom{n}{n}.$$

$$\binom{n}{0} + \binom{n}{2} + \binom{n}{4} + \cdots = \binom{n}{1} + \binom{n}{3} + \binom{n}{5} + \cdots$$

We can also see this by counting subsets: The number of even subsets equals the number of odd subsets.

5. We can apply operations to the Binomial Theorem to get more true statements.

Example:

$$\frac{d}{dx}(1+x)^n = \frac{d}{dx}\sum_{k=0}^{n}\binom{n}{k}x^k$$

$$n(1+x)^{n-1} = \sum_{k=0}^{n}\binom{n}{k}k\,x^{k-1}$$

Now evaluate at $x=1$ to get

$$n\cdot 2^{n-1} = \sum_{k=0}^{n} k\binom{n}{k}$$

$$n\cdot 2^{n-1} = 0\cdot\binom{n}{0} + 1\binom{n}{1} + 2\binom{n}{2} + \cdots + n\binom{n}{n}.$$

This is not very obvious from counting subsets. The algebraic proof is much easier.

6. You can do an infinite number of tricks like this.

Because of the Binomial Theorem, the numbers

$$\binom{n}{k}$$

are called "binomial coefficients"

I've saved their most important property for last.

<u>Theorem</u>: For all relevant k, n we have

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

[You will give <u>two</u> different proofs on HW 3. I'll give a third here.]

Proof: This follows from the Binomial Theorem, using the observation

$$(1+x)^n = (1+x)(1+x)^{n-1}$$

$$= 1 \cdot (1+x)^{n-1} + x(1+x)^{n-1}.$$

$$(1+x)^n = (1+x)^{n-1} + x(1+x)^{n-1}$$

$$\sum_{k=0}^{n} \binom{n}{k} x^k = \sum_{k=0}^{n-1} \binom{n-1}{k} x^k + \underbrace{x \sum_{k=0}^{n-1} \binom{n-1}{k} x^k}$$

$$= \sum_{k=0}^{n-1} \binom{n-1}{k} x^k + \sum_{k=0}^{n-1} \binom{n-1}{k} x^{k+1}$$

$$= \sum_{k=0}^{n-1} \binom{n-1}{k} x^k + \sum_{k=1}^{n} \binom{n-1}{k-1} x^k$$

$$\sum_{k} \binom{n}{k} x^k = \sum_{k} \left[\binom{n-1}{k} + \binom{n-1}{k-1} \right] x^k$$

Do I need to worry about the limits of the sums? Only a tiny bit.

Comparing coefficients on both sides gives

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

[Your two proofs will be more convincing (hopefully).]

We just showed that the binomial coefficients are the entries of "Pascal's Triangle"

$$
\begin{array}{ccccccccccc}
 & & & & & 1 & & & & & \\
 & & & & 1 & & 1 & & & & \\
 & & & 1 & & 2 & & 1 & & & \\
 & & 1 & & 3 & & 3 & & 1 & & \\
 & 1 & & 4 & & 6 & & 4 & & 1 & \\
1 & & 5 & & 10 & & 10 & & 5 & & 1
\end{array}
$$

etc.

$$
\begin{array}{ccccccc}
 & & & \binom{0}{0} & & & \\
 & & \binom{1}{0} & & \binom{1}{1} & & \\
 & \binom{2}{0} & & \binom{2}{1} & & \binom{2}{2} & \\
\binom{3}{0} & & \binom{3}{1} & & \binom{3}{2} & & \binom{3}{3}
\end{array}
$$

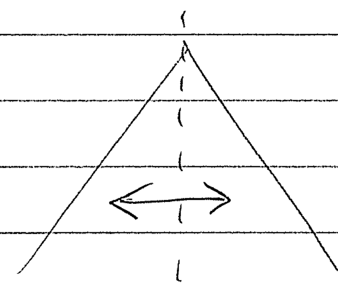
etc.

Each number is the sum of the two above:

$$
\begin{array}{ccc}
\binom{n-1}{k-1} & & \binom{n-1}{k} \\
 & \binom{n}{k} &
\end{array}
$$

Remarks:

1. The symmetry $\binom{n}{k} = \binom{n}{n-k}$ is evident in the picture.



left-right symmetry

2. We conventionally define

$$\binom{n}{k} = 0 \quad \text{if } k < 0 \text{ or } n < k.$$

$$\begin{array}{ccccccccccccccccc}
\cdots & 0 & & 0 & & 0 & & 1 & & 0 & & 0 & & 0 & \cdots \\
\cdots & & 0 & & 0 & & 1 & & 1 & & 0 & & 0 & & \cdots \\
\cdots & 0 & & 0 & & 1 & & 2 & & 1 & & 0 & & 0 & \cdots \\
\cdots & & 0 & & 1 & & 3 & & 3 & & 1 & & 0 & & \cdots
\end{array}$$

3. But what if $\underline{n}$ is negative?

Does Pascal's Triangle go <u>up</u>?

7/16/14

I didn't mean to talk about this today but I did. So here are some notes.

Four Ways to Count:

Let S be a set with n elements. By definition of "set" these elements are

- unordered (there is no "1st" element...)
- distinct (no repeated elements).

Now we want to "choose" k elements from S. There are at least 4 ways to interpret the word "choose".

We might grab all k elements at once or we might pick them one at a time. If we pick them one at a time then we might/might not

- replace each element after we pick it, so we might get it multiple times.
- record the order in which we picked the elements.

We can arrange this information in a table. If we "choose" k things from a set of n, the number of possibilities is given by

	order matters	order irrelevant
with replacement	n^k	?
without replacement	$\binom{n}{k} \cdot k!$	$\binom{n}{k}$

These three entries are pretty straightforward once we know about binomial coefficients.

Bottom-Right: We've been talking about this all week.

↓

Top-Left: We can think of S as a set of symbols (an "alphabet"). Then we want to count the number of words of length k that we can make from this alphabet. (Order of symbols matters and symbols can be repeated.) The answer is

$$\underbrace{\underline{n \text{ choices}}}_{\text{1st letter}} \times \underbrace{\underline{n \text{ choices}}}_{\text{2nd letter}} \times \cdots \times \underbrace{\underline{n \text{ choices}}}_{k\text{th letter}} = n^k$$

Bottom-Left: To choose an ordered set of k elements without repetition, first choose an unordered set of k elements in $\binom{n}{k}$ ways, then order this set in $k!$ ways.

The total number of possibilities is

$$\underbrace{\binom{n}{k}}_{\text{choose an unordered set}} \cdot \underbrace{k!}_{\text{order it.}}$$

Top-Right: ?

This is the hardest one. It requires a clever trick.

First let's do an example. Choose 2 elements from $\{a, b, c\}$ with repetition, but don't record the order. The choices are

a, a   b, b   c, c.
a, b   b, c
a, c

Why are there 6?

I'll show you a bijection to certain kinds of binary strings:

a, a $\longrightarrow$	0011	
a, b $\longrightarrow$	0101	How does
a, c $\longrightarrow$	0110	that work?
b, b $\longrightarrow$	1001	
b, c $\longrightarrow$	1010	
c, c $\longrightarrow$	1100	

Idea: The 0's are "things" and the 1's are "dividers". That is

some # 0's representing a's 1 some # 0's representing b's 1 some # 0's representing c's

00 | 000 | 0

two a's 3 b's 1 c

Clever right? (I didn't invent it)

To encode a choice of k things from n we will use

k 0's (things)
$n-1$ 1's (dividing lines)

So the total # choices is the number of binary strings with k 0's and $n-1$ 1's, which we know to be

$$\binom{k+(n-1)}{k}.$$

The completed table of ways to count is

	ordered	NOT ordered	
	n^k	$\binom{n+k-1}{k}$	with replacement
	$\binom{n}{k}\cdot k!$	$\binom{n}{k}$	without replacement.

Thinking Problem:

Is there some mystical significance to the equation

$$\binom{-n}{k} = (-1)^k \binom{n+k-1}{k} \quad ?$$

"Choosing negative things without replacement is the same as choosing positive things with replacement."

What ?!

7/16/14

HW 3 due Friday
Quiz 3 on Monday

Last time we proved the

☆ Binomial Theorem: For all real or complex numbers x, y and for all positive integers n we have

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

$$= x^n + n x^{n-1} y + \frac{n(n-1)}{2} x^{n-2} y^2 + \cdots + \frac{n(n-1)}{2} x^2 y^{n-2} + n x y^{n-1} + y^n$$

Today we will apply this. The most important applications are in the theory of probability.

Let p and q be positive real numbers such that

$$p + q = 1$$

Then the Binomial Theorem says:

$$1 = 1^n = (p+q)^n = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k}$$

This has the following interpretation. Suppose you have a <u>biased coin</u> such that

$$\text{Prob}(\text{"heads"}) = p$$
$$\text{Prob}(\text{"tails"}) = q$$

If you flip the coin n times, what is the probability that you get "heads" exactly k times?

<u>Example</u>: If you flip the coin 3 times the probability of getting the <u>sequence</u> HTH is

$$\begin{aligned}\text{Prob}(HTH) &= \text{Prob}(H)\,\text{Prob}(T)\,\text{Prob}(H)\\ &= pqp\\ &= p^2 q.\end{aligned}$$

To compute the probability of getting exactly 2 heads, we sum over the ways it can happen.

Prob (getting heads twice)

$= \text{Prob}(\{HHT, HTH, THH\})$

$= \text{Prob}(HHT) + \text{Prob}(HTH) + \text{Prob}(THH)$

$= p^2q \quad + \quad p^2q \quad + \quad p^2q$

$= 3p^2q.$

In general, the probability of getting exactly k heads in n tosses is

$$\binom{n}{k} p^k q^{n-k}$$

Example: Suppose in a certain population each birth has

$$\text{Prob(boy)} = 1/3 \qquad \left(\frac{1}{3} + \frac{2}{3} = 1\right)$$
$$\text{Prob(girl)} = 2/3$$

If a certain family has 4 children, how many boys are they likely to have?

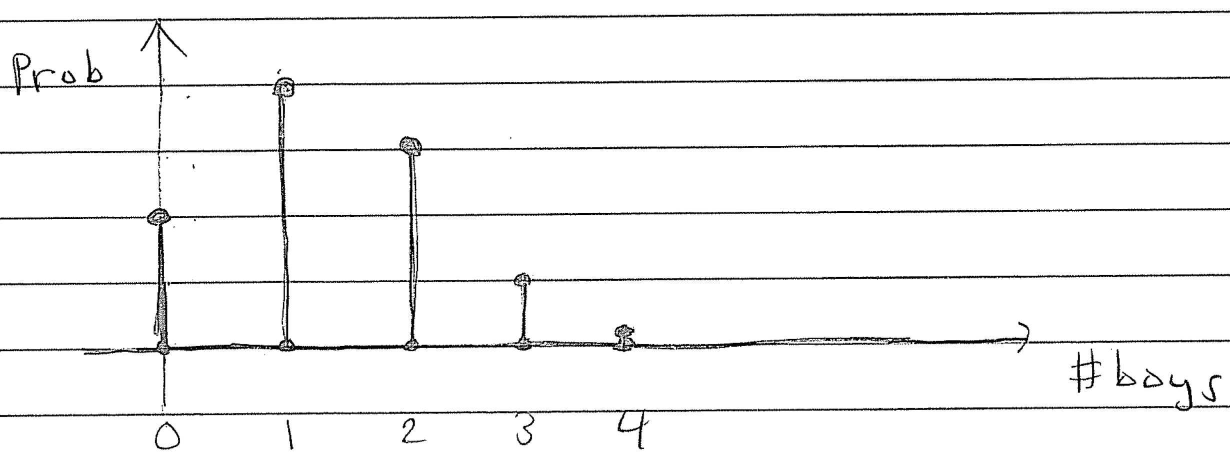
We'll compute the full distribution

# boys	0	1	2	3	4
Prob	$\binom{4}{0}\left(\frac{1}{3}\right)^0\left(\frac{2}{3}\right)^4$	$\binom{4}{1}\left(\frac{1}{3}\right)^1\left(\frac{2}{3}\right)^3$	$\binom{4}{2}\left(\frac{1}{3}\right)^2\left(\frac{2}{3}\right)^2$	$\binom{4}{3}\left(\frac{1}{3}\right)^3\left(\frac{2}{3}\right)^1$	$\binom{4}{4}\left(\frac{1}{3}\right)^4\left(\frac{2}{3}\right)^0$
	$1 \cdot \frac{1^0 \cdot 2^4}{3^4}$	$4 \cdot \frac{1^1 \cdot 2^3}{3^4}$	$6 \cdot \frac{1^2 \cdot 2^2}{3^4}$	$4 \cdot \frac{1^3 \cdot 2^1}{3^4}$	$1 \cdot \frac{1^4 \cdot 2^0}{3^4}$
	$\frac{16}{81}$	$\frac{32}{81}$	$\frac{24}{81}$	$\frac{8}{81}$	$\frac{1}{81}$

Notice that

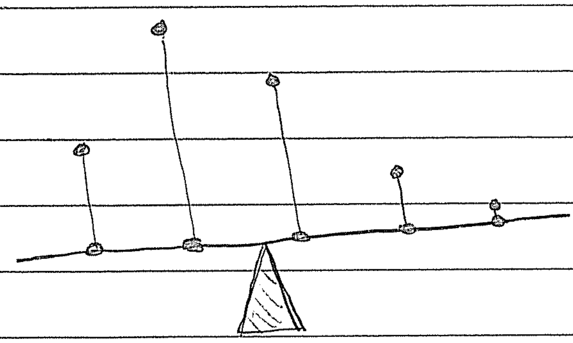
$$\frac{16}{81} + \frac{32}{81} + \frac{24}{81} + \frac{8}{81} + \frac{1}{81} = \frac{81}{81} = 1$$

as it should be. We can think of probability as a distribution of mass



Then the "average" or "expected outcome" is the same as the "center of mass".

Where is it? Where does it balance?



Archimedes tells us the answer. His "law of the lever" says that

mass × (distance from center)

should balance. In our case, the "center of mass" is

$$0 \cdot \frac{16}{81} + 1 \cdot \frac{32}{81} + 2 \cdot \frac{24}{81} + 3 \cdot \frac{8}{81} + 4 \cdot \frac{1}{81}$$

$$= \frac{0 + 32 + 48 + 24 + 4}{81} = \frac{108}{81} = \frac{4}{3}$$

$$= 1.33333\ldots$$

In a family with 4 children we expect

1.333... boys.

Is that surprising? NO. It's just

$$p \cdot n$$

Prob(boy) # of children

Theorem: Flip a biased coin (Prob(H) = p, Prob(T) = q, $p+q=1$) n times. The *expected* number of heads is

$$pn$$

Proof: By Archimedes' principle, the expected number of heads is

$$E(\#\text{heads}) = \sum_{k=0}^{n} k \cdot \text{Prob}(k \text{ heads})$$

"distance" "mass"

Since $\text{Prob}(k \text{ heads}) = \binom{n}{k} p^k q^{n-k}$ we have

$$E(\# \text{heads}) = \sum_{k=0}^{n} k \binom{n}{k} p^k q^{n-k}$$

What tricks do we have for evaluating this sum? Here's the key trick:

$$k\binom{n}{k} = \frac{k\, n!}{k!(n-k)!} = \frac{n!}{(k-1)!(n-k)!}$$

$$= n \cdot \frac{(n-1)!}{(k-1)!(n-k)!} = n \cdot \frac{(n-1)!}{(k-1)!\,[(n-1)-(k-1)]!}$$

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

In summary, we have

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

Plugging this in to our sum gives

$$E(\# \text{heads}) = \sum_{k=0}^{n} k \binom{n}{k} p^k q^{n-k}$$

$$= \sum_{k=1}^{n} k \binom{n}{k} p^k q^{n-k}$$

$$= \sum_{k=1}^{n} n \binom{n-1}{k-1} p^k q^{n-k}$$

$$= pn \sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1} q^{n-k}$$

$$= pn \sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1} q^{(n-1)-(k-1)}$$

$$= pn \sum_{j=0}^{n-1} \binom{n-1}{j} p^{j} q^{(n-1)-j}$$

$$= pn\,(p+q)^{n-1}$$

$$= pn\,(1)^{n-1}$$

$$= pn.$$

There may be trickier tricks, but that one let us practice our skills.

Thinking Problem: What is the expected value of

$$(\#\text{heads} - pn)^2 \quad ?$$

7/17/14

HW 3 due tomorrow
Quiz 3 on Monday

Last time we talked about binomial random variables.

Consider a biased coin with

$$P(H) = p$$
$$P(T) = q$$

(and hence $p+q=1$ with $p,q \in [0,1]$).
If we flip the coin n times, the probability of getting "heads" exactly k times is

$$\binom{n}{k} p^k q^{n-k}.$$

Note: These probabilities sum to 1 because of the Binomial Theorem

$$\sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} = (p+q)^n = 1^n = 1$$

//

We also saw that the "expected" number of heads in n tosses is

$$\sum_{k=0}^{n} k \cdot \binom{n}{k} p^k q^{n-k} = np ,$$

which makes sense. (If we make n flips and heads occur "p of the time", then we expect pn heads.)

As a special case, consider a fair coin

$$p = q = \frac{1}{2}$$

Then we have

$$\sum_{k=0}^{n} k \cdot \binom{n}{k} \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{n-k} = \frac{1}{2} n$$

$$\frac{1}{2^n} \sum_{k=0}^{n} k \cdot \binom{n}{k} = \frac{1}{2} n$$

$$\sum_{k=0}^{n} k \cdot \binom{n}{k} = n \cdot 2^{n-1}$$

Do you recognize this?

But what *is* probability?

Here is a brief history.

- Probability "theory" began with (1654) correspondence between Pascal & Fermat.

- First textbook (1812) by Laplace

- Modern definition of probability is due to Kolmogorov (1933)

Here are Kolmogorov's axioms.

A *probability space* is a set S (called the *sample space*) together with a function

$$P : \wp(S) \longrightarrow \mathbb{R}$$

Subsets $E \subseteq S$ are called *events*. We call the real number $P(E)$ the *probability* of the event E. It must satisfy 3 axioms:

(1) For all events $E \subseteq S$ we have

$$0 \leq P(E) \leq 1$$

(2) $P(S) = 1$

(3) If events E, F are mutually exclusive (i.e. if $E \cap F = \emptyset$) then we have

$$P(E \cup F) = P(E) + P(F)$$

That's it. //

From these 3 axioms we can derive many theorems. For example, we have

(4) $P(E^c) = 1 - P(E)$.

Proof: Since $E \cup E^c = S$ and $E \cap E^c = \emptyset$ we have

$$\begin{aligned} 1 &= P(S) && (2) \\ &= P(E \cup E^c) \\ &= P(E) + P(E^c) && (3) \end{aligned}$$

//

(5) $P(\emptyset) = 0$

Proof: Since $\emptyset = S^c$, we have

$$
\begin{aligned}
P(\emptyset) &= P(S^c) \\
&= 1 - P(S) && (4) \\
&= 1 - 1 && (2) \\
&= 0
\end{aligned}
$$

//

(6) If $E \subseteq F$ then $P(E) \leq P(F)$.

Proof: If $E \subseteq F$ then we can use Boolean algebra to show that

$$F = E \cup (E \cap F^c) \quad \& \quad \emptyset = E \cap (E \cap F^c)$$

Since (1) says $0 \leq P(E \cap F^c)$, we get

$$
\begin{aligned}
0 &\leq P(E \cap F^c) \\
P(E) &\leq P(E) + P(E \cap F^c) \\
&= P(F) && (3)
\end{aligned}
$$

//

(7) For any events $E_1, E_2, \dots, E_n$ such that $E_i \cap E_j = \emptyset$ for all $i \neq j$, we have

$$P\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} P(E_i)$$

Proof: This follows from (3) using induction, which we'll discuss soon. //

Finally, we have the important

☆ Principle of Inclusion-Exclusion:

For any events $E, F \subseteq S$ we have

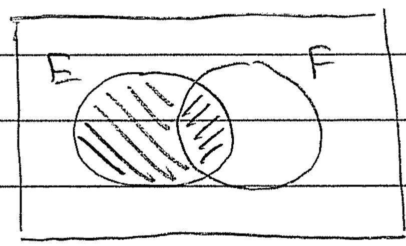
$$\boxed{P(E \cup F) = P(E) + P(F) - P(E \cap F)}$$

Proof: First use Boolean algebra to see

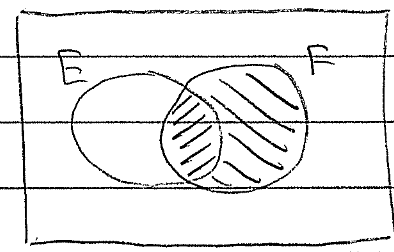
$$E = (E \cap F) \cup (E \cap F^c) \ \& \ \emptyset = (E \cap F) \cap (E \cap F^c)$$

$$F = (E \cap F) \cup (E^c \cap F) \ \& \ \emptyset = (E \cap F) \cap (E^c \cap F)$$

Pictures:



$(E\cap F)\cup(E\cap F^c)$



$(E\cap F)\cup(E^c\cap F)$

So axiom (3) says

$$P(E) = P(E\cap F) + P(E\cap F^c)$$
$$P(F) = P(E\cap F) + P(E^c\cap F)$$

Now observe that we have a "disjoint union"

$$E\cup F = (E\cap F^c)\sqcup(E\cap F)\sqcup(E^c\cap F).$$

Notation: We will write "$A = B\sqcup C$" to mean that

"$A = B\cup C$ and $\emptyset = B\cap C$".

In this case we say that A is the <u>disjoint union</u> of B and C

Thus (7) implies

$$P(E \cup F) = P(E \cap F^c) + P(E \cap F) + P(E^c \cap F).$$

Finally, add $P(E \cap F)$ to both sides to get

$$P(E \cup F) + P(E \cap F)$$

$$= \underbrace{P(E \cap F^c) + P(E \cap F)} + \underbrace{P(E^c \cap F) + P(E \cap F)}$$

$$= P(E) + P(F).$$

The P.I.E. is quite useful.

Given a single element $x \in S$ we write

$$P(x) := P(\{x\})$$

for the probability of the elementary event $\{x\}$.

We will often define a probability space by specifing the numbers $P(x)$ for all $x \in S$.

Then the probability of an event $E \subseteq S$ is

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

Important Special Case:

Let S be finite. If all elementary events are eqully likely then we have

$$P(x) = \frac{1}{\# S} \text{ for all } x \in S.$$

In this case, the probability of an event $E \subseteq S$ is

$$\begin{aligned}
P(E) &= \sum_{x \in E} P(x) \\
&= \sum_{x \in E} \frac{1}{\# S} \\
&= \underbrace{\frac{1}{\# S} + \frac{1}{\# S} + \cdots + \frac{1}{\# S}}_{\# E \text{ times}} \\
&= \frac{\# E}{\# S}.
\end{aligned}$$

In other words, if all possible outcomes are equally likely, then

$$P(E) = \frac{\# E}{\# S} = \frac{\# \text{ favorable outcomes}}{\text{total } \# \text{ possible outcomes}}$$

You are probably familiar with is concept.

Example: Flip a fair coin n times.

The sample space is $S = \{H, T\}^n$.
For any sequence $x \in S$ we have

$$P(x) = \frac{1}{\# S} = \frac{1}{2^n}.$$

Let E be the event "exactly k heads".
Then

$$P(E) = \frac{\# E}{2^n} = \frac{\# \text{ ways to get } k \text{ heads}}{2^n}$$

$$= \frac{\binom{n}{k}}{2^n}.$$

This agrees with our previous formula

$$P(k \text{ heads}) = \binom{n}{k}\left(\frac{1}{2}\right)^{k}\left(\frac{1}{2}\right)^{n-k}$$

$$= \binom{n}{k}\left(\frac{1}{2}\right)^{n}$$

$$= \binom{n}{k}\cdot\frac{1}{2^n}$$

Remark: The events $x \in \{H,T\}^n$ are NOT equally likely when the coin is biased.

Another Example: Roll two fair 6-sided dice.

If we regard the dice as different (say, one is red, one is blue) then the sample space is

$$S = \{1,2,3,4,5,6\}^2$$

and every elementary event is equally likely:

$$P(x) = \frac{1}{\#S} = \frac{1}{36} \quad \forall\, x \in S.$$

What is the probability of "rolling a 6"?

Define $E :=$ "rolling a 6"
$= \{(1,5), (2,4), (3,3), (4,2), (5,1)\}$

Then $P(E) = \#E / \#S$

$= 5/36$.

On average we will "roll a 6" in 5 out of every 36 rolls.

Thinking Problem: What if the two dice look identical and I can't tell the difference?

Answer: This is no problem. Your sample space in this case will have

$$\binom{6+2-1}{2} = \binom{7}{2} = 21 \text{ elements},$$

but now the elements are NOT equally likely.

Explicitly: We need to find a notation for unordered pairs with repetition allowed.
I propose we use this notation:

$$S = \{ \{1,1\}, \{1,2\}, \{1,3\}, \{1,4\}, \{1,5\}, \{1,6\},$$
$$\{2,2\}, \{2,3\}, \{2,4\}, \{2,5\}, \{2,6\},$$
$$\{3,3\}, \{3,4\}, \{3,5\}, \{3,6\},$$
$$\{4,4\}, \{4,5\}, \{4,6\},$$
$$\{5,5\}, \{5,6\},$$
$$\{6,6\} \}$$

I realize that $\{1,1\} = \{1\}$ but don't worry about it; it's just notation.
In this language the event "rolling a 6" is

$$E := \text{"rolling a 6"} = \{ \{1,5\}, \{2,4\}, \{3,3\} \}$$

But the following computation is WRONG

$$P(\text{"rolling a 6"}) = P(E) = \frac{\#E}{\#S} = \frac{3}{21}$$

this is WRONG.

Why is it wrong?

Because in this language, the outcomes $x \in S$ are <u>not</u> equally likely, so

$$P(E) = \frac{\#E}{\#S} \text{ is NOT valid.}$$

The correct computation is

$$P(E) = P(\{\{1,5\}, \{2,4\}, \{3,3\}\})$$

$$= P(\{1,5\}) + P(\{2,4\}) + P(\{3,3\})$$

$$= \frac{2}{36} + \frac{2}{36} + \frac{1}{36} = \frac{5}{36}$$

and the only reason I know these probabilities is because of the first calculation we did.

Moral: There may be many ways to encode an experiment. One notation may be better than another.

7/19/14

Friday's lecture was extemporized.
Here is my recollection of what we did.

First we discussed the solutions to HW 3.

Then we recalled the basic language of probability.

Then we discussed "urn problems".

Example: An "urn" contains 6 red balls and 3 blue balls. The balls feel identical to the touch. We reach in (without looking) and grab 4 balls. What is the probability we got 2 red balls?

To encode the experiment we need to make a choice: Did we record the order in which we chose the four balls?

Let's say NO.

Then the sample space is

S = all subsets of size 4 from the collection of $6+3=9$ balls

and $\#S = \binom{9}{4} = \frac{9 \cdot 8 \cdot 7 \cdot 6}{4 \cdot 3 \cdot 2 \cdot 1} = 126$

By symmetry considerations I assume that the outcomes are <u>equally likely</u> so that for any event $E \subseteq S$ we have

$$P(E) = \frac{\#E}{\#S} = \frac{\#E}{126}.$$

Now let E = "we got 2 red balls". What is $\#E$?

It helps to name the balls

$$\underbrace{r_1, r_2, r_3, r_4, r_5, r_6}_{\text{red balls}}, \underbrace{b_1, b_2, b_3}_{\text{blue balls}}$$

Then $E =$ "we got 2 red balls"

$$
\begin{aligned}
= \{ & \{r_1, r_2, b_1, b_2\}, \{r_1, r_2, b_1, b_3\}, \{r_1, r_2, b_2, b_3\}, \\
& \{r_1, r_3, b_1, b_2\}, \{r_1, r_3, b_1, b_3\}, \{r_1, r_3, b_2, b_3\}, \\
& \{r_1, r_4, b_1, b_2\}, \{r_1, r_4, b_1, b_3\}, \{r_1, r_4, b_2, b_3\}, \\
& \qquad \vdots \qquad\qquad\qquad \vdots \qquad\qquad\qquad \vdots \\
& \{r_5, r_6, b_1, b_2\}, \{r_5, r_6, b_1, b_3\}, \{r_5, r_6, b_2, b_3\} \}
\end{aligned}
$$

I couldn't write them all out.
How many are there? There are $\binom{3}{2} = 3$ columns corresponding to the ways to choose 2 blue balls and there are $\binom{6}{2} = 15$ rows corresponding to the ways to choose 2 red balls.

We conclude that

$$P(E) = \frac{\# E}{\# S} = \frac{\binom{3}{2}\binom{6}{2}}{\binom{9}{4}} = \frac{3 \cdot 15}{126} = \frac{45}{126}$$

There is a 35.7% of getting 2 red balls.

Using the same idea we can compute the full distribution. Grab 4 balls and let X be the number of red balls we get. Then

k	0	1	2	3	4
$P(X=k)$	$\frac{\binom{3}{4}\binom{6}{0}}{\binom{9}{4}} = 0$	$\frac{\binom{3}{3}\binom{6}{1}}{\binom{9}{4}} = \frac{6}{126}$	$\frac{\binom{3}{2}\binom{6}{2}}{\binom{9}{4}} = \frac{45}{126}$	$\frac{\binom{3}{1}\binom{6}{3}}{\binom{9}{4}} = \frac{60}{126}$	$\frac{\binom{3}{0}\binom{6}{4}}{\binom{9}{4}} = \frac{15}{126}$

They should add to 1. Do they?

$$\frac{0}{126} + \frac{6}{126} + \frac{45}{126} + \frac{60}{126} + \frac{15}{126} = \frac{126}{126} = 1 \checkmark$$

What is the expected value?
Wait, let's guess first.

The ratio of red balls in the urn is

$$\frac{6 \text{ red balls}}{9 \text{ total balls}} = \frac{6}{9} = \frac{2}{3}.$$

So if I grab 4 balls, I expect $2/3$ of them to be red, hence I expect

$$\frac{2}{3} \cdot 4 = \frac{8}{3} = 2.6666\ldots \text{ red balls.}$$

Now let's check. By Archimedes' law of the lever, the expected value of X is

$$E(X) = \sum_{k=0}^{4} k \cdot P(X=k)$$

$$= 0 \cdot \frac{0}{126} + 1 \cdot \frac{6}{126} + 2 \cdot \frac{45}{126} + 3 \cdot \frac{60}{126} + 4 \cdot \frac{15}{126}$$

$$= \frac{0 + 6 + 90 + 180 + 60}{126} = \frac{336}{126} = \frac{8}{3} \checkmark$$

Our guess was correct. (If not, I would have said it's the math that's wrong, not our guess. Any reasonable theory of probability should have given $8/3$.)

The general "urn problem":

There are R red and B blue balls in an "urn". We reach in and grab n balls. Let X be the number of red balls we get.

By previous considerations we know that

$$P(X=k) = \frac{\binom{R}{k}\binom{B}{n-k}}{\binom{R+B}{n}}$$

Do these numbers add to 1? Let's check.

$$\sum_k P(X=k) = 1 \quad ?$$

$$\sum_k \frac{\binom{R}{k}\binom{B}{n-k}}{\binom{R+B}{n}} = 1 \quad ?$$

$$\sum_k \binom{R}{k}\binom{B}{n-k} = \binom{R+B}{n} \quad ?$$

Yes! This is a famous identity called "Vandermonde's convolution".

We essentially just proved Vandermonde's convolution by thinking about balls in an urn.

Let's do another trick. What's the expected value of X?

The ratio of red balls in the urn is

$$\frac{\#\text{ red balls}}{\text{total }\#\text{ balls}} = \frac{R}{R+B}.$$

So if we grab n balls we expect that

$n \cdot \frac{R}{R+B}$ will be red.

In other words,

$$\sum_k k\,P(X=k) = n \cdot \frac{R}{R+B}$$

$$\sum_k k \frac{\binom{R}{k}\binom{B}{n-k}}{\binom{R+B}{n}} = n \cdot \frac{R}{R+B}.$$

How would we ever prove that strange identity algebraically?

You don't need to, but here's how I would do it. We want to show

$$\sum_k k \binom{R}{k}\binom{B}{n-k} = n \cdot \frac{R}{R+B} \binom{R+B}{n}$$

First I would use the fact that

$$k\binom{R}{k} = R\binom{R-1}{k-1}$$

(remember this?)

Then we have

$$\sum_k k\binom{R}{k}\binom{B}{n-k} = \sum_k R\binom{R-1}{k-1}\binom{B}{n-k}$$

$$= R\sum_k \binom{R-1}{k-1}\binom{B}{n-k}$$

$$= R\sum_k \binom{R-1}{k-1}\binom{B}{(n-1)-(k-1)}$$

$$= R\sum_j \binom{R-1}{j}\binom{B}{(n-1)-j}$$

Now I use "Vandermonde's convolution" to conclude that

$$\sum_j \binom{R-1}{j}\binom{B}{(n-1)-j} = \binom{(R-1)+B}{n-1}$$

(Do you want an algebraic proof of Vandermonde convolution? NOT TODAY!)

Plugging this in gives

$$\sum_k k\binom{R}{k}\binom{B}{n-k} = R\binom{R+B-1}{n-1}$$

$$= R\,\frac{n}{R+B}\binom{R+B}{n}$$

$$= n \cdot \frac{R}{R+B}\binom{R+B}{n}$$

as desired :)

Which proof do you like better:

algebraic or probabilistic?