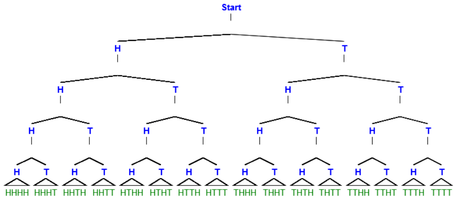
Sample space of sequences: coin

Flip a fair coin n times independently.
Sample space Ω_n: all possible sequences of heads and tails.



$|\Omega_n| = 2^n$.
$\Omega_n = \Omega_1 \times \Omega_1 \times \ldots \times \Omega_1 = \Omega_1^n$.
Ω_1: base sample space of flipping the fair coin once.
Ω_n: hyper sample space of flipping n times independently.
Population of sequences.

Sample space of sequences: die

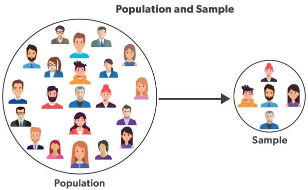
Roll a fair die n times independently.
Sample space Ω_n: all possible sequences of numbers.

	1	2	3	4	5	6
1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

$|\Omega_n| = 6^n$.
$\Omega_n = \Omega_1 \times \Omega_1 \times \ldots \times \Omega_1 = \Omega_1^n$.
Ω_1: base sample space of rolling the fair die once.
Ω_n: hyper sample space of rolling n times independently.
Population of sequences.

Sample space of sequences: population

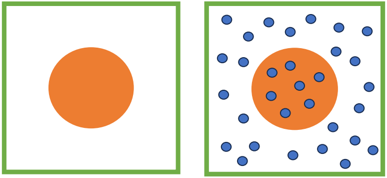
Randomly sample a person from a population of N (e.g., 300 million) people.
Repeat random sampling n (e.g., 1000) times independently.
Sample space Ω_n: all possible sequences of people.



$|\Omega_n| = N^n$ (e.g., $300m^{1000}$).
$\Omega_n = \Omega_1 \times \Omega_1 \times ... \times \Omega_1 = \Omega_1^n$.
Ω_1: base sample space, the population of people.
Ω_n: hyper sample space, the hyper-population of sequences.
Population of sequences.

Sample space of sequences: region

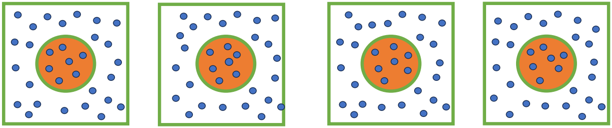
Randomly sample a point from a region.
Repeat the above n times independently.
Sample space Ω_n: all possible sequences of points.



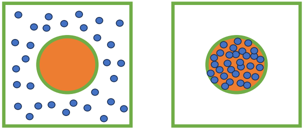
$\Omega_n = \Omega_1 \times \Omega_1 \times ... \times \Omega_1 = \Omega_1^n$.
Ω_1: base sample space, unit square $[0, 1]^2$.
Ω_n: hyper sample space, unit hyper-cube $[0, 1]^{2n}$.
$(x_1, y_1, x_2, y_2, ..., x_n, y_n)$: a point in Ω_n.
Population of sequences.

Population of sequences

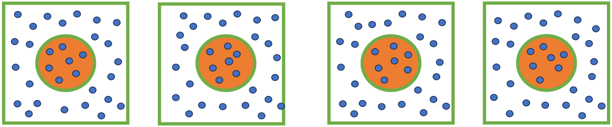
Equally likely outcomes in Ω_1 + independent repetitions
$=$ equally likely sequences in Ω_n.
m: number of times A happens.
m fluctuates over all sequences.
Among all equally likely possibilities, 99.999$\%$ are like below,
where m/n is close to $P(A)$.



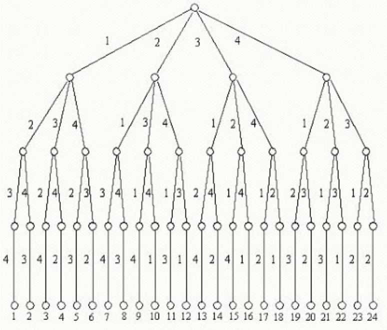
.00000001$\%$ are like below, where m/n are far from $P(A)$.



All sequences in Ω_n are equally likely.
Among all equally likely possibilities, 99.999% are like below, where m/n is close to $P(A)$.



Can prove $P(|\frac{m}{n} - P(A)| \leq .01) \to 1$ as $n \to \infty$.
A representative sequence: $|m(\text{sequence})/n - P(A)| \leq .01$.
A non-representative sequence: $|m/n - P(A)| > .01$.
Among all possible sequences, the proportion of representative sequences $\to 1$ as $n \to \infty$.
(1) Population setting: count the number of sequences.
(2) Region setting: measure the volume of set of sequences.



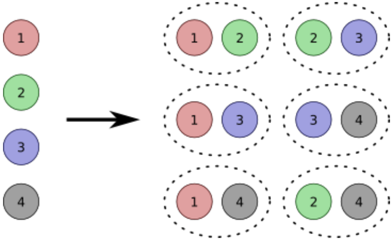
n different cards. Choose k of them. Order matters. Number of different sequences:

$$P_{n,k} = n(n-1)...(n-k+1).\ P_{4,2} = 4 \times 3 = 12.$$

$$P_{n,n} = n!.$$

How many different ways to permute things.

Combination



n different balls. Choose k of them. Order does NOT matters. Number of different combinations:

$$\binom{n}{k} = \frac{P_{n,k}}{k!} = \frac{n(n-1)\ldots(n-k+1)}{k!} = \frac{n!}{k!(n-k)!}.$$

$$\binom{4}{2} = \frac{4 \times 3}{2} = 6.$$

Combination

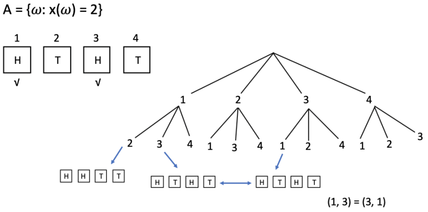
Each combination corresponds to $k!$ permutations.

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

$$\binom{4}{2} = \frac{4 \times 3}{2} = 6.$$

Coin flipping

Example 4: Coin flipping



HHHH, THHH, HTHT, TTHT,
HHHT, HHTT, THHT, THTT,
HHTH, TTHH, HTTH, HTTT,
HTHH, THTH, TTTH, TTTT

$|A_2| = \binom{4}{2} = \frac{4\times 3}{2} = 6.$

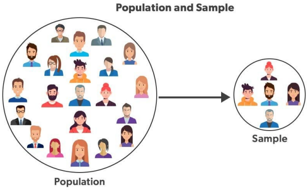
Do not confuse order of picking blanks with order of coin flippings.

In general, flip a fair coin n times independently,

$$P(A_k) = P(\{\omega : X(\omega) = k\}) = P(X = k) = \frac{\binom{n}{k}}{2^n}.$$

Survey sampling

Population of N people, M males.
Repeat random sampling n times independently



$\rightarrow N^n$ equally likely sequences.
For a sequence ω, $X(\omega) =$ number of males in ω.
$A_m = \{\omega : X(\omega) = m\}$: sequences with m males.
$|A_m| = \binom{n}{m} M^m (N-M)^{n-m}$. n blanks. Choose m blanks for males, the rest $n-m$ blanks for females. Each male blank has M choices. Each female blank has $N-M$ choices.

Survey sampling

Population of N people. M males.
Sample a person, $p = M/N$ = Prob(male).

$$P(A_m) = P(X = m) = \frac{|A_m|}{|\Omega_n|}$$
$$= \frac{\binom{n}{m} M^m (N-M)^{n-m}}{N^n}$$
$$= \binom{n}{m} p^m (1-p)^{n-m}.$$

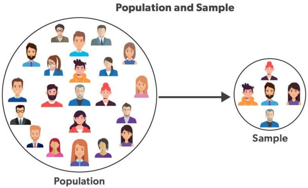
Most sequences are representative, $X/n \approx M/N = p$.

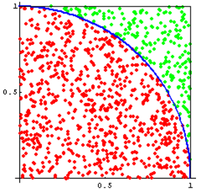
Binomial distribution, probability mass function

Flip a coin n times independently, $p =$ probability of head.

$$p(x) = P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$$
$$x = 0, 1, ..., n.$$

$p(x)$: probability mass function, probability distribution.





Survey sampling, poll before election, $p = M/N$.
Monte Carlo, $p = \pi/4$.