Probability Theory & Bayes Theorem

Mahit Kumar Paul
Lecturer, Dept. of CSE
RUET, Rajshahi-6204

mahit.cse@gmail.com

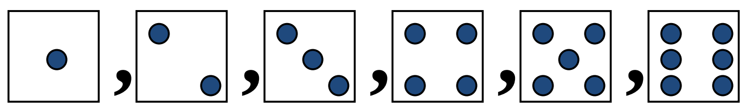
Sample Space, Sample Points, Events

- **Sample space** Ω is the set of all possible **sample points** $\omega \in \Omega$
  - **Example 0**. Tossing a coin: $\Omega = \{H, T\}$
  - **Example 1**. Casting a die: $\Omega = \{1,2,3,4,5,6\}$
  - **Example 2**. Number of customers in a queue: $\Omega = \{0,1,2,...\}$
  - **Example 3**. Call holding time (e.g. in minutes): $\Omega = \{x \in \Re \mid x > 0\}$
- **Events** $A, B, C, ... \subset \Omega$ are measurable subsets of the sample space Ω
  - **Example 1**. "Even numbers of a die": $A = \{2,4,6\}$
  - **Example 2**. "No customers in a queue": $A = \{0\}$
  - **Example 3**. "Call holding time greater than 3.0 (min)": $A = \{x \in \Re \mid x > 3.0\}$

Examples of Sample Space

1. Tossing a coin – Outcomes $\boldsymbol{\Omega}$ ={**Head, Tail**}

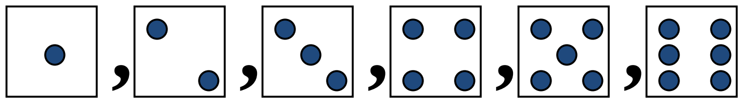
2. Rolling a die – outcomes

   $\boldsymbol{\Omega}$= {  }

   ={1, 2, 3, 4, 5, 6}

Example of Events

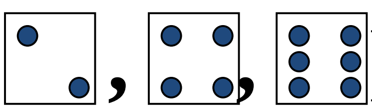
Rolling a die – Outcomes

$\Omega = \{$  $\}$

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

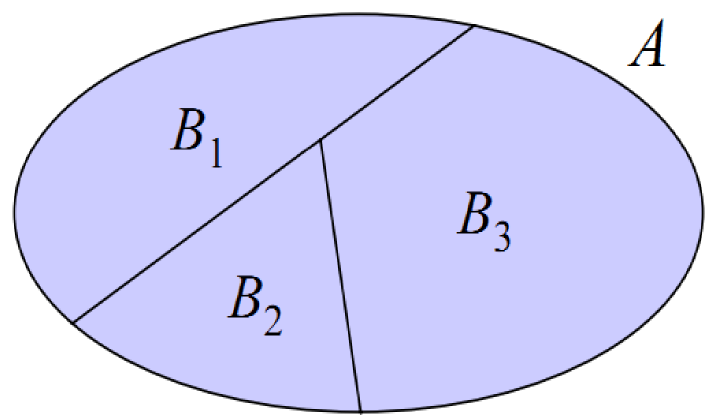
E = The event that an even number is rolled

$= \{2, 4, 6\}$

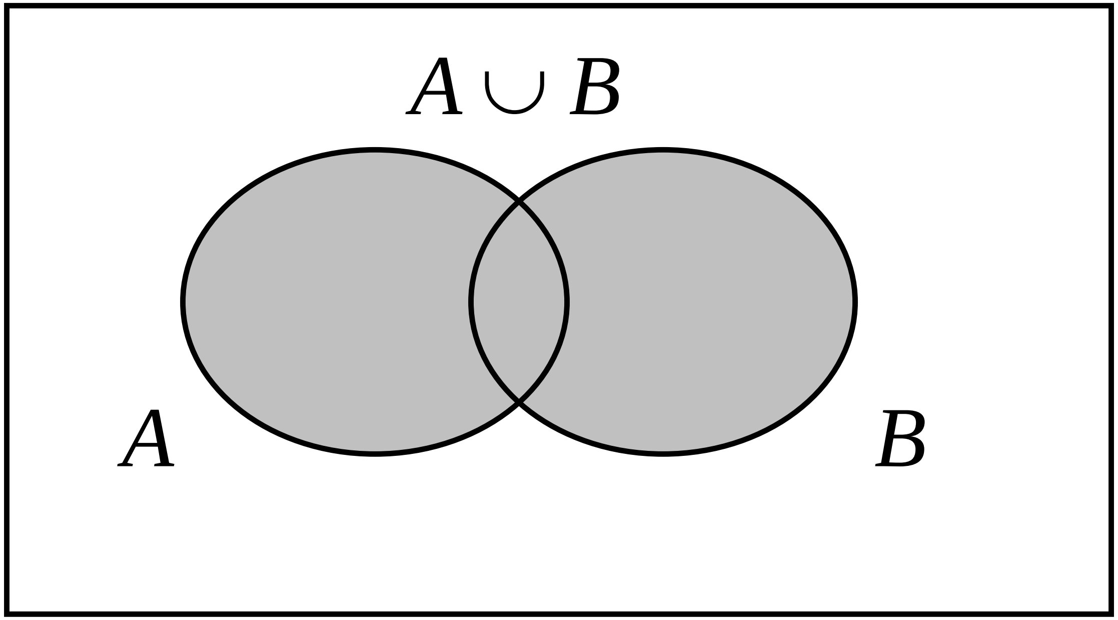
$= \{$  $\}$

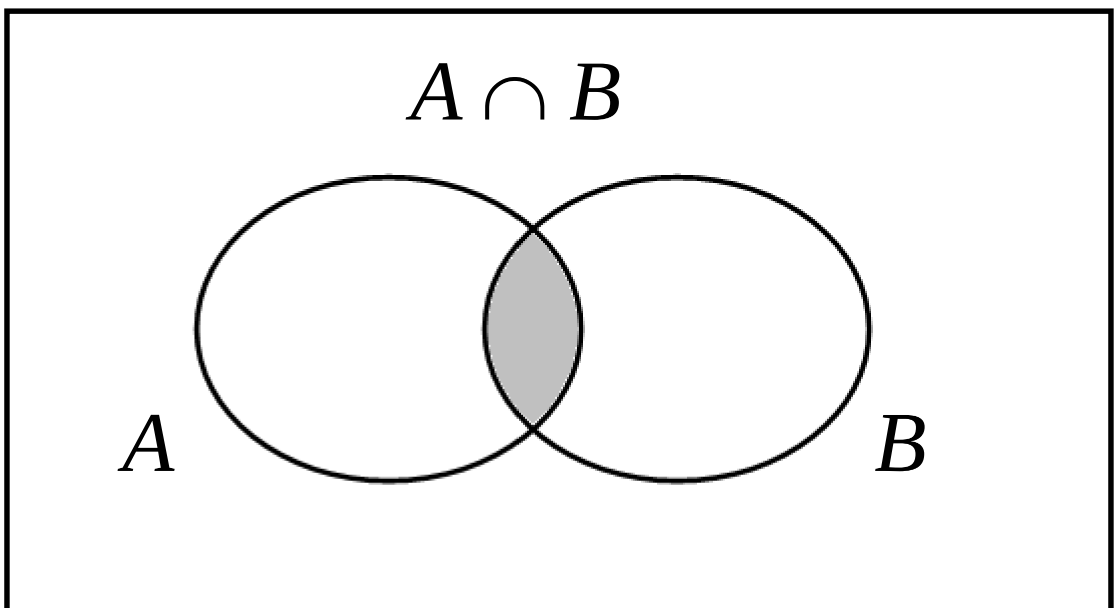
Combination of Events

- **Union** "A or B": $A \cup B = \{\omega \in \Omega \mid \omega \in A \text{ or } \omega \in B\}$
- **Intersection** "A and B": $A \cap B = \{\omega \in \Omega \mid \omega \in A \text{ and } \omega \in B\}$
- **Complement** "not A": $A^c = \{\omega \in \Omega \mid \omega \notin A\}$
- Events A and B are **disjoint** if
  - $A \cap B = \varnothing$
- A set of events $\{B_1, B_2, \ldots\}$ is a **partition** of event A if
  - (i) $B_i \cap B_j = \varnothing$ for all $i \neq j$
  - (ii) $\cup_i B_i = A$

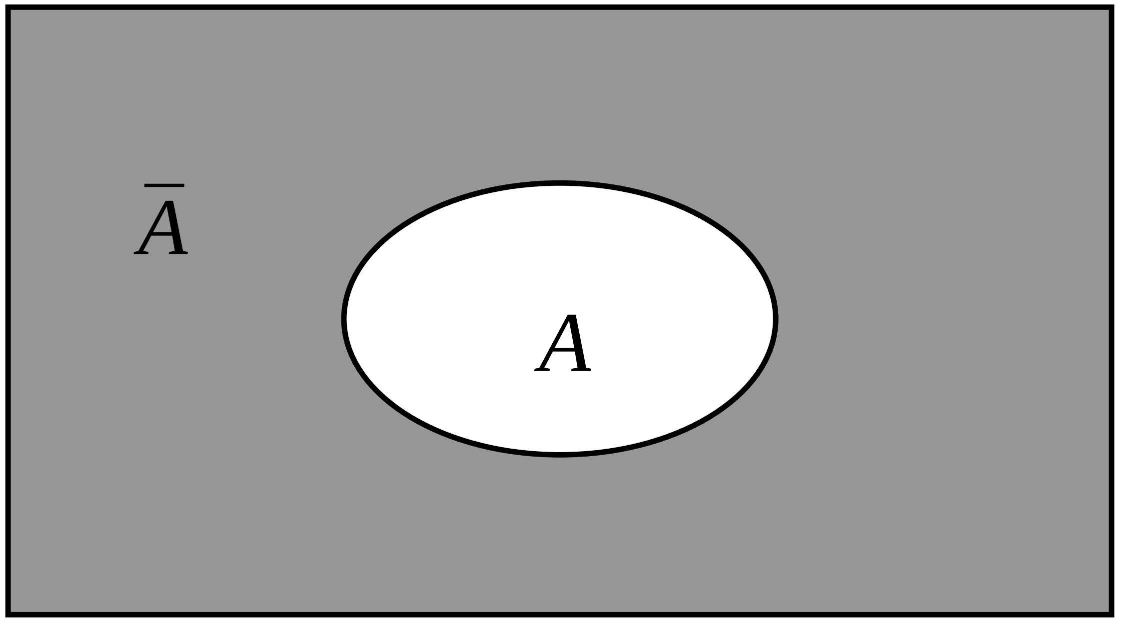


Union & Intersection





Complement



Notes

In problems you will recognize that you are working with:

1. **Union** if you see the word **or**,
2. **Intersection** if you see the word **and**,
3. **Complement** if you see the word **not**.

Probability of an Event 'E'

Suppose that the **sample space** $\boldsymbol{S = \{o_1, o_2, o_3, \ldots\ o_N\}}$ has a finite number, N, of observation points.

Also each of the observation is **equally likely** (because of symmetry).

Then for any event $\boldsymbol{E}$, *probability* is

$$P(E) = \frac{n(E)}{n(S)} = \frac{no.\,of\ observation\ points\ in\ E}{total\ no.\,of\,observation\ points}$$

Note: the symbol $n(A) =$ no. of elements of A

Probability of an Event 'E'…

Applies only to the special case when

1. The sample space has a **finite no. of observations**.
2. Each outcome is **equi-probable**.

Note: If this is not true a more general definition of probability is required.

Example

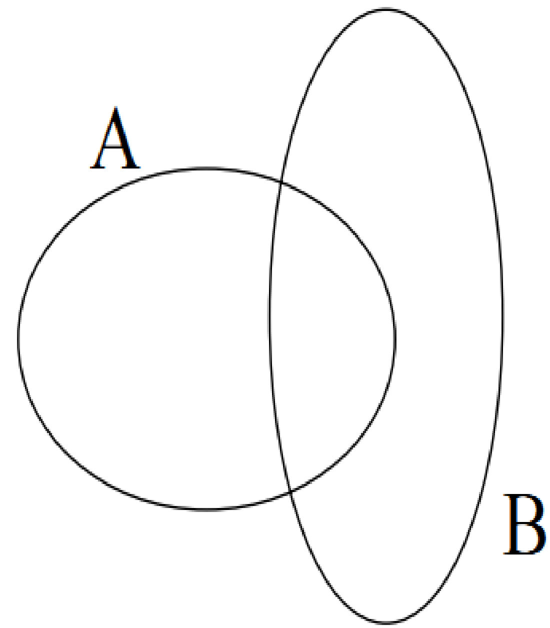
Jar #	Red	White	Blue
1	3	4	1
2	1	2	3
3	4	3	2

For Jar#1

$$P(Red) = 3/8$$

Basic Properties

- **Probability** of event A is denoted by $P(A)$, $P(A) \in [0,1]$
  - Probability measure P is thus
    a real-valued set function defined on the set of events $\mathfrak{I}$, $P: \mathfrak{I} \to [0,1]$
- **Properties**:
  - (i) $0 \le P(A) \le 1$
  - (ii) $P(\varnothing) = 0$
  - (iii) $P(\Omega) = 1$
  - (iv) $P(A^c) = 1 - P(A)$
  - (v) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
  - (vi) $A \cap B = \varnothing \Rightarrow \; P(A \cup B) = P(A) + P(B)$
  - (vii) $\{B_i\}$ is a partition of $A \Rightarrow \; P(A) = \Sigma_i P(B_i)$
  - $(viii)$ $A \subset B \Rightarrow P(A) \le P(B)$



Joint Probability

- A joint probability is a statistical measure that calculates the likelihood of two events **occurring together** and **at the same point of time**.
- Joint probability is the probability of event Y occurring at the same time event X occurs.
- Notation for joint probability takes the form:

$$\mathbf{P(X \cap Y)} \text{ or } \mathbf{P(X \text{ and } Y)} \text{ or } \mathbf{P(XY)} \text{ or } \mathbf{P(X,Y)}$$

which reads as the joint probability of X and Y

- The symbol "∩" in a joint probability is referred to as an **intersection**. The probability of event X and event Y happening is the same thing as the point where X and Y intersect. Therefore, **joint probability** is also called the **intersection of two or more events**.

Joint Probability…

Example: From a deck of 52 cards, the joint probability of picking up a card that is both red and 6 is

$$P(6 \cap \text{red}) = 2/52 = 1/26$$

Can be also calculated as:

$$\begin{aligned} P(6 \cap \text{red}) &= P(6)*P(\text{red}) \\ &= (4/52) * (26/52) \\ &= (1/13)*(1/2) \\ &= 1/26 \end{aligned}$$

If the probability of one event doesn't affect the other, you have an independent event. **Then multiply the probability of one by the probability of another.

Joint Probability…

Clubs:	A♣	2♣	3♣	4♣	5♣	6♣	7♣	8♣	9♣	10♣	J♣	Q♣	K♣
Diamonds:	A♦	2♦	3♦	4♦	5♦	6♦	7♦	8♦	9♦	10♦	J♦	Q♦	K♦
Hearts:	A♥	2♥	3♥	4♥	5♥	6♥	7♥	8♥	9♥	10♥	J♥	Q♥	K♥
Spades:	A♠	2♠	3♠	4♠	5♠	6♠	7♠	8♠	9♠	10♠	J♠	Q♠	K♠

	6	others	
Red	2	24	26
Black	2	24	26
Total	4	48	52

	6	others	
Red	1/26	6/13	½
Black	1/26	6/13	½
Total	1/13	12/13	1

Total Probability

- Let $\{B_i\}$ be a partition of the sample space Ω
- It follows that $\{A \cap B_i\}$ is a partition of event A. Thus

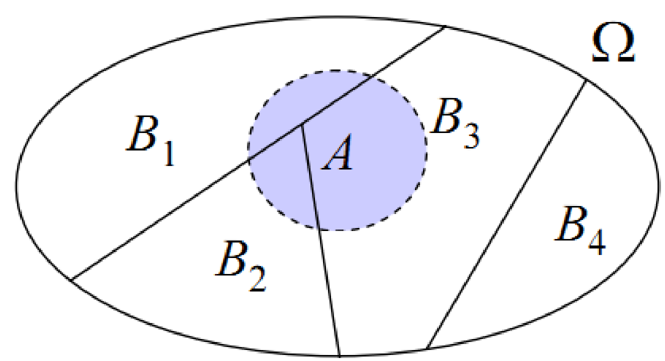
[redacted]

$$P(A) = \sum_i P(A \cap B_i)$$

- Assume further that $P(B_i) > 0$ for all i. Then

$$P(A) = \sum_i P(B_i) P(A \mid B_i)$$

- This is the **theorem of total probability**



i.e. P(A) = P(B1) + P(B2) + P(B3)

Conditional Probability

We will start through example:

Let an ordinary six sided die is to be rolled once.

What will be the sample space?

Equally likely samples

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

What is the probability that a 2 is rolled?

$$P(2) = 1/6$$

Conditional Probability…

Suppose, the die came up with an even number. And, no other info is given.

What will be the reduced sample space?

Equally likely samples

$S\_r = \{2,4,6\}$

Given an even number is rolled, what is the probability that it is 2?

It's a conditional probability case.

$$P(2|\text{Even}) = 1/3$$

Conditional Probability…

Frequently, before observing the outcome of a random experiment you are given information regarding the outcome

- Assume that $P(B) > 0$
- **Definition**: The **conditional probability** of event A **given** that event B occurred is defined as

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

- It follows that

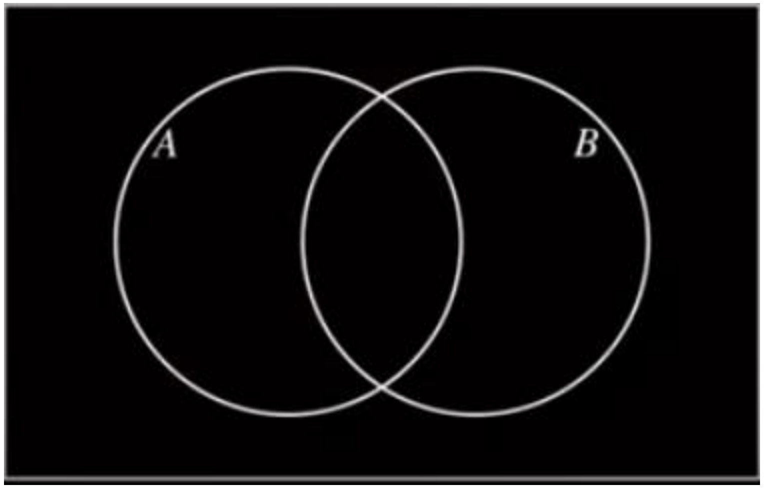
$$P(A \cap B) = P(B)P(A \mid B) = P(A)P(B \mid A)$$

Conditional Probability…

How does this definition make sense?

Here, rectangle is total sample space.

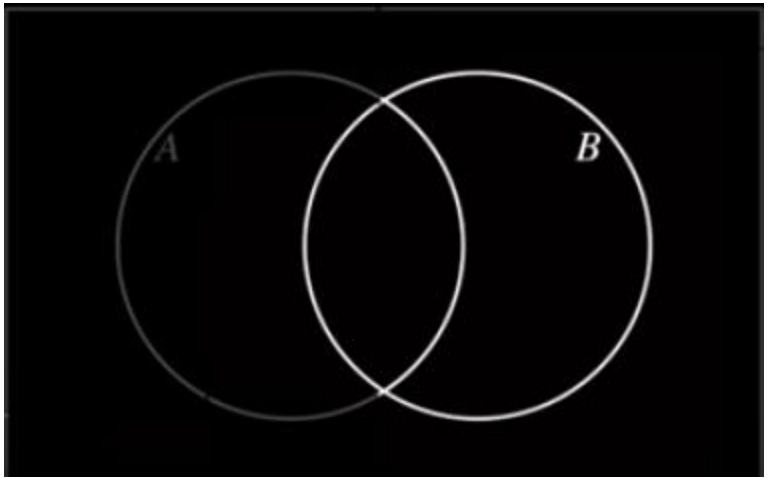
A and B are two individual event sets.



Conditional Probability…

We would like to know the probability of A given that B has occurred.

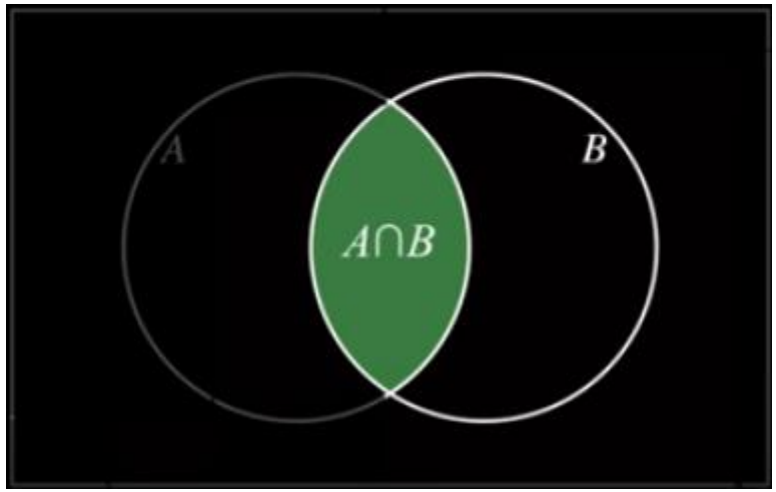
So, the sample space has been reduced to just event B.



If event B has occurred i.e. we are in circle B, then what is the probability that A has occurred.

Conditional Probability…

Event A shares the green with B. i.e. A ∩ B. That's definition of conditional probability make sense.



Conditional Probability…

Finding solution using classical probability:

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

Back to rolling a die. Consider the two events:

$$A = \{1, 2, \underline{3, 4, 5}\}$$

$$B = \{\underline{3, 4, 5}, 6\} \leftarrow$$ Reduced sample space

$$P(A|B) = \frac{3}{4}$$

What is the conditional probability of A, given B?

so the probability of event A, given event B has occurred, is 3 out of 4.

Conditional Probability…

Finding solution using Conditional Probability:

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

$$A \cap B = \{3, 4, 5\}$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{3/6}{4/6} = \frac{3}{4} \qquad P(A \cap B) = \frac{3}{6}$$

Note: Here the unconditional probability of A is 5/6

Conditional Probability…

Example Where the sample points are not equally likely.

A certain $1 lottery ticket has the following payouts:

Payout	0	1	10	100
Probability	0.689	0.301	0.009	0.001

Given the payout is more than $1, what is the probability it is $10?

Conditional Probability...

Finding solution using classical probability:

Payout	0	1	10	100
Probability	0.689	0.301	0.009	0.001

Reduced sample space

Result is:

$$\frac{0.009}{0.009+0.001} = 0.9$$

Conditional Probability…

Finding solution using Conditional probability:

A: Payout is more than \$1 $\quad A = \{10,100\}$

B: Payout is \$10 $\quad B = \{10\}$

P(A) = (0.009+0.001)/1 = 0.01

P(B) = (0.009)/1 = 0.009

$$P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{0.009}{0.010} = 0.9 \qquad B \cap A = \{10\}$$

P(B∩A) = (0.009)/1 = 0.009

Conditional Probability…

A card is drawn from a well-shuffled ordinary 52 card deck. What is the probability it is a jack?

Clubs:	A♣	2♣	3♣	4♣	5♣	6♣	7♣	8♣	9♣	10♣	J♣	Q♣	K♣
Diamonds:	A♦	2♦	3♦	4♦	5♦	6♦	7♦	8♦	9♦	10♦	J♦	Q♦	K♦
Hearts:	A♥	2♥	3♥	4♥	5♥	6♥	7♥	8♥	9♥	10♥	J♥	Q♥	K♥
Spades:	A♠	2♠	3♠	4♠	5♠	6♠	7♠	8♠	9♠	10♠	J♠	Q♠	K♠

J: Drawing a jack $\qquad P(J) = \frac{4}{52} = \frac{1}{13}$

Conditional Probability…

A card is drawn from a well-shuffled ordinary 52 card deck. If the card is red, what is the probability it is a jack?

**Finding solution using classical probability:

Clubs:	A♣	2♣	3♣	4♣	5♣	6♣	7♣	8♣	9♣	10♣	J♣	Q♣	K♣
Diamonds:	A♦	2♦	3♦	4♦	5♦	6♦	7♦	8♦	9♦	10♦	J♦	Q♦	K♦
Hearts:	A♥	2♥	3♥	4♥	5♥	6♥	7♥	8♥	9♥	10♥	J♥	Q♥	K♥
Spades:	A♠	2♠	3♠	4♠	5♠	6♠	7♠	8♠	9♠	10♠	J♠	Q♠	K♠

R: Drawing a red card

$$P(J|R) = \frac{2}{26} = \frac{1}{13}$$

Conditional Probability…

Finding solution using Conditional probability:

$$P(J|R) = \frac{P(J \cap R)}{P(R)} = \frac{2/52}{26/52} = \frac{1}{13}$$

$$P(J \cap R) = 2/52$$

In this example, **Unconditional** and **Conditional**, both probabilities are 1/13. i.e. they are same.

Conditional Probability…

In this example,

$$P(J|R) = P(J)$$

We say that the events:

J: The card is a jack
R: The card is red

are *independent*.

Conditional Probability…

- Three jars contain colored balls as described in the table below.
  - One jar is chosen at random and a ball is selected. If the ball is red, what is the probability that it came from the 2nd jar?

Jar #	Red	White	Blue
1	3	4	1
2	1	2	3
3	4	3	2

Conditional Probability…

- We will define the following events:
  - J_1 is the event that *first* jar is chosen
  - J_2 is the event that *second* jar is chosen
  - J_3 is the event that *third* jar is chosen
  - R is the event that a *red* ball is selected

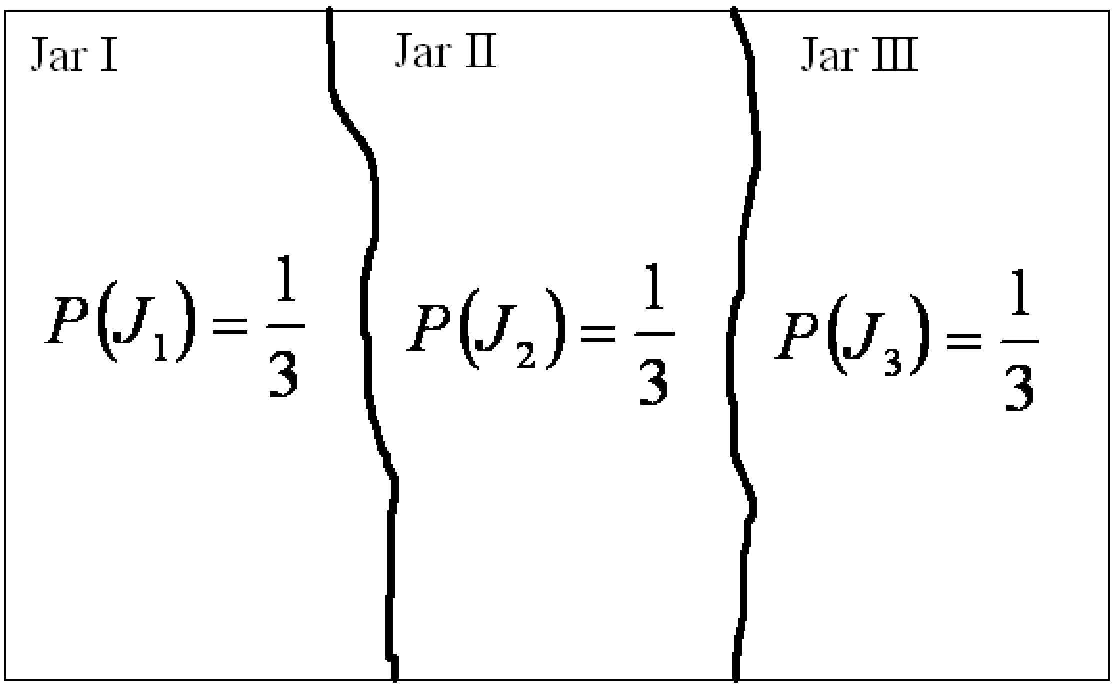
Conditional Probability…

- The events J_1 , J_2 , and J_3 are mutually exclusive
  - Why?
    - *You can't chose two different jars at the same time*

- Because of this, our sample space has been divided or *partitioned* along these three events

Venn Diagram

- Let's look at the Venn Diagram



Computing Probabilities

$$P(J_2|R) = \frac{P(J_2 \cap R)}{P(R)}$$

$$= \frac{P(J_2 \cap R)}{P(J_1 \cap R) + P(J_2 \cap R) + P(J_3 \cap R)}$$

$$= \frac{\frac{1}{6}}{\frac{3}{8} + \frac{1}{6} + \frac{4}{9}} = \frac{12}{71}$$

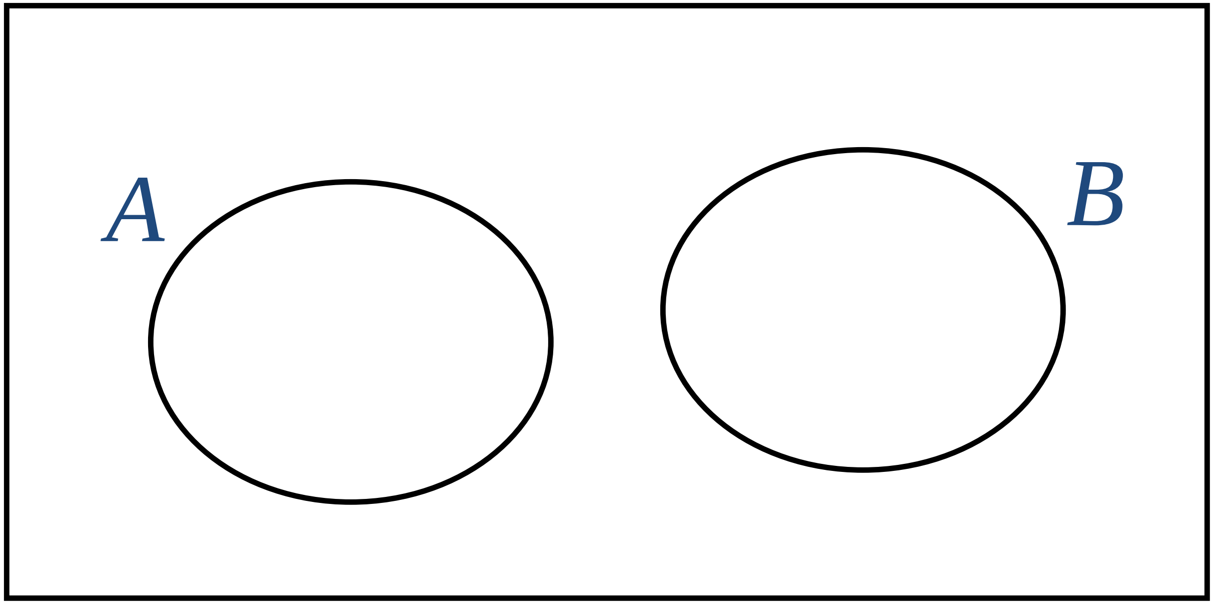
$$= 0.1690140845 \approx 0.17$$

Mutually Exclusive/Disjoint

Two events ***A*** and ***B*** are called **Mutually Exclusive/ Disjoint** if:

$$A \cap B = \phi$$

- They have **no outcomes in common**.
- They can't occur at the same time.



Mutually Exclusive/Disjoint…

Example: What is the probability of choosing one card from a standard deck and getting either a Queen of Hearts or Ace of Hearts?

Since you can't get both cards with one draw, add the probabilities:

P(Queen of Hearts or Ace of Hearts) = P(Queen of Hearts) + P(Ace of Hearts) = 1/52 + 1/52
= 2/52.

Clubs:	A♣	2♣	3♣	4♣	5♣	6♣	7♣	8♣	9♣	10♣	J♣	Q♣	K♣
Diamonds:	A♦	2♦	3♦	4♦	5♦	6♦	7♦	8♦	9♦	10♦	J♦	Q♦	K♦
Hearts:	A♥	2♥	3♥	4♥	5♥	6♥	7♥	8♥	9♥	10♥	J♥	Q♥	K♥
Spades:	A♠	2♠	3♠	4♠	5♠	6♠	7♠	8♠	9♠	10♠	J♠	Q♠	K♠

*The probability of two disjoint events A or B happening is:

P(A or B) = P(A) + P(B).

Independence of Events

Two events *A* and *B* are called **independent** if occurrence of one does not affect the others.

$$P(A \cap B) = \mathrm{P(A)^*P(B)}$$

Example: From a deck of 52 cards, the probability of picking up a card that is both red and 6 is:

$$\begin{aligned} \mathrm{P}(6 \cap \mathrm{red}) &= \mathrm{P}(6)^*\mathrm{P}(\mathrm{red}) \\ &= (4/52) * (26/52) \\ &= 1/26 \end{aligned}$$

Clubs:	A♣	2♣	3♣	4♣	5♣	6♣	7♣	8♣	9♣	10♣	J♣	Q♣	K♣
Diamonds:	A♦	2♦	3♦	4♦	5♦	6♦	7♦	8♦	9♦	10♦	J♦	Q♦	K♦
Hearts:	A♥	2♥	3♥	4♥	5♥	6♥	7♥	8♥	9♥	10♥	J♥	Q♥	K♥
Spades:	A♠	2♠	3♠	4♠	5♠	6♠	7♠	8♠	9♠	10♠	J♠	Q♠	K♠

Independence of Events…

Note

$$if\ P(A) \neq 0\ and\ P(B) \neq 0\ then$$

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

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

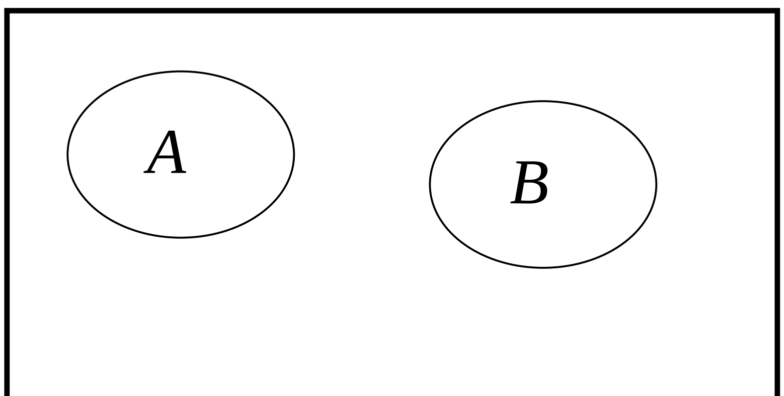
Thus in the case of independence the Conditional Probability of an event is not affected by the knowledge of the other event

Independent Events Versus Mutually Exclusive Events

Mutually Exclusive

Two Mutually Exclusive events are Independent only in the special case where

$$P(A) = 0 \; and \; P(B) = 0; \; Also \; P(A \cap B) = 0$$



Mutually Exclusive events are highly Dependent. A and B **cannot** occur simultaneously. If one event occurs the other event does not occur.

Dependent Events

- When two events are dependent, one event influences the probability of the other.

- A dependent event is an event that relies on **another event to occur first**.

- If A and B are dependent events then

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

Dependent Events…

Example: Suppose you have 52 candidates for a committee. Four are persons aged 18 to 21. If you randomly select one person, and then (without replacing the first person) randomly select a second person, what is the probability both persons will be between 18 and 21.

Let, choosing 1^{st} one is event A and 2^{nd} one is event B.

Step-1: 4 are aged between 18 and 21. Hence, $P(A) = 4/52 = 1/13$

Step-2: $P(B|A) = 3/51 = 1/17$

Step-3: $P(A \cap B) = P(A) * P(B|A) = 1/13 * 1/17 = 1/221$

Bayes Theorem

For a joint probability distribution **P(A∩B)** over events A and B, the conditional probability of A given B is defined as

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \text{if P(B)} \neq 0$$

Also, probability of B given A is defined as

$$P(B|A) = \frac{P(B \cap A)}{P(A)}, \text{if P(A)} \neq 0$$

$$\text{or}, P(B \cap A) = P(B|A) * P(A)$$

$$\text{or}, P(A \cap B) = P(B|A) * P(A)$$

Therefore,

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

Bayes Theorem…

Likelihood of Evidence B If A is True

Prior Probability of A

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

Posterior Probability of A Given Evidence B

The Probability of B Being True

Bayes Theorem-Example…

What is the probability of two girls given at least one is girls?

$$p(2g|1g) = \frac{p(1g|2g) * p(2g)}{p(1g)}$$

$$= \frac{1 * \frac{1}{4}}{\frac{3}{4}} = \frac{1}{3}$$

Combination: gg, gb, bg, bb

Bayes Theorem-Example…

- Three jars contain colored balls as described in the table below.
  - One jar is chosen at random and a ball is selected. If the ball is red, what is the probability that it came from the 2nd jar?

Jar #	Red	White	Blue
1	3	4	1
2	1	2	3
3	4	3	2

Bayes Theorem-Example…

$$P(J_2|R) = \frac{P(R|J_2) * P(J_2)}{P(R)}$$

$$= \frac{P(R|J_2)*P(J_2)}{P(R\cap J_1)+P(R\cap J_2)+P(R\cap J_3)}$$

$$= \frac{P(R|J_2)*P(J_2)}{P(R|J_1)*P(J_1)+P(R|J_2)*P(J_2)+P(R|J_3)*P(J_3)}$$

$$= \frac{\frac{1}{6}*\frac{1}{3}}{\frac{3}{8}*\frac{1}{3}+\frac{1}{6}*\frac{1}{3}+\frac{4}{9}*\frac{1}{3}} = \frac{12}{71}$$

$$= 0.1690140845 \approx 0.17$$

Bayes Theorem-Example…

Similarly,

$$P(J_1|R) = \frac{27}{71} = 0.38$$

$$P(J_3|R) = \frac{32}{71} = 0.45$$

Note: From the table, it is seen that the probability obtained using Bayes Theorem makes sense. Because, according to data point, Jar-3 should have highest probability, then Jar-1 should have 2nd highest and Jar-2 should have the lowest probability.

Chain Rule

In probability theory, the **chain rule**, also called the **general product rule** permits the calculation of any member of the joint distribution of a set of random variables using only conditional probabilities. The rule is useful in the study of Bayesian networks, which describe a probability distribution in terms of conditional probabilities.

Let, consider a bunch of events $\boldsymbol{A_n, A_{n-1}, \dots, A_1}$ and if $\boldsymbol{P(A_{n-1} \cap \cdots \cap A_1) > 0}$, then the Joint Probability using the concept of Conditional Probability can be sated as Equation 1 which is called **Chain Rule**.

$$P(A_n \cap \cdots \cap A_1) = P(A_n | A_{n-1} \cap \cdots \cap A_1) . P(A_{n-1} \cap \cdots \cap A_1) \qquad \dots \dots \dots (1)$$

If n=2 then,

$$P(A_2 \cap A_1) = P(A_2 | A_1) . (A_1)$$

That is Equation 1 is the generalization of the **Joint Probability**.

Chain Rule…

Thus the Equation 1 can be written as following:

$$\mathrm{P}\left(\bigcap_{k=1}^{n} A_k\right) = \prod_{k=1}^{n} \mathrm{P}\left(A_k \,\middle|\, \bigcap_{j=1}^{k-1} A_j\right)$$

With four variables, the chain rule generates the below product of conditional probabilities:

$$P(A_4 \cap A_3 \cap A_2 \cap A_1)$$
$$= P(A_4|A_3 \cap A_2 \cap A_1).P(A_3|A_2 \cap A_1).P(A_2|A_1).P(A_1)$$

Chain Rule…

Jar#	Black	White
1	1	2
2	1	3

Suppose we pick a Jar at random and then select a ball from that Jar. Then what is the probability of choosing the 1st jar and white ball from it?

Let event A be choosing the first Jar: P(A) = 1/2.
Let event B be the chance we choose a white ball.

The chance of choosing a white ball, given that we've chosen the first Jar, is P(B|A) = 2/3.

Event A, B would be their **intersection**: choosing the first Jar and a white ball from it. The probability can be found by the chain rule for probability:

$$P(A \cap B) = P(B \cap A) = P(B|A).P(A) = \frac{2}{3} * \frac{1}{2} = \frac{1}{3}$$

Naïve Bayes Classifier

- **Naive Bayes classifiers** are a family of simple "probabilistic classifiers" based on applying Bayes' theorem with strong (naive) independence assumptions between the features.

- Naïve Bayes is a Conditional Probability model: Given a problem instance to be classified, represented by a vector $\boldsymbol{X} = < x_1, x_2, \dots x_n >$, where n = features (**independent** variables) of that object.

- The main aim of **Naïve Bayes** is to assign the instances' probabilities for each of the C_k classes as: $p(C_k | x_1, x_2, \dots x_n)$

- Using Bayes' theorem, the **Conditional Probability** can be decomposed as:

$$P(C_k | \boldsymbol{X}) = \frac{P(\boldsymbol{X} | C_k) P(C_k)}{P(\boldsymbol{X})} \qquad \dots \dots \dots (2)$$

Naïve Bayes Classifier…

- In Equation 2, the denominator does not depend on C_k. The values of the features x_i are given, so that the denominator is effectively constant.

- The numerator of Equation 2 can be written as:

$$P(\boldsymbol{X}|C_k)P(C_k) = P(x_1, x_2, \ldots, x_n|C_k)P(C_k) \quad \ldots\ldots\ldots(3)$$

Naïve Bayes Classifier…

Naïve Bayes Assumption: Features x_1 and x_2 are Conditionally Independent given the class label C_k :

$$P(x_1, x_2|C_k) = P(x_1|C_k)P(x_2|C_k)$$

More generally,

$$P(x_1, \dots, x_n\,|C_k) = \prod_{i=1}^{n} P(x_i|C_k)$$

Naïve Bayes Classifier…

Thus, from Equation 2 & 3:

$$
\begin{aligned}
P(C_k|\boldsymbol{X}) &\propto P(\boldsymbol{X}|C_k)P(C_k) \\
&\propto P(x_1, x_2, \dots, x_n|C_k)P(C_k) \\
&\propto P(C_k)P(x_1, x_2, \dots, x_n|C_k) \\
&\propto P(C_k)\prod_{i=1}^{n} P(x_i|C_k)
\end{aligned}
$$

This means that under the above independence assumptions, the conditional distribution over the class variable $\boldsymbol{C_k}$ is:

$$P(C_k|\boldsymbol{X}) = \frac{P(C_k)\prod_{i=1}^{n} P(x_i|C_k)}{P(\boldsymbol{X})}$$

Naïve Bayes Classifier-Example

age	income	student	credit_rating	buys_computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
30…40	high	no	fair	yes
>40	medium	no	fair	yes
>40	low	yes	fair	yes
>40	low	yes	excellent	no
31…40	low	yes	excellent	yes
<=30	medium	no	fair	no
<=30	low	yes	fair	yes
>40	medium	yes	fair	yes
<=30	medium	yes	excellent	yes
31…40	medium	no	excellent	yes
31…40	high	yes	fair	yes
>40	medium	no	excellent	no

Naïve Bayes Classifier-Example…

Class:

C_1: buys_computer = ‘yes’

C_2: buys_computer = ‘no’

Data sample (**Test Sample**):

X = (age<=30, Income=medium, Student=yes, Credit_rating=Fair)

Naïve Bayes Classifier-Example…

$$P(yes|X) = \frac{P(X|yes)*P(yes)}{P(X)}$$

$$P(yes|X) = \frac{P(\text{age<=30 ,income =medium, student=yes,credit\_rating=fair}|yes) * P(yes)}{P(X)}$$

$$P(yes|X) = \frac{\begin{array}{c}\text{P(age=“<30” | buys\_computer=“yes”)}*\\ \text{P(income=“medium” | buys\_computer=“yes”)}*\\ \text{P(student=“yes” | buys\_computer=“yes)}*\\ \text{P(credit\_rating=“fair” | buys\_computer=“yes”)}*P(yes)\end{array}}{P(X)}$$

$$= \frac{0.222 * 0.444 * 0.667 * 0.0.667 * 0.643}{P(x)} = \frac{\mathbf{0.028}}{P(x)}$$

Naïve Bayes Classifier-Example…

$$P(no|X) = \frac{P(X|no)*P(no)}{P(X)}$$

$$P(no|X) = \frac{P(\text{age<=30 ,income =medium, student=yes,credit\_rating=fair}|no) * P(no)}{P(X)}$$

$$P(no|X) = \frac{\text{P(age="<30" | buys\_computer="no")}* \text{P(income="medium" | buys\_computer="no")}* \text{P(student="yes" | buys\_computer="no")}* \text{P(credit\_rating="fair" | buys\_computer="no")}*P(no)}{P(X)}$$

$$= \frac{0.6 * 0.4 * 0.2 * 0.4 * 0.357}{P(x)} = \frac{\mathbf{0.007}}{P(x)}$$

Naïve Bayes Classifier-Example…

Class Decision:

X belongs to class "buys_computer = yes"

Naïve Bayes Classifier-Example…

Compute $P(X|C_i)$ for each class:

P(age=“<30” | buys_computer=“yes”) = 2/9=0.222

P(age=“<30” | buys_computer=“no”) = 3/5 =0.6

P(income=“medium” | buys_computer=“yes”)= 4/9 =0.444

P(income=“medium” | buys_computer=“no”) = 2/5 = 0.4

P(student=“yes” | buys_computer=“yes)= 6/9 =0.667

P(student=“yes” | buys_computer=“no”)= 1/5=0.2

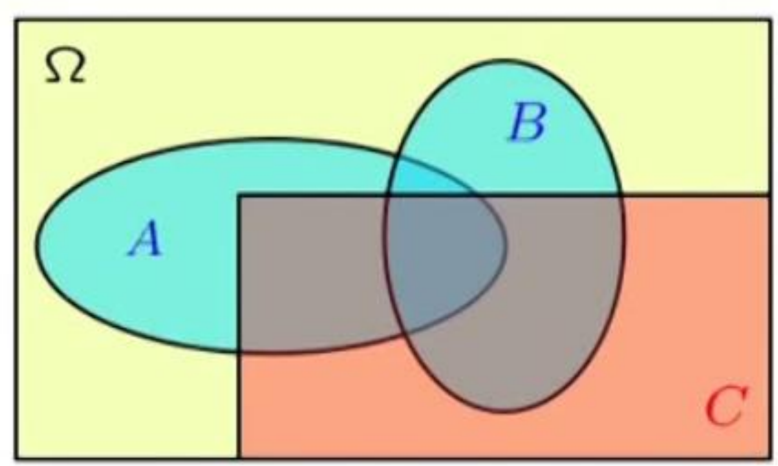
P(credit_rating=“fair” | buys_computer=“yes”)=6/9=0.667

P(credit_rating=“fair” | buys_computer=“no”)=2/5=0.4

Conditional Independence

Definition: Two events A and B are conditionally independent given an event C with P(C) > 0 if

$$P(A \cap B|C) = P(A|C)P(B|C)$$



The gist of conditional independence: **Knowing C makes A and B independent**.

Conditional Independence…

Occurrence of B and C together, doesn't hamper the occurrence of A and Vice Versa.

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

$$= \frac{P(A \cap B \mid C) \cdot P(C)}{P(B \cap C)}$$

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

$$= P(A \mid C)$$

i.e. if C is known, it doesn't hamper the independence of A and B.

Conditional Independence…

The following two equations are equivalent statements of the definition of conditional independence.

$$P(A \cap B|C) = P(A|C)P(B|C)$$

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

References

[1] https://www.youtube.com/watch?v=bgCMjHzXTXs
[2] https://brilliant.org/wiki/bayes-theorem/
[3] http://www.statisticshowto.com/probability-of-a-and-b/
[4] https://www.investopedia.com/terms/j/jointprobability.asp
[5] https://en.wikipedia.org/wiki/Chain_rule_(probability)
[6] https://en.wikipedia.org/wiki/Naive_Bayes_classifier
[7] http://www.cs.cmu.edu/~aarti/Class/10701_Spring14/slides/BayesClassifier.pdf
[8] https://www.youtube.com/watch?v=XQoLVl31ZfQ
[9] https://www.youtube.com/watch?v=k6Dw0on6NtM
[10] https://www.youtube.com/watch?v=w423ypsUHf0
[11]https://www.youtube.com/watch?v=7B3cDe39lwY&list=PLUl4u3cNGP60hI9ATjSFgLZpbNJ7myAg6&index=34

THANKS