Disclaimer: Please note that any topics that are not covered in today's lecture will be covered in the next lecture.

Content

- Conditional Probability
- Multiplication Rule
- Marginal and Joint Probability
- Law of Total Probability
- Baye's Theorem
    - Prior, Posterior and Likelihood Probabilities

WhatsApp Autocomplete Example

Conditional probability is a very important concept to understand.

In our daily life, all of you see direct examples of conditional probability. Lets look at one of them.

When typing a message on WhatsApp, we often encounter suggested words after typing a few.

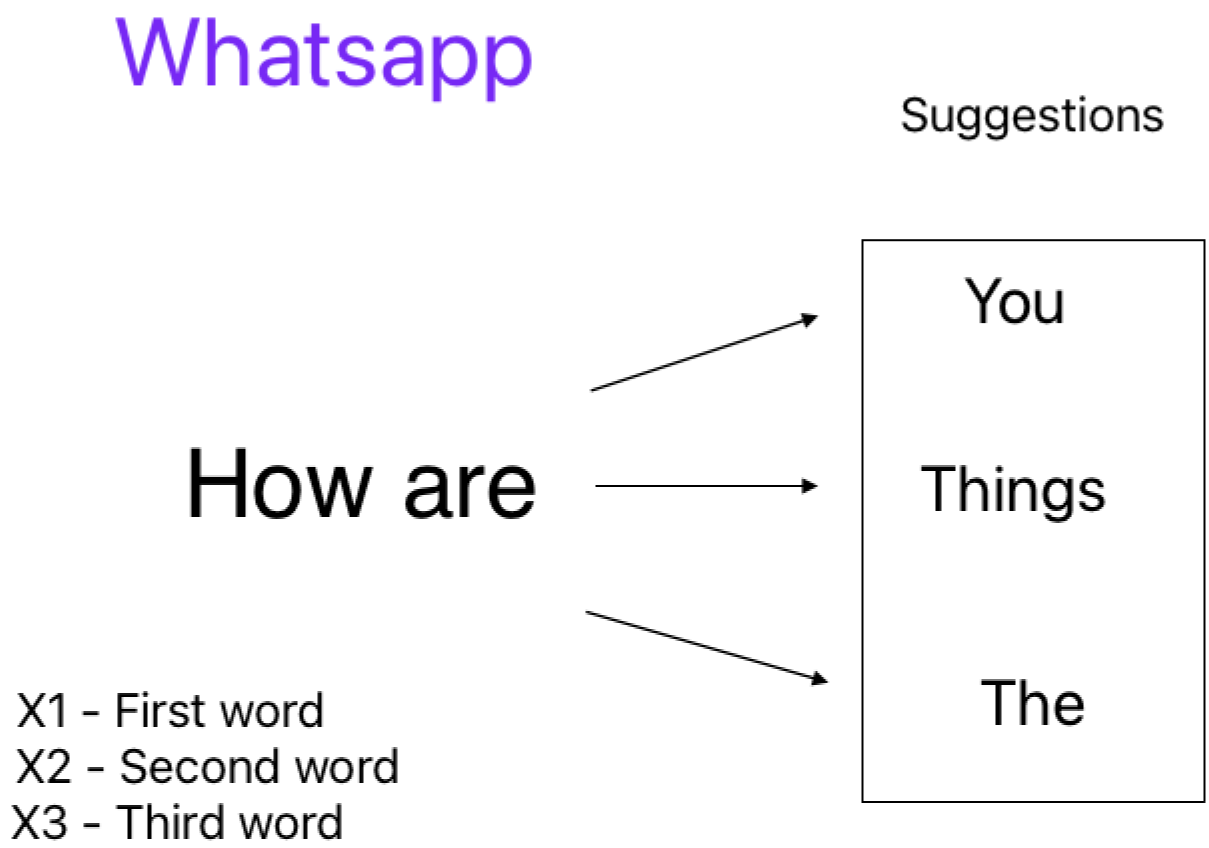
For instance, after typing `"How are"`, we might see suggestions like **"you"**, **"things"**, and **"the"**.

While these suggestions aren't guaranteed to be the next word you'll type but they're highly probable choices.

Is that magic? How did they know which words you may want to use next?

Let's assign a simple notations

- Let x_1 represents the first word
- Let x_2 represents the second word
- Let x_3 represents the third word


Whatsapp
Suggestions
How are
You
Things
The
X1 - First word
X2 - Second word
X3 - Third word

Now, you have given the following information to the keyboard:

- x_1 = "How" x_2 = "are"

Now internally, the algorithm needs to compute the probability for a word w that belongs in the dictionary, given the information about words x_1 and x_2.

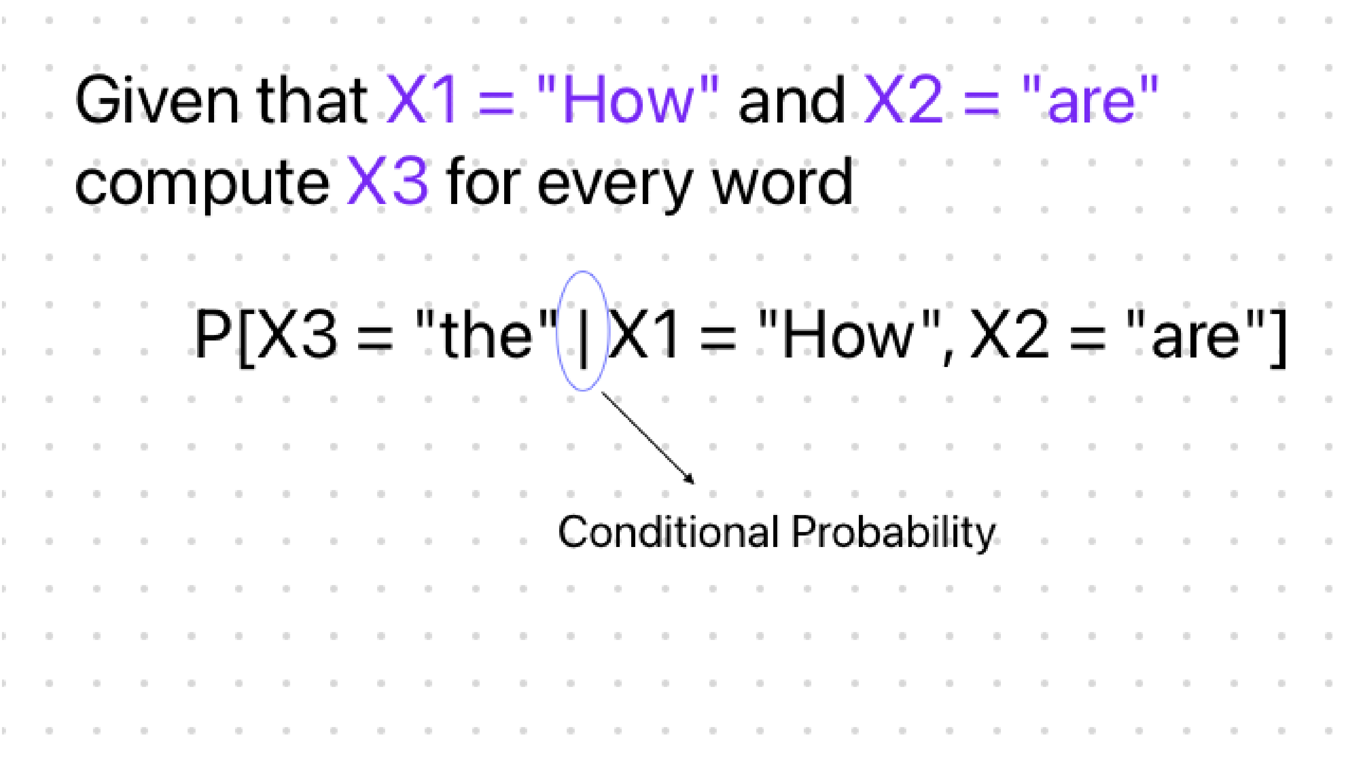
Consider this structure: $P(A|B)$

- Here, A represents the event whose probability we are trying to find
- B represents the events that have already happened / information given to us
- The vertical line | represents conditional probability

Therefore, we can represent it as:
$P(x_3 = w | x_1 = "How" \ and \ x_2 = "are")$

> Read it as:
>
> - Probability of the word x_3 given that we have seen the words x_1 and x_2.

It then presents its findings, i.e. the words that are most likely to occur (having maximum probability) given that we have seen the words x_1 and x_2.

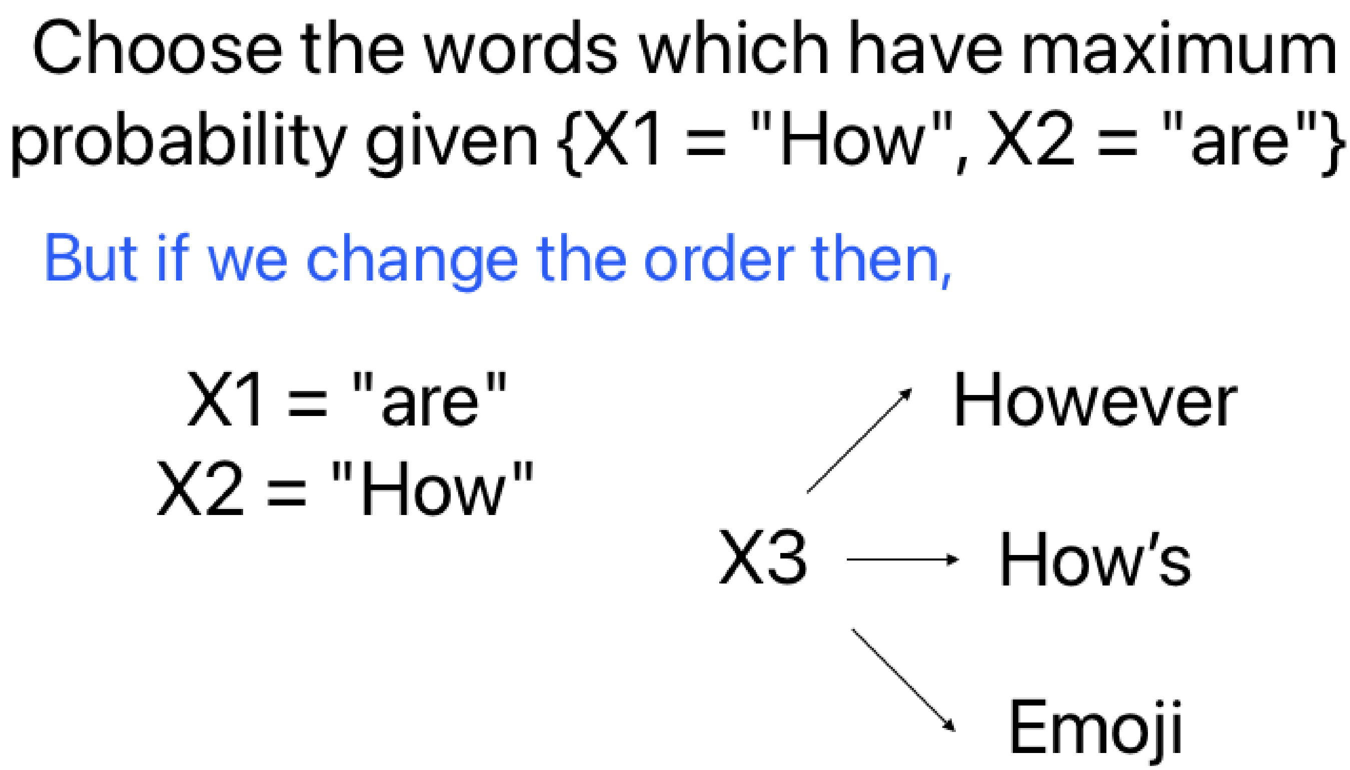


Note:

- The sequence is also important here.
- `you`, `things`, `the` are the top suggestions when $x_1=$$x_1=$ "How" and $x_2=$$x_2=$ "are".
- It would suggest different words if the case was $x_1=$$x_1=$ "are" and $x_2=$$x_2=$ "How"

Since this is not a sequence of words used very often, it might not give good suggestions here.

Auto complete is another example.



Conditional Probability

Probability of Event A, given Event B has already happened, is equivalent to the probability of A∩B , divided by probability of event B

i.e. $P(A|B) = \frac{P(A∩B)}{P(B)}$$P(A|B) = \frac{P(A∩B)}{P(B)}$

This equation is known as the **Conditional Probability Formula**

Multiplication Rule

Let's analyse this further,

From the above formula we will get:

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

In probability and statistics, this is known as the **Product / Multiplication Rule**.

Similarly, we can expand $P(B∩A) = P(B|A).P(A)$$P(B∩A) = P(B|A).P(A)$

Marginal and Joint Probabilities

Experiment: Sachin Tendulkar batting for India

Let's define the events happening here:

- $W$$W$: Sachin's team winning the match
- $C$$C$: Sachin scoring a century

Won	False	True	All
century			
False	160	154	314
True	16	30	46
All	176	184	360

1) Marginal Probability

Let's answer a few questions based on this contingency table

> **Q1.What is the probability that Sachin's team wins the match?**

We need to find $P(W) = \frac{No \ of \ matches \ won \ by \ Sachin}{Total \ no \ of \ matches} = \frac{184}{360}$$P(W) = \frac{No \ of \ matches \ won \ by \ Sachin}{Total \ no \ of \ matches} = \frac{184}{360}$

> **Q2.What is the probability of Sachin scoring a century?**

$P(C) = \frac{No \ of \ matches \ with \ century}{Total \ no \ of \ matches} = \frac{46}{360}$$P(C) = \frac{No \ of \ matches \ with \ century}{Total \ no \ of \ matches} = \frac{46}{360}$

Similarly, we can calculate $P(W^C)$$P(W^C)$ and $P(C^C)$$P(C^C)$ as well.

All of these probability values are known as **Marginal Probability**

- It is the probability of an event irrespective of the outcome of other variable.
- For instance, consider $P(W)$$P(W)$
  - It denotes the total probability of Sachin's team winning the match, considering both possibilities that Sachin may or may not score a century.
- It is not conditioned on another event. It may be thought of as an **unconditional probability.**
- Other example:
  - Probability that a card drawn is a 4 : P(four)=1/13.
    - This includes the possibility of the 4 being a spades, heart, club or diamond.
  - Probability that a card drawn is spades : P(spades)=1/4.

2) Joint Probability

Now let's look at the second type of probability values, by answering the following questions.

> **Q1.What is the probability that Sachin's team wins AND he scores a century?**

We need to find $P(W∩C) = \frac{30}{360}$$P(W∩C) = \frac{30}{360}$

> **Q2.What is the probability that Sachin scored a century AND his team wins?**

We need to find $P(C∩W)$$P(C∩W)$

This will be the same as $P(C∩W) = P(W∩C) = \frac{30}{360}$$P(C∩W) = P(W∩C) = \frac{30}{360}$

> **Q3.What is the probability that Sachin scores a century AND his team loses?**

$P(W^C ∩ C) = \frac{16}{360}$$P(W^C ∩ C) = \frac{16}{360}$

Similarly, we can find $P(W^C \cap C^C)$ and $P(W \cap C^C)$

Note:

- Here we calculated the likelihood of two events occurring **together** and at the same point in time.
- This type of probability value is known as **Joint Probability**.
- And it is represented as we saw: $P(A \cap B)$
  - Where, A and B are 2 events.
  - It is read as `Probability that event A and B happen at same time.`
- Other Example: the probability that a card is a four and red = P(four and red) = 2/52

The third kind of probability value, we've just studied, i.e. **Conditional Probability**.

Let's answer a few questions on this also

> **Q1.What is the probability that Sachin's team wins the match given that he scored a century?**

Since it is given that he scores a century, our subset reduces to the second row.

Now since we want to find the prob of team winning among these matches, our probability becomes: $P(W|C) = \frac{30}{46}$

> **Q2.What is the probability that Sachin scores a century, given that his team has won the match?**

As per the given extra information, our subset reduces to the second column.

So among these 184 matches, where India won, Sachin scored a century in only 30 matches.

Therefore $P(C|W) = \frac{30}{184}$

Similarly, we can be asked to calculate other conditional probabilities such as: $P(W|C^C)$, $P(W^C|C)$, $P(C|W^C)$, etc.

> **Q.How can we find the values of Marginal Probability?**

Law of Total Probability

- If we re-arrange the formula of conditional probability $P(A|B) = \frac{P(A \cap B)}{P(B)}$$P(A|B) = \frac{P(A \cap B)}{P(B)}$, we will get get:

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

This is known as **Law of Total Probability**

Total Probability Law Generic Formula

- Mathematically, The Law of Total Probability is stated as follows:

$P(A) = \sum_{i=1}^n P(A \mid B_i) P(B_i) $$P(A) = \sum_{i=1}^n P(A \mid B_i) P(B_i) $

Let's have a look into example

`Example: Email Spam Detection`

The Law of Total Probability helps combines the information from multiple scenarios or conditions to arrive at a comprehensive probability estimate, making it a valuable tool in various data science and machine learning applications.

Formulas learned so far

1) Conditional Probability:

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

2) Multiplication Rule:

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

3) Law of Total Probability:

- $P(A) = \sum_{i=1}^n P(A \mid B_i) P(B_i)$$P(A) = \sum_{i=1}^n P(A \mid B_i) P(B_i)$

Let's jump to new concept

Baye's Theorem

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

This equation that we used here is known as the **Bayes Theorem.**

Quick Derivation of Bayes Theorem

From the questions we have solved so far,

> **Q1. Can we say that $P(A∩B) = P(B∩A)$?**

We know that $A∩B$ and $B∩A$ represent the same subset, i.e. the common elements between A and B.

And, from the Multiplication Rule we can expand them as:

- $P(A∩B) = P(A|B).P(B)$
- $P(B∩A) = P(B|A).P(A)$

Since the LHS of both these equations is same, we can equate the RHS also.

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

Dividing both sides by $P(B)$,

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

This is exactly the equation of Baye's Theorem.

Let's take a closer look at the Bayes equation.

Prior, Posterior and Likelihood Probabilities

It consists of 4 parts:

- **Posterior probability** (updated probability after the evidence is considered)
- **Prior probability** (the probability before the evidence is considered)
- **Likelihood** (probability of the evidence, given the belief is true)
- **Marginal probability** (probability of the evidence, under any circumstance)

$$\underbrace{\boxed{P(A|B)}}_{\text{posterior}} = \underbrace{\boxed{P(A)}}_{\text{prior}} \times \frac{\boxed{P(B|A)}\ \text{likelihood}}{\boxed{P(B)}\ \text{marginal}}$$

The equation: Posterior = Prior x (Likelihood over Marginal probability)

To understand this better, let's think in a different context.

Consider 2 events:

- **Hypothesis** (which can be true or false), and
- **Evidence** (which can be present or absent).

Therefore, we can write bayes theorem as follows:

$$P(Hypothesis \mid Evidence) = P(Hypothesis) \times \frac{P(Evidence \mid Hypothesis)}{P(Evidence)}$$

Let's understand the different terms here.

- **Posterior probability**
  - The Bayes' Theorem lets you calculate the posterior (or "updated") probability.
  - It is the conditional probability of the **hypothesis being true, if the evidence is present.**
  - $P(Hypothesis \mid Evidence)$$P(Hypothesis \mid Evidence)$
- **Prior Probability**
  - Can be perceived as your **belief in the hypothesis before seeing the new evidence.**

  - Therefore, if we have a strong belief in the hypothesis already, the prior probability will be large.
  - $P(Hypothesis)$$P(Hypothesis)$

- **Likelihood**
  - The prior is multiplied by a fraction.
  - Think of this as the "strength" of the evidence.
  - The posterior probability is greater when the top part (numerator) is big, and the bottom part (denominator) is small.
  - The numerator is the likelihood.
  - It is the conditional probability of the **evidence being present, given the hypothesis is true.**