Sets of Sets, Combinations

Set of n items Select r items	Order matters Permutation "Arrange"	Order does not matter Combination "Choose"
No repetition	$\mathrm{P(n, r)} = \frac{\mathrm{n!}}{\mathrm{(n\text{-}r)!}}$	$C(n, r) = \frac{\mathrm{n!}}{(n-r)!\,r!}$
Repetition allowed	n^r	$C(n-1+r, r)$

Problem 1)

a) Suppose that we flip a coin 8 times. How many sequences are there?
b) How many of the sequences contain exactly 1 head?
c) How many of the sequences contain all heads?
d) How many of the sequences contain 4 heads?

a) For each coin flip there are 2 possibilities. This is an arrangement with 8 selections and repetition.
n=2, r=8

$\boldsymbol{n^r}$**=**2^8**=256**

b) We have to choose 1 spot in the sequence to be heads.
n=8, r=1

$$\boldsymbol{C}(\mathbf{8}, 1) = \frac{8!}{(8\text{-}1)!1!} = \frac{8!}{7!0!} = \frac{8}{1!} = 8$$

c) There is only one sequence that is all heads. Formally, we have to choose 0 spots in the sequence to be heads.
n=8, r=0

$$\boldsymbol{C}(\mathbf{8}, \mathbf{0}) = \frac{8!}{(8\text{-}0)!0!} = \frac{8!}{8!0!} = 1$$

d) We have to choose 4 spots in the sequence to be heads.
n=8, r=4

$$\boldsymbol{C}(\mathbf{8}, 4) = \frac{4!}{(8-4)!4!} = \frac{8 \times 7 \times 6 \times 5}{4 \times 3 \times 2 \times 1}$$

Cancelling we have

$$= \frac{8 \times 7 \times 6 \times 5}{4 \times 3 \times 2 \times 1} = 7 \times 2 \times 5 = \mathbf{70}$$

Note: The number of sequences of length n with exactly r heads is given by C(n,r). Each of these is mutually exclusive so they add to 2^n.

$$\boldsymbol{C(n,0) + C(n,1) + C(n,2) + \cdots + C(n,n-1) + c(n,n) = 2^n}$$

Problem 2) You are going on a trip and want to choose 5 books to take with you from your library which has 2001 books. Your library does not have any duplicates.

The size of the set you are selecting from is n=2001. You are selecting r=5 objects. This is a combination.

$$C(2001,5) = \frac{2001!}{(2001-5)!5!} = \frac{2001!}{(1996)!5!} = \frac{2001 \times 2000 \times 1999 \times 1998 \times 1997}{5 \times 4 \times 3 \times 2 \times 1}$$

$$= 667 \times 50 \times 1999 \times 1998 \times 1997$$

$$= 266{,}000{,}333{,}499{,}900$$

Problem 3) You are going on a trip that will last 14 days. You have mini tubes of toothpaste which come in three flavors regular, mint, and cherry. How many ways can you choose 14 mini-tubes?

The size of the set you are selecting from is n=3. You are selecting r=14 objects. This is a combination with repetition. (Note that we must have repeat choices.) So we use C(n-1+r, r) to compute the count.

$$\boldsymbol{C}(3-1+14,14) = C(16,14)$$

$$= \frac{16!}{(16\text{-}14)!14!} = \frac{16!}{(2)!14!} = \frac{16 \times 15}{2 \times 1}$$

$$= 8 \times 15$$

$$= 120$$

Problem 4) Suppose that you have 20 people and are going to form team of 4. How many ways are there to form the teams? When selecting a team, the order of selection doesn't matter.

Task 1: Form the first team: 20 choose 4
Task 2: Form the second team: 16 choose 4
Task 3: Form the second team: 12 choose 4
Task 4: Form the second team: 8 choose 4
Task 5: Form the second team: 4 choose 4

$$C(20,4) = \frac{20!}{16!\,4!} = \frac{20 \times 19 \times 18 \times 17}{4 \times 3 \times 2 \times 1} = 5 \times 19 \times 3 \times 17$$
= 4,845

$$C(16,4) = \frac{16}{12!\,4!} = \frac{16 \times 15 \times 14 \times 13}{4 \times 3 \times 2 \times 1} = 4 \times 5 \times 7 \times 13$$
= 1,820

$$C(12,4) = \frac{12}{8!\,4!} = \frac{12 \times 11 \times 10 \times 9}{4 \times 3 \times 2 \times 1} = 11 \times 5 \times 9$$
= 495

$$C(8,4) = \frac{8!}{4!\,4!} = \frac{8 \times 7 \times 6 \times 5}{4 \times 3 \times 2 \times 1} = 2 \times 7 \times 5$$
= 70

$$C(2,4) = \frac{4!}{0!\,4!} = \frac{4 \times 3 \times 2 \times 1}{4 \times 3 \times 2 \times 1}$$
= 1

Total is 4845 × 1820 × 495 × 70 × 1
= 305,540,235,000

Problem 5) Suppose that you have 2m items and you are going to select m. If order doesn't matter and repetition is allowed, how many ways are there to make the selection? Express this using big O notation.

In this case n=2m and r=m.
The number is C(2m-1+m, m) = C(3m-1, m) = $\frac{(3m-1)!}{(2m-1)!(m)!}$

In big O notation, = $\frac{(3m-1) \times (3m-2) \times \ldots \times (2m)}{m \times (m-1) \times \ldots \times 1}$ is approximately
= (2m) × (2m) ×…× (2m)
Is $O((2m)^m)$
Is $O(2^m m^m)$
This is clearly worse than exponential. This kind of growth is an example of combinatorial growth.

Why to Count:

1) **You want to determine the complexity of an algorithm.**
2) **You want to determine a probability.**

Complexity:

There are a number of algorithms that use a brute force approach. Try all possible solutions and see if any of them work or find the best solution. In those cases, we can use a counting argument to find the complexity.

Example: Suppose that we have a graph with n nodes. Is there a fully connected subgraph?

Problem: Given k nodes, how long does it take to determine if they are fully connected? We need to select every pair of nodes and see if there is an edge. We don't care about the order of the nodes, so this is k choose 2.

$$C(k,2) = \frac{k(k-1)}{2} = O(k^2)$$

In our particular case, since we have a fixed value of k, we will treat this as O(1)

Algorithm 1: Use four nested loops:

```
for n1 in nodes:     O(n)
     for n2 in nodes:     O(n)
          if n2 == n1: skip    O(1)
          for n3 in nodes:          O(n)
               if n3==n1 or n3==n2: skip      O(1)
               for n4 in nodes:     O(n)
                    if n4==n1 or n4==n2 or n4==n3: skip
                    check(n1, n2, n3, n4)     O(1)
```

So we get $O(n^4)$, but this algorithm isn't very efficient. It will try sets of nodes many times. We can improve the algorithm.

```
for (n1, n2, n3, n4) in generate(nodes, 4):
     check(n1, n2, n3, n4)
```

It is possible to efficiently generate all the subsets of a given size of which we know that there are n choose 4 of them.

$$C(n,4) = \frac{n(n-1)(n-2)(n-3)}{4\times3\times2\times1} \approx \frac{n^4}{4!} = O(n^4)$$

So, even though this algorithm is faster, it has the same complexity (order or growth as our previous algorithm.) We definitely prefer it to the previous algorithm, but would prefer to find something that is algorithmically faster if possible.

Probability:

In many situations, we can compute probabilities through the use of counting. In order to use counting in this way, we need to have elementary events that are equally likely.

Step 1: Count all the possible elementary events.
Step 2: Count the elementary events that you are interested in.
Step 3: Divide to find the probability in the range of 0 to 1.

Situation: We have a single random event

Examples:

Roll a die one time.
Flip a coin.
Spin a roulette wheel.

Problem: What is the probability that the result on a six-sided die is prime?

6 events are {1, 2, 3, 4, 5, 6}
Prime events are {2, 3, 5}
Probability is 3/6 = 0.5

Situation: We have a sequence of independent random events.

Examples:

Roll a die two times.
Flip a coin 10 times.
Spin a roulette wheel four times.

Problem: What is the probability that if we roll a die twice that the sum is divisible by 3.

This is a sequence with repetition so we have 6x6 = 36 events.

The easiest way to count the sequences that are divisible by 3 is to use a table with each square representing one of the 36 sequences and then mark the squares where the sum is divisible by 3 with an x.

	1	2	3	4	5	6
1		x			x	
2	x			x		
3			x			x
4		x			x	
5	x			x		
6			x			x

There are 12 marks so the probability is 12/36 =1/3.

Problem: What is the probability that if we roll two indistinguishable dice that the sum is divisible by 3.

Thought experiment… if we put a tiny dot on one of the dice and two dots on the other die, we could sequence the results…. This should not change the probability.

In fact, this helps us. If we try to use combinations with repetition for the counting, we are selecting from n=6 with r=2 times with repetition is

$$C(6-1+2,2) \; = \; C(7,2) = \frac{7\times 6}{2\times 1} = 21$$

And they are (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)

The problem for using these to compute probabilities is that they are not all equally likely. Each of the possibilities where the dice match are half as likely as the other results.

The standard approach is to treat multiple dice as if they were a sequence of dice throws and then count the sequences.

Situation: We are selecting unique items from a larger set and order does not matter

Examples:

Dealing a poker hand from a deck

Picking numbers for a lottery

Problem: Suppose that we pick 5 numbers from 1 to 75. What is the probability that we match for the jackpot?

There are 75 from which we choose 5 without repetition.

$C(75,5) = \frac{75\times74\times73\times72\times71}{5\times4\times3\times2\times1} = 5 \times 74 \times 73 \times 9 \times 71$

= 17,259,390

There is exactly one choice where we match all 5 numbers so the probability of the jackpot is about 1 in 17 million. This is pretty small. The sad thing is that there are lotteries out there that have even lower probabilities of winning the jackpot. The big lotteries are in the range of 1 in 200 to 300 million. If we play once a week, we have about a 1 in 4 million chance of winning. For comparison, about 25 people are killed per year in the USA due to lightning or about a 1 in 12 million chance. The chance of being hit by lightning is about 10 times more likely at 1 in 1million.

While we can count permutations or combinations in these cases, it is almost always easier to do the counting with combinations.

Situation: We are selecting unique items from a larger set and order does not matter

Examples:

Dealing a blackjack from a deck

Calling bingo numbers.

Problem: What is the probability that the first 5 numbers called are 1, 2, 3, 4, and 5?

There are 75 numbers, so there are 75! Permutations.

If the first 5 numbers are 1 through 5, there are 5! permutations, the remaining numbers can be arranged in 70! Permutations, so the total is 5! Times 70!.

The probability is 5! Times 70! Divided by 75!.