Sample Spaces, Events, Probability

Foundations of Data Analysis

February 4, 2021

Brain Teaser

You have cards numbered 1 through 10. If you pick five cards at random, what is the probability that you selected the cards 1, 2, 3, 4, 5 (not necessarily in that order)?

Sets

Definition

A **set** is a collection of unique objects.

Here “objects” can be concrete things (people in class, schools in ACC), or abstract things (numbers, colors).

Examples:

$A = \{3, 8, 31\}$
$B = \{\text{apple, pear, orange, grape}\}$
Not a valid set definition: $C = \{1, 2, 3, 4, 2\}$

Sets

- Order in a set does not matter!

$$\{1,2,3\} = \{3,1,2\} = \{1,3,2\}$$

- When x is an element of A, we denote this by:

$$x \in A.$$

- If x is not in a set A, we denote this as:

$$x \notin A.$$

- The “empty” or “null” set has no elements:

$$\emptyset = \{\,\}$$

Sample Spaces

Definition

A **sample space** is the set of all possible outcomes of an experiment. We'll denote a sample space as Ω.

Examples:

- Coin flip: $\Omega = \{H, T\}$
- Roll a 6-sided die: $\Omega = \{1, 2, 3, 4, 5, 6\}$
- Pick a ball from a bucket of red/black balls: $\Omega = \{R, B\}$

Some Important Sets

- Integers:

$$\mathbb{Z} = \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$$

- Natural Numbers:

$$\mathbb{N} = \{0, 1, 2, 3, \ldots\}$$

- Real Numbers:

$\mathbb{R} =$ “any number that can be written in decimal form”

$$5 \in \mathbb{R}, \quad 17.42 \in \mathbb{R}, \quad \pi = 3.14159\ldots \in \mathbb{R}$$

Building Sets Using Conditionals

- Alternate way to define natural numbers:

$$\mathbb{N} = \{x \in \mathbb{Z} : x \geq 0\}$$

- Set of even integers:

$$\{x \in \mathbb{Z} : x \text{ is divisible by } 2\}$$

- Rationals:

$$\mathbb{Q} = \{\, p/q : p, q \in \mathbb{Z}, q \neq 0\}$$

Subsets

Definition

A set A is a **subset** of another set B if every element of A is also an element of B, and we denote this as $A \subseteq B$.

Examples:

- $\{1, 9\} \subseteq \{1, 3, 9, 11\}$
- $\mathbb{Q} \subseteq \mathbb{R}$
- $\{$apple, pear$\} \nsubseteq \{$apple, orange, banana$\}$
- $\emptyset \subseteq A$ for any set A

Events

Definition

An **event** is a subset of a sample space.

Examples:

- You roll a die and get an even number: $\{2,4,6\} \subseteq \{1,2,3,4,5,6\}$
- You flip a coin and it comes up "heads": $\{H\} \subseteq \{H,T\}$
- Your code takes longer than 5 seconds to run: $(5,\infty) \subseteq \mathbb{R}$

Set Operations: Union

Definition

The **union** of two sets A and B, denoted $A \cup B$ is the set of all elements in either A or B (or both).

When A and B are events, $A \cup B$ means that event A *or* event B happens (or both).

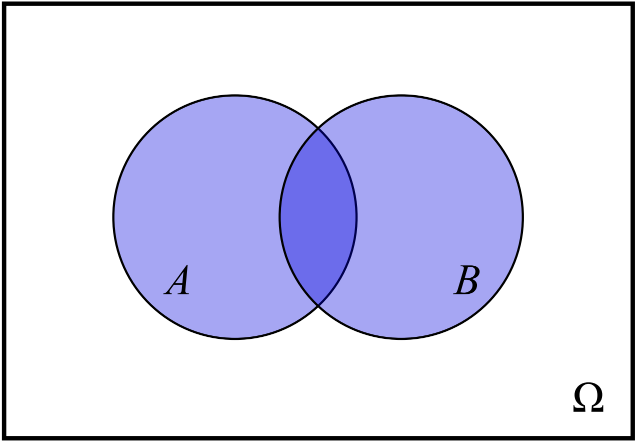
Example:

$A = \{1, 3, 5\}$ “an odd roll”

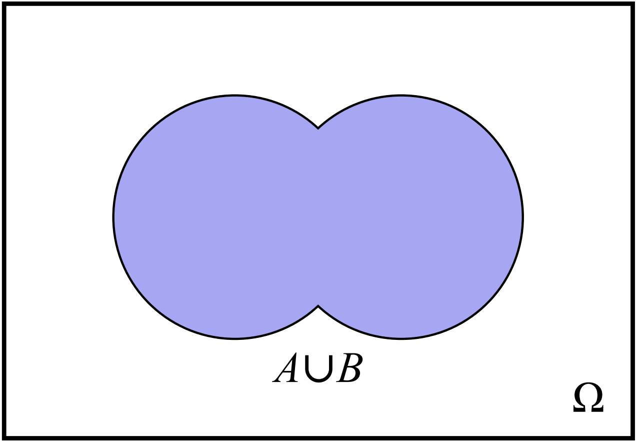
$B = \{1, 2, 3\}$ “a roll of 3 or less”

$A \cup B = \{1, 2, 3, 5\}$

Venn Diagram: Union



Venn Diagram: Union



Set Operations: Intersection

Definition

The **intersection** of two sets A and B, denoted $A \cap B$ is the set of all elements in both A and B.

When A and B are events, $A \cap B$ means that both event A *and* event B happen.

Example:

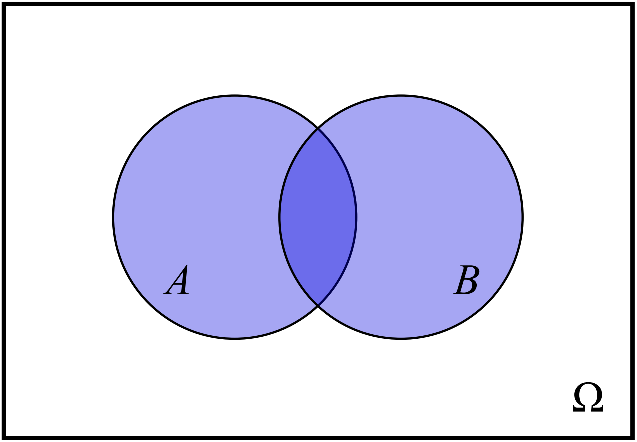
$A = \{1, 3, 5\}$ "an odd roll"

$B = \{1, 2, 3\}$ "a roll of 3 or less"

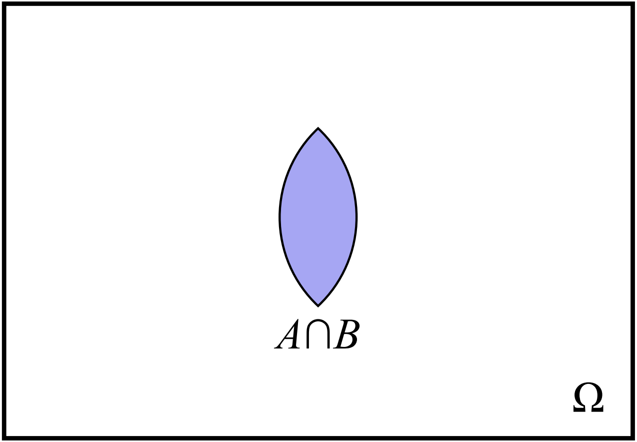
$A \cap B = \{1, 3\}$

Note: If $A \cap B = \emptyset$, we say A and B are **disjoint**.

Venn Diagram: Intersection



Venn Diagram: Intersection



Set Operations: Complement

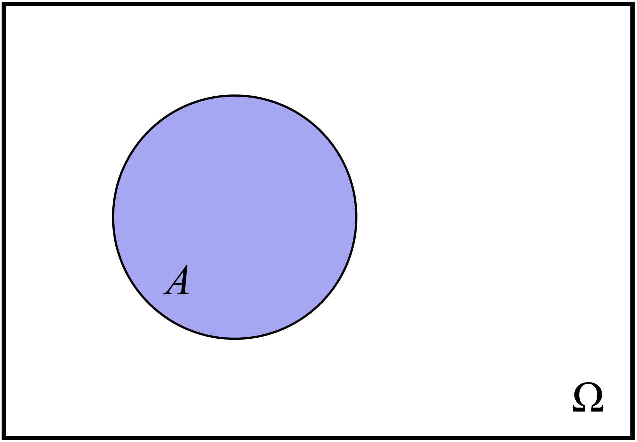
Definition

The **complement** of a set $A \subseteq \Omega$, denoted A^c, is the set of all elements in Ω that are not in A.

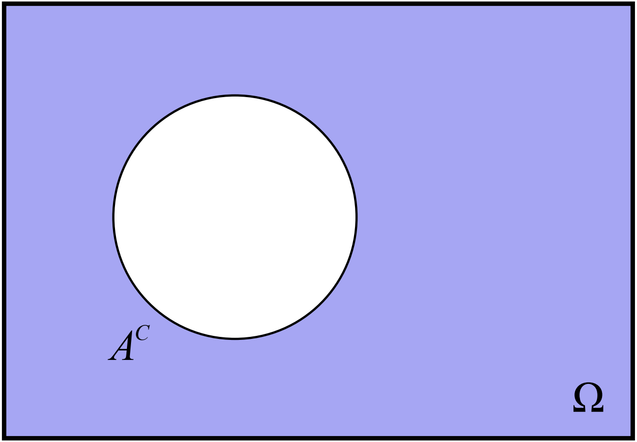
When A is an event, A^c means that the event A does not happen.

Example:
$A = \{1, 3, 5\}$ "an odd roll"
$A^c = \{2, 4, 6\}$ "an even roll"

Venn Diagram: Complement



Venn Diagram: Complement



Set Operations: Difference

Definition

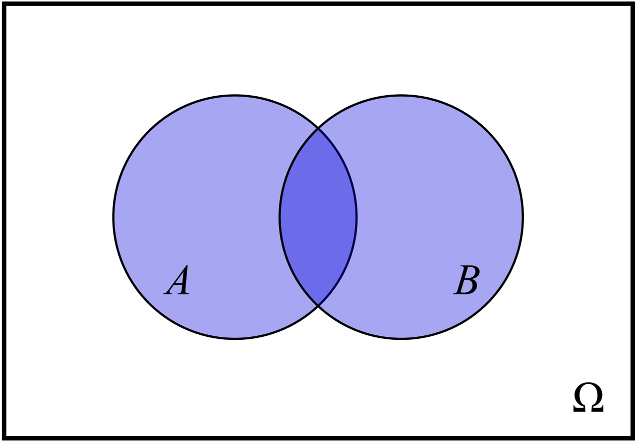
The **difference** of a set $A \subseteq \Omega$ and a set $B \subseteq \Omega$, denoted $A - B$, is the set of all elements in Ω that are in A and are not in B.

Example:

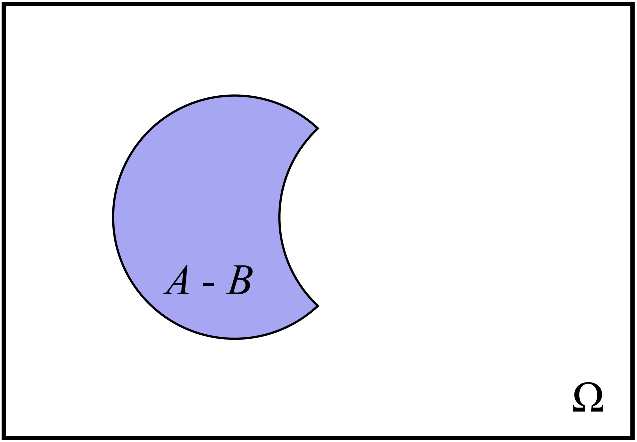
$A = \{3, 4, 5, 6\}$
$B = \{3, 5\}$
$A - B = \{4, 6\}$

Note: $A - B = A \cap B^c$

Venn Diagram: Difference



Venn Diagram: Difference



De Morgan's Laws

Complement of union or intersection:

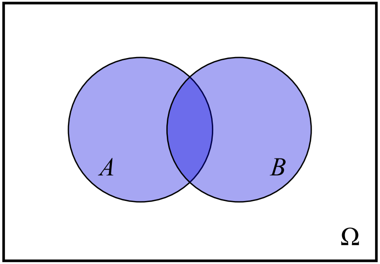
$$(A \cup B)^c = A^c \cap B^c$$

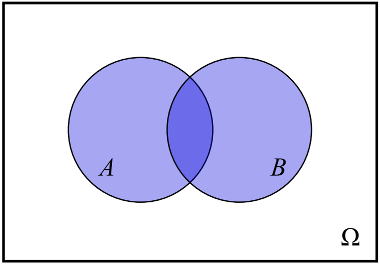
$$(A \cap B)^c = A^c \cup B^c$$

What is the English translation for both sides of the equations above?

Venn Diagram: De Morgan

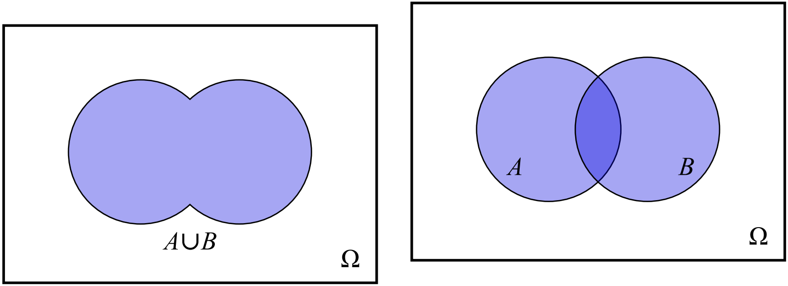
$$(A \cup B)^c = A^c \cap B^c$$





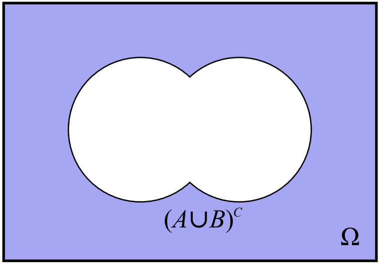
Venn Diagram: De Morgan

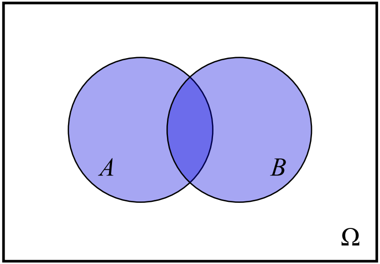
$$(A \cup B)^c = A^c \cap B^c$$



Venn Diagram: De Morgan

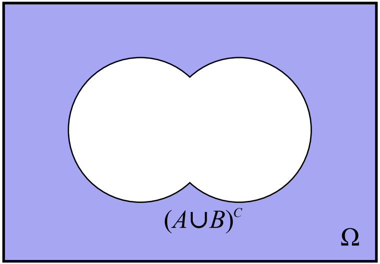
$$(A \cup B)^c = A^c \cap B^c$$

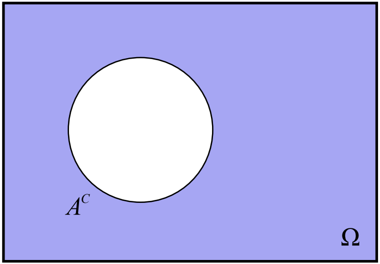




Venn Diagram: De Morgan

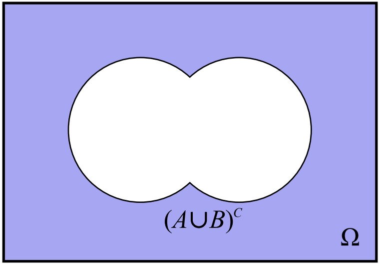
$$(A \cup B)^c = A^c \cap B^c$$

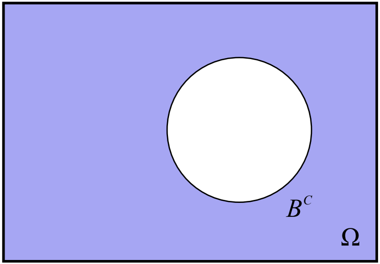




Venn Diagram: De Morgan

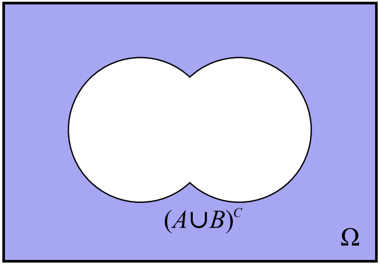
$$(A \cup B)^c = A^c \cap B^c$$





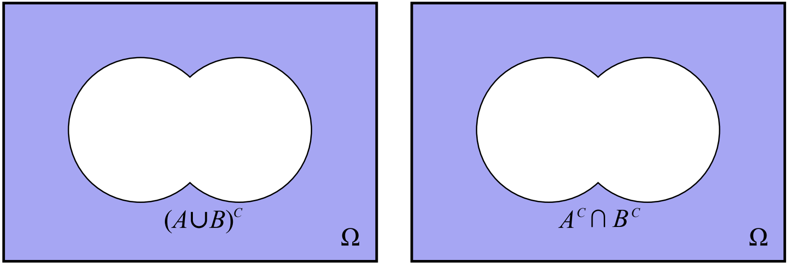
Venn Diagram: De Morgan

$$(A \cup B)^c = A^c \cap B^c$$



Venn Diagram: De Morgan

$$(A \cup B)^c = A^c \cap B^c$$



Exercises

Check whether the following statements are true or false. (Hint: you might use Venn diagrams.)

- $A - B \subseteq A$
- $(A - B)^c = A^c \cup B$
- $A \cup B \subseteq B$
- $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$

Probability

Definition

A **probability function** on a finite sample space Ω assigns every event $A \subseteq \Omega$ a number in $[0, 1]$, such that

1. $P(\Omega) = 1$
2. $P(A \cup B) = P(A) + P(B)$ when $A \cap B = \emptyset$

$P(A)$ is the **probability** that event A occurs.

Equally Likely Outcomes

The number of elements in a set A is denoted $|A|$.

If Ω has a finite number of elements, and each is equally likely, then the probability function is given by

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

Example: Rolling a 6-sided die

- $P(\{1\}) = 1/6$
- $P(\{1, 2, 3\}) = 1/2$

Repeated Experiments

If we do two runs of an experiment with sample space Ω, then we get a new experiment with sample space

$$\Omega \times \Omega = \{(x, y) : x \in \Omega, y \in \Omega\}$$

The element $(x, y) \in \Omega \times \Omega$ is called an **ordered pair**.

Properties:
Order matters: $(1, 2) \neq (2, 1)$
Repeats are possible: $(1, 1) \in \mathbb{N} \times \mathbb{N}$

More Repeats

Repeating an experiment n times gives the sample space

$$\begin{aligned}\Omega^n &= \Omega \times \cdots \times \Omega \quad (n \text{ times}) \\ &= \{(x_1, x_2, \ldots, x_n) : x_i \in \Omega \text{ for all } i\}\end{aligned}$$

The element $(x_1, x_2, \ldots, x_n)$ is called an n-**tuple**.

If $|\Omega| = k$, then $|\Omega^n| = k^n$.

Probability Rules

Complement of an event A:

$$P(A^c) = 1 - P(A)$$

Union of two overlapping events $A \cap B \neq \emptyset$:

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

Exercise

You are picking a number out of a hat, which contains the numbers 1 through 100. What are the following events and their probabilities?

- The number has a single digit
- The number has two digits
- The number is a multiple of 4
- The number is not a multiple of 4
- The sum of the number's digits is 5

Permutations

A **permutation** is an ordering of an n-tuple. For instance, the n-tuple $(1, 2, 3)$ has the following permutations:

$$(1, 2, 3), (1, 3, 2), (2, 1, 3)$$
$$(2, 3, 1), (3, 1, 2), (3, 2, 1)$$

The number of unique orderings of an n-tuple is n **factorial**:

$$n! = n \times (n-1) \times (n-2) \times \cdots \times 2$$

How many ways can you rearrange $(1, 2, 3, 4)$?

Binomial Coefficient or "n choose k"

The **binomial coefficient**, written as $\binom{n}{k}$ and spoken as "n choose k", is the number of ways you can select k items out of a list of n choices.

Formula:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Binomial Coefficient or "n choose k"

Example: You have cards numbered 1 through 10. If you pick five cards at random, what is the probability that you selected the cards 1, 2, 3, 4, 5 (not necessarily in that order)?

Answer:

We'll use the formula $P(A) = \frac{|A|}{|\Omega|}$.

There is only one combination that gives us cards 1,2,3,4,5, so $|A| = 1$.

The total number of possible 5 card selections is

$$|\Omega| = \binom{10}{5} = \frac{10!}{5!(10-5)!} = 252$$

So, finally the probability is

$$P(A) = \frac{|A|}{|\Omega|} = \frac{1}{252} \approx 0.00397 = 0.397\%$$