CS 237: Probability in Computing

Wayne Snyder
Computer Science Department
Boston University

Lecture 5:

- Conditional Probability concluded: The Law of Total Probability
- Counting principles and combinatorics;
  - Counting considered as sampling and constructing outcomes; selection with and without replacement;
  - Counting sequences:
    - Enumerations and Cross-products;
    - Permutations;
    - K-Permutations
    - Permutations with Duplicates
    - Circular Permutations

Law of Total Probability

Conditional Probability addresses the probability of one event in the context of one other event. Sometimes, it gets more complicated, because there are many different events. The **Law of Total Probability** can help!

Consider this problem (from the textbook):

I have three bags that each contain 100 marbles:

Bag 1 has 75 red and 25 blue marbles;

Bag 2 has 60 red and 40 blue marbles;

Bag 3 has 45 red and 55 blue marbles;

I choose a bag at random, and then a marble at random from that bag. What is the probability that I get a red marble?

Events: A = Marble is red

B1 = Bag 1 was chosen

B2 = Bag 2 was chosen in the first step

B3 = Bag 3 was chosen

Law of Total Probability

I have three bags that each contain 100 marbles:

Bag 1 has 75 red and 25 blue marbles;

Bag 2 has 60 red and 40 blue marbles;

Bag 3 has 45 red and 55 blue marbles;

I choose a bag at random, and then a marble at random from that bag. What is the probability that I get a red marble?

Events: A = Marble is red   B2 = Bag 2 was chosen in the first step

B1 = Bag 1 was chosen   B3 = Bag 3 was chosen

$P(A \mid B1) = 75/100 = 0.75$ $P(A \mid B2) = 60/10 = 0.6$

$P(A \mid B3) = 45/100 = 0.45$

Law of Total Probability

The Law of Total Probability:

$$P(A) = \sum_i P(A \cap B_i) = \sum_i P(A|B_i)P(B_i).$$

This is essentially conditional probability where the conditions form a partition of the sample space.

Or, simply think of it as "case analysis." (you've probably be doing this already without calling it anything special!).

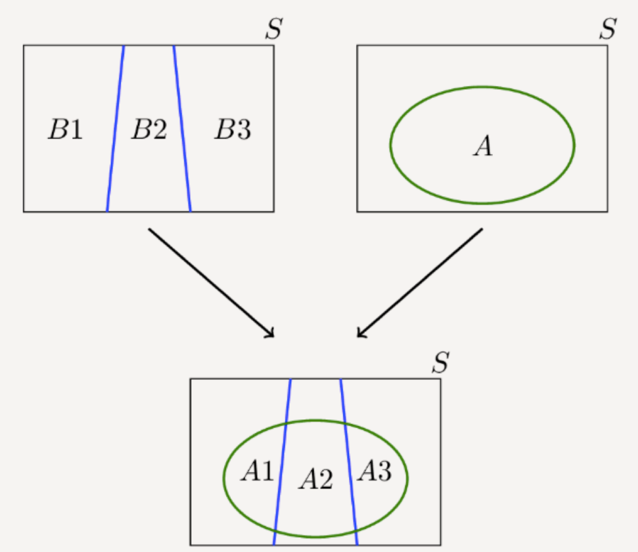


Fig.1.24 - Law of total probability.

Law of Total Probability

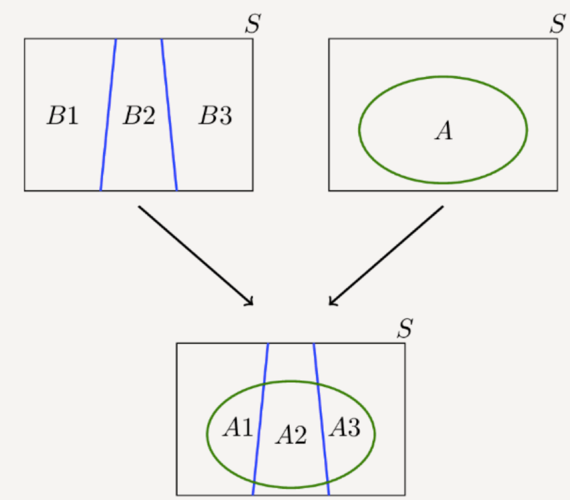


Fig.1.24 - Law of total probability.

I have three bags that each contain 100 marbles:

Bag 1 has 75 red and 25 blue marbles;

Bag 2 has 60 red and 40 blue marbles;

Bag 3 has 45 red and 55 blue marbles;

I choose a bag at random, and then a marble at random from that bag. What is the probability that I get a red marble?

Events: A = Marble is red

B1 = Bag 1 was chosen

B2 = Bag 2 was chosen in the first step

B3 = Bag 3 was chosen

$P(A \mid B1) = 75/100 = 0.75$

$P(A \mid B2) = 60/10 = 0.6$

$P(A \mid B3) = 45/100 = 0.45$

$P(A) = P(A \mid B1) * P(B1) + P(A \mid B2) * P(B2) + P(A \mid B3) * P(B3)$

$= 0.74 / 3 + 0.6 / 3 + 0.45 / 3 = 0.6$

Base Rate Fallacy

Conditional Probability is an excellent tool for evaluating what can happen under various conditions, but it is sensitive to extreme conditions.

A group of police officers have breathalyzers displaying false drunkenness in 5% of the cases in which the driver is sober. However, the breathalyzers never fail to detect a truly drunk person. One in a thousand drivers is driving drunk. Suppose the police officers then stop a driver at random, and force the driver to take a breathalyzer test. It indicates that the driver is drunk. We assume you don't know anything else about him or her.

What is the probability he or she really IS drunk? (-Wikipedia)

95%

50%

20%

2%

Base Rate Fallacy

Conditional Probability is an excellent tool for evaluating what can happen under various conditions, but it is sensitive to extreme conditions.

A group of police officers have breathalyzers displaying false drunkenness in 5% of the cases in which the driver is sober. However, the breathalyzers never fail to detect a truly drunk person. One in a thousand drivers is driving drunk. Suppose the police officers then stop a driver at random, and force the driver to take a breathalyzer test. It indicates that the driver is drunk. We assume you don't know anything else about him or her.

What is the probability he or she really is drunk? (-Wikipedia)

1 driver is drunk, and it is 100% certain that for that driver there is a *true* positive test result, so there is 1 *true* positive test result.

2%

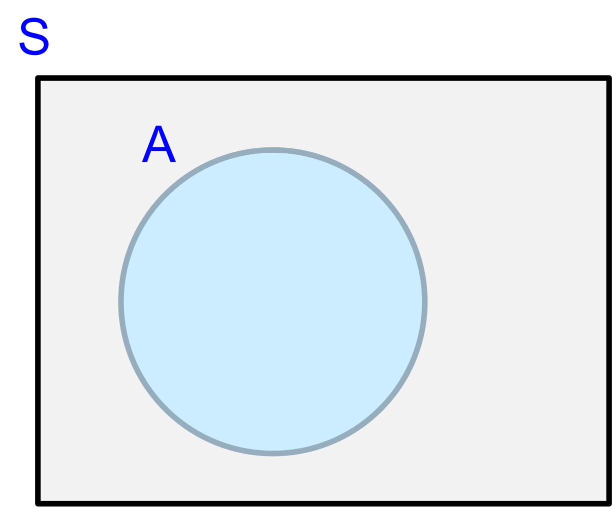
999 drivers are not drunk, and among those drivers there are 5% *false* positive test results, so there are 49.95 % *false* positive test results.

Therefore, the probability that one of the drivers among the 1 + 49.95 = 50.95 positive test results really is drunk is 1/50.95 = 0.0197.

Finite Combinatorics

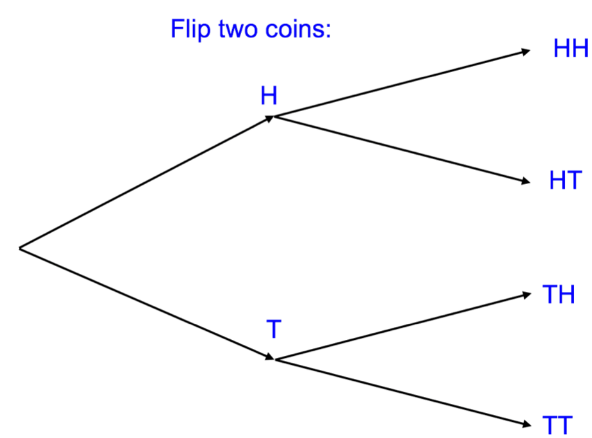
Recall the rule for finite, equiprobable probability spaces:

$$P(A) = \frac{|A|}{|S|}$$



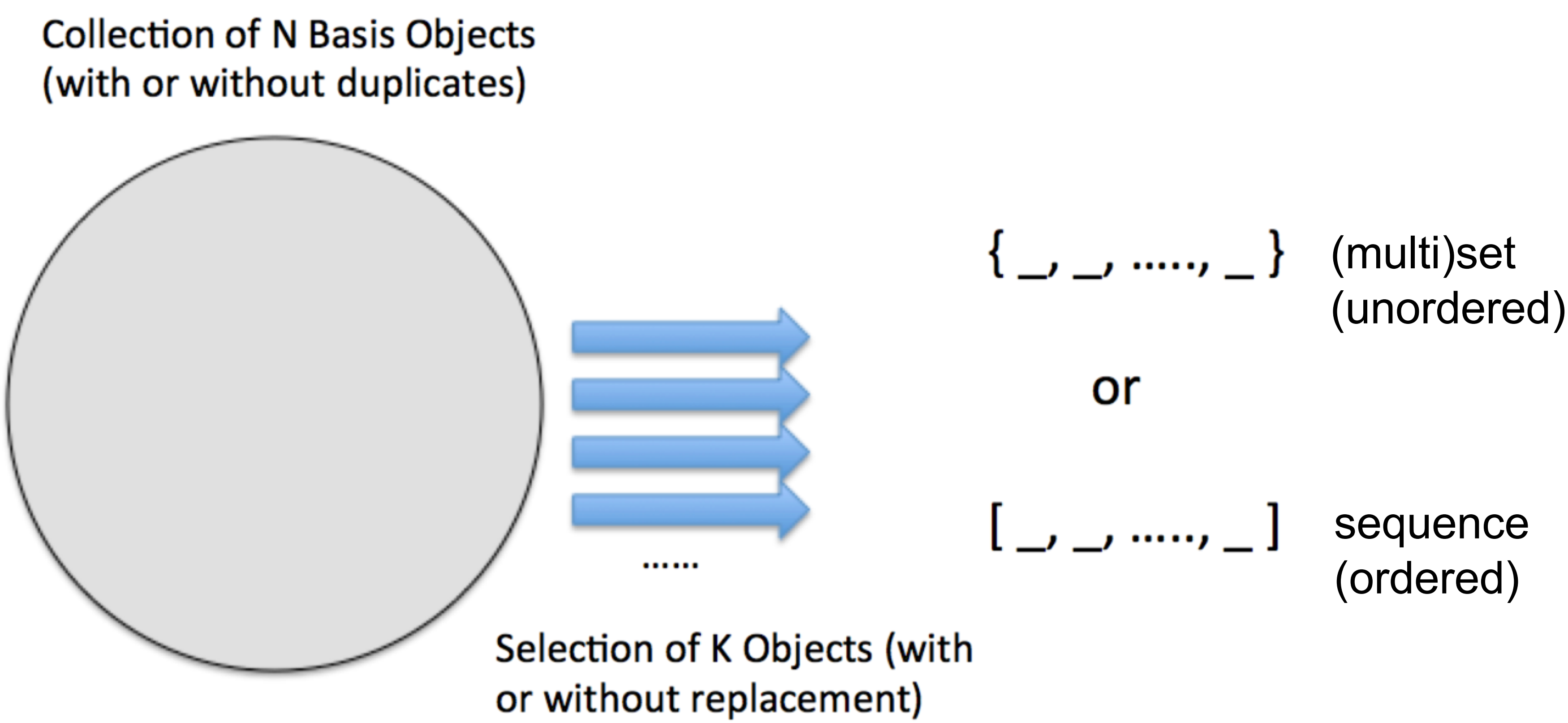
To work with this definition, we will need to calculate the number of elements in A and S and we will analyze this according to how we "constructed" the sample points in S and in A during the random experiment.

Compare with how we "construct" the sample space using a tree diagram!

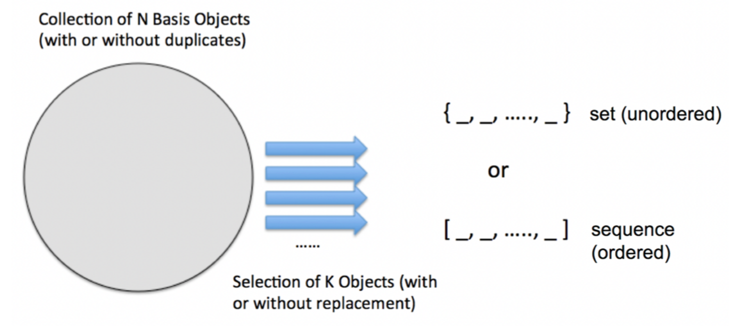


Finite Combinatorics

The way in which we "construct" the sample space almost always follows what we might characterize as a sampling process:



Finite Combinatorics



The important issues to note are (and you
Should them out in this order):

(A) Is the selection done with or without replacement?

Examples of with replacement:

How many **enumerations** of .....

Choose letters for a password…

Flip a coin or toss a die ….

Examples of without replacement:

How many **permutations** of .....

Choose a committee of 5 people from a group of 100 people…..

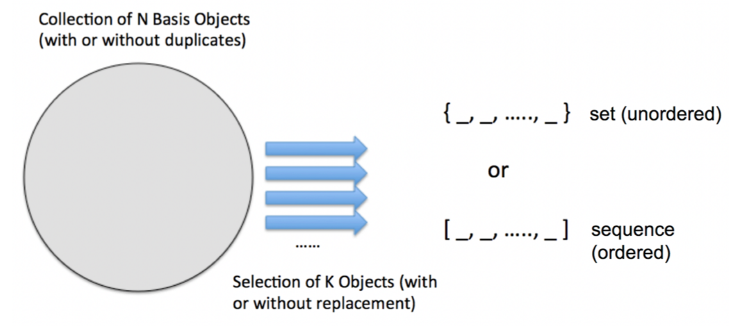
Deal five cards for a hand of Poker…..

Choose letters for a password, with no repeated letters…..

When the issue might be unclear, the problem statement will specify, e.g.,

"Suppose you have a bag of 3 blue and 2 red balls and you choose 2 with replacement… "

Finite Combinatorics



The important issues to note are (and you Should them out in this order):

(A) Is the outcome ordered or unordered?

Ordered outcomes are **sequences:**

Enumerations and Permutations

Strings of characters

Rows of seats

Unordered outcomes are **sets (no duplicates):**

Hands in card games // these are Combinations, covered next lecture!

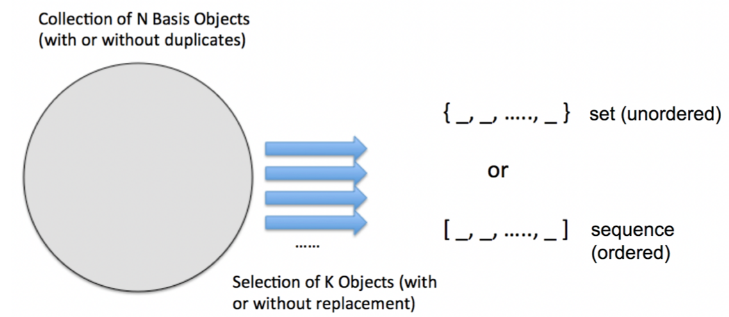
Groups of people

When it might be unclear, the problem statement will say something specific about what you are creating:

"How many sequences of ...." "How many permutations of ...

"Two

Finite Combinatorics



The important issues to note are (and you probably want to figure them out in this order):

(i) Is the selection done with or without replacement?

Examples:

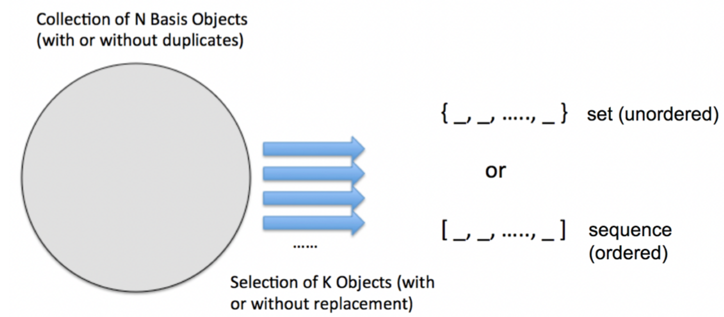
Selecting 5 cards for a poker hand is done without replacement (you keep the cards and don't put them back in the deck); choosing a committee of 3 from a group of 10 people is without replacement; in many cases, as with balls in a sack, it is part of the problem statement.

(ii) Is the outcome ordered or unordered?

Examples: If the two numbers showing on the dice are 2 and 5, put them in sequence [2,5] (duplicates allowed); put the 5 letters into a word (a sequence); put the 5 cards into a hand (a set).

Note: in the case of [2,5] you may speak of the "first roll" or the "second roll" but in { 2,5 } you may only make statements about the collection without specifying an order ("at least one of the rolls is 5"). Words are sequences, and hands in card games are sets; otherwise you need to use context or it will be clear from the problem statement.

Finite Combinatorics



We will organize this along the dimensions of

- ordered vs unordered and
- selection with replacement vs without.

and we will consider the role of duplicates when appropriate.

These problems have names you should be familiar with from CS 131:

	Selection Without Replacement	Selection With Replacement
Ordered Outcome (Sequence or String)	**Permutations**	**Enumerations**
Unordered Outcome (Set or Multiset)	**Combinations**	(We will not study this possibility...)

For each of these I will provide a canonical problem to illustrate; I STRONGLY recommend you memorize these problems and the solution formulae, and when you see a new problem, try to translate it into one of the canonical problems.

Finite Combinatorics

	Selection Without Replacement	Selection With Replacement
Ordered Outcome (Sequence or String)	Permutations	Enumerations
Unordered Outcome (Set or Multiset)	Combinations	(We will not study this possibility...)

Enumerations

The simplest situation is where we are constructing a sequence with replacement, such as where the basis objects are literally replaced, or consist of information such as symbols, which can be copied without eliminating the original.

Canonical Problem: You have N letters to choose from; how many words of K letters are there?

Formula: N^K

Example: How many 10-letter words all in lower case? 26^{10}

A more general version of this involves counting cross-products:

Generalized Enumerations: Suppose you have K sets $S_1, S_2, \ldots, S_k$. What is the size of the cross-product
$S_1 \times S_2 \times \ldots \times S_k$?

Solution: $|S_1| * |S_2| * \ldots * |S_k|$

Example: Part numbers for widgets consist of 3 upper-case letters followed by 2 digits. How many possible part numbers are there? $26*26*26*10*10 = 1{,}757{,}600$

Finite Combinatorics

	Selection Without Replacement	Selection With Replacement
Ordered Outcome (Sequence or String)	Permutations	Enumerations
Unordered Outcome (Set or Multiset)	Combinations	(We will not study this possibility...)

Permutations

Next in order of difficulty (and not yet very difficult) are permutations, where you are constructing a sequence, but without replacement. This explains what happens when the basis set is some physical collection which can not (like letters) simply be copied from one place to another.

The most basis form of permutation is simply a rearrangement of a sequence into a different order. The number of such permutations of N objects is denoted **P(N,N).**

Canonical Problem 1(a): Suppose you have N students S_1, S_2, ..., S_n. In how many ways can they ALL be arranged in a sequence in N chairs?

Formula: P(N,N) = N* (N-1) * ... * 1 = N!

Example: How many permutations of the word "COMPUTER" are there?

Answer: 8! = 40,320

Finite Combinatorics

	Selection Without Replacement	Selection With Replacement
Ordered Outcome (Sequence or String)	Permutations	Enumerations
Unordered Outcome (Set or Multiset)	Combinations	(We will not study this possibility...)

K-Permutations

If we do not simply rearrange all N objects, but consider selecting K <= N of them, and arranging these K, we have a K-Permutation indicated by P(N,K).

Canonical Problem 1(b): Suppose you have N students S_1, S_2, ..., S_n. In how many ways can K of them be arranged in a sequence in K chairs?

Formula:

$$P(N,K) = \overbrace{N * (N-1) * \cdots * (N-K+1)}^{\text{K terms}} = \frac{N * (N-1) * \cdots * (N-K+1) * (N-K) * \cdots * 1}{(N-K) * \cdots * 1} = \frac{N!}{(N-K)!}$$

Example: How many passwords of 8 lower-case letters and digits can be made, if you are not allowed to repeat a letter or a digit?

Answer: The "not allowed to repeat" means essentially that you are doing this "without replacement." So we have P(36,8) = 36! / 28! = 1,220,096,908,800.

Note: The usual formula at the extreme right is extremely inefficient. The first formula is the most efficient, if not the shortest to write down!

$$P(N,K) = N * (N-1) * \cdots * (N-K+1)$$

Finite Combinatorics

Counting With and Without Order

	Selection Without Replacement	Selection With Replacement
Ordered Outcome (Sequence or String)	Permutations	Enumerations
Unordered Outcome (Set or Multiset)	Combinations	(We will not study this possibility...)

Before we discuss combinations, let us first consider the relationship between ordered sequences and unordered collections (sets or multisets)
For example, consider a set

Set = unordered, no duplicates

$$A = \{ S, E, T \}$$

of 3 letters (all distinct). Obviously there is only one such set.

But there are 3! = 6 different sequences (=permutations) of all these letters:

S E T
S T E
E S T
E T S
T S E
T E S

Finite Combinatorics

	Selection Without Replacement	Selection With Replacement
Ordered Outcome (Sequence or String)	Permutations	Enumerations
Unordered Outcome (Set or Multiset)	Combinations	(We will not study this possibility...)

The Ordering Principle

If A is an unordered collections (set) consisting of N distinct elements, then there are N! ordered collections (sequences) of A.

Question: If A = { S, E, T }, how many sets of 2 distinct letters can we choose from A? Note: N = 2.

Answer: Hm…. Let's just count: { S, E }, { S, T }, { E, T }… there are M = 3.

Question: How many sequences of two distinct letters can we choose from A?

Answer: Again, let's just count:
All orderings of { S, E } gives us SE, ES // 2! ways to order each
All orderings of { S, T} gives us ST, TS
All orderings of { E, T } gives us ET, TE

So: there are 3*2! = 6 possible sequences derived from these three sets.

Finite Combinatorics

The Unordering Principle

	Selection Without Replacement	Selection With Replacement
Ordered Outcome (Sequence or String)	Permutations	Enumerations
Unordered Outcome (Set or Multiset)	Combinations	(We will not study this possibility...)

If there are M ordered collections (sequences) of the N elements in A, then there are M/N! unordered collections (sets) of A.

When all elements are distinct, as in our previous example, then obviously, M/N! = N!/N! = 1.

The basic idea here is that we are correcting for the overcounting when we assumed that the ordering mattered. Therefore we divide by the number of permutations.

This principle also applies to only a part of the collection:

Example: Suppose we have 4 girls and 5 boys, and we want to arrange them in 9 chairs, but we do not care what order the girls are in. How many such arrangements are there?

Answer: There 9! permutations, but if we do not care about the order of the (sub)collection of 4 girls, then there are 9!/4! = 15,120 such sequences.

	Selection Without Replacement	Selection With Replacement
Ordered Outcome (Sequence or String)	Permutations	Enumerations
Unordered Outcome (Set or Multiset)	Combinations	(We will not study this possibility...)

Finite Combinatorics

Permutations with Repetitions

As another example of the Unordering Principle, let us consider what happens if you want to form a permutation P(N,N), but the N objects are not all distinct. An example may clarify:

Example: How many distinct (different looking) permutations of the word "FOO" are there?

If we simply list all 3! = 6 permutations, we observe that because the 'O' is duplicated, and we can not tell the difference between two occurrences of 'O's, there are really only 3 distinct permutations. This should be clear if we distinguish the two occurrences of 'O' with subscripts:

$F\,O_1O_2$	FOO	FOO
$F\,O_2O_1$	FOO	
$O_1F\,O_2$	OFO	OFO
O_1O_2F	OOF	OOF
$O_2F\,O_1$	OFO	
O_2O_1F	OOF	

Sequences: O_1O_2
O_1O_2

Multiset: { O, O }

There are 2! sequences, so 6/2! = 6/2 = 3.

Finite Combinatorics

	Selection Without Replacement	Selection With Replacement
Ordered Outcome (Sequence or String)	Permutations	Enumerations
Unordered Outcome (Set or Multiset)	Combinations	(We will not study this possibility...)

Permutations with Repetitions

If you have N (non-distinct) elements, consisting of m (distinct) elements with multiplicities $K_1, K_2, \ldots, K_m$, that is, $K_1 + K_2 + \ldots + K_m = N$, then the number of distinct permutations of the N elements is

$$\frac{N!}{K_1! * K_2! * \cdots K_m!}$$

Example: How many distinct (different looking) permutations of the word "MISSISSIPPI" are there?

Solution: There are 11 letters, with multiplicities:

M: 1
I: 4
S: 4
P: 2

Therefore the answer is $$\frac{11!}{1! * 4! * 4! * 2!} = \frac{39,916,800}{1 * 24 * 24 * 2} = 34,650$$

Finite Combinatorics

Circular Permutations

	Selection Without Replacement	Selection With Replacement
Ordered Outcome (Sequence or String)	Permutations	Enumerations
Unordered Outcome (Set or Multiset)	Combinations	(We will not study this possibility...)

A related idea is permutations of elements arranged in a circle. The issue here is that (by the physical arrangement in a circle) we do not care about the exact position of each elements, but only "who is next to whom." Therefore, we have to correct for the overcounting by dividing by the number of possible rotations around the circle.

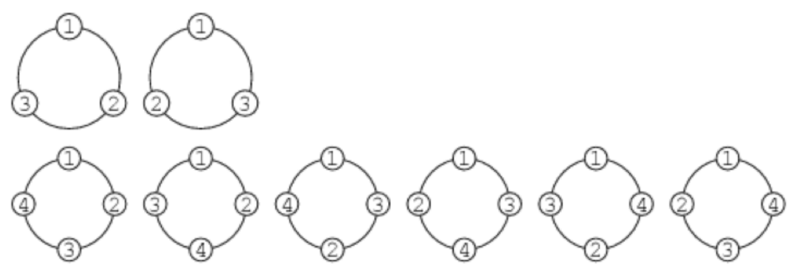
Example: There are 6 guests to be seated at a circular table. How many arrangements of the guests are there?

Hint: The idea here is that if everyone moved to the left one seat, the arrangement would be the same; it only matters who is sitting next to whom. So we must factor out the rotations. For N guests, there are N rotations of every permutation.

Solution: There are 6! permutations of the guests, but for any permutation, there are 6 others in which the same guests sit next to the same people, just in different rotations.

Formula: There are $$\frac{N!}{N} = (N-1)!$$

circular permutations of N distinct objects.



Finite Combinatorics

Application of Enumerations and Permutations

The Birthday Problem: What is the probability that at least two students in a class of size K have the same birthday? Assume all birthdays are equally likely throughout the year and each year has 365 days.

Our class has 120 students. What is the probability that two people in the class have the same birthday?

Finite Combinatorics

Application of Enumerations and Permutations

	Selection Without Replacement	Selection With Replacement
Ordered Outcome (Sequence or String)	Permutations	Enumerations
Unordered Outcome (Set or Multiset)	Combinations	(We will not study this possibility...)

The Birthday Problem: What is the probability that at least two students in a class of size K have the same birthday? Assume all birthdays are equally likely throughout the year and each year has 365 days.

Solution: There are 365 possibilities for each student. Thus, the sample space has 365^K points (it is an enumeration!). The possibility that no two students share a birthday is P(365,K) (it is a K-permutation).

Using the inverse method, we compute $$1.0 - \frac{P(365,K)}{365^K}$$

For K = 120 (our class), we have

$$1.0 - \frac{\begin{array}{l}5988035315899628681690785355501387087938634583409444426238157897061547078707911888\\7477106916446498086324568862301736636690983217378682095824395749333107762215569645\\0374414943397638317625086729731699993749587824201991495317862940613896478540378931\\2000000000000000000000000000000\end{array}}{\begin{array}{l}2986370829311944964358274839590110328083609355668702859405175850733854777961555956\\8756410421132727649968822828558979444400029791296101007363166254729362144213820917\\5498009240750978365040858616868848384987386084028835479982958096136133239782838523\\50408358032298733312527583283468857189291156828403472900390625\end{array}} = 0.99999999975608\ldots$$