DATA ANALYSIS BOOTCAMP

PROBABILITY

SET THEORY

SOME DEFINITIONS

Data set is a collection of some items (elements), the order doesn't matter.
Often used capital letters to denote a set.

A = { x | x condition}   or A = { x : x condition} | and : are "such that"

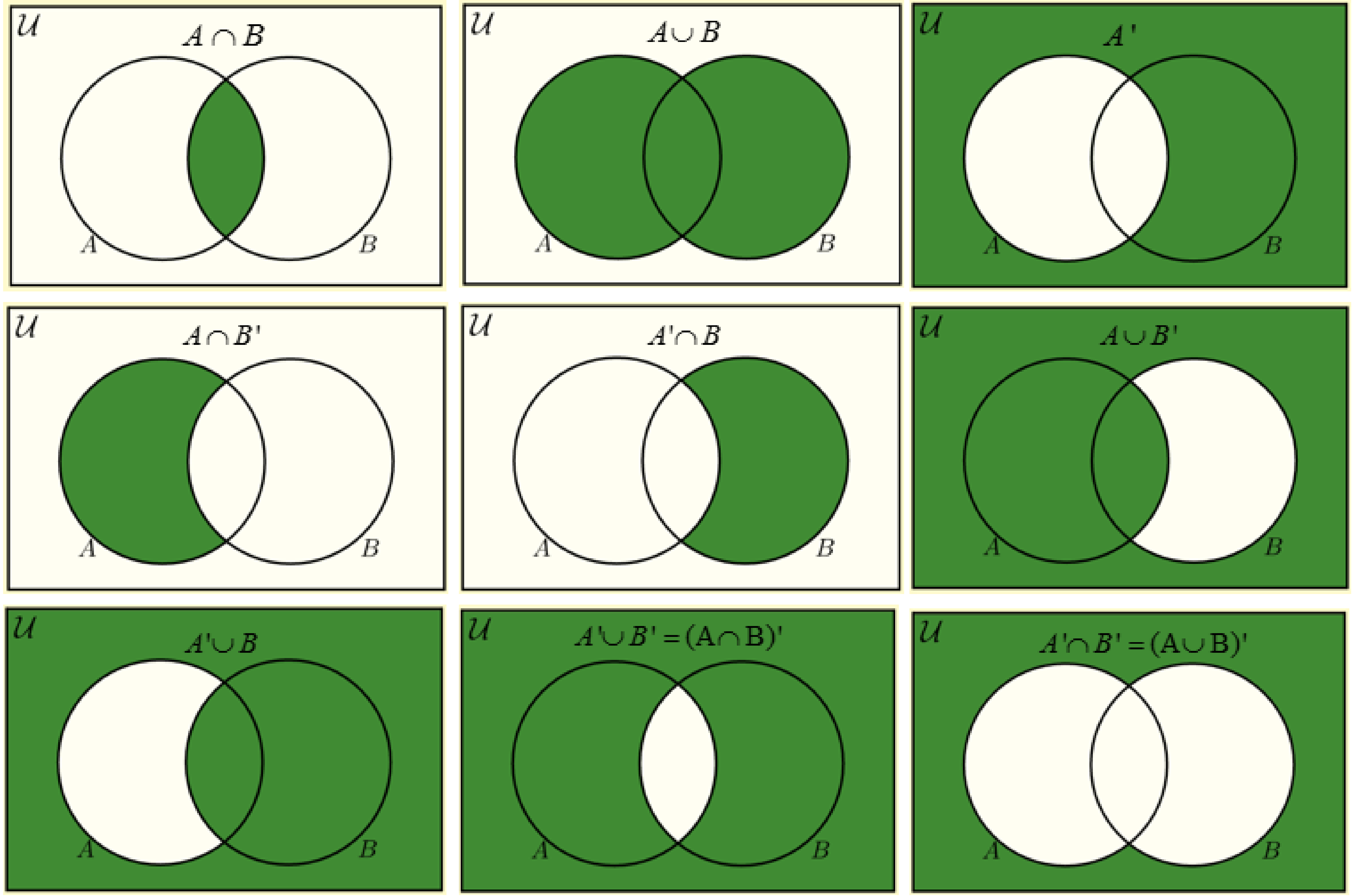
A is a **Subset** of set B if every element of A is also an element of B. B is the **superset**.

Null set / Empty set is the set with no elements.

Two sets are mutually **exclusive or disjoint** if they have no element in common.

Universal set is the set of all things we could possibily consider in the context.

UNION, INTERSECTION, NOT



EXERCISES

1 - Partition the whole universal set.

2 - If the universal set is given by S ={1,2,3,4,5,6}, and A ={1,2}, B = {2,4,5}, C = {1,5,6} are three sets, find the following sets:

- A U B
- A ∩ B
- not A
- not B

CARTESIAN PRODUCT, INCLUSION-EXCLUSION

A Cartesian Product of two sets A and B, written in A x B, is the set containing **ordered pairs** from A and B. That is, if C = A x B, then each element of C is of the form (x,y), where:
$x \in A$ and $y \in B$.

$$A \times B = \{(x,y) \mid x \in A \text{ and } y \in B\}$$

Inclusion-Exclusion principle: $|A \cup B| = |A| + |B| - |A \cap B|$

Infinite / finite set: difference between countable VS uncountable set (N VS R)

EXERCISES

1 - In a party,

- there are 10 ppl with white shirts and 8 ppl with red shirts;
- 4 ppl have black shoes and white shirts;
- 3 ppl have black shoes and red shirts;
- the total number of ppl with white or red shirts or black shoes is 21.

How many ppl have black shoes?

2 - Write all the possible partitions of $S = \{1,2,3\}$

3 - Let A, B, C be three sets. For each of the following sets, draw a Venn diagram and shade the area representing the given set.

- $A \cup B \cup C$
- $A \cap B \cap C$
- $A \cup (B \cap C)$
- $A - (B \cap C)$
- $A \cup \text{not } (B \cap C)$

RANDOM EXPERIMENTS AND PROBABILITY

A **random experiments** is a process by which we observe something uncertain. After the experiments the result of the random experiments, the **outcome**, is known The **sample space S** is the set of all possible outcomes, and the **event** is a subset of the sample space.

We assign a probability measure **P(A)** to an event. This is a value between 0 and 1 that shows how likely the event is.
- P(A) close to 0 -> very unlikely
- P(A) close to 1 -> very likely

Stuff to remember:
- P(A) >= 0 , for all A
- P (S) = 1
- The sum probability of all disjoint probability is their sum.

INCLUSION-EXCLUSION PRINCIPLE

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

What with three sets?

EXERCISE

1 - Suppose we have the following information:

- there is a 60% chance that it will rain today.
- there is a 50% chance that it will rain tomorrow.
- there is a 30% chance that it does not rain on either day

Find the following probability:

- that it will rain today or tomorrow
- that it will rain today and tomorrow
- that it will rain today but not tomorrow
- that either will rain today or tomorrow, but not both.

2 - In a single toss of 2 fair (evenly-weighted) six-sided dice, find the probability that their sum will be at most 8?

LAPLACE'S RULE

IF THE SAMPLE SPACE IS FINITE AND ALL ELEMENTAR EVENTS ARE EQUALLY LIKELY, THEN

$$P(A) = \frac{\text{\# ELEMENTS IN A}}{\text{\# ELEMENTS TOTAL}}$$

THIS IS ACTUALLY AN EXTREMELY SIMPLIFIED VERSION OF LAPLACE'S RULE, WHICH STATES THAT, UNDER VERY GENERAL CONDITIONS, THE PROBABILITY OF A FURTHER SUCCESS OVER A RUN OF N TRIALS WITH S SUCCESSES SO FAR IS S/N

EXERCISES

A DECK OF CARDS HAS 52 CARDS, DIVIDED IN 4 SUITS, WITH 13 CARDS EACH

IF YOU PICK A CARD AT RANDOM, WHAT IS THE PROBABILITY OF GETTING

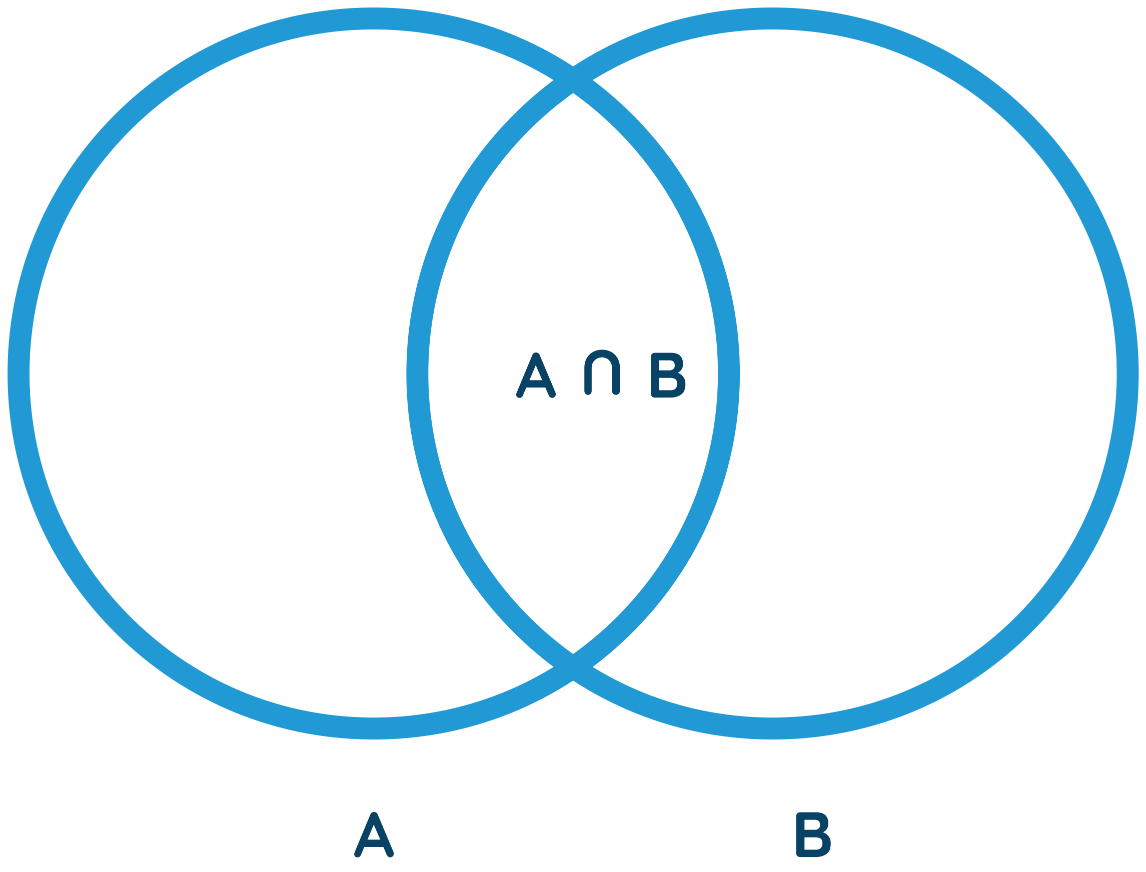
- THE JACK OF HEARTS
- ANY OF THE 4 JACKS
- ANY CARD FROM THE HEARTS SUIT

AFTER TAKING A CARD, YOU NOTICE IT'S THE JACK OF HEARTS. YOU PUT IT IN YOUR POCKET AND TAKE ANOTHER CARD. WHAT IS THE PROBABILITY OF GETTING

- THE JACK OF HEARTS
- THE KING OF HEARTS
- ANY JACK
- ANY KING
- ANY CARD FROM THE HEARTS SUIT
- ANY CARD FROM THE DIAMONDS SUIT

INTERSECTION PROBABILITY

WHEN THERE IS BOTH A AND B



How do we know if they are independent?

P (A ∩ B) = P(A) * P(B)   (Definition)

If they are not indipendent than A or B may be the cause of the other. In that case the probability of intesecting changes.

P (A ∩ B) = P(B | A) * P(A)
P (A ∩ B) = P(A | B) * P(B)

BAYESIAN PROBABILITY

WHEN THERE IS AN A BEFORE B - CONDITIONAL PROBABILITY

Suppose

A is the hypothesis (event that came first)

B is the event that happened after, as a consiquence

P (B|A) is the probability of getting B once you already have A

$P(B \mid A) = P(A \cap B) / P(A)$

If they are not indipendent than A or B may be the hypothesis

BAYESIAN THEOREM

IF YOU GOT B, WHAT'S THE PROBABILITY OF HAVING STARTED WITH A?

P (A)	Prior Probability - probability of A in the first palce (before B)
P (B | A)	Likelihood - probability of event B after having event A
P (B)	Marginal Probability - probability of event B at all, with or without A
P (A | B)	Posterior Probability - probability of having gotten B with A as hypothesis

$$P(A \mid B) = \frac{P(B \mid A) * P(A)}{P(B)}$$

DISJOINT VS INDEPENDANCE

Disjoint:

A and B cannot occur at the same time

- $A \cap B = 0$
- $P(A \cup B) = P(A) + P(B)$

Independant:

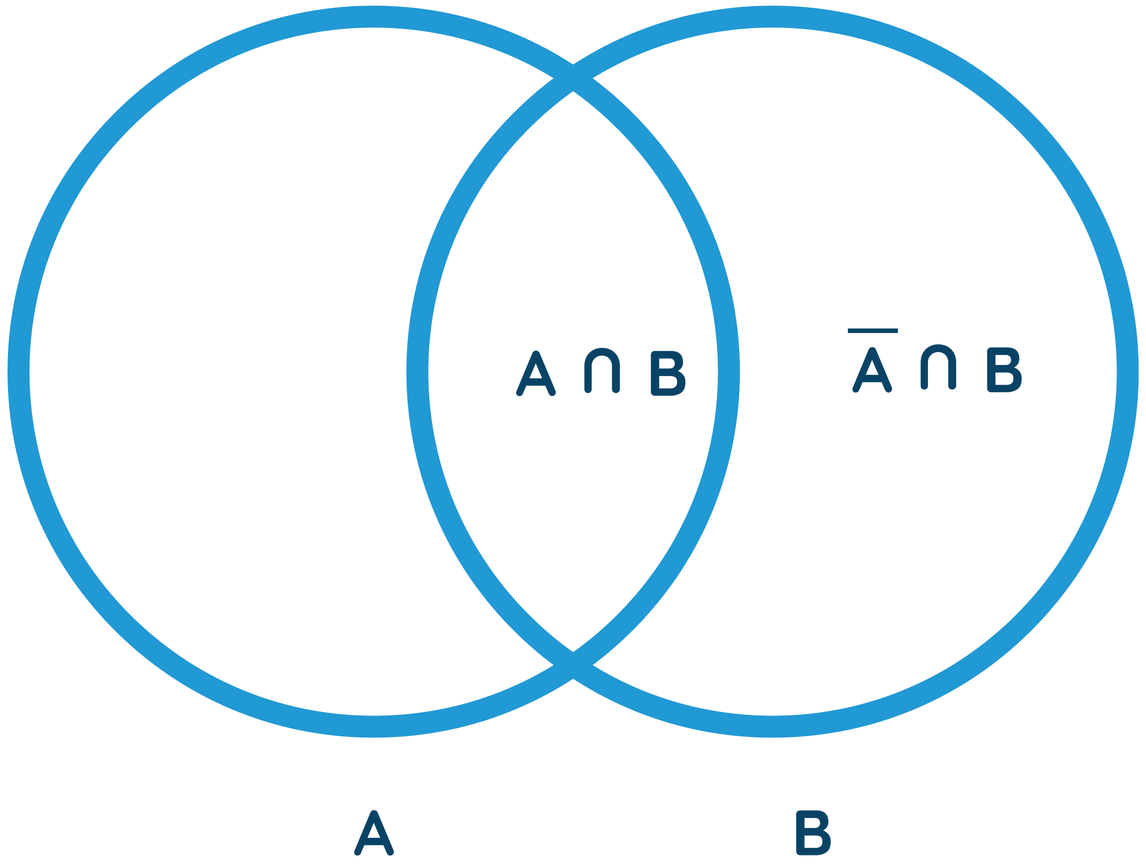
A does not give any information on B

- $P(A|B) = P(A), P(B|A) = P(B)$
- $P(A \cap B) = P(A)P(B)$

BAYESIAN THEOREM

HOW DO YOU GET MARGIN PROBABILITY P(B)?

$$P(B) = P(A \cap B) + P(\bar{A} \cap B)$$
$$= P(B \mid A) * P(A) + P(B \mid \bar{A}) * P(\bar{A})$$



Summary

Universal manipulations (always true)

$P(S) = 1$

$P(\{\}) = 0$

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

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

$P(A|B) = P(A \cap B)/P(B)$ (Definition)

$P(A|B) = P(B|A)P(A)/P(B)$ (Bayes Law)

$P(B) = P(B|A)P(A) + P(B|\bar{A})P(\bar{A})$ (Law of total probability)

If A,B are independent (written $A \perp\!\!\!\perp B$)

$P(A \cap B) = P(A)*P(B)$ (Definition)

$P(A|B) = P(A)$

If A,B are disjoint ($A \cap B = \{\}$)

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

Laplace's rule (Only if samples equiprobable)

$P(A) = |A| / |S|$