PRINCIPLES OF DATA SCIENCE

2021 - 2022

Assoc. Prof. Dr. Salha Alzahrani

Department of Computer Science

TU

جامعة الطائف

TAIF UNIVERSITY

CHAPTER 5: IMPOSSIBLE OR IMPROBABLE – A GENTLE INTRODUCTION TO PROBABILITY

- BASIC DEFINITIONS
- PROBABILITY
- BAYESIAN VERSUS FREQUENTIST
- COMPOUND EVENTS
- CONDITIONAL PROBABILITY
- THE RULES OF PROBABILITY

BASIC DEFINITIONS

Basic definitions

- A procedure is an act that leads to a result. For example,
  - throwing a dice
  - visiting a website.
- An event is a collection of the outcomes of a procedure, such as
  - getting a heads on a coin flip
  - leaving a website after only 4 seconds.
- A simple event is an outcome/event of a procedure that cannot be broken down further.
  - For example, rolling two dice can be broken down into two simple events: rolling die 1 and rolling die 2.

- The sample space of a procedure is the set of all possible simple events. For example, an experiment is performed, in which a coin is flipped three times in succession.
  - What is the size of the sample space for this experiment?
  - The answer is eight, because the results could be any one of the possibilities in the following sample space—{HHH, HHT, HTT, HTH, TTT, TTH, THH, or THT}.

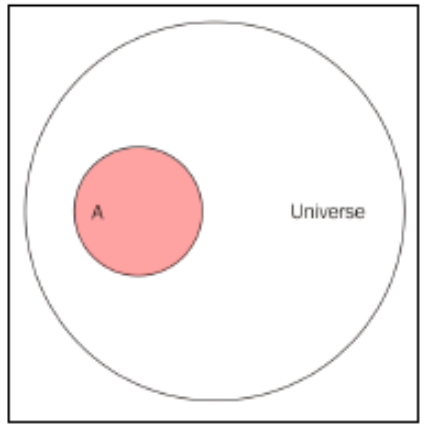
PROBABILITY

Probability

- The probability of an event represents the frequency, or chance, that the event will happen.
- For notation, if *A* is an event, *P(A)* is the probability of the occurrence of the event.
- We can define the actual probability of an event, *A*, as follows:

$$P(A) = \frac{\text{number of ways A occur}}{\text{size of sample space}}$$

- Here, *A* is the event in question. Think of an entire universe of events where anything is possible, and let's represent it as a circle. We can think of a single event, *A*, as being a smaller circle within that larger universe, as shown in the following diagram:



Probability

- The maximum probability of any event is 1. This can be understood as the red circle grows so large that it is the size of the universe (the larger circle).

- **Example 1:** Let's now pretend that our universe involves a research study on humans, and the *A* event is people in that study who have cancer. If our study has 100 people and *A* has 25 people, the probability of A or P(A) is 25/100.

- **Example 2:** Let's say we have two coins and we want the probability that we will roll two heads. First, let's define *A*. It is the event in which two heads occur. The number of ways that *A* can occur is 1. Second, The sample space of the experiment is *{HH, HT, TH, TT},* where each two-letter word indicates the outcome of the first and second coin simultaneously. The size of the sample space is four. So, *P(getting two heads) = 1/4.*

BAYESIAN VERSUS FREQUENTIST

Bayesian versus Frequentist

- In a Frequentist approach, the probability of an event is calculated through experimentation. It uses the past in order to predict the future chance of an event.
- The basic formula is as follows:

$$P(A) = \frac{\text{number of times A occurred}}{\text{number of times the procedure was repeated}}$$

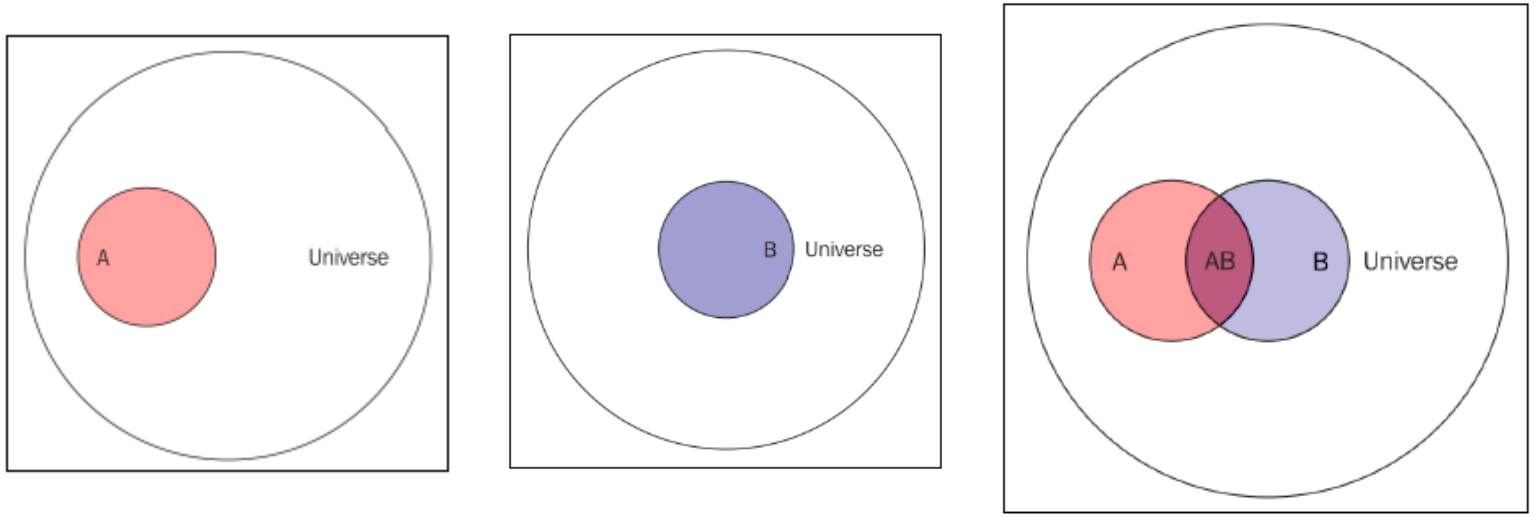
- Basically, we observe several instances of the event and count the number of times A was satisfied. The division of these numbers is an approximation of the probability.

- The Bayesian approach differs by dictating that probabilities must be discerned using theoretical means. Using the Bayes approach, we would have to think a bit more critically about events and why they occur.
- The relative frequency of an event is how often an event occurs divided by the total number of observations.

COMPOUND EVENTS

Compound events

- Sometimes, we need to deal with two or more events. These are called compound events.
- A compound event is any event that combines two or more simple events.
- When this happens, we need some special notation. Given events A and B:
  - The probability that A and B occur is P(A ∩ B) = P(A and B)
  - The probability that either A or B occurs is P(A U B) = P(A or B)



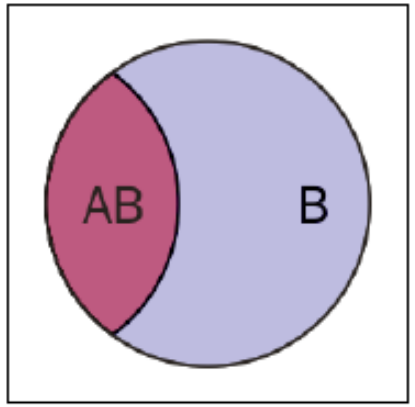
CONDITIONAL PROBABILITY

Conditional probability

- A conditional probability of A given B or P(A|B), it is asking you to calculate the probability of an event given that another event has already happened.

- The formula can be given as follows:

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



THE RULES OF PROBABILITY

The rules of probability

- **The addition rule:** The addition rule is used to calculate the probability of *either or* events.

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

- **Mutual exclusivity:** We say that two events are mutually exclusive if they cannot occur at the same time. This means that A∩B= or just that the *intersection* of the events is the empty set.

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

- **The multiplication rule:** The multiplication rule is used to calculate the probability of *and* events.

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

- **Independence:** Two events are independent if one event *does not* affect the outcome of the other, that is P(B|A) = P(B) and P(A|B) = P(A). If two events are independent, then: P(A ∩ B) = P(A) · P(B|A) = P(A) · P(B)

- **Complementary events:** The complement of A is the *opposite* or *negation* of A.

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

s.zahrani@tu.edu.sa