HCAI5DS02 – Data Analytics and Visualization.
Lecture – 02
Foundations of Probability.
Science of Understanding and Quantifying Uncertainty.

Siman Giri

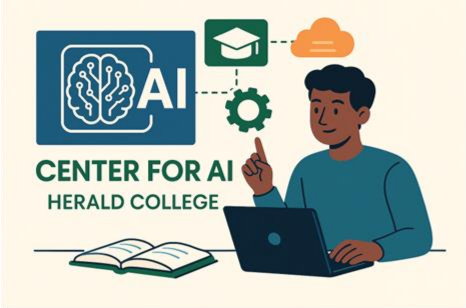


image generated via copilot.

1. A Review of Probability for Data Analytics.

{A big Picture.}

1.1 Probability: A Big Picture …

- **Let's do some Magic!!!!**
  - **Scenario:**
    - I ask you to draw from a **well shuffled deck of 52 standard playing cards**:
    - You are to draw a card, to look at it, and not to show me what it is.

Can I predict what you draw? What are my options?

1.1 Probability: A Big Picture …

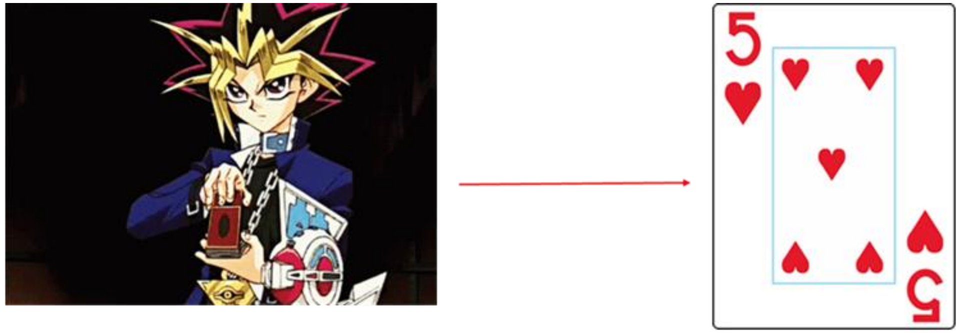


Fig: Card, you picked …

- Now suppose I make one of the following claims:
    1. You have drawn either a red card or a black card
    2. You have drawn a red card
    3. You have drawn a heart
    4. You have drawn the five of hearts
- Am I Correct, Let's analyze aforementioned Claims!!!

1.1 Probability: A Big Picture …

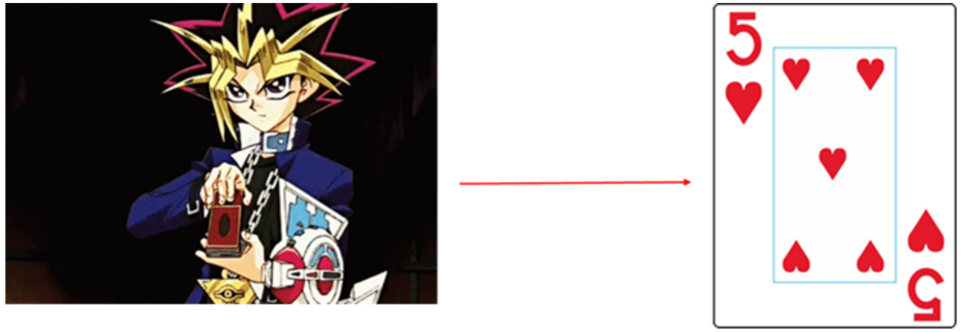


Fig: Card, you picked …

- Now suppose I make one of the following claims:
  1. You have drawn either a red card or a black card
  2. You have drawn a red card
  3. You have drawn a heart
  4. You have drawn the five of hearts
- Am I Correct, Looks like I will be right, If I made any of the above claims.
- Big Question is:
  - How can I measure the level of certainty associated with each of these statements?

1.1 Probability: A Big Picture …

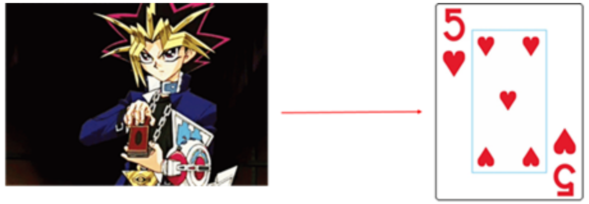


Fig: Card, you picked …

- Now suppose I make one of the following claims:
  1. You have drawn either a red card or a black card
  2. You have drawn a red card
  3. You have drawn a heart
  4. You have drawn the five of hearts
- Am I Correct, Looks like I will be right, If I made any of the above claims.
- Big Question is:
  - How can I measure the level of certainty associated with each of these statements?

Claim	Description	Probability {Certainty}
"Red or black card"	This is always true	$\mathbf{Pr(Claim\ 1)} \coloneqq \frac{52}{52} \times 100\% = 100\%$
"Red card"	26 red cards out of 52	$\mathbf{Pr(Claim\ 2)} \coloneqq \frac{26}{52} \times 100\% = 50\%$
"Heart"	13 hearts out of 52	$\mathbf{Pr(Claim\ 3)} \coloneqq \frac{13}{52} \times 100\% = 25\%$
"Five of hearts"	Only 1 such card	$\mathbf{Pr(Claim\ 4)} \coloneqq \frac{1}{52} \times 100\% \approx 1.92\%$

- **Probability in Big Picture:**
  - **Probability** is not a tool for **predicting exact outcomes**, but a **framework** for **quantifying** the **likelihood** that a **particular claim or event** is **correct or will occur**.
- **In Lay – man Term:**
  - **Probability** doesn't tell **you what *will* happen** — it tells you **how confident you can be in what *might* happen**.

1.2 Probability: Where do the Numbers Come From?

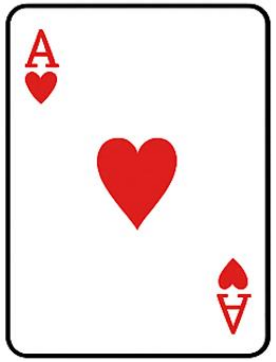
- We are all familiar with the phrases:
  - "the **probability** that a coin will land heads is **0.5**".
  - "the **expected probability** of rolling a 5 on a **fair six-sided die** is $\left\{\frac{1}{6} = 0.1667\right\}$."
- But what does this mean?

Where do the numbers come from?

Claim	Description	Probability {Certainty}
"Red or black card"	This is always true	$\mathbf{Pr(Claim\ 1)} \coloneqq \frac{52}{52} \times 100\% = 100\%$
"Red card"	26 red cards out of 52	$\mathbf{Pr(Claim\ 2)} \coloneqq \frac{26}{52} \times 100\% = 50\%$
"Heart"	13 hearts out of 52	$\mathbf{Pr(Claim\ 3)} \coloneqq \frac{13}{52} \times 100\% = 25\%$
"Five of hearts"	Only 1 such card	$\mathbf{Pr(Claim\ 4)} \coloneqq \frac{1}{52} \times 100\% \approx 1.92\%$

1.3 Probability: Approaches.

1. **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:
        - $\mathbf{P(E)} = \dfrac{\textbf{number of outcomes favourable to E}}{\textbf{number of outcomes in the sample space.}}$
    - 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.



2. **Frequentist Interpretation (aka Empirical Interpretation):**
    - Interprets probability as the long-run frequency of an **event** occurring in repeated trials/experiment or process.
    - The probability of landing heads is 0.50 because, in the long run, if we flip the coin many times, half the flips are expected to result in heads.
        - Backed by **Law of Large Numbers**.

1.3.1 Probability: Approaches.

3. **Subjective {Bayesian} Interpretation:**
   - Interprets probability as **a degree of belief**, updated using prior knowledge and new evidence.
   - The probability of landing heads is 0.5 because we assume prior belief that the coin is fair, and this belief can be updated with more evidence (e.g., results of coin flips).
   - Backed by **Bayes Rule**.
   - {We will discuss Bayes Rule in Future Slides …}

- We know all these methods **assign a number between 0 and 1** to an **event**
  - but can we ensure that this assignment is **consistent, logical, and mathematically valid?**
- **Big Question:**
  - *How can we govern all these assignments under a unified framework?*

1.3.1 Probability: Approaches.

3. **Subjective {Bayesian} Interpretation:**
   - Interprets probability as **a degree of belief**, updated using prior knowledge and new evidence.
   - The probability of landing heads is 0.5 because we assume prior belief that the coin is fair, and this belief can be updated with more evidence (e.g., results of coin flips).
   - Backed by **Bayes Rule**.
   - {We will discuss Bayes Rule in Future Slides …}

- We know all these methods **assign a number between 0 and 1** to an **event**
  - but can we ensure that this assignment is **consistent, logical, and mathematically valid?**
- **Big Question:**
  - *How can we govern all these assignments under a **unified framework?***
    - ***Towards Probability Theory and Kolmogorov Axiomatic Framework of Probability.***

1.4 What is Probability Theory?

- "*Probability theory is nothing but common sense reduced to calculation*."
  — *Pierre Laplace, 1812*
- **Towards Probability Theory …**
  - In general, **Probability** is an **estimate or quantification of uncertainty** attached to an **event** related to some **process**.
    - Where do the **Uncertainty** may arise from:
      - Noisy Measurements in data collection,
      - Natural Variability between samples or individuals,
      - Limited data or finite sample sizes,
    - Probability provides a **consistent framework** for the quantification and manipulation of **uncertainty**.
      - In order to **model the behavior** of a **process** based on observed or empirical outcomes and make inferences about **future events**,
        - we adopt a formal, mathematical interpretation of probability – grounded in the axioms introduced by Andrey Kolmogorov (1933).

1.5 Foundation of Probability Theory

- **Axiomatic or Mathematical Interpretation:**
  - We have **multiple interpretations of probability – Classical, Empirical (frequentist), and Subjective (Bayesian)** each **offering a different perspective** on **how to assign and reason about uncertainty**.
    - While these interpretations differ in philosophy and application, they all rely on a common **mathematical foundation of Probability Theory.**
  - The axiomatic interpretation, developed by **Andrey Kolmogorov**, treats probability as a **mathematical function** defined on a **set of outcomes** (a sample space).
  - His **axiomatic system** defines **probability as a function** satisfying specific logical rules (non – negativity, normalization, and additivity).
  - This rigorous structure allows us to study **all interpretations** within a **unified theory**,
    - enabling clear reasoning, consistent modeling, and broad applicability
      - from games of chance to weather forecasting and machine learning.

A. Kolmogorov

2. Towards Understanding : Commonsense that is Probability.

{Understanding Probability Theory: With in Axiomatic Framework of Kolmogorov.}

2.1 A Probabilistic Space.

- A probability space is a formal **Mathematical Construct** that models a random experiment and consists of:
  - **a random phenomenon {experiment}**,
    - defined by its **sample space**,
      - **outcomes within the sample space**,
        - and **probability measure**
          - {defines how to assign probability to each event governed by Axioms of Probability}.

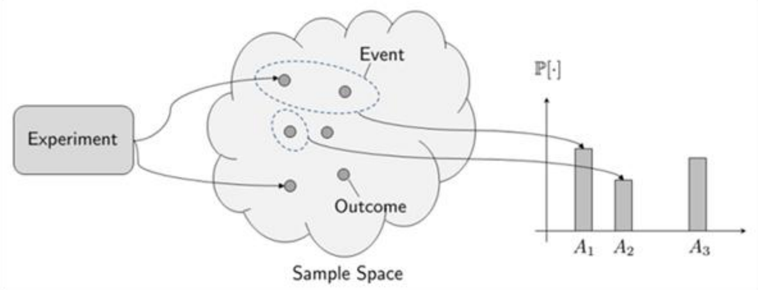


image by: S.Chan from Introduction to Probability for Datascience.

Fig: Elements of Probabilistic Space

- 📐 **Probabilistic Space – A Formal Triplet Notation:**
  - A probability space is denoted as $(\boldsymbol{\Omega}, \boldsymbol{\mathcal{F}}, \mathbf{P})$:
    - $\boldsymbol{\Omega}$: **sample space**
    - $\mathcal{F}$ (or $\mathcal{A}$): σ- **algebra** i.e. collection of events; a set system closed under complements and countable unions).
    - **P: Probability Measure**

2.2 A {Chance} Experiment.

- **{Chance} Experiment aka Random Process:**
  - An experiment or process that leads to **uncertain outcomes (observation),** even though it may be repeatable under the same conditions.
  - A chance experiment or random process by which an **observation** is made is called **Trial**.
  - **Key Features of Chance Experiment:**
    - **Reproducibility**: The experiment can be conducted multiple times under identical or controlled conditions.
    - **Uncertainty**: The **exact outcome** of any single trial cannot be determined in **advance**.
    - **Sample Space**: The set of all **possible outcomes**, known as the sample space, is **well-defined**.
- The **chance experiment provides the context** or **process** that gives rise **to the sample space** and, by extension, **the probability space.**

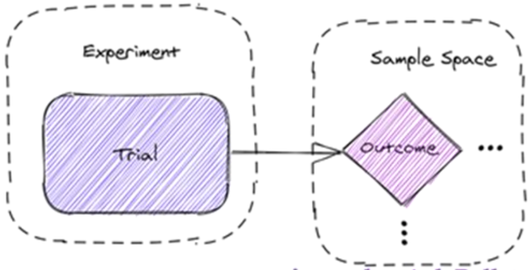


image by: Ash Bellet

Fig: A Chance Experiment and Trial.

2.3 Elements of Probabilistic Space.

- A **probability space** is a mathematical triplet $(\boldsymbol{\Omega}, \boldsymbol{\mathcal{F}}, \mathbf{P})$ that provides the formal framework for probability theory.

1. **A Sample Space (Ω or S):**
   - A **sample space** is the **set** of all **possible outcomes** of a **chance experiment**.
     - It is usually denoted by **Ω** (read as omega) or a capital letter **S**.{ Symbols may be used interchangeably.}
   - **Types of Sample Space:**
     - **Discrete Sample Space:**
       - Contains a **finite or countably infinite** set of outcomes.
       - For example:
         - **Finite Example:** **Tossing a Coin** $\rightarrow \boldsymbol{\Omega} = \{\mathbf{Heads}, \mathbf{Tails}\}$
         - **Countably Infinite Example:** **Rolling a die until a 6 appears** $\rightarrow \boldsymbol{\Omega} = \{\mathbf{1}, \mathbf{2}, \mathbf{3}, \dots\}$
     - **Continuous Sample Space:**
       - Contains an **uncountably infinite set of outcomes**, often represented as **intervals** on $\mathbb{R}$ **or** $\mathbb{R}^{\mathbf{n}}$.
         - Common in measurements like time, position, temperature, etc.
       - For example:
         - **Continuous (Uncountable) Infinite Example:**
           - **YouTube watch time in a Day** $\rightarrow \boldsymbol{\Omega} = \{\mathbf{x} | \mathbf{x} \in \mathbb{R}, \mathbf{0} \le \mathbf{x} \le \mathbf{24}\}$

2.3 Elements of Probabilistic Space.

- A probability space is a mathematical triplet $(\Omega, \mathcal{F}, P)$ that provides the formal framework for probability theory.
  2. **Event in a Sample Space (F) :**
     - An **event F** in a **sample space** $\boldsymbol{S(or\ \Omega)}$ is defined as any **subset** of the sample space. i.e. $\mathbf{F \subseteq S}$
     - Events represent outcomes or groups of outcomes that we are interested in observing.
       - **Events in Discrete Sample Spaces:**
         - **Simple Event**: An event that consists of **exactly one outcome** from the **sample space**.
           - **Example: Single coin flip** $\mathbf{\rightarrow S = \{H, T\}, \longleftarrow}$ **Outcomes**
             - Event $\mathbf{F_1}$: **Getting a Head** $\mathbf{\rightarrow F_1 = \{H\}}$
       - **Events in Continuous Sample Spaces:**
         - Example: **You Tube watch time in a day:** $\mathbf{S = \{x | x \in \mathbb{R}, 0 \leq x \leq 24\}}$
         - Event $\mathbf{F_{wasted\ day}}$: **Watched You Tube for 5 or more hours.** $\mathbf{\rightarrow F_{wasted\ day} = \{x | x \in \mathbb{R}, 5 \leq x \leq 24\}}$
         - This is a continuous event **over an interval**.
       - **Events as a Sets of Functions:**
         - Sometimes, especially in **infinite or functional sample spaces**, outcomes can **be functions**.
           - **Example: Toss a fair coin infinitely many times.** $\mathbf{S = \left\{f : \mathbb{N} \rightarrow \{H, T\}\right\}}$
         - Each function **f** represents **one infinite sequence** of coin toss outcomes.
           - **Event F:** The first two tosses are tails: $\mathbf{F = \{f \in S\} f(1) = T\ adn\ f(2) = T\}}$
         - This is a functional event **defined** on the behavior of a function over its domain.

2.4 Event Space ($\mathcal{F}$ $-$ *Sigma Algebra*) in Probability

- An **event space (aka $\boldsymbol{\sigma}$ $-$ *algebra*)** and written as $\boldsymbol{\mathcal{F}}$ **or** $\boldsymbol{\mathcal{A}}$ is the collection of all possible events associated with a **given sample space**.
  - Events are subsets of the sample space, and the **event space** includes all such subsets that are **relevant** to the probability experiment.
- To make it simpler:

Term	Definition
Sample Space($\boldsymbol{\Omega}$ or S):	The set of all possible outcomes of a random experiment.
Event (F):	A subset of the sample space $F \subseteq S$, representing one or more outcomes.
Event Space ($\mathcal{F}$):	A collection of subsets of the sample space (i.e. events) that satisfies specific closure properties, allowing consistent probability assignments.

2.4.1 Formal Properties of Event Space:

- A **σ – algebra $\mathcal{F}$** over a **sample space Ω** must satisfy the following:
  1. **Empty Set is included**:
     - $\phi \in \mathcal{F}$.
  2. **Sample Space is Included**:
     - $\Omega \in \mathcal{F}$.
  3. **Closure under complementation:**
     - If $A \in \mathcal{F}$, then its complement $A^c \in \mathcal{F}$.
  4. **Closure under Countable Unions:**
     - If $A_1, A_2, \ldots \in \mathcal{F}$, then $\cup_{i=1}^{\infty} A_i \in \mathcal{F}$.
       - (**This also implies closure under countable intersections**.)

- **Example – Event Space for Coin Flip:**
  - Let the sample space be:
    - $\Omega = \{H, T\}$
  - Then a valid event space is:
    - $\mathcal{F} = \{\phi, \{H\}, \{T\}, \Omega\}$
  - This includes:
    - **the empty event,**
    - **each simple event,**
    - **the entire sample space.**

2.4.2 Events that are not Simple.

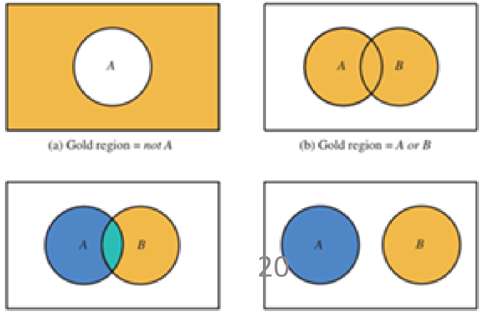
Complex Events

- A **complex event** refers to an event composed of
  - multiple simple events combined using logical operators such as
    - union (∪), intersection (∩), or complement (A’).
  - It may involve **more than** one event but does not
    - specifically require their **simultaneous occurrence.**
- **For example – Rolling a die:**
  - **Event A**: Rolling an even number:
    - $\mathbf{F_A} \coloneqq \{2, 4, 6\}$.
  - **Event B**: Rolling a number greater than 4:
    - $\mathbf{F_B} \coloneqq \{5, 6\}$.
  - **Complex Event**: Rolling an even number **or** a number greater than 4 :
    - $(\mathbf{F_A} \cup \mathbf{F_B}) = \{2, 4, 5, 6\}$

Joint Events

- A **joint event** refers to the simultaneous occurrence of two or more events.
- It is specifically concerned with their **intersection** (A∩B) i.e., the outcomes common to both events.
- For example – Rolling a die:
  - **Event A**: Rolling an even number:
    - $\mathbf{F_A} \coloneqq \{2, 4, 6\}$.
  - **Event B**: Flipping heads on a single coin:
    - $\mathbf{F_B} \coloneqq \{H\}$.
  - **Joint Event**: Rolling an even number and Fliping a head.
    - $(\mathbf{F_A} \cap \mathbf{F_B}) = \{(2, H), (4, H), (6, H)\}$.

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

2.3 Elements of Probabilistic Space.

- A **probability space** is a **mathematical triplet** $(\Omega, \mathcal{F}, P)$ that provides the formal framework for probability theory.
  3. **Probability Measure :**
     - A **probability measure** is a function that maps an events in an **event space** $(\mathcal{F})$ to a real number [0,1] satisfying the **axioms of probability**:
       - $\mathbb{P}: \mathcal{F} \rightarrow [0, 1]$

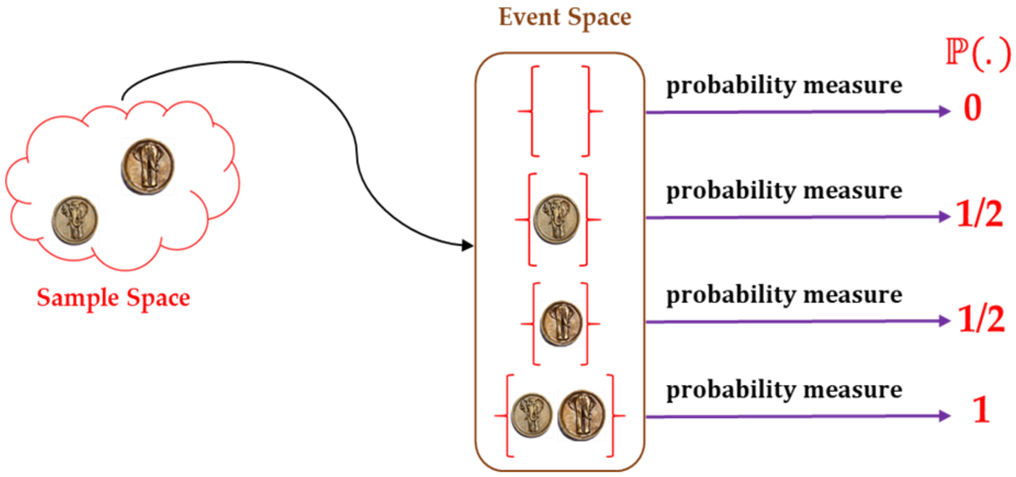


Fig: A Probabilistic Space.

←In a Figure:
Let $(S, \mathcal{F}, \mathbb{P})$ be a probability space, where
S: Sample space (set of all possible outcomes).
$\mathcal{F}$: Event space.
$\mathbb{P}: \mathcal{F} \rightarrow [0, 1]$: A probability measure.

2.5 Kolmogorov's Three Axioms of Probability.

- Let S be the sample space, and let A be any event in S. A probability function P assigns a number to each event A such that:
  1. **Axiom 1: Non – Negativity**
     - For any event A, $\mathbf{P(A) \geq 0}$
     - Interpretation: Probabilities are never negative. You cannot assign a negative likelihood to an event.
  2. **Axiom 2: Normalization (or Unit Measure)**
     - For the entire sample space S, $\mathbf{P(S) = 1}$
     - Interpretation: Something from the sample space must happen **– the total certainty is 100%.**
  3. **Axiom 3: Additivity (for disjoint events)**
     - For two mutually exclusive (disjoint) events A and B, $\mathbf{P(A \cup B) = P(A) + P(B)}$
     - Interpretation: If two events cannot happen at the same time(**disjoint or mutually exclusive events**), the probability that either occurs is the sum of their probabilities.
       - This extend to **any finite or countably infinite number of disjoint events** i.e.
       - For any sequence of disjoint or mutually exclusive events:
         - $\mathbf{A_1, A_2, \ldots}$ **are** $\mathbf{A_i \cap A_j = \phi}$ **for** $\mathbf{i \neq j}$**;** Then $\mathbf{P(\cup_{i=1}^{\infty} A_i = \sum_{i=1}^{\infty} P(A_i)}$.

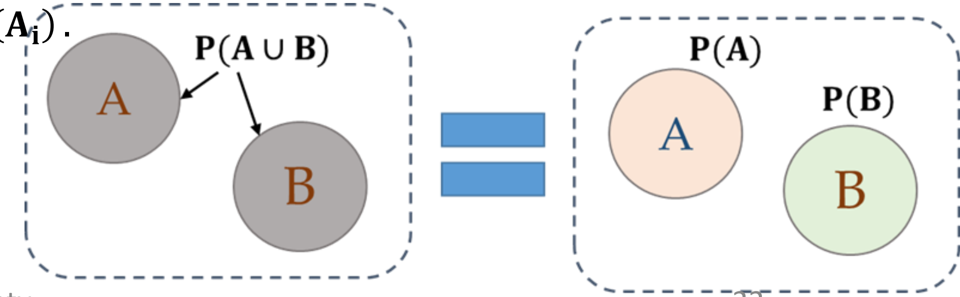


Fig: Understanding Additivity.

3. From Axioms to Application: Rules and Assignments in Probability.

{ Corollaries, Counting, Sampling, and Discrete Outcomes.}

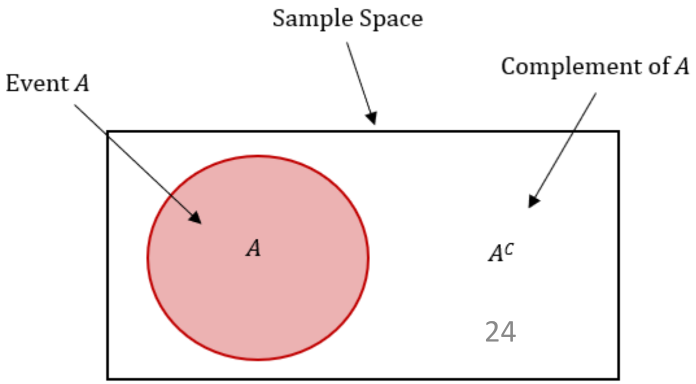
3.1 Corollaries Derived from the Axioms of Probability.

1. **Corollary 1: Probability of the Empty Set is Zero: $\mathbf{P(\phi) = 0}$.**
   - Why?
     - The empty set is disjoint from the **sample space S** and $\mathbf{S \cup \phi = S}$.
     - Using Axiom 3 – Probability of Disjoint events:

$$\mathbf{P(S) = P(S \cup \phi)}$$
$$\mathbf{P(S) = P(S) + P(\phi)}$$
$$\mathbf{P(S) - P(S) = P(S) + P(\phi) - P(S) \ \{\text{Subtract } P(S) \text{on both sides}\}}$$
$$\mathbf{0 = P(\phi)} \square$$

1. **Corollary 2: Probability of the Complement of an Event: $\mathbf{P(A^c) = 1 - P(A)}$.**
   - Why?
     - Let **A be an event** in the **sample space S.**
     - The complement of **A denoted $\mathbf{A^c}$**, consists of all outcomes in **S** that are not in **A**.
       - i.e. **A and $\mathbf{A^c}$** are mutually exclusive (they cannot occur simultaneously) and their union is the entire sample space:
         - $\mathbf{A \cup A^c = S}$
   - Using Axiom 3 and Axiom 2:

$$\mathbf{P(A \cup A^c) = P(A) + P(A^c) \ \{\text{Since A and } A^c \text{ are disjoint adn Axiom 3}}$$
$$\mathbf{P(S) = P(A) + P(A^c) \ \{A \cup A^c = S\}}$$
$$\mathbf{1 = P(A) + P(A^c) \ \{P(S) = 1, \text{Axiom 2}\}}$$
$$\mathbf{P(A^c) = 1 - P(A)} \square$$



3.1 Corollaries Derived from the Axioms of Probability.

3. **Corollary 3: Monotonicity Property i.e. If $A \subseteq B$** (i.e. event A is subset of event B, which means every outcome in A is also in B), then:
   - $\mathbf{P(A) \leq P(B).}$
   - **Proof using Probability Axioms:**
     - Decompose B into disjoint parts: Since $A \subseteq B$ we can write:
       - $\mathbf{B = A \cup (B \backslash A)}$ $\{$ **i. e. $\mathbf{B \backslash A}$ (the set difference)** *means is the part of B not in A.*
         - *Importantly, A and $B \backslash A$ are disjoint they can not happen simultaneously*. $\}$
       - $\mathbf{P(B) = P(A) + P(B \backslash A)}$ $\{$ **Axiom 3 − additive axiom** $\}$
         - *Since* $\mathbf{P(B \backslash A) \geq 0}$ **i. e.** *Axiom 1 − probabilities are never negative.*
       - $\mathbf{P(B) = P(A) + \text{something non} - \text{negative}.}$
       - **Hence**: $\mathbf{P(B) \geq P(A)} \square$

3.1 Corollaries Derived from the Axioms of Probability.

4. Corollary 4: General Additive Properties - For any two events A and B, the probability of their union is:
   - $\mathbf{P(A \cup B) = P(A) + P(B) - P(A \cap B)}$
   - Proof using Probability Axioms:
     1. **Decompose $\mathbf{A \cup B}$ into Disjoint (mutually exclusive) events:**
        - The union $\mathbf{A \cup B}$ can be expressed as the union of **three disjoint (non –overlapping) events**:
          - **only A occurs: $\mathbf{A \backslash B}$ ($\mathbf{i.e.\, A \cup B^c}$),**
          - **only B occurs: $\mathbf{B \backslash A}$ ($\mathbf{i.e.\, B \cap A^c}$),**
          - **Both A and B occur: $\mathbf{A \cap B}$.**
        - Thus, we can write: $\mathbf{A \cup B = (A \backslash B) \cup (B \backslash A) \cup (A \cap B)}$.
     2. **Apply the additivity Axiom:**
        - Since these three events are disjoint, the additivity axiom of probability gives:
          - $\mathbf{P(A \cup B) = P(A \backslash B) + P(B \backslash A) + P(A \cap B) \ \ - [1]}$

3.1 Corollaries Derived from the Axioms of Probability.

4. **Corollary 4: General Additive Properties - For any two events A and B, the probability of their union is:**
   - $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
   - Proof using Probability Axioms:
     3. **Express $P(A)$ and $P(B)$ in terms of disjoint events:**
        - We can write:
          - $A = (A \backslash B) \cup (A \cap B)$, **where** $(A \backslash B)$ **and** $(A \cap B)$ **are disjoint.**
          - $B = (B \backslash A) \cup (A \cap B)$, **where** $(B \backslash A)$ **and** $(A \cap B)$ **are disjoint.**
        - Thus, by the additivity axiom:
          - $P(A) = P(A \backslash B) + P(A \cap B) - [2]$ **and**
          - $P(B) = P(B \backslash A) + P(A \cap B) - [3]$
        - **Solve for** $P(A \backslash B)$ **and** $P(B \backslash A)$**:**
          - **From [2]:** $P(A \backslash B) = P(A) - P(A \cap B)$.
          - **From [3]:** $P(B \backslash A) = P(B) - P(A \cap B)$.
     4. **Substitute back into Equation 1 i.e.** $P(A \cup B) = P(A \backslash B) + P(B \backslash A) + P(A \cap B) \; - [1]$**:**
        - $P(A \cup B) = [P(A) - P(A \cap B)] + [P(B) - P(A \cap B)] + P(A \cap B)$
     5. **Simplifying:**
        - $P(A \cup B) = P(A) + P(B) - P(A \cap B) \; \square$

3.1 Corollaries Derived from the Axioms of Probability.

5. **Corollary 5: Probability is Bounded Between 0 and 1 -** For any event A in a probability space
   - $\mathbf{0 \leq P(A) \leq 1}$.
   - Proof using Probability Axioms:
     1. By Axiom 1 (Non – negativity) of probability, every event satisfies:
        - $\mathbf{P(A) \geq 0 \ -[1]}$.
     2. Since $\mathbf{A \subseteq S}$ (**the sample space**), by corollary of monotonicity:
        - $\mathbf{P(A) \leq P(S)}$
     3. By Axiom 2: $\mathbf{P(S) = 1}$ so,
        - $\mathbf{P(A) \leq 1 \ -[2]}$
     4. By combining $[\mathbf{1}]$ **and** $[\mathbf{2}]$ we have:
        - $\mathbf{0 \leq P(A) \leq 1}$□.

3.2 Assigning Probability to Discrete Events.

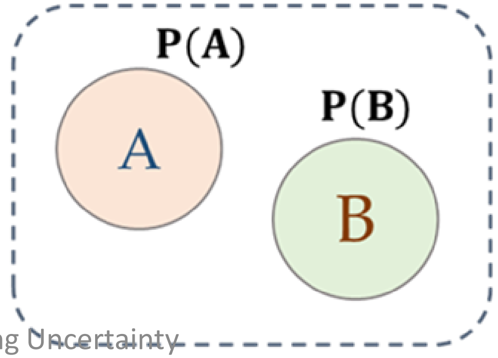
- **Frequentist Approach:**
  - The **Frequentist approach** to probability defines the probability of an event as the long-run relative frequency with which the event occurs in repeated independent trials of an experiment.
    - Probability ➔ relative size of set of event w.r.t the sample space i.e. Out of total Outcomes how many are favorable outcomes i.e.
    - For any event $\mathbf{E} \in \mathcal{F}$.
      - $\mathbf{P(E)} = \dfrac{\text{\# Count number of favourable outocmes} \in \mathbf{E}}{\text{\# Count number of outcomes in the sample spce} \in \boldsymbol{\Omega}}$
- **Example:**
- What is the probability of the coin landing on heads?
  - **Sample space**: $\mathbf{S} \rightarrow \{\mathbf{H}, \mathbf{T}\}$
  - **Event:** $\{\mathbf{H}\} \in \mathcal{F} \rightarrow \{\boldsymbol{\phi}, \mathbf{H}, \mathbf{T}, \mathbf{S}\}$
  - $\mathbf{P(H)} = \dfrac{\text{\# Number of favourable Outcomes}}{\text{\# Number of possible Outocmes}} = \dfrac{1}{2}$

3.2.1 Assigning Probability to Discrete Events.

- **Frequentist Vs. Classical Approach:**
  - The **classical approach** assigns probabilities based on equally likely outcomes and a known sample space.
  - The **Frequentist approach** estimates probabilities through repeated trials and empirical data, **without assuming** equally likely outcomes.
  - In summary, the **Frequentist approach** treats probability as
    - a long-run frequency of an event occurring in repeated trials,
      - emphasizing empirical data and observations
    - rather than **theoretical assumptions** about the **nature of the experiment**.
- Backed by **Law of Large Numbers:**
  - The **Law of Large Numbers (LLN)** is a **fundamental theorem** in **probability theory** that states:
    - *"As the number of trials or observations increases, the average of the observed outcomes will converge to the expected value (the true probability) of the event."*
  - **Example:**
    - If you repeatedly flip a fair coin:
      - In a small number of flips, the proportion of heads may vary significantly
        - e.g., 3 heads out of 4 flips = 0.75.
    - However, as the number of flips increases, the proportion of heads will get closer to 0.50, the expected probability.
    - The **LLN** ensures that observed probabilities **stabilize** with a large number of trials.

3.3 Mutually Exclusive Events.

- **Mutually exclusive events** are events that **cannot occur simultaneously**.
- In other words, the occurrence of one event precludes the occurrence of the other(s). For example:
  - In a single coin toss, the events
    - "**Heads**" and "**Tails**" are mutually exclusive.
  - In rolling a die, the events
    - "**rolling a 3**" and "**rolling a 4**" are mutually exclusive.
- **Definition:**
  - Two events A and B are mutually exclusive if:
      - $\mathbf{A \cap B = \phi}$;
    - where $\mathbf{A \cap B}$ is the intersection of A and B, representing outcomes common to both. Since $\mathbf{A \cap B = \phi}$,
    - the probability of both events occurring together is:
      - $\mathbf{P(A \cap \mathit{B}) = 0}$.



← Fig: Mutually Exclusive Events.

3.4 Probability for Mutually Exclusive Events.

- **Assigning Probability to Mutually Exclusive Events:**
  - When events are mutually exclusive, the probability of their union is:
    - $\mathbf{P(A \cup B) = P(A) + P(B).}$
      - {Based on additivity axiom of probability and also known as Union or Additive law of probability}.
- If we are not sure about the exclusivity of two events A and B , based on inclusion – exclusion we write:
  - $\mathbf{P(A \cup B) = P(A) + P(B) - P(A \cap B)}$
- **Example – Tossing a Coin:**
  - Suppose you toss a fair coin. What is the Probability of getting head or Tail. Let:
    - **A={Heads}, B={Tails}**
  - Since A and B are mutually exclusive:
    - $\mathbf{P(A) = \frac{1}{2}; P(B) = \frac{1}{2}}$
  - The probability of either heads or tails is:
    - $\mathbf{P(A \cup B) = P(A) + P(B) = \frac{1}{2} + \frac{1}{2} = 1.}$

3.5 Probability for 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:

- We have – for individual outcome – Probability is :
  - Probability for Red Ball - R: $P(R) = \frac{5}{8}$;
  - Probability for Blue Ball - B: $P(B) = \frac{3}{8}$;
- Our event of interest is:
  - 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: $\textbf{Events} \coloneqq \{\textbf{Draw 2 red balls and 1 blue ball}\} \coloneqq \{\textbf{RRB}, \textbf{RBR}, \textbf{BRR}\}\{\textbf{\# note order does not matter}\}$
  - Let's Calculate the Probability:

$$P(RRB) = \frac{5}{8} \cdot \frac{5}{8} \cdot \frac{3}{8}$$
$$P(RBR) = \frac{5}{8} \cdot \frac{3}{8} \cdot \frac{5}{8}$$
$$P(BRR) = \frac{3}{8} \cdot \frac{5}{8} \cdot \frac{5}{8}$$

Thus, total Probability: $\textbf{Total probability} = \mathbf{P(RRB) + P(RBR) + P(BRR)}$.

- Example-without 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 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:

- We have – for individual outcome – Probability is :
  - Probability for Red Ball - R: $P(R) = \frac{5}{8}$;
  - Probability for Blue Ball - B: $P(B) = \frac{3}{8}$;
- Our event of interest is:
  - 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: $\textbf{Events} \coloneqq \{\textbf{Draw 2 red balls and 1 blue ball}\} \coloneqq \{\textbf{RRB}, \textbf{RBR}, \textbf{BRR}\}\{\textbf{\# note order does not matter}\}$
  - Let's Calculate the Probability:

$$P(RRB) = \frac{5}{8} \cdot \frac{4}{7} \cdot \frac{3}{6}$$
$$P(RBR) = \frac{5}{8} \cdot \frac{3}{7} \cdot \frac{4}{6}$$
$$P(BRR) = \frac{3}{8} \cdot \frac{5}{7} \cdot \frac{4}{6}$$

Thus, total Probability: $\textbf{Total probability} = \mathbf{P(RRB) + P(RBR) + P(BRR)}$.

Q: What did you Observe?

3.5 Probability for 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:

- We have – for individual outcome – Probability is :
  - Probability for Red Ball - R: $P(R) = \frac{5}{8}$;
  - Probability for **Blue Ball - B:** $\mathbf{P(B)} = \frac{3}{8}$;
- Our event of interest is:
  - 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:
    Events ≔ {Draw 2 red balls and 1 blue ball} ≔ {RRB, RBR, BRR}{# note order does not matter}
  - Let's Calculate the Probability:

$$P(RRB) = \frac{5}{8} \cdot \frac{5}{8} \cdot \frac{3}{8}$$
$$P(RBR) = \frac{5}{8} \cdot \frac{3}{8} \cdot \frac{5}{8}$$
$$P(BRR) = \frac{3}{8} \cdot \frac{5}{8} \cdot \frac{5}{8}$$

Thus, total Probability: **Total probability = P(RRB) + P(RBR) + P(BRR)**.

- Example-without 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 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:

- We have – for individual outcome – Probability is :
  - Probability for Red Ball - R: $P(R) = \frac{5}{8}$;
  - Probability for **Blue Ball - B:** $\mathbf{P(B)} = \frac{3}{8}$;
- Our event of interest is:
  - 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:
    Events ≔ {Draw 2 red balls and 1 blue ball} ≔ {RRB, RBR, BRR}{# note order does not matter}
  - Let's Calculate the Probability:

$$P(RRB) = \frac{5}{8} \cdot \frac{4}{7} \cdot \frac{3}{6}$$
$$P(RBR) = \frac{5}{8} \cdot \frac{3}{7} \cdot \frac{4}{6}$$
$$P(BRR) = \frac{3}{8} \cdot \frac{5}{7} \cdot \frac{4}{6}$$

Thus, total Probability: **Total probability = P(RRB) + P(RBR) + P(BRR)**.

Q: What did you Observe?

3.6 Assigning Probability to an Events {Discrete}: Summary.

- To sum up – "Count" the probability of individual outcome, and sum them up for the probability of a collection of outcomes (= event).
- Thus, we can say that:
  - To evaluate the probability of discrete events, the calculation depends on counting, which in turn depends on the **type of events**: {whether they are disjoint or joint}.
    - **Disjoint events** are always mutually exclusive, meaning they cannot occur simultaneously.
      - $\mathbf{P(A \cup B) = P(A) + P(B).}$
    - **Joint events** can either be **independent (unrelated)** or **dependent (related or conditioned),** depending on whether the occurrence of one event **affects** the probability of the other.
      - $\mathbf{P(A \cap B) = \dfrac{\#\ Count\ outcomes\ of\ A\ and\ B.}{\#Total\ Outcomes\ in\ the\ sample\ space.}}$

3.7 Why Counting Matters?

- Counting is essential in probability because it provides the foundation for calculating probabilities, particularly for **discrete sample spaces**.
  - Counting helps distinguish between **disjoint (mutually exclusive)** and **joint (overlapping)** events:
    - For disjoint events,
      - **P(A∪B) = P(A)+P(B),**
        - and counting is straightforward since there's no overlap.
    - For joint events,
      - **P(A∪B) = P(A) + P(B) – P(A∩B)**
        - {Cautions !! This does not tell you whether relationship exist or not we discuss that in next section},
      - and counting must account for the overlap to avoid double-counting.
  - When events involve multiple stages or combinations (Complex events), counting becomes crucial:

- **For 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 female?
    - 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 a total events in sample space,
  - which can be little complicated depending on the process/experiment and probability desired.
- Hope you know how to count, If **don't do not miss your tutorial, we will discuss on counting principle** – Hint **Permutations and Combinations !!!!**

4. Understanding Conditional Probability & Independence: Quantifying Dependent and Independent Events.

{Conditional Probability and Examination of Independence.}

4.1 Events that are Joint !!!

- **Joint Events and Joint Probability:**
  - Let **A and B** be two events defined on a **common sample space Ω.**
    - A **joint event** refers to the occurrence of both **events A and B**. The Probability of their simultaneous occurrence is denoted as:
      - $\boldsymbol{P(A \cap B)}$ *{ also written as* $\boldsymbol{P(A, B)}$*}* = ***Probability both A and B Occur.***
        - and represents **the likelihood of both A and B occurring together**.
    - This is known as the **Joint Probability of A and B.**
    - This is useful, but **doesn't yet tell us how A and B are related**.
    - Example:
      - Event A: It rains today & Event B: The Traffic is heavy
      - Joint Event $A \cap B$: It rains and traffic is heavy.
    - We are now interested in not just single outcomes, but how two things occur together.
- **Q: "When Events Occur Together: Are they Linked?"**
  - **For example:**
    - **If I told you it's raining, how likely is it that traffic is heavy?**
      - **Do you think rain and traffic always happen together? Could one cause the other?**

4.1.1 Understanding Dependence and Independence.

- **Q: “When Events Occur Together: Are they Linked?”**
  - **Do you think rain and traffic always happen together? Could one cause the other?**
  - While Computing joint probabilities is useful, it is often more important to understand the relationship between events:
    - **Are they related in any meaningful way?**
    - **Does the occurrence of one event influence the likelihood of the other?**
  - This leads to two important types of relationships between events:
- **Dependent Events:**
  - Events A and B are said to be dependent if the occurrence of one affects the probability of the other.
  - That is: $\mathbf{P(A|B) \neq P(A)}$.
- **Independent Events:**
  - Events A and B are independent if knowledge of one event provides no information about the other.
  - Mathematically they satisfy: $\mathbf{P(A \cap B) = P(A) \cdot P(B)}$ **or equivalently** $\mathbf{P(A|B) = P(A)}$

4.1.2 Exploring Dependent Events.

- Simple (or marginal) probability is **inadequate** when the occurrence of one event **influences** the likelihood of another — a situation known as **event dependence**.
  - In such cases, treating events as if they were independent can lead to **incorrect inferences**.
  - We need a **framework** that **accounts** for **how new information alters our belief** about another event.
- This is where **Conditional Probability** comes in.
- **Example: Weather Forecasting**
  - Q: ***"What is the probability that it will rain, given that the sky is cloudy?"***
    - We can not simply compute **P(Rain)** in isolation, because:
      - The information that " *it is cloudy*" provides **additional knowledge**.
    - This change our belief about likelihood of rain.
  - Thus, we must compute the conditional probability:
    - **P(Rain|Cloudy)**
  - This **accounts** for the **dependence** between cloudiness and rain, and **updates our estimate** accordingly.
- **What is Conditional Probability?**

4.2 Conditional Probability: Updating Beliefs with Evidence.

- **Theory of conditional probability** provides the way to measure the probability when two events are dependent with each other, i.e. one event can only occur if another event has already happened.
- For **any event A and B**, if event **A is dependent on B** (or also called **conditioned on B**) and we know $\mathbf{P(B) > 0}$ Then **Conditional Probability of A given B** is :
  - $\mathbf{P(A|B) = \frac{P(A \cap B)}{P(B)}}$ is called conditional probability of A conditioned on B.
- **Relative Frequency Interpretation:**
  - Suppose we **repeat the experiment many times**.
    - Out of all trials, **the fraction in which B occurs** is approximately $\mathbf{P(B)}$.
    - Out of those where **B occurs**, the fraction in which **both A and B occur** is approximately $\mathbf{P(A \cap B)}$.
  - So , the fraction of times that A occurs within the occurrence of B is:
    - $\mathbf{P(A|B) = \frac{\#count\ both\ A\ and\ B\ occur}{\#count\ B\ occur} == \frac{P(A \cap B)}{P(B)}, if\ P(B) > 0}$

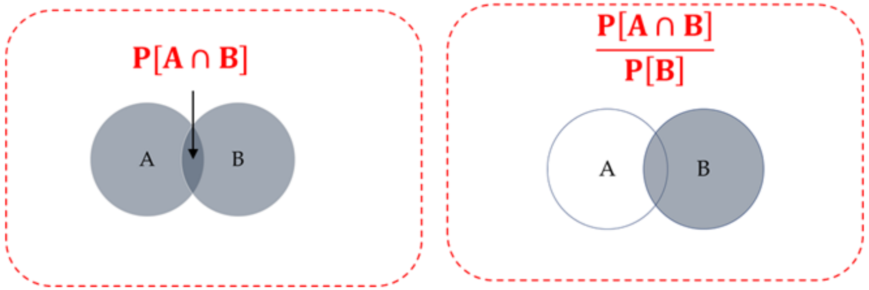


Fig: Illustration of Conditional Probability and comparison to Joint Probability.

4.2.1 Conditional vs. Joint Probability

- Conditional probability refines **our estimate of an event's likelihood based on known evidence**
  - while joint probability **quantifies the simultaneous occurrence of two events in the full context**.

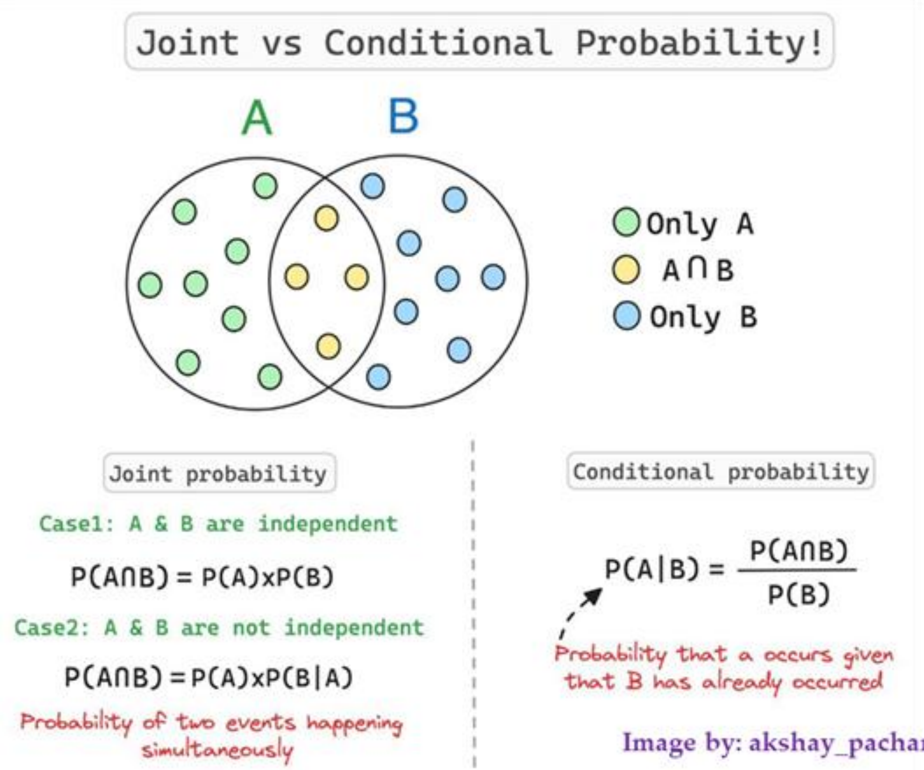


Image by: akshay_pachar

Key Distinction: Conditional vs. Joint Probability.

Concept	Formula	Focus
Conditional Probability	$P(A\|B) = \frac{P(A \cap B)}{P(B)}$	• Focuses on the likelihood of A, given that B has occurred. • **Conditional Probability** isolates the impact of B on A, focusing only on cases where B occurs.
Joint Probability	$P(A \cap B)$	• Measures the Probability that both A and B Occur together in the full sample space. • **Joint Probability** measures the overlap of A and B in the entire sample space.

4.2.2 Axioms of Conditional Probability.

- **Conditional probability defines a new probability measure** over the sample space,
  - where **probabilities are scaled relative to the occurrence of B** and are bounded by following **axioms:**

Axiom	Mathematical Expression	Interpretations
Non - Negativity	$\mathbf{P(A\|B) \geq 0}$	The probability of any event A, given B, is always non negative.
Normalizations	$\mathbf{P(\Omega\|B) = 1}$	Given B has occurred, some outcome in the sample space must occur.
Additivity	$\mathbf{P(A_1 \cup A_2 \cup \cdots \|B) = \sum P(A_i\|B)}$	If events $\mathbf{A_1, A_2, \ldots}$ **are disjoint**, the conditional probability of their union equals the sum of their individual conditional probabilities.
Complement	$\mathbf{P(A^c\|B) = 1 - P(A\|B)}$	Complement of conditional Probability.

- These axioms ensure that **conditional probability behaves like a proper probability measure**, but within the **restricted sample space defined by event B**.

4.3 Example: Conditional Probability in E – Commerce.

- **"Should Daraz recommend smartwatches to smartphone buyers?"**
  - Business Problem:
    - Calculate probability a user buys a smartwatch (W) given they bought a smartphone (P):
      - $\mathbf{P(W|P) = ?}$
  - Why?
    - **To optimize recommendations → Boost cross – selling !!!**
- **Approach – 1 – Simple (marginal) Probability:**
  - $\mathbf{P(W) = \frac{\#\ users\ who\ bought\ W}{\#\ total\ users}}$
    - **For Example:** $\mathbf{\frac{500}{10{,}000} = 5\%\ \{\ Baseline\}}$
  - **Limitation:**
    - ***Ignores user behavior (e.g. smartphone purchases).***

4.3.1 Example: Conditional Probability in E – Commerce.

- **Approach – 2 – Conditional Probability Solution:**
  - Definition:
    - $\mathbf{P(W|P) = \frac{P(W \cap P)}{P(P)}, P(P) > 0}$
  - **Calculation:**
    1. **Numerator:**
       - $\mathbf{P(W \cap P) = \frac{\#users\ who\ bought\ both\ W\ and\ P}{\#total\ users}}$
         - **For example:** $\mathbf{\frac{200}{10{,}000} = 2\ \%}$
    2. **Denominator:**
       - $\mathbf{P(P) = \frac{\#users\ who\ bought\ P}{\#\ total\ users}}$
         - **For example:** $\mathbf{\frac{1000}{10000} = 10\%}$
    3. **Result:**
       - $\mathbf{P(W|P) = \frac{P(W \cap P)}{P(P)} = \frac{2\%}{10\%} = 20\%}$

- **Interpretation & Action:**
  - **High Probability (e.g. 20% vs 5% baseline).**
    - Strong correlation! Recommend smartwatches to smartphone buyers.
  - **Low Probability**
    - Avoid irrelevant recommendations.
- **Business Impact:**
  - Precision marketing → Higher conversion rates.
  - Better user experience.

4.4 When Events Don't Influence Each Other: Independence

- We have introduced the conditional probability for event A conditioned on event B $\{\mathbf{P(A|B)}\}$ to capture the **partial information that event B provides about event A**.
  - An interesting and important special case arises when the occurrence of **B** provides **no information** and does not alter the probability that **A** will occur, i.e. when two events are independent:
  - **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. Mathematically:
    - $\mathbf{P(A \cap B) = P(A) \times P(B)}$ {aka multiplicative rule of probability}
- **Example Scenario – Coin Tosses:**
  - Suppose you toss a coin twice:
    - **A = {Heads on the first toss},**
    - **B = {Heads on the second toss}.**
  - What is the probability of getting A and B:
    - Since the outcome of the first toss does not affect the second:
      - $\mathbf{P(A \cap B) = P(A) \times P(B) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}}$.

4.5 Common Misunderstanding about Independence.

- **Independence is not same as Mutual Exclusivity:**
  - **Independent events:** Can occur simultaneously.
    - For example, flipping a coin and rolling a die.
  - **Mutually exclusive events:** Cannot occur at the same time.
    - For example, rolling a die and getting both a "3" and a "5."
- **Independence is not about disjointness:**
  - **Disjoint events** have no overlap, so **$P(A \cap B) = 0$** .
    - They cannot be independent unless one or both events have **$P(A) = 0$ or $P(B) = 0$**.
  - **Independent events** can overlap.
- **Independence is not about equal probabilities:**
  - Independent events do not require equal probabilities.
  - For example: rolling a die and flipping a coin are independent, even though
    - $\mathbf{P(Die\ Shows\ 4) = \frac{1}{6}}$ and $\mathbf{P(Coin\ shows\ head) = \frac{1}{2}}$.

4.6 Chain Rule of Probability.

- The chain rule of probability express the probability of the intersection of multiple events in terms of conditional probabilities. For n events $\mathbf{A_1, A_2, \cdots, A_n}$ the chain rule is:
  - $\mathbf{P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2) \cdots P(A_n|A_1 \cap A_2 \cap \cdots \cap A_{n-1})}$
- This rule allows us to compute the joint probability of multiple events step by step, starting from the probability of the first event and progressively conditioning on previous events.
- **Independent Events in the Chain Rule:**
  - If the n events $\mathbf{A_1, A_2, \cdots, A_n}$ are independent then, chain rule simplifies to
    - $\mathbf{P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1) \cdot P(A_2) \cdot \cdots \cdot P(A_n) = \prod_{i=1}^{n} P(A_i)}$

5. Understanding Bayes Theorem.

{Deriving Probabilistic Reasoning from Conditional Probability.}

5.1 Challenge with Conditional Probability.

- **The core Problem:**
  - We often want to know how likely something is, given some observed evidence.
- But in most real – world situations:
  - We do not directly observe the probability of the cause.
  - We do observe the effect (evidence) and know how likely that effect is under different causes.
- **Example 1: Medical Diagnosis**
  - Let:
    - **D → Patient has Disease.**
    - **T → Test comes back positive.**
  - Doctors want to know:
    - What is the chance the patient is sick if the test is positive?
      - $\mathbf{P(Disease|Postive\ Test) = \frac{P(Disease \cap\ Postive\ Test)}{P(Positive\ Test)}}$;
    - Looks Simple, so what is the problem?

5.1.1 Challenge with Conditional Probability.

- **Why We *Don't* Know P(Disease ∩ Postive Test) Directly?**
  - Even though **P(Disease ∩ Postive Test)** (i.e. the joint probability of having the disease and testing positive) is conceptually simple,
    - here is why it is typically not directly available.
      - We test people, But Don't know the True Condition:
        - In real populations:
          - You can observe who tested positive from Test Result.
          - But you often do not know who truly has the disease unless you use a perfect Test,
          - i.e. Test with 100% accuracy.
          - These Test may not exist or may be very expensive/invasive not feasible.
        - That means, we can count positives, but not confidently count true positives → so we can not directly observe the joint events.
- Then what do we do in Practice?

5.1.2 Challenge with Conditional Probability.

- **But in Practice we do have:**
  - **$P(\text{Positive Test}|\text{Disease}) \rightarrow$ likelihood:**
    - Observed from controlled clinical Studies (e.g. Clinical Trial)
    - Example: Out of 8,000 people who had the disease, 7,999 tested positive.
    - This gives us the sensitivity of the test.
  - **$P(\text{Disease}) \rightarrow$ Prior:**
    - This is the prevalence of the Disease in real population and represents how common the disease is in the general population
    - which can be observed through large scale epidemiological studies.
  - **$P(\text{Positive Test}) \rightarrow$ Evidence:**
    - This can be computed using Law Total probability.

$$\mathbf{P(Positive\ Test) = P(Positive\ Test|Disease) \cdot P(Disease) + P(Positive\ Test\ |No\ Disease) \cdot P(No\ Disease)}$$

- **$P(\text{Disease} \cap \text{Postive Test})$** could not be measured directly but now could be reconstructed using chain rule of Probability i.e.

$$\mathbf{P(Disease \cap Positive\ Test) = P(Positive\ Test|Disease).P(Disease)}$$

- Based on all above observation, Bayes Rule {or, Theorem} flips the conditional and lets you compute:

$$\mathbf{P(Disease|Positive) = \frac{P(Disease\ \cap\ Positive)}{P(Test\ Positive)}}$$

$$\mathbf{P(Disease|Positive) = \frac{P(Positive\ Test|Disease).P(Disease)}{P(Test\ Positive)}}$$

5.2 Bayes Theorem: Formal Definition.

- **Bayes' rule states:**

$$\text{posterior} \propto \text{likelihood} \times \text{prior} \Rightarrow \text{posterior} = \frac{\text{likelihood} \times \text{prior}}{\text{evidence}}.$$

- Bayes' Rule allows us to *update our beliefs* about *the probability* of *an event* by incorporating new information or *evidence*, making it a powerful tool for adaptive decision-making.
  - This "updating" process is fundamental in probabilistic reasoning and plays a key role in fields like medicine, machine learning, and finance.
- Let's break down how this updating process works with Bayes' Rule based on following scenario:
- Cautions!!!:
  - **Bayes Rule or Theorem is not a new formula**;
    - it is a **re – expression of conditional probability** that:
      - Uses known quantities,
      - Allows us to reason from effect back to cause,
      - Enables belief updating in face of new evidence.

5.3 Understanding Terminologies in Bayes Rule.

- **Example – 2 – Real world – Fraud Detection:**
  - We want to know: What is the probability a transaction is fraudulent, given that triggered an alert?
    - $\mathbf{Posterior} \Rightarrow \mathbf{P(Fraud|Alert)}$
  - We know we can not observe, $\mathbf{P(Fraud \cap Alert)}$ as we can not directly count
    - **how many transactions are both fraudulent and triggered an alert.**
  - But we can observe:
    - $\mathbf{Likelihood\ P(Alert|Fraud)}$ – the probability that a fraudulent transaction triggers an alert,
    - $\mathbf{Prior\ P(Fraud)}$ – the base rate or prevalence of fraud in population.
    - $\mathbf{Evidence\ P(Alert)}$ – the overall probability that any transaction triggers an alert.
      - $\mathbf{\{posterior)\}P(Fraud|Alert)} = \frac{\mathbf{likelihood \times prior}}{\mathbf{evidence}} = \frac{\mathbf{P(Alert|Fraud \times P(Fraud)}}{\mathbf{P(Alert)}} = ?$
- **Question :**
  - How can **each of the three components** above (likelihood, prior, and evidence) be **estimated or obtained** in **practice** in a **real – world fraud detection system**?

What we Can Observe and Estimate:

Component	Meaning	How it's Observed	Suppose
🧪 Likelihood: $\mathrm{P(Alert\|Fraud)}$	The chance that a fraudulent transaction triggers an alert.	Estimated using labeled data from past confirmed fraud cases (e.g., from investigations or audits)	$\mathbf{P(Alert\|Fraud) = 0.98}$ {High detection rate}
📊 Prior $\mathrm{P(Fraud)}$	The overall rate of fraudulent transactions	Estimated from historical transaction data (e.g. 0.5% of all transaction flagged in past were confirmed fraud)	$\mathbf{P(Fraud) = 0.005}$ {0.5% of all transactions are fraudulent}
📈 Evidence $\mathrm{P(Alert)}$	The overall probability that any transaction triggers an alert	Computed from operational logs of the fraud detection system (e.g. 2% of transactions trigger alerts)	$\mathbf{P(Alert) = 0.02}$ {2% of all transactions trigger alerts}

5.4 Putting It All Together

$$P(\text{Fraud}|\text{Alert}) = \frac{P(\text{Alert}|\text{Fraud}) \times P(\text{Fraud})}{P(\text{Alert})} = \frac{0.98 \times 0.005}{0.02} = \frac{0.0049}{0.02} = 0.245 = 24.5\%$$

Interpretations:

- So, only 24.5% of transactions that trigger an alert are actually fraudulent — despite a high detection rate.

Key Insight:

- Even with a strong alert system, if fraud is rare, most alerts may still be false positives.
- Bayes Rule helps us quantify this uncertainty and make informed decisions (e.g. whether to investigate).

5.5 Summarizing Bayes Rule!!!

- The Common Problem we face:
  - We often want to answer questions like:
    - *"Given this evidence, what is the probability that a particular hypothesis is true?"* This is: $\mathbf{P(H|E) = ?}$
  - We do not have direct access to joint probability of hypothesis and evidence i.e. $\mathbf{P(H \cap E)}$.
- What we usually know or we can measure or estimate are likelihood, the prior, and the evidence,
  - and apply a bayes rule such that:

Likelihood:
- How probable the evidence is, assuming the hypothesis is true.
- e.g., If a transaction is actually fraud, how likely is it to trigger an alert?

Prior:
- How probable the hypothesis is before seeing any evidence.
- How common is fraud in general?

$$\boxed{P(H|E)} = \frac{\boxed{P(E|H)} \times \boxed{P(H)}}{\boxed{P(E)}}$$

Posterior:
- How probable the hypothesis is, assuming the evidence is observed.
- e.g., If a transaction triggered an alert, how likely is it to be fraud?

Evidence:
- How probable the evidence is, considering all possible causes.
- What is the overall chance that any transaction (Fraudulent or Not) triggers an alert?
- A Total Probability.

The – End.