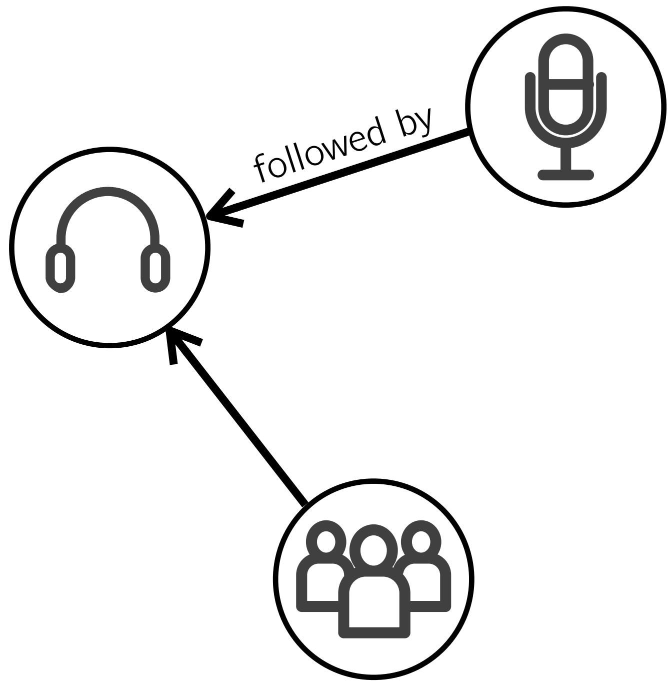
Today

- Problems in combinatorics

Problem

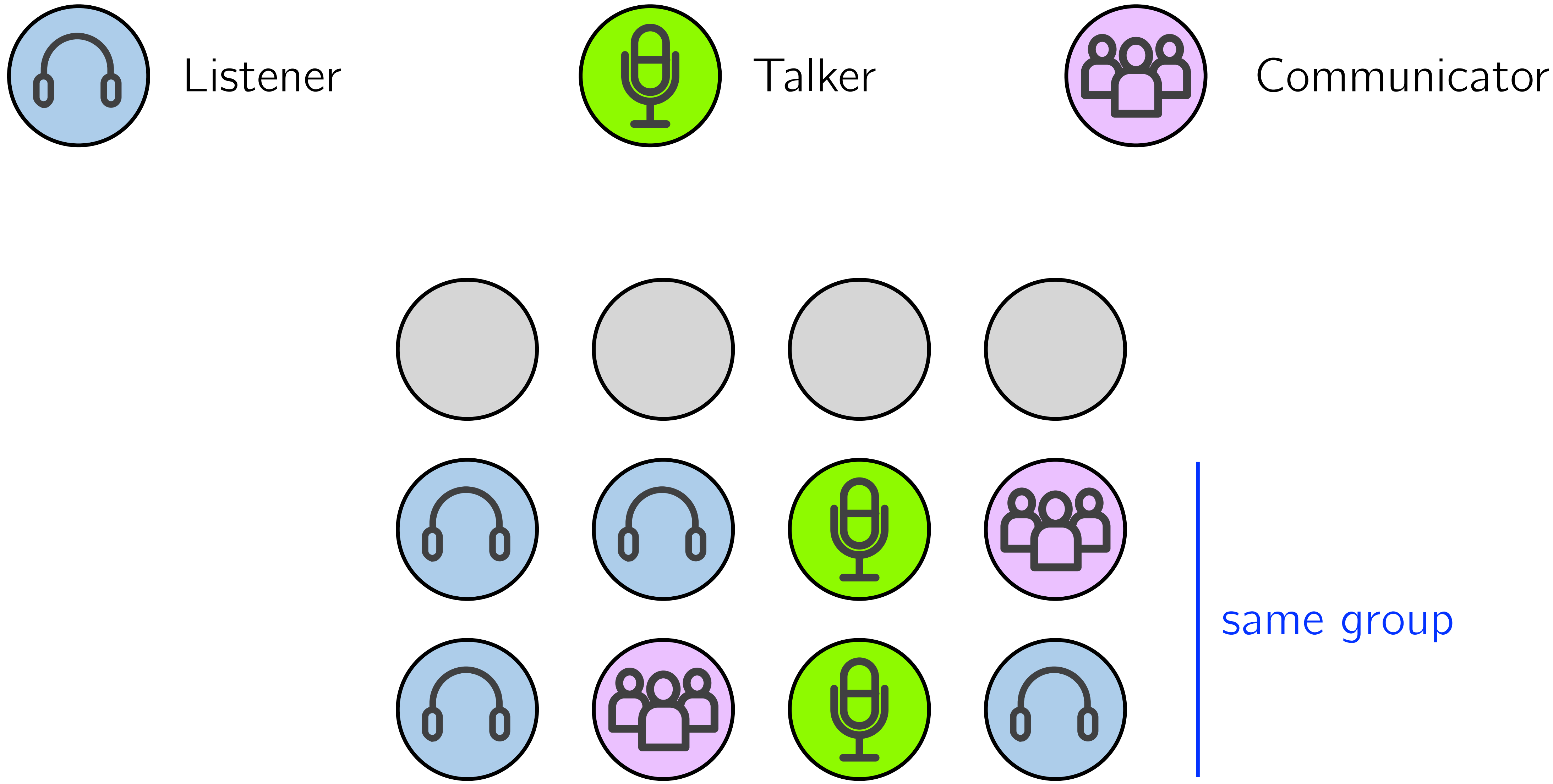
Twitter users (nodes) are often classified as follows:

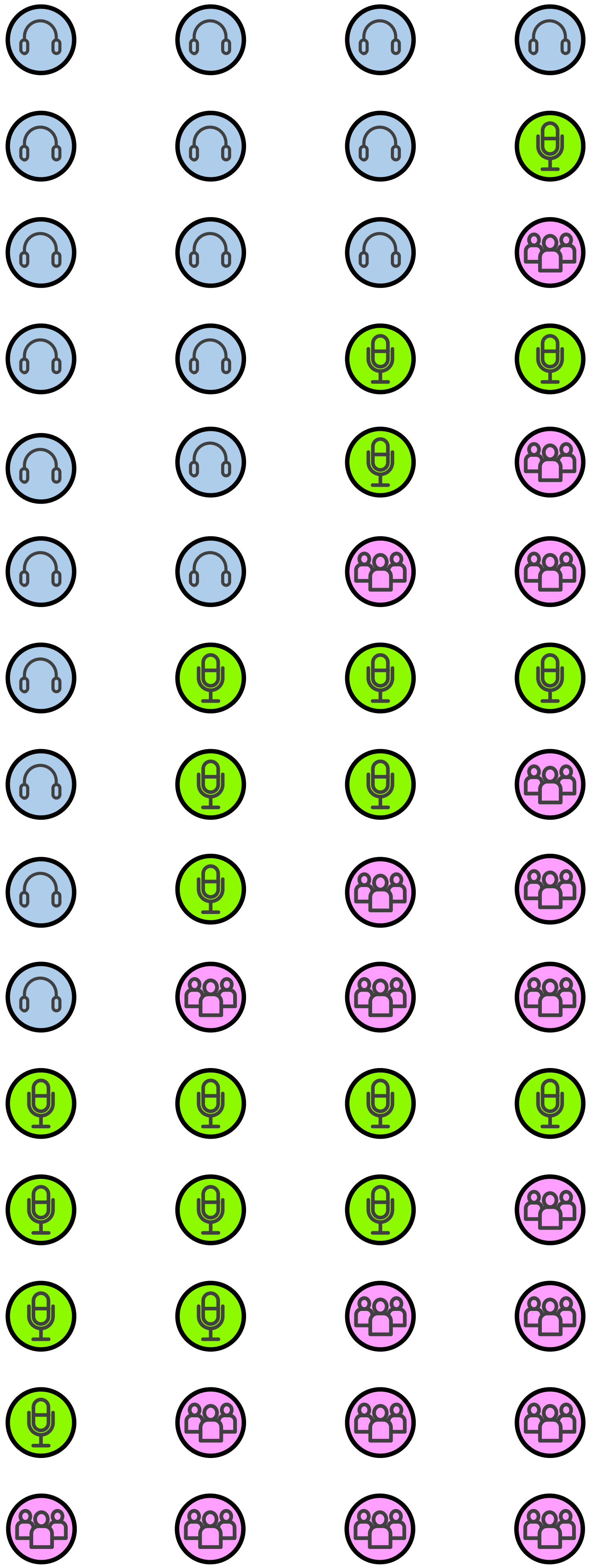
- Listener (greater in-degree than out-degree)
- Talker (greater out-degree than in-degree)
- Communicator (high in-degree and out-degree)



Suppose we have an unlimited supply of users of three different types (listener, talker, communicator) and you want to create a group of size 4 among these three types. How many group configurations are there?

Problem





4 listeners: 1 configuration

3 listeners: 2 configurations

2 listeners: 3 configurations

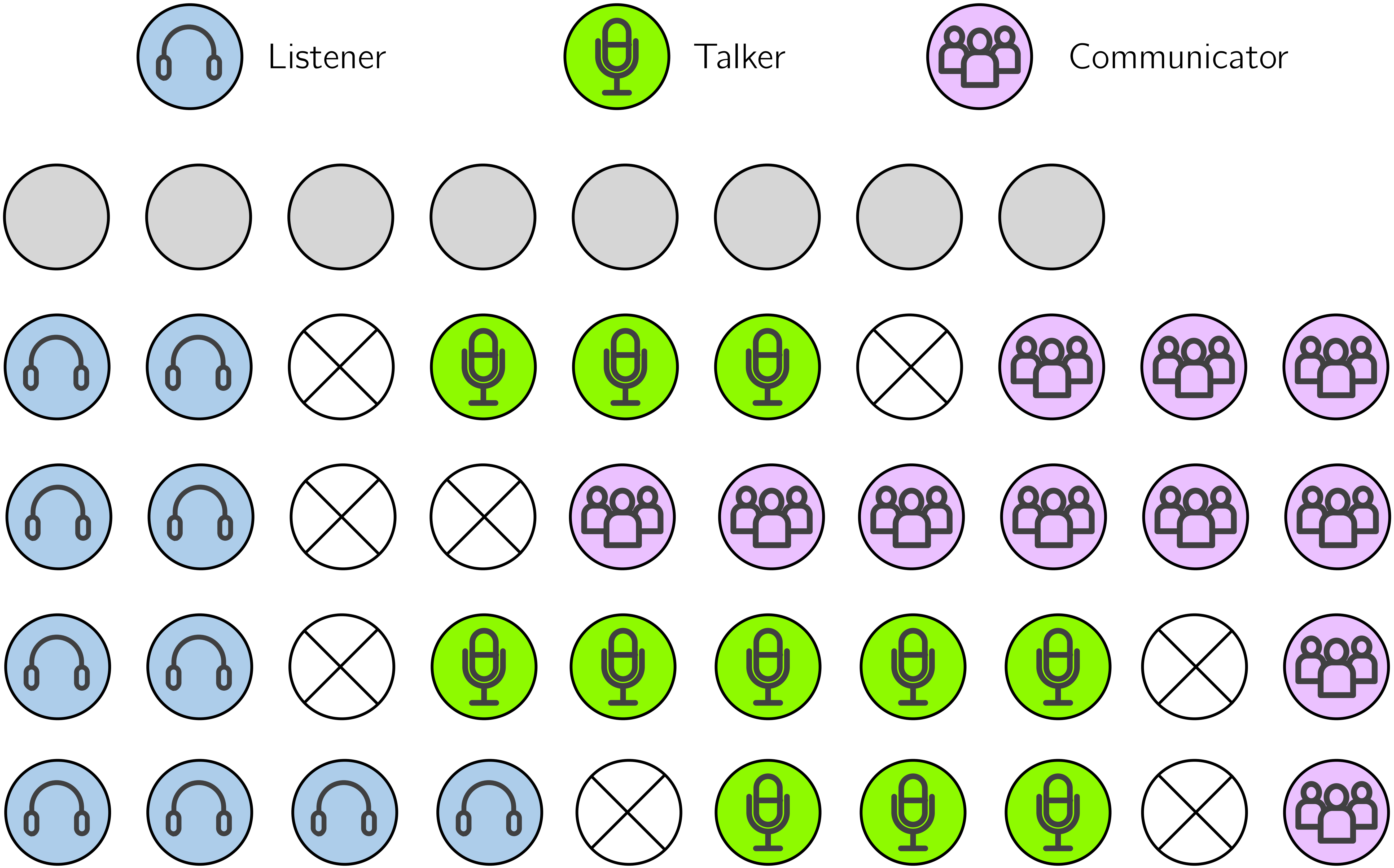
1 listener: 4 configurations

0 listeners: 5 configurations

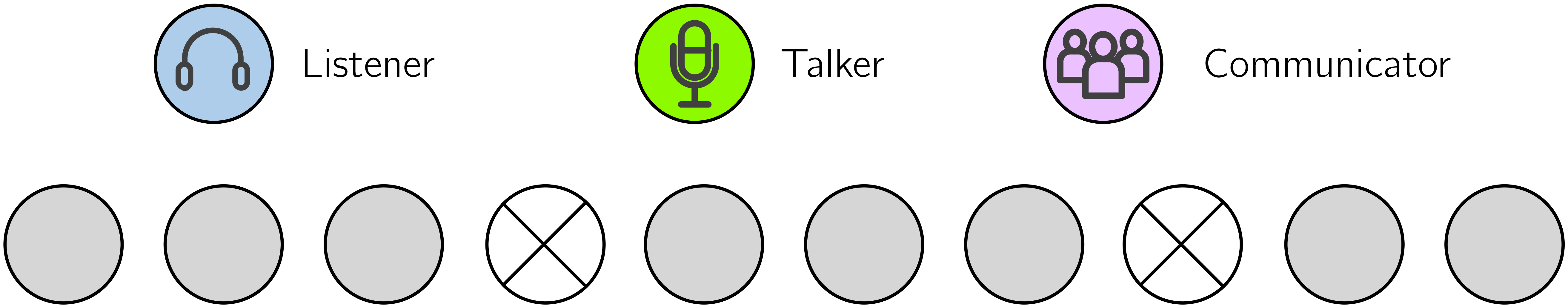
In total: 15 configurations

```python
# lists all types
# L=listener, T=talker, C=communicator
from itertools import combinations_with_replacement
for c in combinations_with_replacement(“TLC", 4):
    print("".join(c))
```

Problem



Problem



- Size of the group (k)
- Number of user types (n)
- Number of possible configurations
- Choose n - 1 types from a sequence of length k + (n - 1)

$$\binom{k+n-1}{n-1} = \binom{10}{2} = 45$$

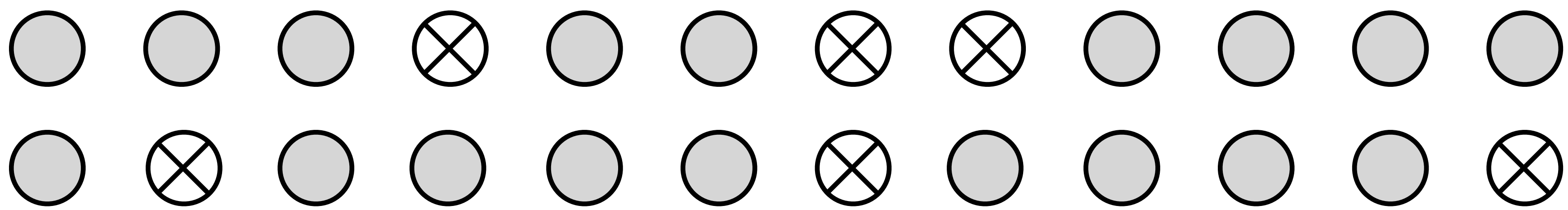
Combinations of size k from n options

Problem

How many non-negative numbers below 10000 exist, such that the sum of their digits is equal to 9?

Split weight $k = 9$ among $n = 4$ digits

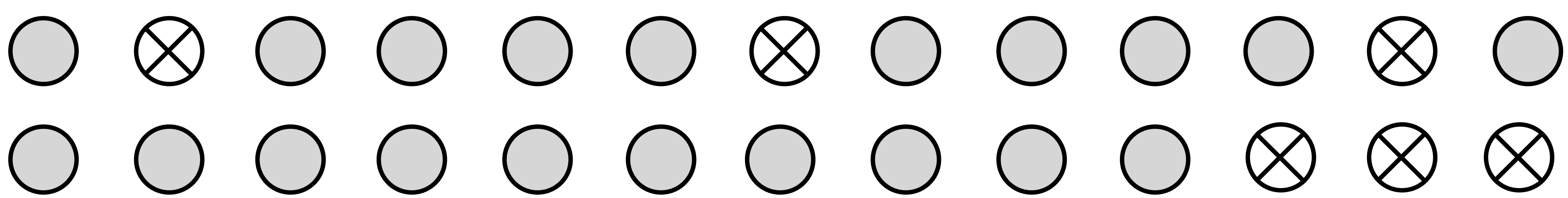
3204 =
1440 =



$$\binom{9+4-1}{4-1} = 220$$

How many non-negative numbers below 10000 exist, such that the sum of their digits is equal to 10?

1441 =
10000 =



Similarly for 01000, 00100, 00010

$$\binom{10+4-1}{4-1} - 4 = 282$$

Summary

	With repetitions	Without repetitions
Ordered	k-tuples: n^k	k-permutations: $\dfrac{n!}{(n-k)!}$
Unordered	k-combinations with repetitions: $\binom{k+n-1}{n-1}$	k-combinations: $\binom{n}{k}$

Example:

Ordered	`(a,a) (a,b) (a,c)`
	`(b,a) (b,b) (b,c)`
	`(c,a) (c,b) (c,c)`
		`(a,b) (a,c)`
		`(b,a) (b,c)`
		`(c,a) (c,b)`

Unordered	`{a,a} {a,b} {a,c}`
	`{b,b} {b,c} {c,c}`
		`{a,b}`
		`{a,c}`
		`{b,c}`

Restrictions:

Ordered	n, k	$k \leq n$
Unordered	n, k	$k \leq n$

Next

- Random experiments (intuition about the notion of probability)
- Definition of an event
- Calculating probabilities

Random experiments

Example 1: Toss a coin once

- Outcomes: head(h) or tail(t) (not known in advance)
- Sample space: set of all possible outcomes $\Omega = \{h,t\}$

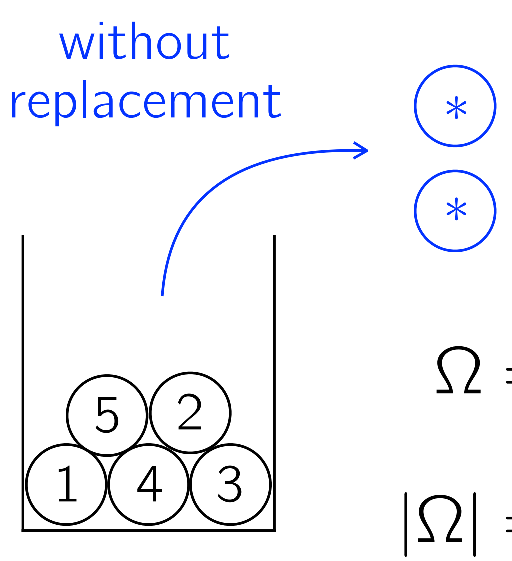
Example 2: Toss a coin twice

$$\Omega = \{(h,h), (h,t), (t,h), (t,t)\}$$

Example 3: Roll a dice

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

Example 3: Bucket of balls



$$\Omega = 2\text{-combinations}$$

$$|\Omega| = \binom{5}{2} = \frac{5!}{2!3!} = 10$$

Event:
Description of a situation that can happen when we perform a random experiment

The notion of an event

Toss a coin twice

- $\Omega = \{(h,h), (h,t), (t,h), (t,t)\}$
- Event A (some outcomes satisfy event A)
- A = "both times get the same result"
- $A = \{(h,h), (t,t)\}$
- B = "get at least one head"
- $B = \{(h,h), (h,t), (t,h)\}$

Roll a dice

- $\Omega = \{1, 2, 3, 4, 5, 6\}$
- A = "get more than 3 points"
- $A = \{4, 5, 6\}$

Events are subsets of the sample space

Operations on events

- A, B - events associated to some random experiment
- $A \subset \Omega$ and $B \subset \Omega$
- Then $A \cap B$ - event that both A and B happen

Roll a dice

- $\Omega = \{1, 2, 3, 4, 5, 6\}$
- A = “points are less than 4” = $\{1, 2, 3\}$
- B = “points are even” = $\{2, 4, 6\}$
- $A \cap B = \{2\}$
- C = “points are greater than 5” = $\{6\}$

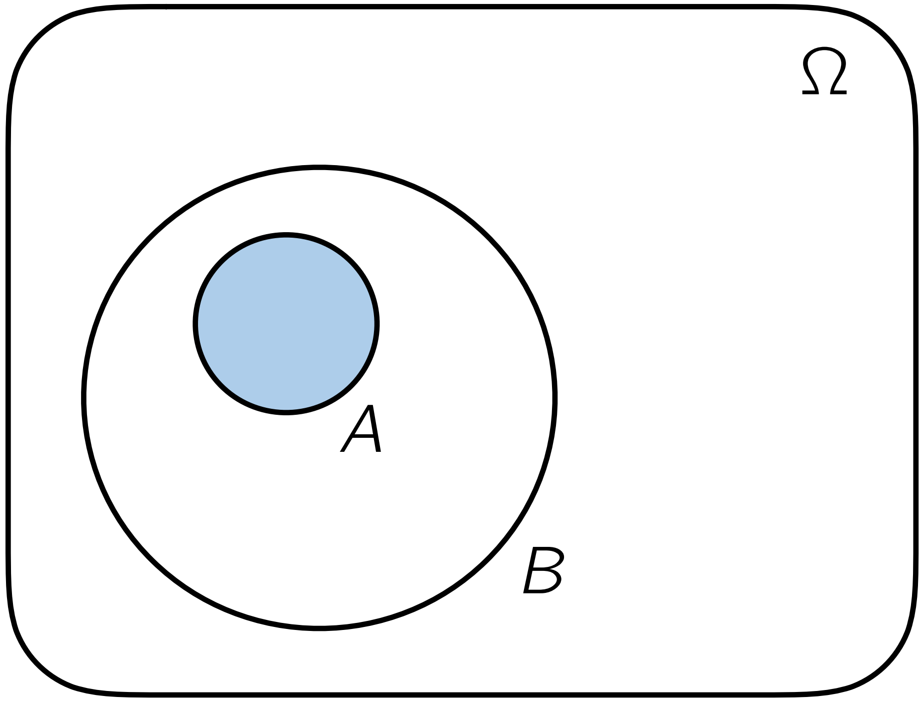
Operations on events

- $A \cap C = \emptyset$ (an impossible event)
- $A \cup B$ (at least of the two events happen)
- $\bar{A} = \Omega - A$ (event A does not happen)
- $A \subset B$

Key assumptions:

$P(\Omega) = 1$

$P(\emptyset) = 0$



If A happens, then B also happens

Calculating probabilities

- How often an event occurs? (empirical approach)
- Suppose
    - Finite number of outcomes
    - All outcomes are equally likely
- Define the probability of an event as

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

- Good definition?
- Several outcomes, each with the same probability
- The sum of the probabilities of all outcomes equals to one
  (at least one happens)
- For events with several outcomes, probability is proportionally larger

Example

Toss a coin twice

- $\Omega = \{(\text{h,h}), (\text{h,t}), (\text{t,h}), (\text{t,t})\}$
- A = "we get tail at least once"
- $P(A) = P(\{(\text{h,t}), (\text{t,h}), (\text{t,t})\}) = 3/4$
- B = "we get exactly one tail"
- $P(B) = P(\{(\text{h,t}), (\text{t,h})\}) = 1/2$

Example

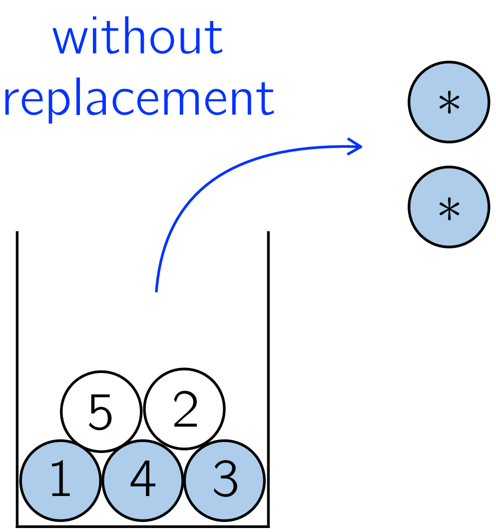
Roll a dice

- $\Omega = \{1, 2, 3, 4, 5, 6\}$ - fair dice
- A = “points are greater than 4”
- $P(A) = P(\{5, 6\}) = 2/6$

If we have an event that is explicitly described as a subset of the set of all outcomes, then finding of this probability is rather simple

Probabilities and combinatorics

Bucket of balls



- A = "both balls have color blue"
- Ω is the set of all 2-combinations of 5 possible options
- The size of all possible outcomes

$$|\Omega| = \binom{5}{2} = \frac{5!}{2!3!} = 10$$

- A are 2-combinations of 3 possible options, so

$$|A| = \binom{3}{2} = 3$$

and

$$P(A) = \frac{3}{10} = 0.3$$

Example

Toss a coin five times

- A = "get exactly 2 heads"
- Ω is the set of all sequences (of length 5) of heads and tails
- The size of all possible outcomes

$$|\Omega| = 2^5 = 32$$

- Which two elements of the sequence are heads?

$$\underline{\text{t}} \quad \underline{\text{h}} \quad \underline{\text{t}} \quad \underline{\text{h}} \quad \underline{\text{t}}$$

$$|A| = \binom{5}{2} = 10$$

- Therefore

$$P(A) = \frac{10}{32} = 0.31$$

event is explicitly described in terms of outcomes.

Probabilities and operations on events

- A, B - events associated to some random experiment
- $A \subset \Omega$ and $B \subset \Omega$
- Know probabilities of some events $\mathrm{P}(A)$ and $\mathrm{P}(B)$
- Need find probability of some other event
- Probability of event A or B occurring?

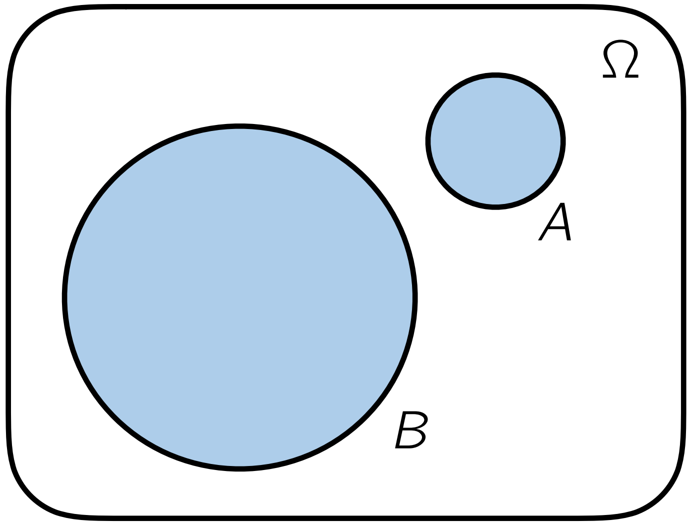
$$
\begin{aligned}
P(A \cup B) &= \frac{|A \cup B|}{|\Omega|} \\
&= \frac{|A| + |B| - |A \cap B|}{|\Omega|} \\
&= \frac{|A|}{|\Omega|} + \frac{|B|}{|\Omega|} - \frac{|A \cap B|}{|\Omega|} \\
&= P(A) + P(B) - P(A \cap B)
\end{aligned}
$$

- If A and B are mutually exclusive ($A \cap B = \emptyset$), then

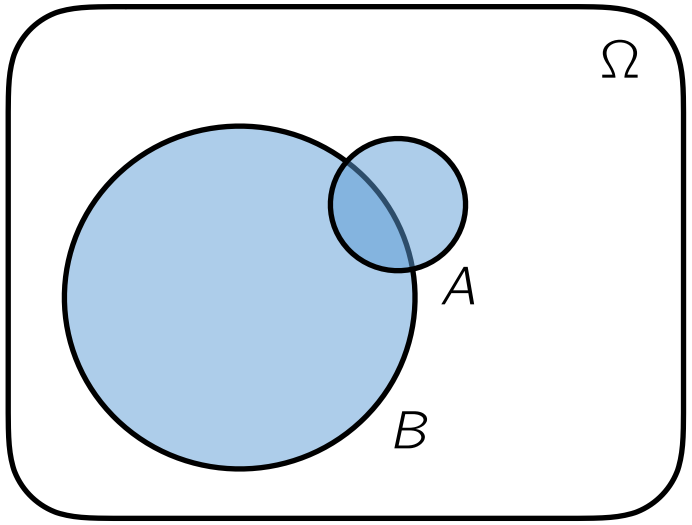
$$P(A \cup B) = P(A) + P(B)$$

Probabilities and operations on events

- A, B - events associated to some random experiment
- $A \subset \Omega$ and $B \subset \Omega$
- Know probabilities of some events $\text{P}(A)$ and $\text{P}(B)$
- Need find probability of some other event
- Probability of event A and B occurring?
- In general, not possible!



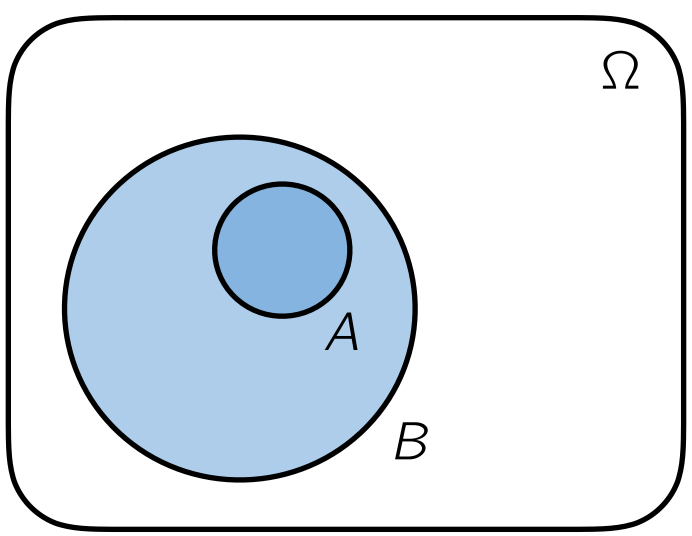
$P(A \cap B) = 0$



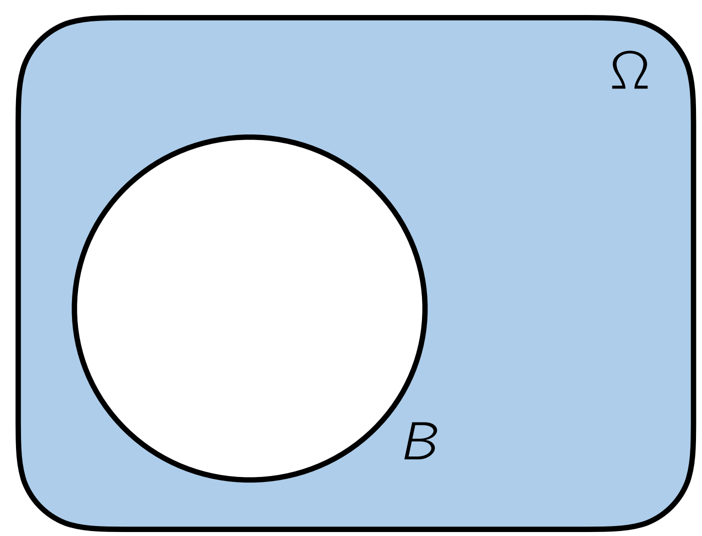
$0 < P(A \cap B) \leq P(A)$ and $P(A \cap B) \leq P(B)$

Probabilities and operations on events

- A, B - events associated to some random experiment
- $A \subset \Omega$ and $B \subset \Omega$
- Know probabilities of some event P(B)
- Probability of event A occurring if A⊂B?
- Probability of event A not occurring?



$P(A) \leq P(B)$

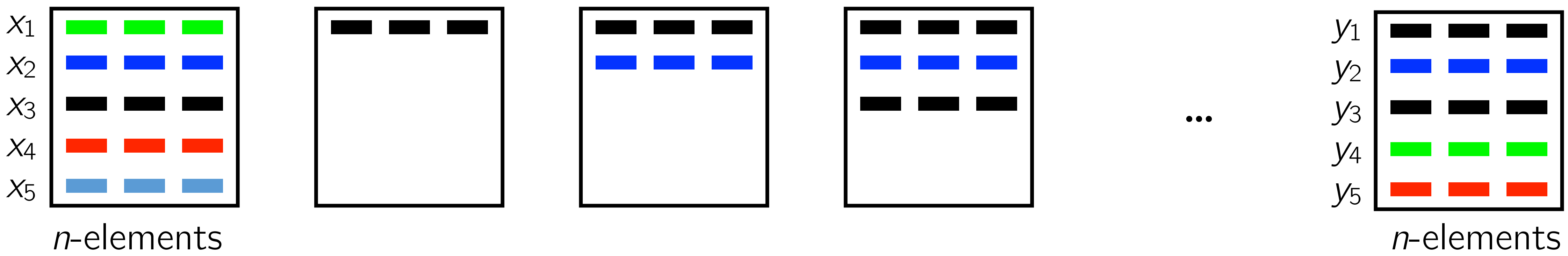


$P(\bar{A}) = 1 - P(A)$

Often useful in practice!

Problem

- Consider a table dataset
- Sampling with replacement (bagging)



- What is the probability of including a particular element in the new dataset?
- A = "first element is included in new dataset"
- $\mathrm{P}(A)$ hard to calculate!
- Consider an outcome $(y_1, y_2, \cdots, y_n) \in \{x_1, x_2, \cdots, x_n\}^n$
- Size of all possible outcomes $|\Omega| = n^n$
- So $|\bar{A}| = (n-1)^n$ and $P(A) = 1 - \dfrac{(n-1)^n}{n^n} = 1 - \left(1 - \dfrac{1}{n}\right)^n \approx 1 - \dfrac{1}{e} \approx \dfrac{2}{3}$

Allows for out-of-bag cross-validation

Outcomes with non-equal probabilities

- Example: probability of heads is not the same as probability of tails
- Probability spaces with arbitrary probabilities attached to outcomes
- Set of outcomes $\Omega = \{\omega_1, \cdots, \omega_n\}$ and $|\Omega| = n$
- An event of exactly one outcome occurs with probability $p_i = P(\{\omega_i\})$
- Probabilities satisfy $p_i \geq 0$ and $\sum_{i=1}^{n} p_i = 1$
- Consider $A \subset \Omega$
- What is P(A)?
- Note that $\omega_i \in A$ are mutually disjoined

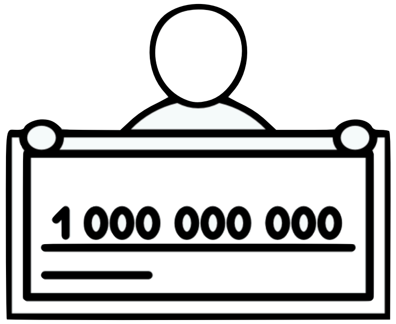
$$A = \{\omega_1, \omega_2, \cdots, \omega_k\} = \{\omega_1\} \cup \{\omega_2\} \cup \cdots \cup \{\omega_k\}$$

- Therefore

$$P(A) = P(\{\omega_1\}) + P(\{\omega_2\}) + \cdots + P(\{\omega_k\}) = \sum_{\omega_i \in A} p_i$$

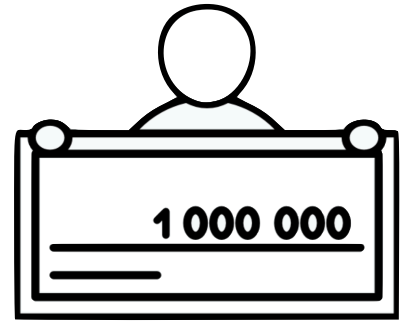
Example

Lottery



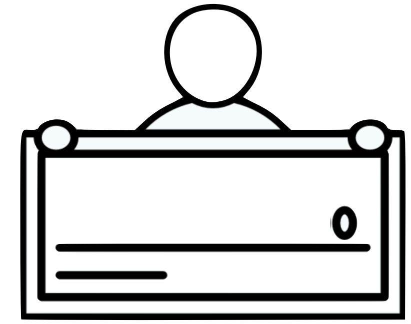
ω_1 : big prize

$p_1 = 0.01$



ω_2 : small prize

$p_2 = 0.1$



ω_3 : no prize

$p_3 = 1 - p_1 - p_2 = 0.89$

- A = "winning a prize" = $\{\omega_1, \omega_2\}$
- $P(A) = p_1 + p_2 = 0.11$

Operations on events still hold:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(\emptyset) = 0$$

$$P(\Omega) = 1$$

If $A \subset B$ then $P(A) \leq P(B)$

Exercises

Lottery

In a 6 number lottery one is trying to guess a subset of 6 numbers among 44 without repetitions. How many ways are there to pick different numbers

$$\binom{44}{6} = 7059052$$

After the lottery, the organizers are asked, how many possible ways are there to guess correctly four numbers?

$$\binom{6}{4}\binom{44-6}{2} = 10545$$

Exercises

Coin tossing

A fair coin is tossed 3 times. What is the probability that the first and second tossing coincide?

There are 4 outcomes that satisfy the condition: ttt, tth, hht, hhh

Total number of outcomes $2^3 = 8$. Thus $p = 1/2$

The coin is now tosses 10 times. What is the probability that at least one tail occurs?

$$p = 1 - \frac{1}{2^{10}}$$

What is the probability that the first and second tossing coincide if the coin is tossed 30 times?

There are $2 \cdot 2^{28}$ outcomes that satisfy the condition

Total number of outcomes $2^{30} = 8$. Thus $p = 1/2$

Exercises

Roll a pair dices

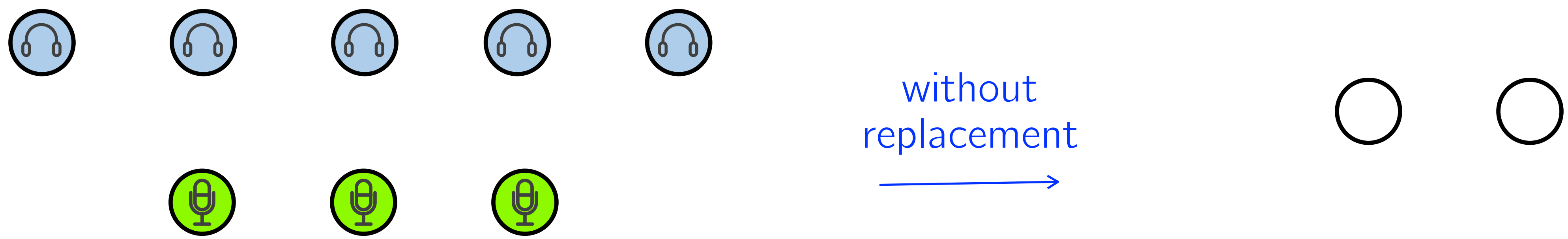
A pair of fair dices is rolled. What is the probability that the sum of points is greater than 10?

There are 3 outcomes that satisfy the condition: 5+6, 6+5, 6+6

Total number of outcomes $6^2 = 36$. Thus $p = 1/12$

Exercises

Consider two types of users



There are 5 listeners and 3 talkers. Two users where chosen randomly (without replacement). What is the probability of both users being talkers?

$$\frac{\binom{5}{2}}{\binom{8}{2}}$$

What is the probability of one user is a talker and the other is a listener?

$$\frac{\binom{5}{1}\binom{3}{1}}{\binom{8}{2}}$$

Summary

- Rule of sum
- Rule of product
- Permutations and combinations (with and without repetitions)
- Random experiments and probability spaces
- Applications to compute probabilities