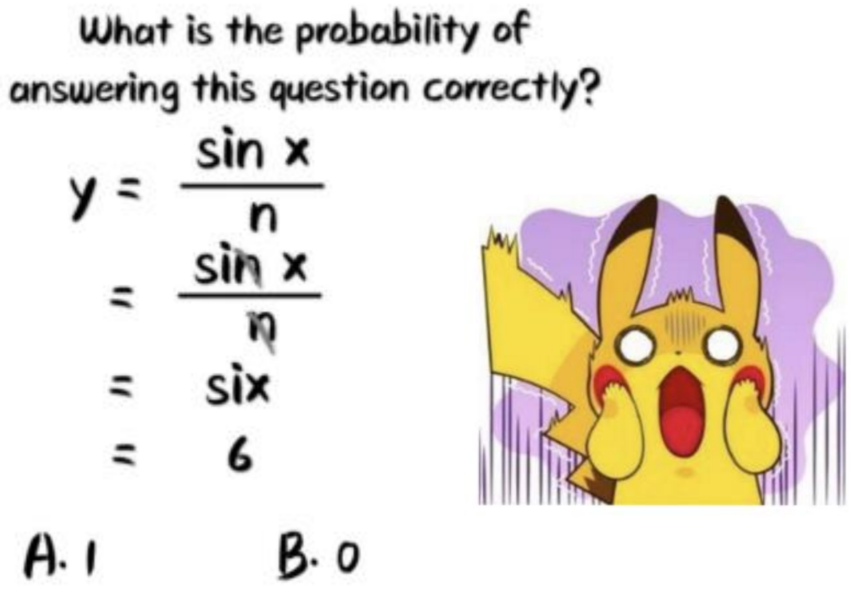
5CS037-Concepts and Technologies of AI.
Week03, Lecture03

What, Why and How of Probability.

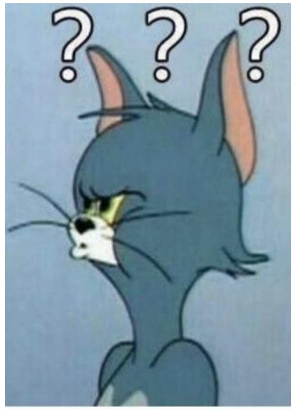
Introduction to Probability for Machine Learning.

Siman Giri

1. What, Why and How of Probability.



1.1 What is Probability?

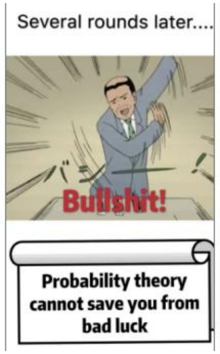


- We make decisions based on uncertainty everyday.
  - Should you buy an **extended warranty** for your new mac-book?
  - Should you allow **45 min.** to get to your 9:00 AM class or **35 min.** is enough?
  - If an artificial heart has four key parts, how likely is that **at least one** will fail?
- Thus, in a **process** where **multiple** possible **outcomes** are ***possible***; How do you make decision?
  - When the process is repeated a large number of times, each outcome occurs with a ***relative frequency***, or ***probability***
  - If a particular outcome occurs more often, we say it is more probable.
- Is probability **the mathematics**; used to ***quantify*** the ***likelihood*** of an **event** may occur….
  - Probability is the systematic study of uncertainty.

1.2 Why Probability?

- The term probability is the measure of **one's belief** in the **occurrence** of a **future event**.
  - This is a meaningful and practical interpretation of probability but we further seek a ***clearer understanding of its context*** and ***how it is measured*** or ***it is estimated*** and ***how it helps in making inferences?***
- Probability arises in two contexts
  - In actual repeated experiments
    - Example: You record the color of 1,000 cars driving by. 57 of them are green. You estimate the probability of a car being green as 57/1,000 = 0.057.
  - In idealized conceptions of a repeated process
    - Example: You consider the behavior of an unbiased six-sided die. The expected probability of rolling a 5 is 1/6 = 0.1667.
    - Example: You need a ***model*** for how people's heights are ***distributed***. You choose a normal distribution to represent the expected relative probabilities.

Probability theory provides a mathematical framework for representing and quantifying uncertain quantities.

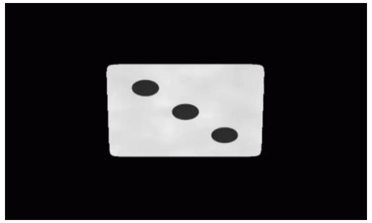


1.3 How are Probability determined?

- Probability statements are frequently encountered in newspapers, magazines, advertising etc.
- What is the basis for these probability statements?
    - In most cases, **reasonable estimates** are based on **observation and analysis**.
- In general, if a probability is stated, it is based on one of the following approaches:
    - *The classical approach*: This approach is appropriate only for modelling chance experiments with equally likely equipment.
    - *The relative frequency approach*: An estimate is based on an accumulation of experimental results. This estimate, usually derived empirically, presumes a replicable chance experiment.
    - *The subjective approach*: In this case, the probability represents an individual's judgment based on facts combined with personal evaluation of other information.

Summary!!!

- **To estimate (quantify) the uncertainty of an event related to some process; we can modeled the behavior of a process based on observed outcomes, such that an inference could be made on future events.**
- So far we have introduced following terms:
  - **process, event**, **outcomes**, *relative frequency, likelihood,* estimate, ***model, distributed***
  - **Objective of this week-define above terms in the context of probability theory.**
- **When we speak about probability, we often refer to the probability of an event of uncertain nature taking place.**
  - For e.g. What is the probability of rain next Tuesday?
- **Therefore, in order to discuss probability theory formally, we must first clarify what the *possible events are* to which we would like to *attach probability*.**

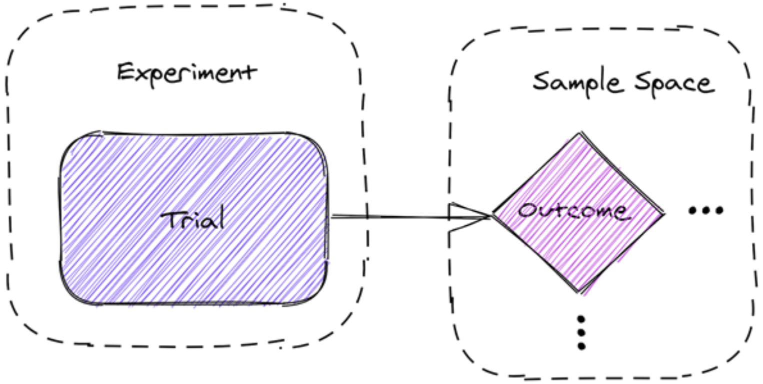


2. Elements of Probabilistic Space.

Experiments(Chance), Sample Space, Events and Probability measure.

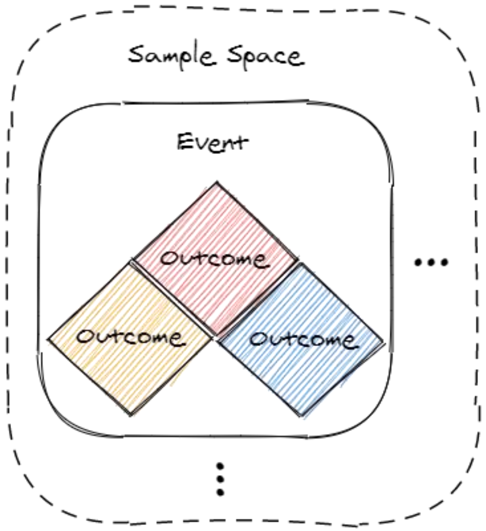
2.1 Experiment(chance).

- An experiment is the process by which an observation is made.
- A chance experiment is any activity or situation in which there is uncertainty about which of the two or more possible outcomes will result.

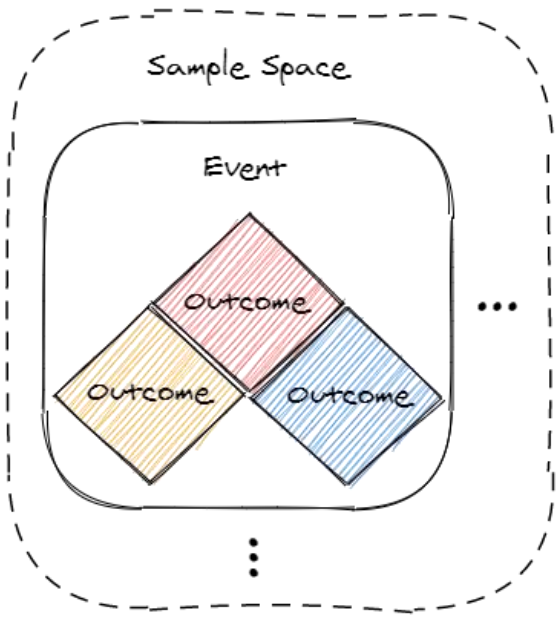


2.2 Probability Space.

- A probability space is a triple (Ω, E,P) consisting of:
  - A sample space Ω/S, which contains all possible outcomes of the experiment.
  - A set of events E,
  - A probability measure P that assigns probabilities to the events in E.



2.2 Probability Space.



Definition: Sample Space, S

A Sample Space is the set of all possible outcomes of an experiment. A sample spaces may be discrete or continuous.

Continuous sample spaces are usually intervals of $\mathbb{R}$ or $\mathbb{R}^n$ used to model time, position, temperature, etc.

For Example:

- Coin Flip: S = {Heads, Tails}(Discrete)
- Flipping two coins: S = {(H,H),(H,T),(T,H),(T,T)}
- YouTube hours in a day: S = $\{x | x \in \mathbb{R}, 0 \leq x \leq 24\}$ (Continuous)

Definition: Event, E

An event in a discrete sample space S is a collection of sample points(outcomes) i.e. any subset of S.($E \subseteq S$)

A simple event is an event consisting of exactly one outcome.

For Example:

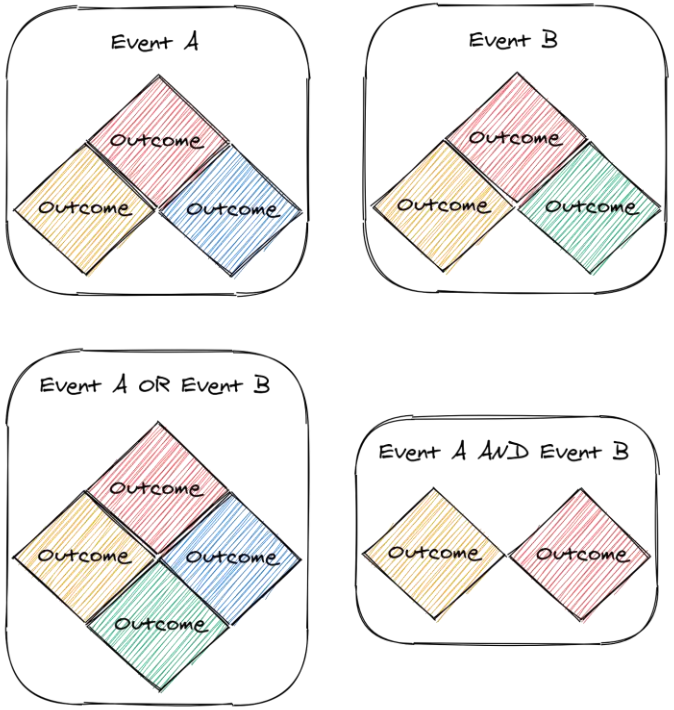
- Coin flip is heads: E = {H}
- At least 1 head on 2 coin flips: E = {(H,H),(H,T),(T,H)}
- Wasted Day(YouTube ≥ 5 hours): E = $\{x | x \in \mathbb{R}, 5 \leq x \leq 24\}$

In the world of probability: events are binary i.e. they either happen or they do not.

2.3 Events and Sets.

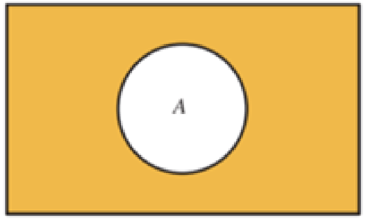
Let **A** and **B** be two **Events** then:

1. The event not A consists of all the experimental outcomes that are not in event A. $\{A^c, A' or\ \bar{A}\}$
2. The event A or B consists of all experimental outcomes that are in at least one of the two events. $\{A \cup B\}$
3. The event A and B consists of all experimental outcomes that are in both of the events A and B. $\{A \cap B\}$
4. The events that have **no common outcomes** are said to be **disjoint** or **mutually exclusive.**

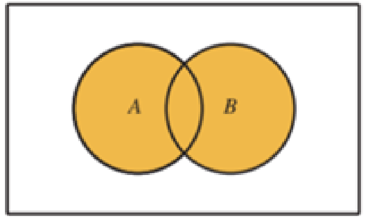


In Figure →

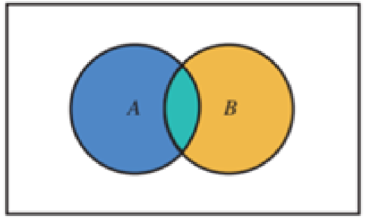
a) not A $\{A^c\}$
b) A or B $\{A \cup B\}$
c) A and B $\{A \cap B\}$
d) Disjoint.



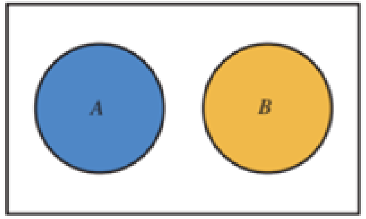
(a) Gold region = *not A*



(b) Gold region = *A or B*

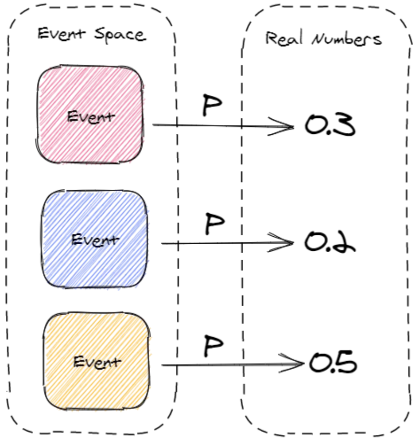


(c) Green region = *A and B*

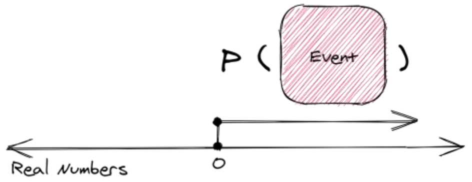


(d) Two disjoint events

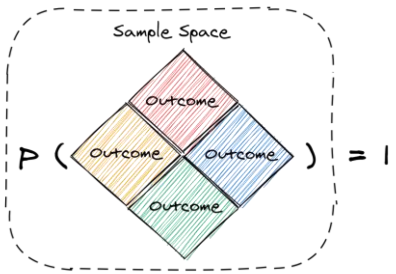
2.4 Probability Measure.



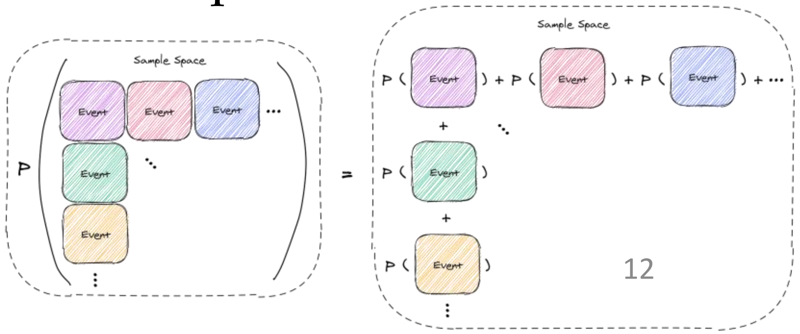
- Suppose S is a sample space associated with an experiment.
- To every event in S, we assign a number ,P(E), called the probability measure of E, so that the following **axioms** hold:
  - **Non-negativity**
    - For any event $E \in \mathcal{S}$, $P(E) \geq 0$



  - **All possible outcomes**
    - Probability of the entire sample space is 1, $P(\mathcal{S}) = 1$



  - **Additivity of disjoint events**
    - For all events $E_1, E_2 \in \mathcal{S}$ that are mutually exclusive ($E_1 \cap E_2 = \emptyset$), the probability that both events happen is equal to the sum of their individual probabilities, $P(E_1 \cup E_2) = P(E_1) + P(E_2)$



2.5 Choosing an Sample Space.

- different **elements** of the **sample space** should be **distinct** and **mutually exclusive** so that when the **experiment is carried out**, there is a **unique outcome**.
- the sample space chosen for a probabilistic model must be **collectively exhaustive**, in the sense that no matter what happens in the experiment, we always **obtain an outcome** that has **been included** in the **sample space**.
- the sample space should have **enough detail** to **distinguish** between all **outcomes of interest** to the modeler, while avoiding **irrelevant details**.

3.Calculating the Probability of an Discrete Event.

3.1 Calculating Probability – Classical Approach.

Flip a Coin

Q] What is the probability of the coin landing on heads?

$$\underbrace{\Pr(H) = \frac{\#\, Number\ of\ favorable\ cases}{\#\, Number\ of\ all\ possible\ cases}}_{\text{Experiment}}$$

Sample Spaces

$$\Pr(H) = \frac{1}{2}$$

Events

- **classical approach to Probability**
  - when the outcomes in the sample space of a **chance experiment are equally likely**, the probability of an event E, denoted by P(E), is the ratio of number of outcomes in the sample space:
    - $\boldsymbol{P(E) = \frac{number\ of\ outcomes\ favourable\ to\ E}{number\ of\ outcomes\ in\ the\ sample\ space.}}$
  - As per the definition:
    - probability measures consist of **counting the number of events.**
- This approach may only valid till events in a sample case are equally likely.

3.2 Calculating Probability of "**or**"

- The equation for calculating the probability of either event A ***or*** event B happening, written {P(A or B)\P(A∪B)} is deeply analogous to counting the number of events of **mutually exclusive events or disjoint events (i.e. A∩B = ∅).**

Definition: Probability of or for mutually exclusive events:

If two events: A, B are mutually exclusive then the probability of A ***or*** B occurring is:

$$P(A \; or \; B) = P(E) + P(F)$$

This property applies regardless of how you calculate the probability of A or B. Moreover, the idea extends to more than two events. Lets say you have n events $E_1, E_2, \ldots, E_n$ where each event is mutually exclusive of one another (in other words, no outcome is in more than one event). Then:

$$P(E_1 \; or \; E_2 \; or \ldots or \; E_n) = P(E_1) + \cdots + P(E_n) = \sum_{i=1}^{n} P(E_i)$$

3.2 Calculating Probability of "or"

- Unfortunately, not all events are mutually exclusive. If you want to calculate P(C or D) where the events C and D are ***not*** mutually exclusive you can ***not*** simply add the probabilities.
- **Why?**
  - If two events are not mutually exclusive, simply adding their probabilities double counts the probability of any outcome which is in both events.
  - There is a formula for calculating ***or*** of two non-mutually exclusive events: it is called the "inclusion exclusion" principle.

Definition: Inclusion-Exclusion Principle:

If two events: A, B

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

  - the inclusion exclusion principle also applies for mutually exclusive events. If two events are mutually exclusive P(E and F)=0 since its not possible for both E and F to occur. As such the formula P(E)+P(F)−P(E and F) reduces to P(E)+P(F).

3.4 Probability: with and without replacement.

- Example-with replacement:
  - In a bag, there are 5 red balls and 3 blue balls. A ball is drawn from the bag, its color is noted, and then it's placed back in the bag. This process is repeated three times. What is the probability of drawing exactly 2 red balls and 1 blue ball?
  - Solution:
    - Individual Probability:
      - $P(red) = \frac{5}{8}; P(blue) = \frac{3}{8}.$
    - We want to calculate the probability of drawing exactly 2 red balls and 1 blue ball in 3 draws. There are three ways this can happen: RRB, RBR, and BRR, where R represents a red ball and B represents a blue ball.
      - $P(RRB) = \frac{5}{8} . \frac{5}{8} . \frac{3}{8}$
      - $P(RBR) = \frac{5}{8} . \frac{3}{8} . \frac{5}{8}$
      - $P(BRR) = \frac{3}{8} . \frac{5}{8} . \frac{5}{8}$
    - Thus; total probability:
      - $Total\ probability = P(RRB) + P(RBR) + P(BRR).$

3.5 Probability: with and without replacement.

- Example-without replacement (Home Task):
  - In a bag, there are 5 red balls and 3 blue balls. A ball is drawn from the bag, its color is noted, and then it's not placed back in the bag. This process is repeated three times. What is the probability of drawing exactly 2 red balls and 1 blue ball?
  - Solution:

3.2 Calculating Probability: Example.

- Solve it:
  - We have a committee of $n = 10$ people and we want to choose a **chairperson, a vice-chairperson** and a **treasurer**. Suppose that 6 of the members of the committee are male and 4 of the members are female. What is the probability that the three executives selected are all male?
    - Solutions:
    - Favorable events: {C, VC, T: Male.}
    - Sample space: {(M1,M2,M3),(M2, M1,M3),……(F3,F4,F1)}
    - P = ?
- Solving for probability requires counting of an total events in sample space, which can be little complicated depending on the process/experiment and probability desired.
  - Let's take a small detour and revise some important concepts for **Counting Theory.**

A. Counting

A.1 Basic Counting Rules.

- Addition and Multiplication Principles:
- **“Addition Principle”:**
  - If the **outcome** of an **experiment** can either be drawn from **event** of **set A** or **set B**, where **none of the outcomes** in **set A** are the same as any of the **outcomes** in **set B** (called **mutual exclusion**), then there are
    - **|A or B|=|A|+|B| possible outcomes** of the **experiment**.
  - To illustrate the addition principle:
  - if a restaurant offers (presuming no course is counted twice):
    - **5 main courses with chicken**,
    - **6 main courses with mutton**,
    - **and 12 vegetarian main courses**,
  - then the **total possible number** of **main courses** is
    - **5 + 6 + 12 = 23.**

A.1 Basic Counting Rules.

- Addition and Multiplication Principles:
- **“Multiplication Rule”:**
  - If an **experiment** has **two parts**, where the **first part** can result in one of **m outcomes** and the **second part** can result in one of **n outcomes** regardless of the **outcome** of the **first part**, then the **total** number of **outcomes** for the **experiment** is $\boldsymbol{m \times n}$.
  - To illustrate the multiplication principle:
    - if a fair coin is tossed (2 possible outcomes)
    - and then a fair 6-sided die is rolled (6 possible outcomes),
  - the total number of possible results of flipping a coin and then rolling a die is
    - $\mathbf{2 \times 6 = 12.}$

A.2 Counting Rules: Combinatorics.

- Many **counting problems** can be approached from the basic building blocks described in the first section: **Counting Rules.**
- However **some counting problems** are so **ubiquitous in the world of probability** that it is worth knowing a few higher level **counting abstractions**.
  - **Permutations and Combinations**.
- Example:
  - Determine the number of ways we can order the objects{A,B,C,D}.

- Solution:
  - There are 4 letters to be arranged into 4 locations.
    - For the **first letter** there are **4** choices.
    - For the **second letter** there are only **3** choices.
    - Continuing this way, there are **2** choices for the **third letter** and only **1** choice for **fourth letter**.
    - By the **multiplication rule**: total number of **arrangement possible** is:
    - $\mathbf{4 \times 3 \times 2 \times 1 = 24 = 4!}$
- In particular, the number of ways of **rearranging n** distinct items is **n!**.
- **What if we have to choose k distinct items from a list of n possibilities, and where the order of our choices matters.**

A.2 Counting Rules: Permutations.

- Proposition (Permutations):
  - The **number of ways** of **choosing k ordered items** from a **list of n distinct** possibilities (where **the order of the k** items **matters**) is equal to
    - $\frac{\boldsymbol{n!}}{\boldsymbol{(n-k)!}} = \boldsymbol{n.(n-1).\ldots.(n-k+1).}$
- Proof(optional):
  - There are
    - n possibilities for the first item,
    - n − 1 for the second item (any possibility but the one already chosen),
    - n − 2 for the third item (any possibility but the two already chosen),
    - … , and n − k + 1 possibilities for the kth item.
    - This yields a total number of possibilities of
    - $n \cdot (n-1) \cdot \cdots \cdot (n-k+1).$
  - For the formula, we have:

    $$n\cdot(n-1)\cdot\cdots\cdot(n-k+1)\cdot(n-k)! = n\cdot(n-1)\cdot\cdots\cdot(n-k+1)\cdot(n-k)\cdot\cdots\cdot 1 = n!.$$
  - Thus, $n \cdot (n-1) \cdot \cdots \cdot (n-k+1) = n!/(n-k)!$.

A.2.1 Permutations: Example.

- Question:
  - A new company logo has four design elements, which must all be different colors chosen from red, orange, yellow, green, blue, and purple. How many different logos are possible?
- Solution:
  - We are choosing $\boldsymbol{k = 4}$ design elements from a list of $\boldsymbol{n = 6}$ elements, and the order matters.
  - From previous result on permutations the choice is:
    - $\frac{\boldsymbol{n}!}{(\boldsymbol{n}-\boldsymbol{k})!} = \frac{\mathbf{6}!}{(\mathbf{6}-\mathbf{4})!} = \frac{\mathbf{6}!}{\mathbf{2}!} = \frac{\mathbf{6}\times\mathbf{5}\times\mathbf{4}\times\mathbf{3}\times\mathbf{2}!}{\mathbf{2}!} = \mathbf{6} \times \mathbf{5} \times \mathbf{4} \times \mathbf{3} = \mathbf{360}$
    - **360 distinct choices** are possible.

A.3 Counting Rules: Combinations.

- In certain other types of counting problems, **the order of the list** of the **k items** we choose from the list of **n** does not matter.
- We can also give a formula for counting in this way:
- Proposition (Combinations):
  - The number of ways of choosing **k unordered** items from a list of **n distinct possibilities** is equal to
    - $\binom{n}{k} = nCk = \frac{n!}{k!(n-k)!} = \frac{n(n-1)\ldots\ldots(n-k+1)}{k.(k-1)\ldots\ldots 1}$
- Proof(Optional):
  - From our calculation above,
    - we know that the number of ways to choose k ordered items from a list of n distinct possibilities is:
      - $\frac{n!}{(n-k)!}$
  - If instead we want to count unordered lists, we can simply observe that for any unordered list, there are k! ways to rearrange the k elements on the list.
  - Thus we have counted each unordered list k! times, so the number of unordered lists is
      - $\frac{1}{k!} . \frac{n!}{(n-k)!} = \frac{n!}{k!(n-k)!}.$

A.3.1 Combinations: Example.

- **Question:**
  - How many 3-element subsets of {1, 2, 3, 4, 5, 6, 7, 8, 9} are there?
- **Solution**:
  - Since subsets are not ordered, we are simply counting the number of ways to choose 3 unordered elements from the given set of 9.
  - From our discussion of combinations, the number of such subsets is :
    - $\binom{9}{3} = \frac{n!}{k!(n-k)!} = \frac{9!}{3!(9-3)!} = \frac{9!}{3!(6!)} = \frac{9\times8\times7}{3\times2\times1} = 84.$

4.Conditional Probability

4.1 Conditional Probability: Intuition.

- Conditional probability provides us with a way to reason about the outcome of **an experiment**, based on **partial information**.
- Here are some examples of situations we have in mind:
  - In an experiment involving two successive rolls of a die, you are told that the sum of the two rolls is 9. How likely is it that the first roll was a 6?
  - In a word guessing game, the first letter of the word is a "t". What is the likelihood that the second letter is an "h"?
  - How likely is it that a person has a disease given that a medical test was negative?
- In more precise terms, **given an experiment**, a **corresponding sample space**, and a **probability law** can **we answer following**:
  - "**what** is the **chance** of an **event E happening given that** I have already **observed some** other **event F**"
- thus seek to **construct a new probability law**, which **takes into account** this **knowledge** and which, for any **event F**, gives us the conditional probability of **E given F**, denoted by **P(E | F)**.
- **It is a critical idea in** *machine learning and probability* **because it allows us to** *update our probabilities* **in the** *face of new evidence*.

4.2 Conditional Probability: Definition.

Definition: Conditional Probability.

Let E and F be two events with $P(F) > 0$. The conditional probability of the event E given that the event F has occurred, denoted by $P(E|F)$ is :

$$P(E|F) = \frac{P(E \cap F)}{P(F)}.$$

The Conditional Paradigm.

Axioms	Axioms - Conditioned on P(E|F)
$0 \leq P(E) \leq 1$	$0 \leq P(E\|F) \leq 1$
$P(S) = 1$	$P(S\|F) = 1$
$P(E \text{ or } F) = P(E) + P(F)$	$P(E \text{ or } G\|F) = P(E\|F) + P(F\|G)$
$P(E^c) = 1 - P(E)$	$P(E^c\|F) = 1 - P(E\|F)$

4.3 Conditional Probability: Example.

- Imagine Daraz wants to figure out the probability that a user will buy a smart watch (let's say E), based on knowing that they bought smart phone (let's say F).
- Solution:
  - To answer, lets start with simpler question, what is the probability that a user bought a smart watch (E):
    - $\boldsymbol{P(E) = \frac{\#\ people\ who\ bought\ smart\ watch}{\#people\ on\ Daraz}}$
  - Now for second question; What is the probability that a user will buy smart watch given they bought smart phone.
    - $P(E|F) = \frac{P(E\ and\ F)}{P(F)} \approx \frac{\#\ who\ bought\ E\ and\ F/\#people\ in\ daraz}{\#who\ bought\ F/\#people\ in\ Daraz}$;{Definition of Prob.}
    - $P(E|F) \approx \#\ of\ people\ who\ bought\ E\ and\ F/\ \#of\ people\ whou\ bought\ F.$

4.3 Conditional Probability: Example.

- In any champions league game night:
  - For a specific married couple the probability that the husband watches the show is 80%, the probability that his wife watches the show is 65%, while the probability that they both watch the show is 60%.
  - If the husband is watching the show, what is the probability that his wife is also watching the show:
- Solution:
  - **Let B = event that husband watches the match; P(B) = 0.85.**
  - **Let A = event wife watches the match; P(A) = 0.65.**
  - **And; P(A∩B)=0.60.**
  - **Thus**:
    - $\boldsymbol{P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0.60}{0.80} = 0.75}$.

5. Independence

5.1 Independence: Intuition.

- We have introduced the conditional probability **P(E | F)** to capture the **partial information that event F provides about event E**.
- An interesting and important special case arises when the occurrence of **F** provides **no information** and does not alter the probability that **E** has occurred, i.e. in such case:
  - **Two events** are said to be **independent** if **knowing** the **outcome** of **one event** does not **change your belief** about **whether or not** the **other event** will occur.
- For example, you might say that two separate dice rolls are independent of one another: the outcome of the first dice gives you no information about the outcome of the second -- and vice versa.

5.2 Independence: Definition.

Definition: Independence

Two events E and F are said to be independent if:

$$P(E|F) = P(E)$$

The events E and F are independent if and only if:

$$P(E \cap F) = P(E).P(F)$$

- Independence is often easy to grasp intuitively.
  - For example, if the occurrence of two events is governed by distinct and non interacting physical processes, such events will turn out to be independent.
  - On the other hand, independence is not easily visualized in terms of the sample space.
  - A common first thought is that two events are independent if they are disjoint, but in fact the opposite is true:
    - two disjoint events A and B with
      - P(A) > 0 and P(B) > 0 are never independent,
      - since their intersection A ∩ B is empty and has probability 0.

5.3 Independence: Example.

- Consider an experiment involving two successive rolls of a 4-sided die in which all 16 possible outcomes are equally likely and have probability 1/16.
  - (a) Are the events A_i = {1st roll results in i}, B_j= {2nd roll results in j}, independent?
  - b) Are the events A = {1st roll is a 1}, B = {sum of the two rolls is a 5}, independent?
- Solution (a):We have:
  - P(A∩B) = P(the result of the two roll is (i,j))=1/16.
  - $\mathrm{P}(A_i) = \dfrac{number\ of\ elements\ A_i}{total\ number\ of\ possible\ outcomes} = \dfrac{4}{16},$
  - $\mathrm{P}(B_j) = \dfrac{number\ of\ elements\ B_i}{total\ number\ of\ possible\ outcomes} = \dfrac{4}{16},$
  - We observe that: $\mathrm{P}(A_i \cap B_j) = \mathrm{P}(A_i)\mathrm{P}(B_j)$;
  - Independence of A_i and B_j are verified.

5.3 Independence: Example.

- Consider an experiment involving two successive rolls of a 4-sided die in which all 16 possible outcomes are equally likely and have probability 1/16.
  - b) Are the events A = {1st roll is a 1}, B = {sum of the two rolls is a 5}, independent?
- Solution (b):We have:
  - P(A∩B) = P(the result of the two roll is (i,j))=1/16.
  - P(A) = $\frac{number\ of\ elements\ A}{total\ number\ of\ possible\ outcomes} = \frac{4}{16},$
  - P(B) = $\frac{number\ of\ elements\ B}{total\ number\ of\ possible\ outcomes} = \frac{4}{16},$
  - We observe that: P(A ∩B) = P(A)P(B);
  - Independence of A and B are verified.

5.4 Independence: Conditional.

- Conditional Independence:
  - We noted earlier that the conditional probabilities of events, conditioned on a particular event, form a legitimate probability law.
  - We can thus talk about independence of various events with respect to this conditional law.
  - In particular, given an event G, the events E and F are called conditionally independent if
    - $P(E \cap F \mid G) = P(E \mid G)P(F \mid G)$
  - In words, this relation states that if G is known to have occurred, the additional knowledge that F also occurred does not change the probability of E.

5.5 Independent Events Vs. Mutually Exclusive Event.

- **Independent Event-Example**:
  - when rolling two dice, the outcome of the first die does not affect the outcome of the second die. The events "getting a 3 on the first die" and "getting a 4 on the second die" are independent events.
- ***Mutually exclusive event-Example***:
  - when drawing a card from a standard deck of playing cards, the events "drawing a red card" and "drawing a black card" are mutually exclusive because a card cannot be both red and black at the same time.
- **Moral of the story**:
  - **Independent events have no influence on each other’s outcomes, and their probabilities multiply when both events occur.**
  - **Mutually exclusive events cannot happen simultaneously, and there probabilities add up when either event occurs.**

3.3 Calculating Probability of "and"

- In slide 3.2 we discussed how to calculate Probability of "or"
- Let's see how to calculate **Probability of "and":**
- The probability of the ***and*** of two events, say E and F, written P(E and F), is the probability of both events happening:

Definition: Probability of ***and*** for independent events.
If two events: E, F are independent then the probability of E ***and*** F occurring is:

P(E and F)=P(E)·P(F)

This property applies regardless of how the probabilities of E and F were calculated and whether or not the events are mutually exclusive.
The independence principle extends to more than two events. For n events $E_1 \dots E_n$ that are ***mutually*** independent of one another -- the independence equation also holds for all subsets of the events:

$$P(E_1 and\ E_2 adn \dots E_n) = \prod_{i=1}^{n} P(E_i)$$

Proof(Optional):
If E is independent of F then $P(E\ and\ F) = P(E).P(F)$.

$$P(E|F) = \frac{P(E\ and\ F)}{P(F)} \ \{\text{Definition of conditional probability}\}$$

$$P(E) = \frac{P(E\ and\ F)}{P(F)} \ \{\text{Definition of independence.}\}$$

$$P(E\ and\ F) = P(E).P(F) \ \{\text{Rearranging terms.}\}$$

3.3 Calculating Probability of "and"

- Let's see how to calculate **Probability of "and" with dependent events.**
- ***Definition***: The chain rule.
  - The formula in the definition of conditional probability can be re-arranged to derive a general way of calculating the probability of the ***and*** of any two events:
    - **P(E and F) = P(E | F)·P(F)**
  - Of course there is nothing special about E that says it should go first. Equivalently:
    - **P(E and F) = P(F and E) = P(F | E)·P(E)**
  - We call this formula the "chain rule." Intuitively it states that the probability of observing events E ***and*** F is the probability of observing F, multiplied by the probability of observing E, given that you have observed F. It generalizes to more than two events:
    - $\boldsymbol{P(E_1 and \ldots \ldots and\ E_n) = P(E_1).P(E_2|E_1).\ldots \ldots P(E_n|E_1 \ldots E_{n-1})}$

6. Rules of Probabilities.

4.1 Axioms of Probability!!!

- Discussed in slide-12:

Axioms of Probability:

- **Axiom1**: $0 \leq P(E) \leq 1$ {All probabilities are numbers between 0 and 1.}
- **Axiom2**: $P(S) = 1$ {All outcomes must be from the sample space.}
- **Axiom3**: If E and F are mutually exclusive then:
  - $\boldsymbol{P(E \text{ or } F) = P(E) + P(F)}$ {Mutually Exclusive events}

4.2 Corollaries of Probability!!!

- Following are the extension to the rule of probability directly provable from the three axioms:

Provable Identities.

Corollaries1:
$P(E^c) = 1 - P(E)$ {The probability of event E not happening.}
Corollaries2:
If $E \subseteq F$, then $P(E) \leq P(F)$ {Events which are subsets}

Proof: $P(E^c) = 1 - P(E)$

We know:

P(S) = P(E or E^c) E or E^c covers every outcome in sample space.

P(S) = $P(E) + P(E^c)$ Events E or E^c are mutually exclusive.

1 = $P(E) + P(E^c)$ Axioms 2.

$P(E^c) = 1 - P(E)$. By Rearranging.

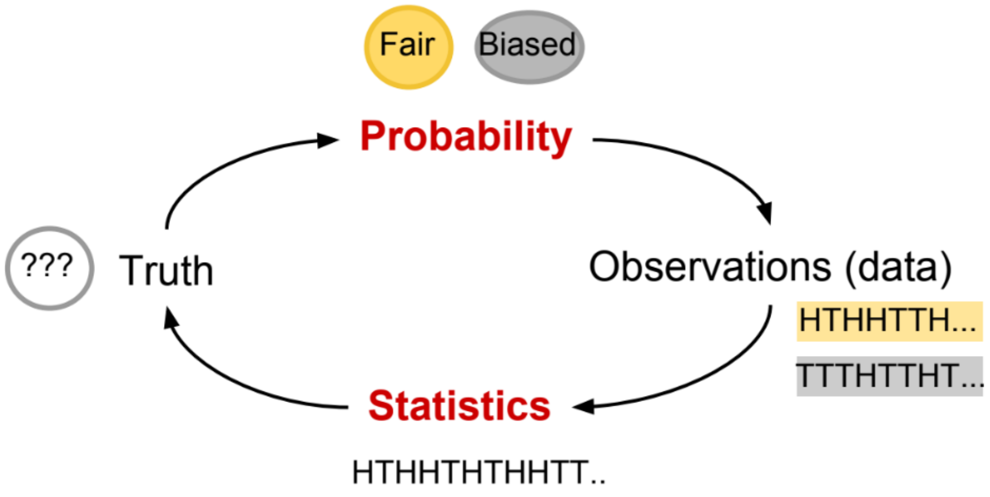
4.3 Properties of Probability Rules

Some properties of Probability Rules:

Consider a probability law, and let A, B and C be events.

1. If $A \subset B$ then $P(A) \leq P(B)$
2. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
3. $P(A \cup B) \leq P(A) + P(B)$.
4. $P(A \cup B \cup C) = P(A) + P(A^c \cap B) + P(A^c \cap B^c \cap C)$

Statistics Vs. Probability!!!



Probability: Given Model, Predict Data
Statistics: Given Data, predict Model

Thank You any Question!!!