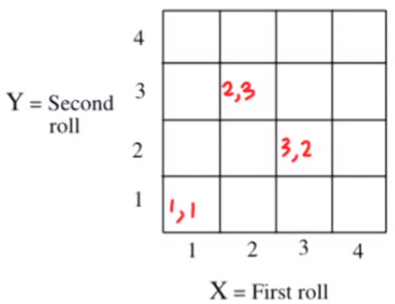
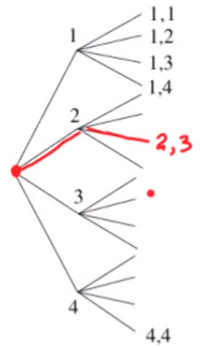
Introduction to probability

Sample space is a set describes the outcomes and the likelihood of occurring each of them
Sample space can be discrete and finite as any set

Ex: Two consecutive rolls of A tetrahedral (4 faces) die
We have 16 possible outcomes

discrete sample space

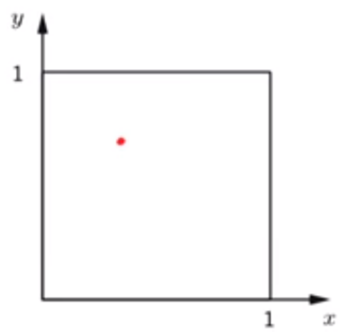




To calculate the probability of any event we use Sequential description (tree)

continuous sample space

- (x, y) such that $0 \leq x, y \leq 1$



Finite discrete sample space	Infinite continuous sample
means that I have a definite number of outcomes	A continuous sample space is based on the same principles, **but it has an infinite number of items in the space**.

Event :subset of sample space that we assign the probability of a certain element (or a group of elements) to it

AXIOMS

- COMPLEMENT :A'+A=1
- Probability of SAMPLE SPACE occurring = 1
- Disjoint events :P(A ∪ B)= P(A)+P(B)
- IF P(A∩B) ≠ 0 :P(A ∪ B)= P(A)+P(B)-P(A∩B)

Python and probability

A Bernoulli trial is one of the simplest experiments you can conduct. It's an experiment where you can have one of two possible outcomes.

Dependent vs independent

Probability of Compound Events

Independent Events

$$P(A \text{ and } B) = P(A) \times P(B)$$

Dependent Events

$$P(A \text{ and } B) = P(A) \times P(B \mid A)$$

Mutually Exclusive

$$P(A \text{ or } B) = P(A) + P(B)$$

Mutually Inclusive

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

Distribution ?

A probability distribution is **a statistical function that describes all the possible values and likelihoods that a random variable can take within a given range**
Types of distributions

- Bernoulli Distribution.
- Uniform Distribution.
- Binomial Distribution.
- Normal Distribution.
- Poisson Distribution.
- Exponential Distribution

Binomial distribution vs normal distribution

Binomial	normal
Binomial distribution describes the distribution of binary data from a finite sample. Thus it gives the probability of getting r events out of n trials from bernoulli trial	Normal distribution describes continuous data which have a symmetric distribution, with a characteristic 'bell' shape.

Probability mass functions (discrete variables)

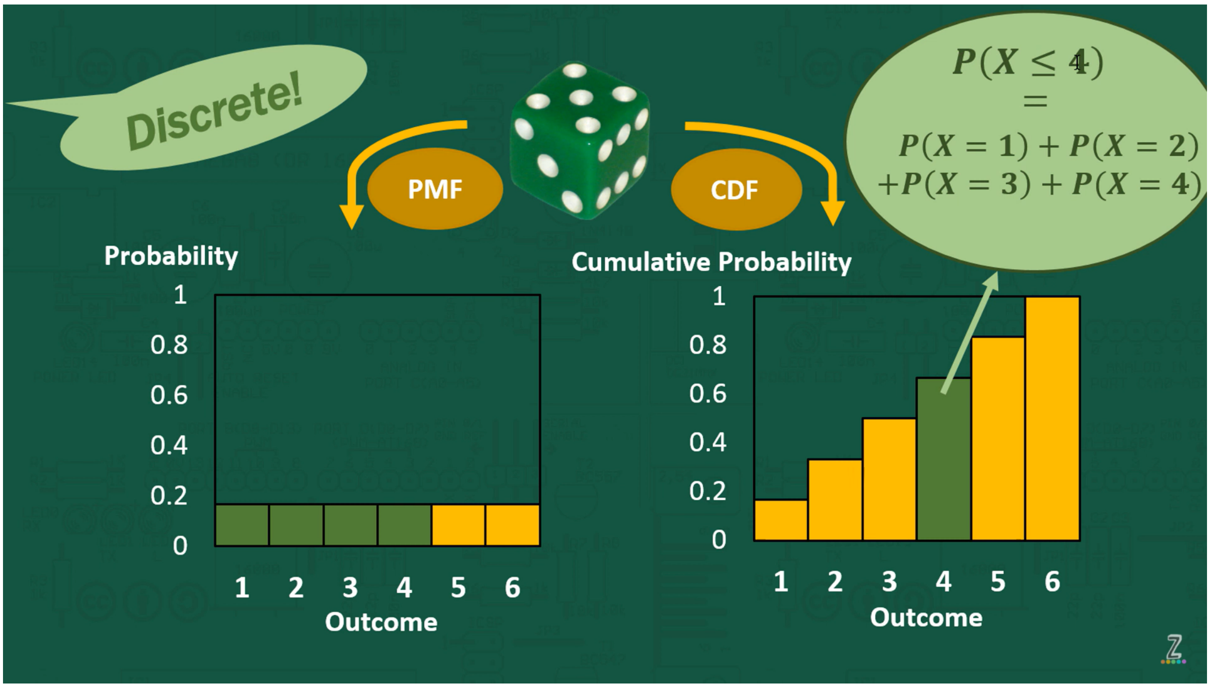
Probability mass function :A discrete random variable has a finite number of possible outcomes. The probability mass function allows you to calculate the probability of getting a particular outcome for a discrete random variable. The binomial probability mass function allows you to calculate the probability of getting k heads from n coin flips with p probability of getting heads.

We import the `binom` object from `scipy.stats`.

- `binom.pmf()` calculates the probability of having exactly `k` heads out of `n` coin flips
- `binom.cdf()` calculates the probability of having `k` heads or less out of `n` coin flips.
- `binom.sf()` calculates the probability of having more than `k` heads out of `n` coin flips.

    `Cdf is the complement of survival function`

Cumulative distribution function (discrete and continuous)



Counting Principles

Permutation

Basically Permutation is an arrangement of objects in a particular way or order. While dealing with permutation one should be concerned about the selection as well as arrangement. In Short, ordering is very much essential in permutations. In other words, the permutation is considered as an ordered combination.

For example:

- If we have a licence plate with 2 letters and 3 digits

With repetition 26;26;10;10;10
Without repetition 26;25;10;9;8
i= stage
Product of the number of choices we have in each stage is the counting principle

- billiard balls

when we want to select **all** of the billiard balls the permutations are:

16! = 20,922,789,888,000

n(n-1)(n-2)........1= n!

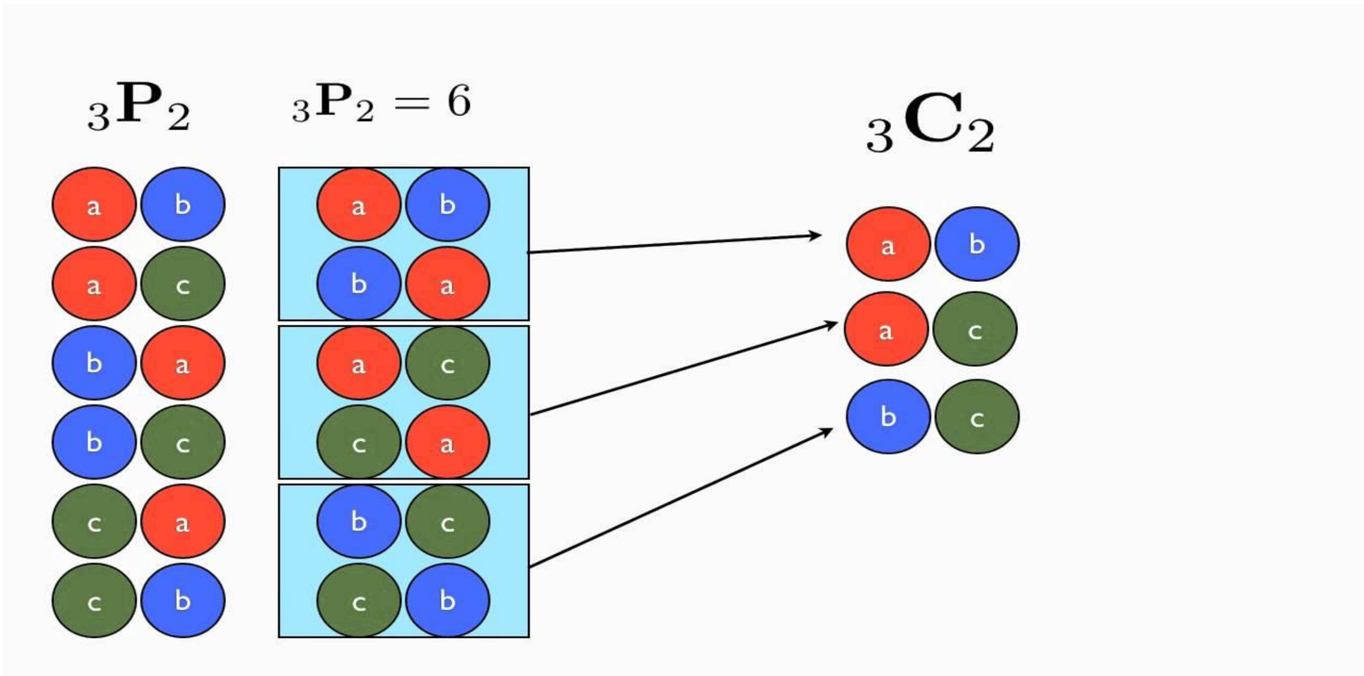
n	n-1	n-2	…...	1

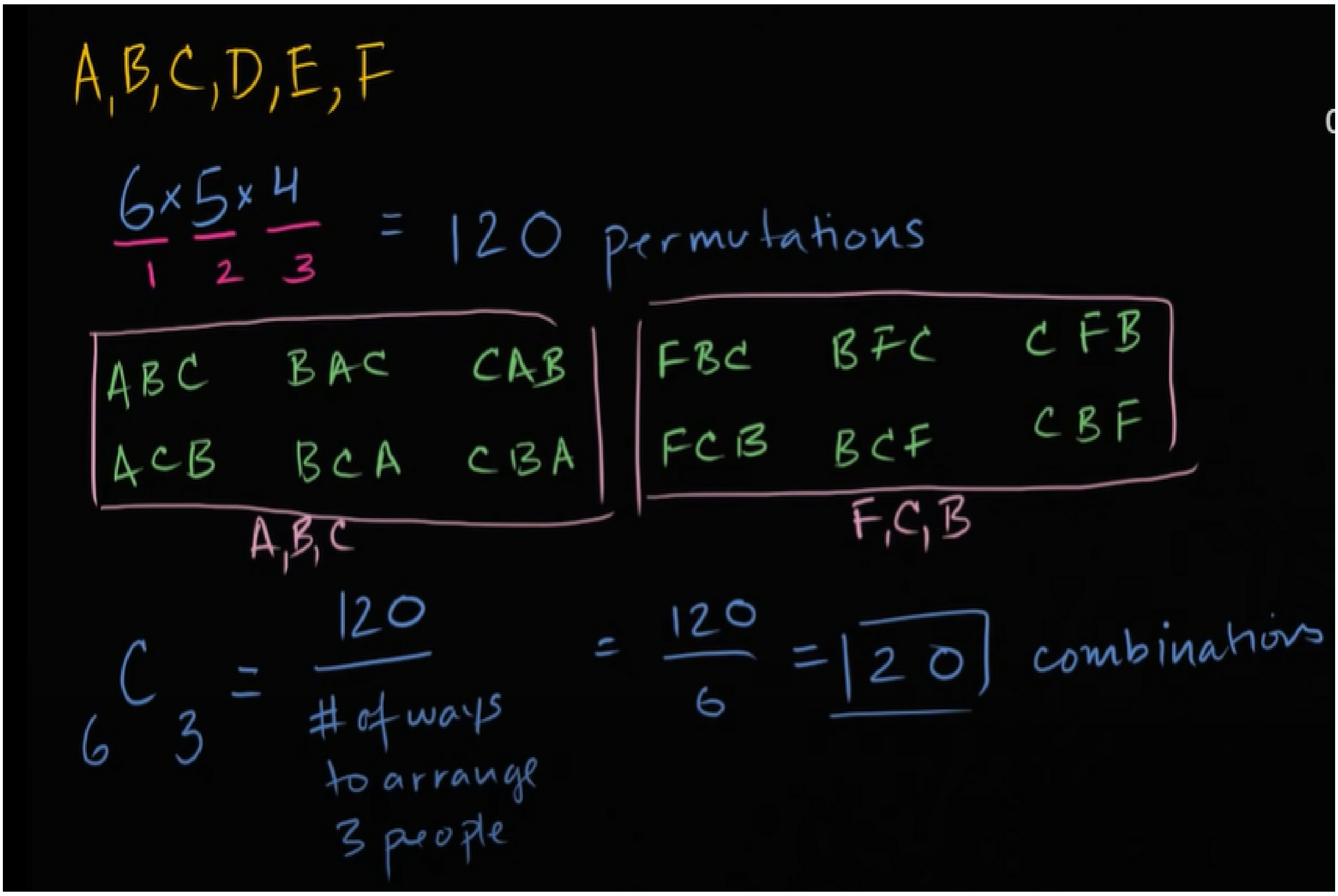
But when we want to select just 3 we don't want to multiply after 14. How do we do that? There is a neat trick: we divide by **13!**

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

combination

The combination is a way of selecting items from a collection, such that (unlike permutations) the order of selection does not matter. In smaller cases, it is possible to count the number of combinations. Combination refers to the combination of n things taken k at a time without repetition. To refer to combinations in which repetition is allowed, the terms k-selection or k-combination with repetition are often used.





In this problem we will cover:

Permutation	**Combination**
Permutation means the selection of objects, where the order of selection matters	**The combination means the selection of objects, in which the order of selection does not matter.**
In other words, it is the arrangement of r objects taken out of n objects.	**In other words, it is the selection of r objects taken out of n objects irrespective of the object arrangement.**
The formula for permutation is ${}_nP_r = n!/(n-r)!$	**The formula for combination is** ${}_nC_r = n!/[r!(n-r)!]$

Enumerating Gene Orders solved by 10688

Aug. 2, 2012, 2 a.m. by Rosalind Team

Topics: Combinatorics, Genome Rearra

←

Rearrangements Power Large-Scale Genomic Changes click to expand

Problem

A **permutation** of length n is an ordering of the positive integers $\{1, 2, \ldots, n\}$. For example, $\pi = (5, 3, 2, 1, 4)$ is a permutation of length 5.

Given: A positive integer $n \leq 7$.

Return: The total number of permutations of length n, followed by a list of all such permutations (in any order).

Sample Dataset

```
3
```

Sample Output

```
6
1 2 3
1 3 2
2 1 3
2 3 1
3 1 2
3 2 1
```

You can access the problem from: http://rosalind.info/problems/perm/

How to solve it ?

1. The main tool we use for this is a library specifically designed to iterate over objects in different ways: itertools
2. the question we need to find out how many tuples are in that tuple. We do this using the Python len tool which returns the length of something: