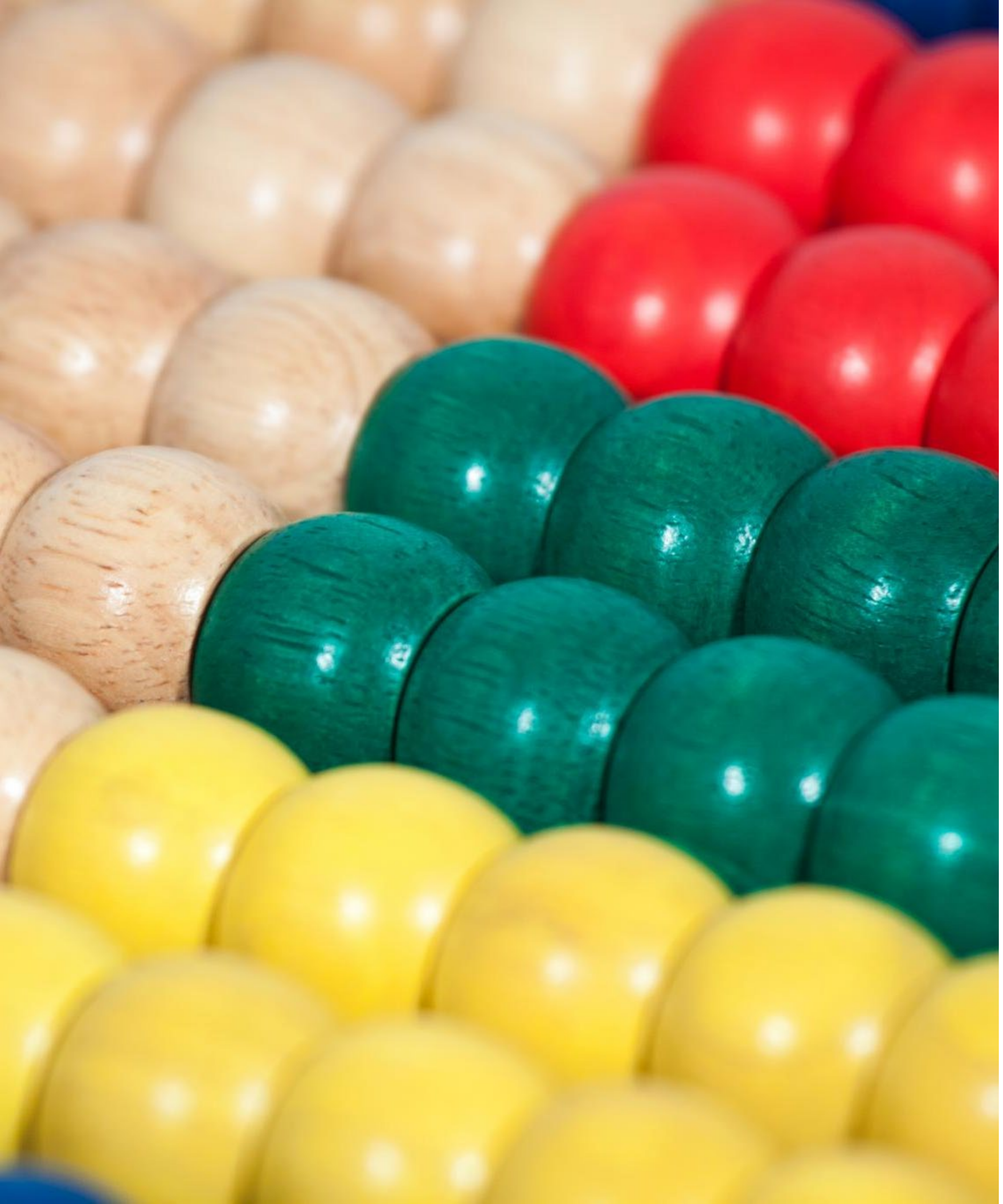
BASIC OF PROBABILITY

"there are many instances where chance is involved in determining the outcome of a decisions."

TOPIC TO DISCUSS

1. DEFINITION AND CONCEPT OF PROBABILITY
2. TERMINOLOGY IN PROBABILITY
3. FUNDAMENTAL RULES IN PROBABILITY
4. PROBABILITY THEORY
5. PROBABILITY DISTRIBUTION



DEFINITION

- Probability is a fundamental concept in mathematics and statistics that deals with the likelihood of different outcomes occurring.
- The probability value will be in the **range 0 to 1**.
- A value of **0** means the **event will not occur**.
- A probability of **1** means the **event will occur**.
- Anything **between 0 and 1** reflects the **uncertainty of the event occurring**.

PROBABILITY IN HEALTHCARE

EXAMPLE:

1. If a patient is diagnosed with a confirmed case of the flu based on lab results, we can say the **probability that they currently have the flu is 1 since we know for sure that they are infected.**
2. If a new medication requires a minimum dosage of 50 mg to be effective, but a patient receives only 10 mg, we can say **the probability that the medication will successfully treat the condition is zero**—it can't happen.
3. A person with a family history of cancer and certain lifestyle habits may have a **40% probability** of developing cancer in their lifetime. This is based on statistical data, but it does not guarantee **whether they will or will not get cancer (uncertainty)**.

TERMINOLOGY IN PROBABILITY

1. Key Events in Statistics

Events are **outcomes or combinations of outcomes** from a statistical experiment. Understanding events is crucial for data analysis.

2. Understanding Sample Space

Sample space encompasses **all possible outcomes** of a statistical experiment. It forms the foundation of statistical analysis.

3. Interpreting Outcomes

Outcomes are **specific results of an event** within the sample space. They are vital for making informed decisions in healthcare research.

HOW TO RELATE THOSE TERMINOLOGIES IN HEALTHCARE DOMAIN?

1. KEY EVENT IN STATISTICS

Example: A hospital conducts a **study on patient recovery after surgery**.

- **Event**: A patient recovers within 2 weeks after surgery.
- **Event Combination**: A patient recovers **within 2 weeks AND has no complications**.

Understanding events helps doctors and researchers analyze **treatment effectiveness**.

2. SAMPLE SPACE IN HEALTHCARE

Example: A **diagnostic test** for a disease can have multiple possible outcomes.

- **Sample Space**:
  - Patient tests **positive** and has the disease (True Positive).
  - Patient tests **positive** but does NOT have the disease (False Positive).
  - Patient tests **negative** but has the disease (False Negative).
  - Patient tests **negative** and does NOT have the disease (True Negative).

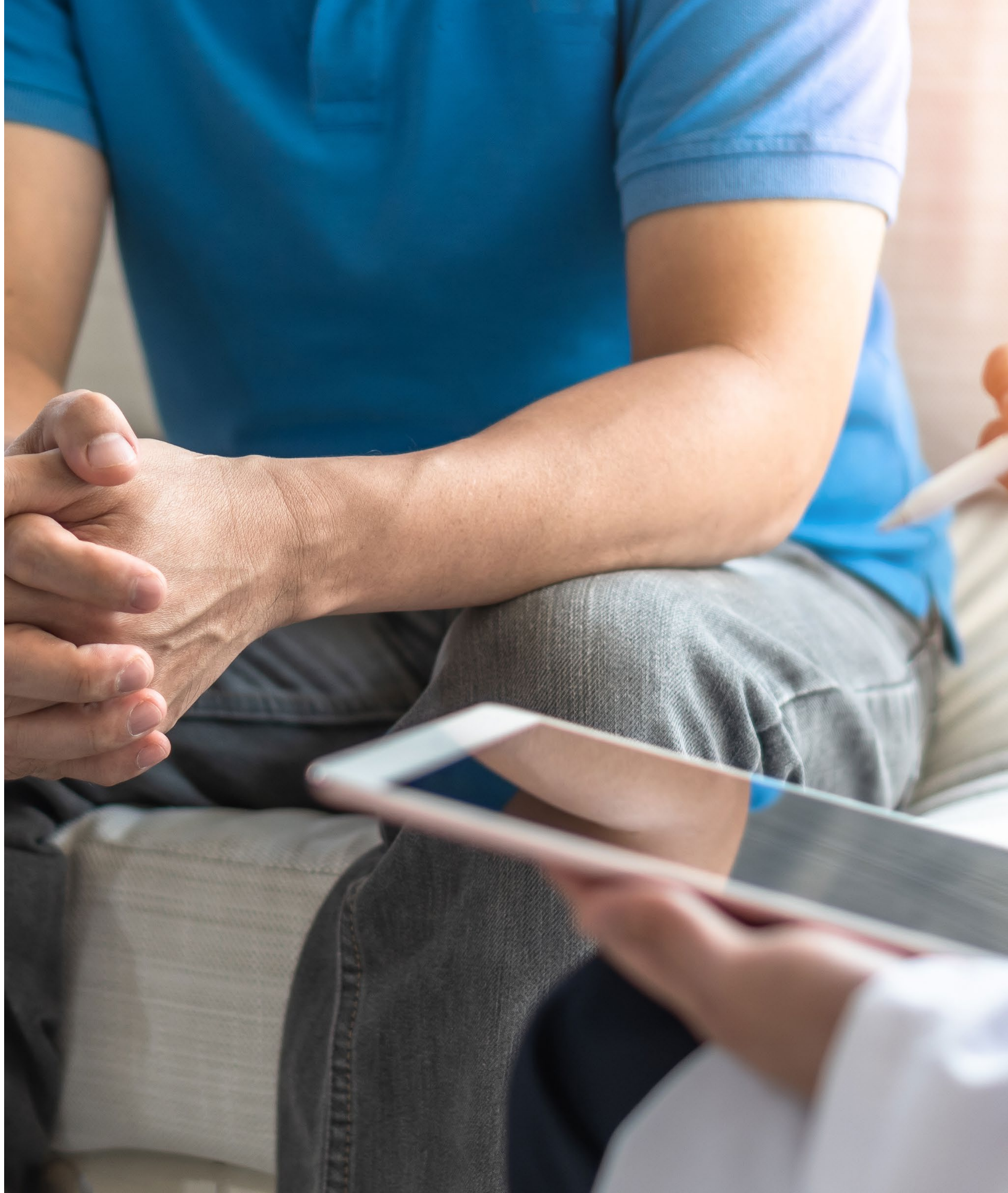
Understanding the **sample space** allows doctors to assess **test accuracy** and **make better medical decisions**.

3. INTERPRETING OUTCOMES

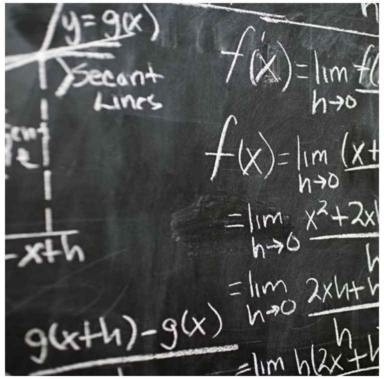
Example: A **clinical trial** for a new medication records different patient responses.

- **Outcome 1**: The patient's condition improves.
- **Outcome 2**: The patient experiences side effects but improves.
- **Outcome 3**: The patient does not respond to the medication.

By analyzing **outcomes**, researchers determine **whether the new drug is effective**.

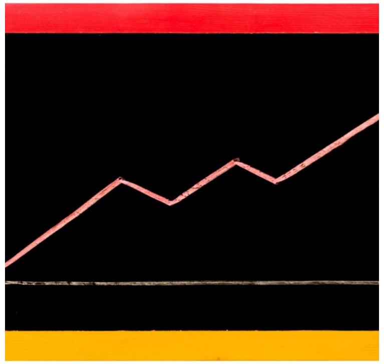


TYPES OF PROBABILITY (THEORETICAL, EXPERIMENTAL, SUBJECTIVE)



1. Theoretical Probability

Theoretical probability is determined through logical reasoning and is calculated based on all possible outcomes.



2. Experimental Probability

Experimental probability is based on actual experiments and observed results, providing empirical data for outcomes.

3. Subjective Probability

Subjective probability relies on personal beliefs or opinions regarding the likelihood of an event occurring.

HOW TO RELATE IT IN HEALTHCARE DOMAIN?

THEORETICAL PROBABILITY

Example:

If a hospital knows that a certain genetic condition affects 1 in 1,000 people, the theoretical probability of a randomly selected patient having the condition is 1/1000 (0.001 or 0.1%).

Explanation:

This is calculated based on known statistical data and does not require real-world testing.

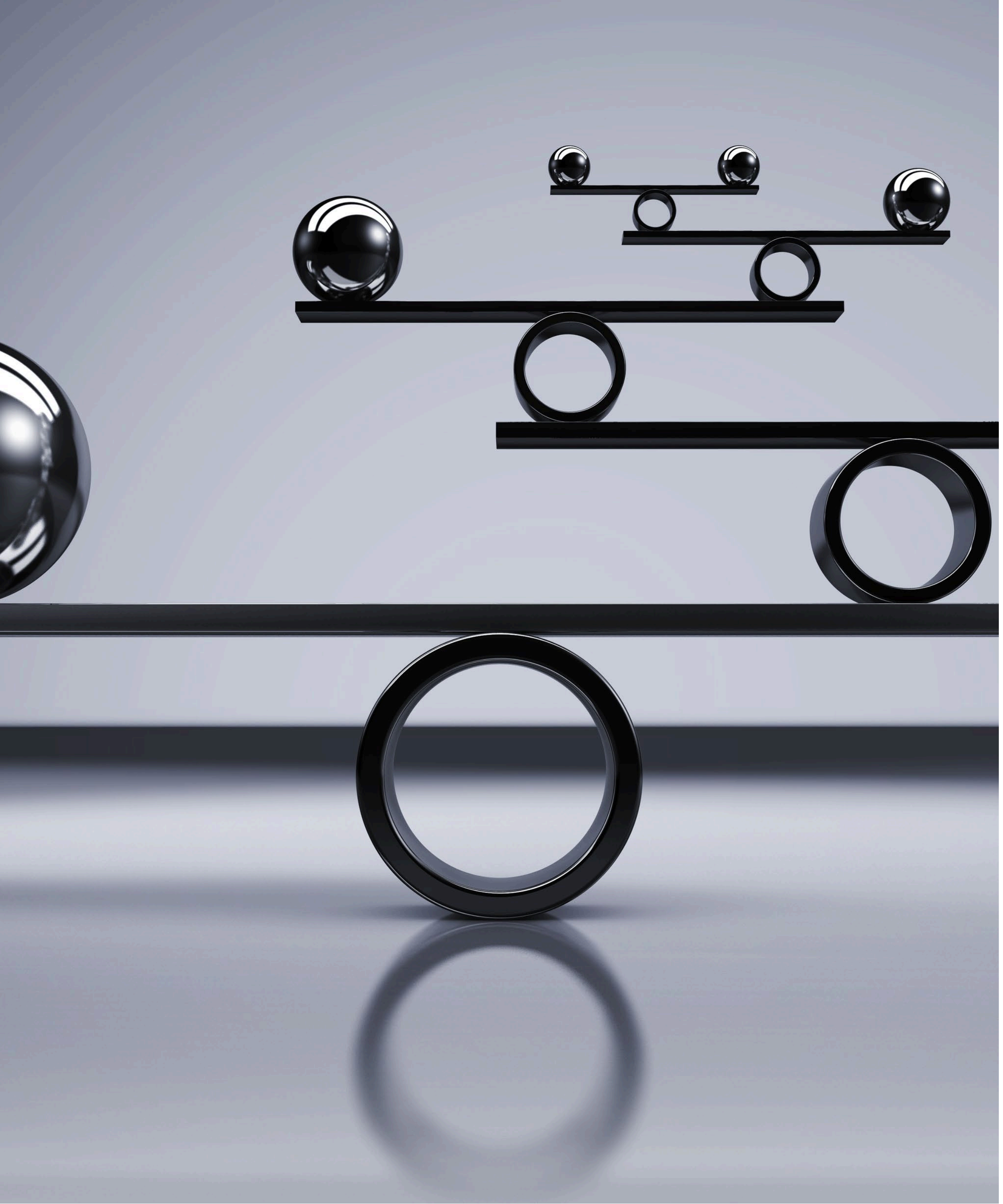
Why it matter? **Theoretical probability** helps in risk assessment and predictive modeling.

EXPERIMENTAL PROBABILITY

Example: A clinical trial tests a new vaccine on 10,000 people, and 9,500 participants develop immunity. The experimental probability of the vaccine being effective is P(Immunity) = 9500/10000 = 0.95 (or 95%)

Explanation: This probability is based on actual observed results, not just theoretical predictions.

Why it matter? **Experimental probability** is crucial in medical trials and drug effectiveness studies.



SUBJECTIVE PROBABILITY

Example: A doctor estimates that a 70-year-old smoker has a high chance of developing lung disease based on their experience and knowledge, even without precise statistics.

Explanation: This is based on personal judgment, intuition, or expert opinion, rather than hard data.

Why it matter? **Subjective probability** influences clinical decisions when data is limited.

FUNDAMENTAL RULES OF PROBABILITY

ADDITION RULE & MULTIPLICATION RULES

Understanding these rules is essential for effective statistical analysis and making informed decisions based on probabilities.

ADDITION RULES

The **addition rule** is used to calculate the probability of **at least one of two events happening**.

Example:
A hospital tracks patients who have **hypertension (H)** and **diabetes (D)**. The probabilities are:

- $P(H) = 0.30$ (30% of patients have hypertension)
- $P(D) = 0.25$ (25% of patients have diabetes)
- $P(H \cap D) = 0.10$ (10% of patients have both conditions)

Using the **addition rule**, the probability of a patient having **at least one of these conditions** is:

$$P(H \cup D) = P(H) + P(D) - P(H \cap D)$$

$$P(H \cup D) = 0.30 + 0.25 - 0.10 = 0.45$$

Interpretation:
There is a **45% probability** that a randomly selected patient has **either hypertension or diabetes** (or both).

In healthcare situation: The **addition rule** helps doctors and researchers estimate **disease prevalence and risk factors**.

MULTIPLICATION RULES

The **multiplication rule** is used to find the probability of **two independent events occurring together**.

Example:
Suppose a hospital finds that:

- The probability of a patient **recovering from surgery within a week** is $P(R)=0.80$ (80%).
- The probability of a patient **not experiencing post-surgical complications** is $P(C)=0.90$ (90%).
- If recovery and complications are **independent events**, then, the probability that a patient **both recovers within a week and has no complications** is:

$$P(R \cap C)=P(R)\times P(C)$$

$$P(R \cap C)=0.80\times 0.90=0.72$$

Interpretation:
There is a **72% probability** that a patient will **recover within a week and have no complications**.

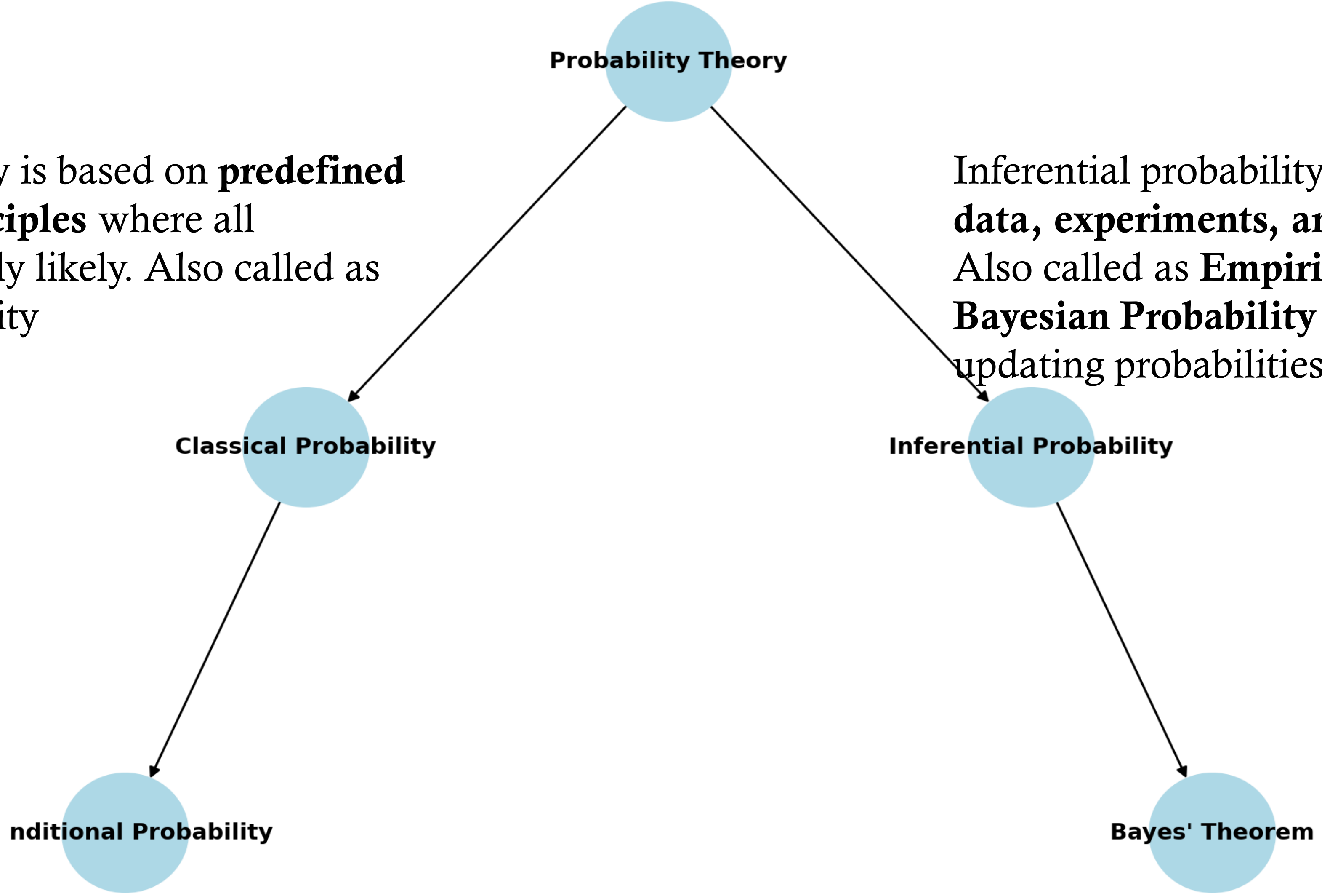
In healthcare situation: The **multiplication rule** is useful for predicting **treatment success rates** and **evaluating risks in patient care**.

PROBABILITY THEORY

1. THEORETICAL PROBABILITY THEORY
2. INFERENTIAL PROBABILITY THEORY

Flowchart: Where Conditional Probability & Bayes' Theorem Fit in Probability Theory

Classical probability is based on **predefined mathematical principles** where all outcomes are equally likely. Also called as theoretical probability



Inferential probability is based on **observed data, experiments, and statistical methods**. Also called as **Empirical Probability or Bayesian Probability** because it involves updating probabilities based on evidence.

Conditional Probability

Bayesian Theorem

Key Differences Between Classical & Inferential Probability

Feature	Classical Probability	Inferential Probability
Definition	Based on known sample spaces and logic	Based on observed data and experiments
Nature	Theoretical	Data-driven
Formula	$P(A) = \frac{\text{Favorable Outcomes}}{\text{Total Outcomes}}$	$P(A) = \frac{\text{Observed Occurrences}}{\text{Total Trials}}$
Real-World Example	Probability of rolling a 6 on a fair die	Probability of vaccine success based on clinical trial data
Medical Example	Probability of inheriting a genetic disease	Probability of having a disease given a positive test result (Bayes' Theorem)
Use Cases	Gambling, genetics, dice games	AI, clinical trials, disease prediction

1. CONDITIONAL PROBABILITY

Definition of Conditional Probability

The probability of an event occurring given that another event has already occurred.

Formula:

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

Where

$P(A|B)$ = Probability of event A occurring given that B has occurred.

$P(A \cap B)$ = Probability that both events A and B occur.

$P(B)$ = Probability of event B occurring.

EXAMPLE: MEDICAL DIAGNOSTIC ON FLU & FEVER

A doctor is checking patients for the flu. Not everyone with a fever has the flu, but the flu often causes a fever. The doctor knows the following:

- 30% of all patients have the flu → $P(\text{Flu})=0.30$
- 80% of flu patients have a fever → $P(\text{Fever}|\text{Flu})=0.80$
- 50% of all patients (flu or not) have a fever → $P(\text{Fever})=0.50$

Question: If a Patient has a fever, what is the probability that they have the Flu?

The formula for Conditional Probability is: $P(\text{Flu}|\text{Fever})=\dfrac{P(\text{Fever} \cap \text{Flu})}{P(\text{Fever})}$

We already know:

$P(\text{Fever} \cap \text{Flu})$ = Probability of having both a fever and the flu

$P(\text{Fever} \cap \text{Flu})=P(\text{Fever}|\text{Flu}) \times P(\text{Flu}) = 0.80 \times 0.30 = 0.24$

$P(\text{Fever}) = 0.50$

Now, applying the conditional probability formula:

$$P(\text{Flu}|\text{Fever}) = \frac{P(\text{Fever} \cap \text{Flu})}{P(\text{Fever})} = \frac{0.24}{0.5} = 0.48$$

Interpretation

1. If a patient **has a fever**, the chance that they **have the flu is 48%**.
2. **Why is it not 80%?** Because **some people have a fever from other causes**, like infections or heat exhaustion.

MORE EXAMPLE WITH DATA

A hospital is studying the probability of patients testing positive for a certain disease given that they show specific symptoms. The hospital has collected data from 1000 patients regarding whether they tested positive for the disease and whether they exhibited a high fever.

Data Set: hypothetical data for 1000 patients

Fever (Symptom)	Disease Positive (Yes)	Disease Negative (No)	Total
Yes	120	280	400
No	50	550	600
Total	**170**	**830**	**1000**

Question: A patient comes into the hospital and has a high fever. What is the probability that they have the disease?

SOLUTION:

This requires us to calculate P(Disease|Fever), which is the probability of having the disease given that the patient has a fever.

Using the conditional probability formula:

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

- $P(A|B)$ = Probability of having the disease given that the patient has a fever.
- $P(A\cap B)$ = Probability of both having a fever and testing positive for the disease.
- $P(B)$ = Probability of having a fever.

From the dataset:

- $P(A\cap B) = 120/1000 = 0.12$ (120 patients have both the disease and a fever)
- $P(B) = 400/1000 = 0.40$ (400 patients have a fever)

Now, calculating:

$P(\text{Disease}|\text{Fever})$ or $P(A|B) = 0.12/0.40 = 0.30$

CONCLUSION:

If a patient has a fever, the probability that they have the disease is **30%**.

TEST YOUR UNDERSTANDING

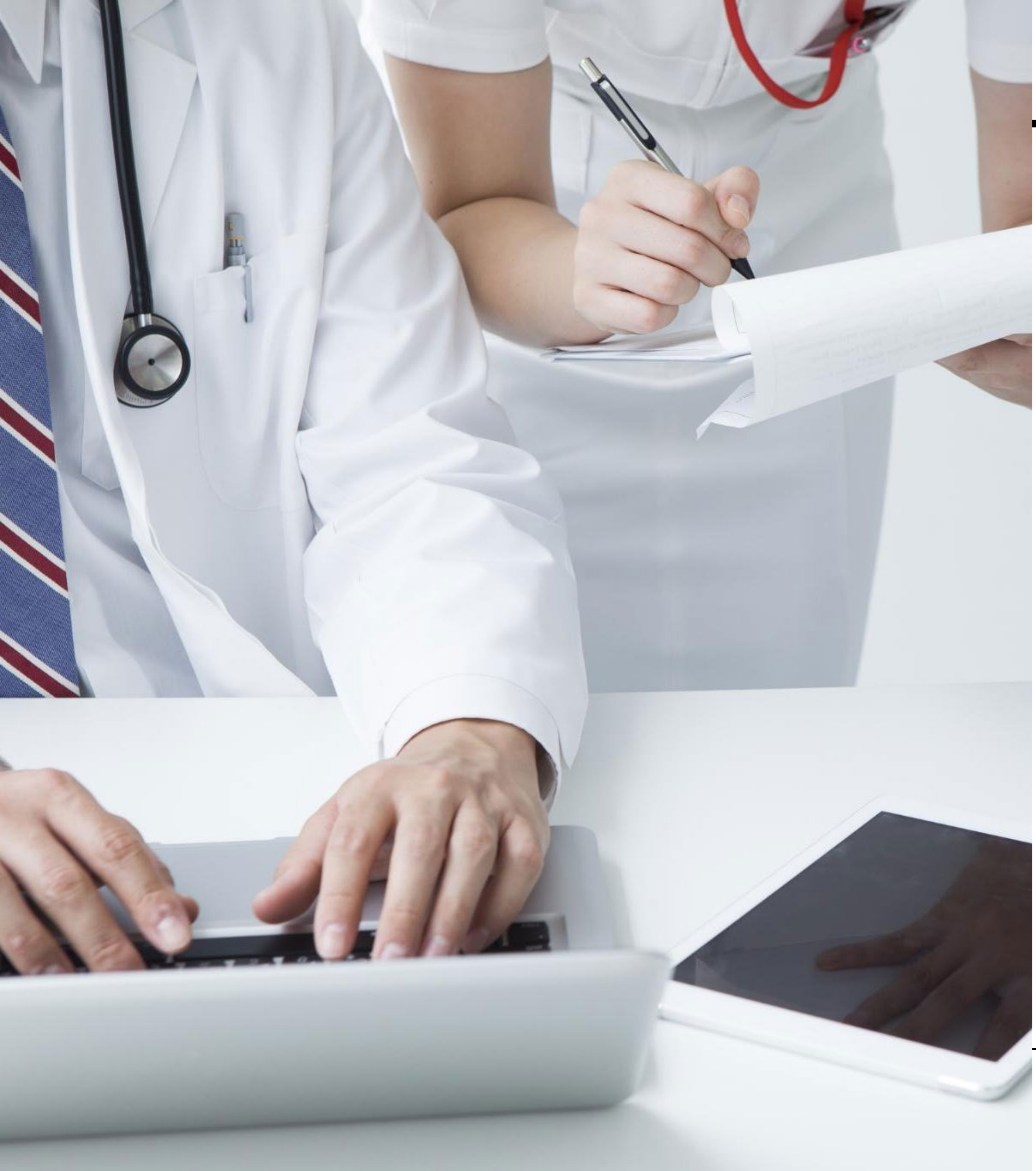
A clinic is investigating the probability of a patient having diabetes given that they have high blood sugar levels. They have collected data from 1200 patients, recording whether they have diabetes and whether their blood sugar levels were high.

Data Set: hypothetical data for 1200 patients

High Blood Sugar	Diabetes (Yes)	Diabetes (No)	Total
Yes	180	320	500
No	90	610	700
Total	**270**	**930**	**1200**

HOMEWORK QUESTIONS:

1. Calculate P(Diabetes|HighBloodSugar) using the provided dataset. **[0.36 or 36%]**
2. What is P(Diabetes), the probability of having diabetes regardless of blood sugar levels? **[0.225 or 22.5%]**
3. What is P(HighBloodSugar), the probability of a patient having high blood sugar? **[0.42 or 42%]**
4. Compare P(Diabetes|HighBloodSugar) with P(Diabetes). Does having high blood sugar increase the likelihood of diabetes? Explain. **[The probability of diabetes in the general population is 22.5%, but for those with high blood sugar, it rises to 36%. This means high blood sugar is associated with a greater risk of diabetes.]**



2. BAYES' THEOREM

Understanding Bayes' Theorem

A mathematical formula that allows to reverse conditional probabilities. It helps calculate the probability of a cause (A) given that we observed an effect (B).

Formula:

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

P(A|B) = Probability of event A given B.

P(B|A) = Probability of event B given A.

P(A) = Prior probability of A occurring.

P(B) = Total probability of B occurring..

EXAMPLE: HEART DISEASE AND SMOKING

A hospital studies the probability that a patient **develops heart disease given that they are a smoker**.

- **P(Heart Disease) = 20%** (20% of patients have heart disease or 0.20)
- **P(Smoker ∩ Heart Disease) = 60%** (60% of heart disease patients are smokers or 0.60)
- **P(Smoker) = 30%** (30% of all patients are smokers or 0.30)

Using conditional probability:

$$\boldsymbol{P(Heart\ Disease|Smojer)\ or}$$

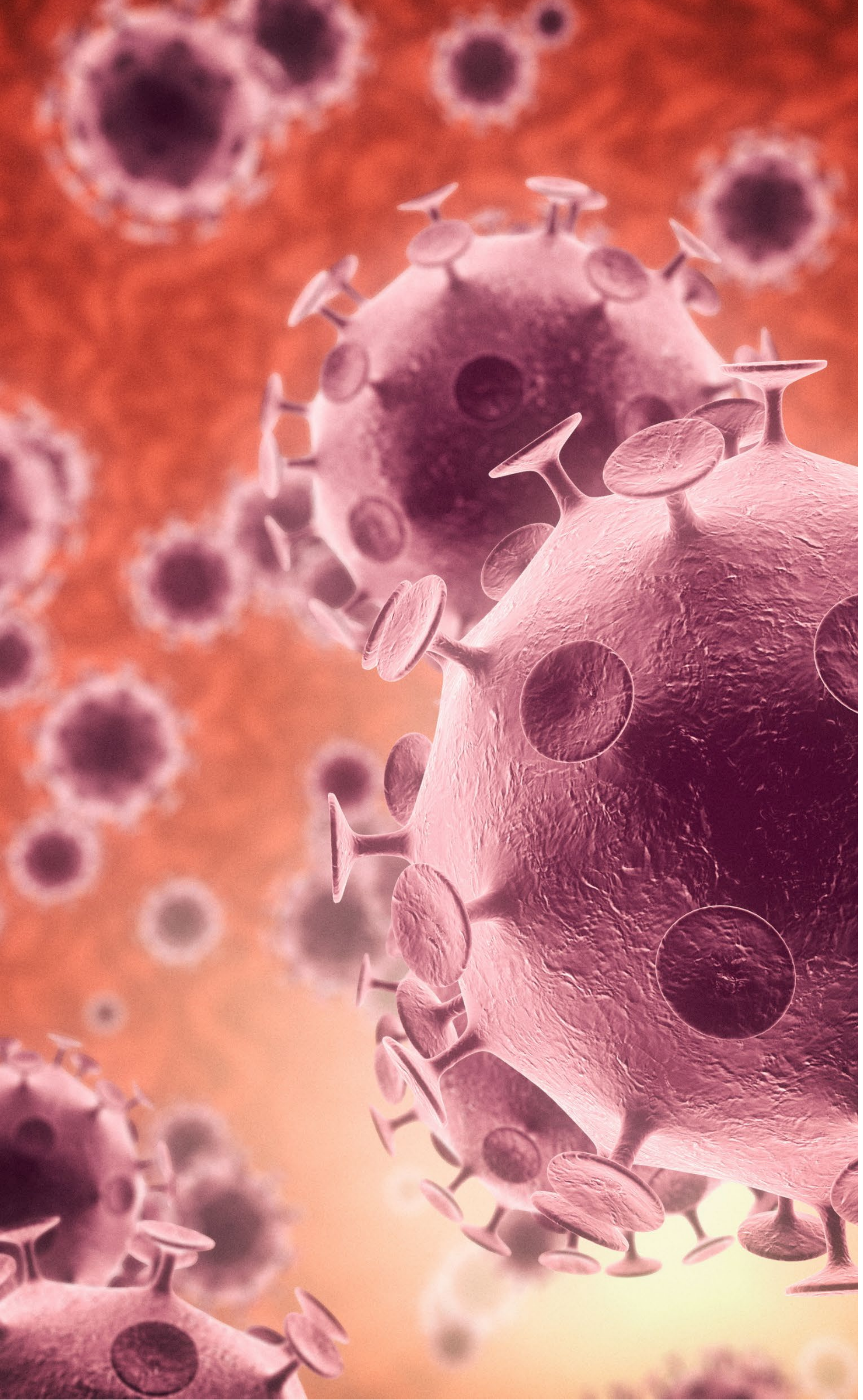
$$\text{probabiity develop heart disease given a smoker}$$

$$= \frac{\textbf{P(Smoker∩Heart Disease) x P(Heart Disease)}}{\textbf{P(Smoker)}}$$

$$= \frac{\textbf{0.60 x 0.20}}{\textbf{0.30}} = \mathbf{0.40}\ (\boldsymbol{or}\ \mathbf{40\%})$$

Interpretation**:** A smoker has a **40% probability of having heart disease**, which is **higher than the general population of patient having heart disease (20%)**.

MORE EXAMPLE WITH DATA

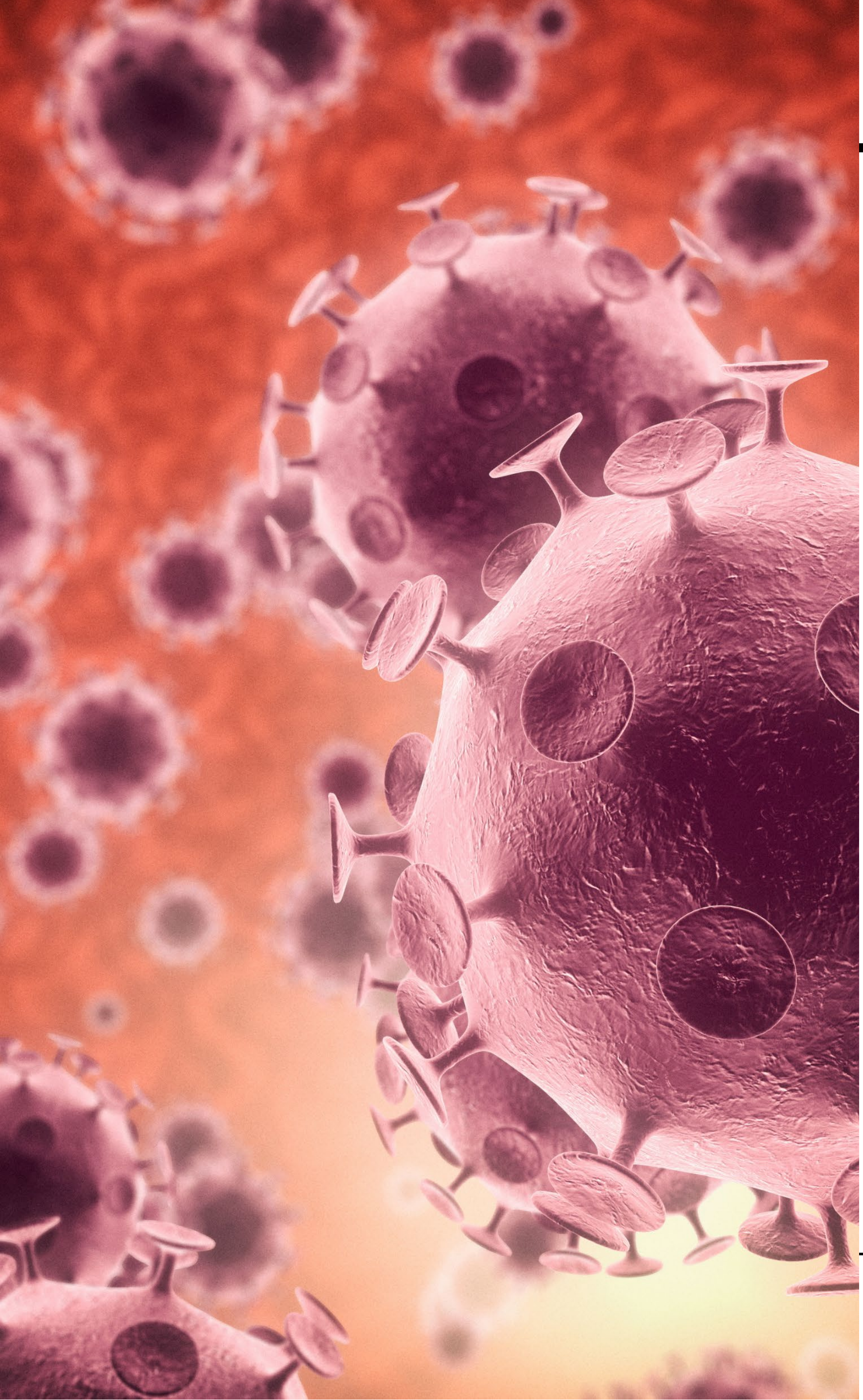


Problem Statement:

A hospital is testing for a rare disease that affects **1%** of the population. A diagnostic test for the disease is **90% accurate** when detecting the disease (true positive rate), but it also has a **5% false positive rate** (i.e., it incorrectly diagnoses healthy patients as having the disease 5% of the time). A patient has tested positive for the disease. What is the probability that they actually have the disease?

Data Given:

- **P(Disease) = 0.01** (1% of people have the disease)
- **P(No Disease) = 0.99** (99% of people do not have the disease)
- **P(Positive Test | Disease) = 0.90** (True positive rate)
- **P(Positive Test | No Disease) = 0.05** (False positive rate)



Solution:

- Firstly, need to calculate **P(Positive Test)** that is Total probability of testing positive. This can be calculated using the Law of Total Probability:
- P(PositiveTest) = [P(PositiveTest|Disease) x P(Disease)] + [P(PositiveTest|NoDisease) x P(NoDisease)]

  $= (0.90\times0.01)+(0.05\times0.99)$

  $= 0.009+0.0495$

  $= 0.0585$

- Then, can calculate P(Disease | Positive Test) using the Bayes Theorem

$$= \frac{P(Positive\ Test|Disease)xP(Disease)}{P(Positive\ Test)} = \frac{0.90\ x\ 0.01}{0.0585} = 0.154 \sim 15.4\%$$

Explanation: Even though the test was **90% accurate**, the probability that the patient **has** the disease after testing positive is **only 15.4%**. This is because the disease is rare, and the number of false positives contributes significantly to the total positive tests.

TEST YOUR UNDERSTANDING

TAKE HOME QUESTION:

1. What happens to the probability P(Disease | Positive Test) if the disease becomes more common in the population (whereby it is now 10% instead of 1%)?

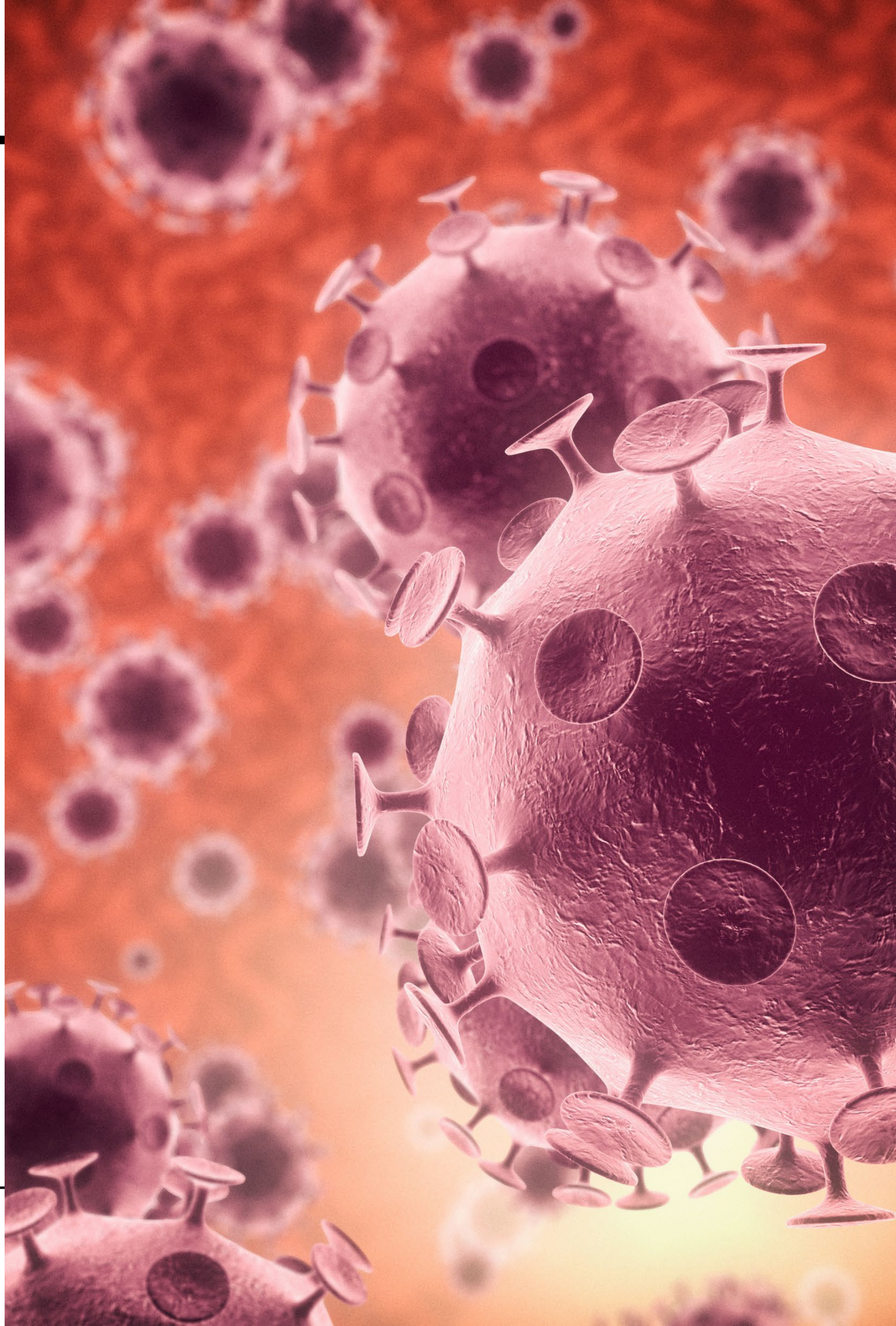
[0.667 or 66.7%]

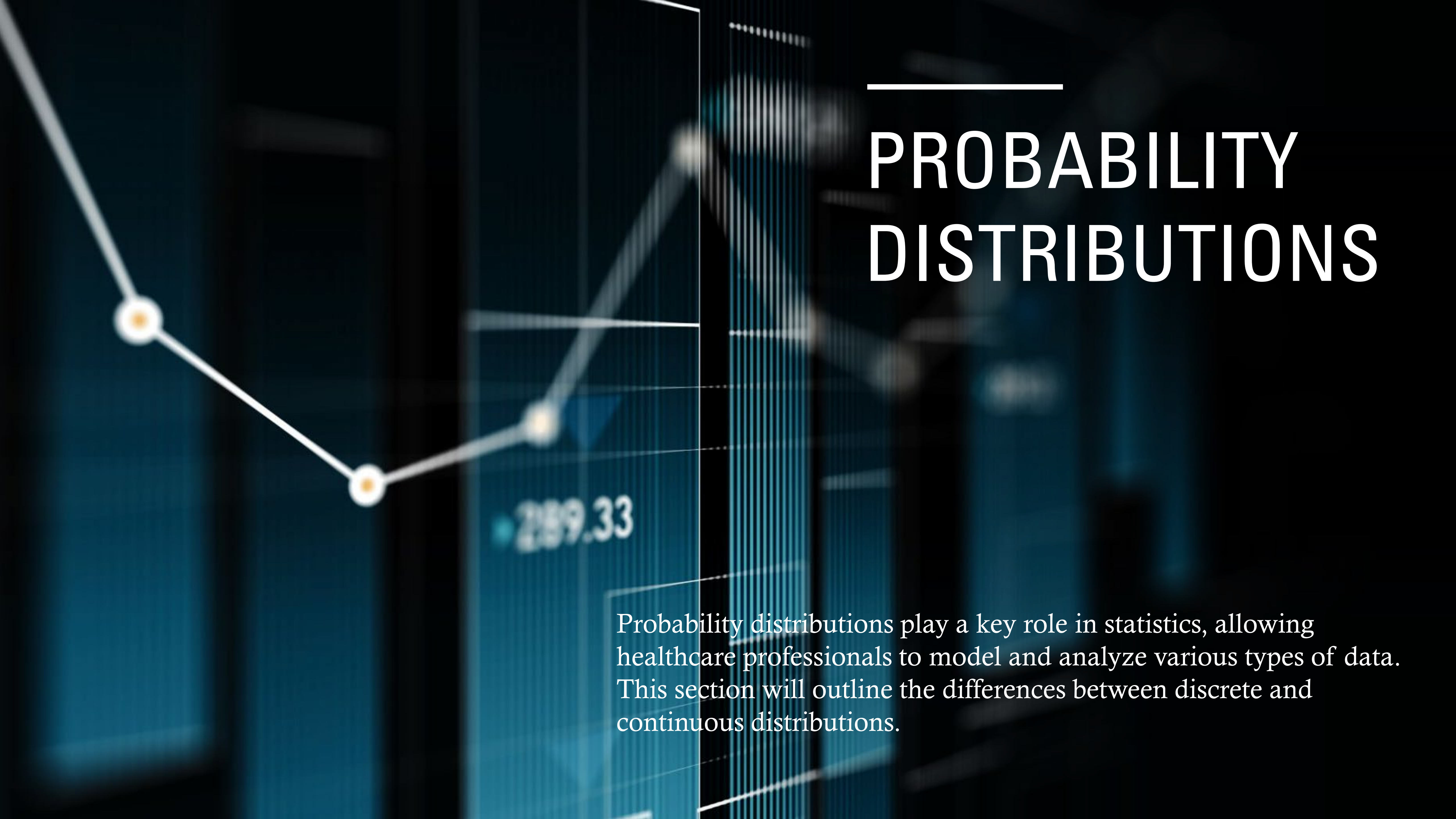
Summary: If the disease is more common in the population (10% prevalence), the probability that a patient has the disease after testing positive increases significantly to 66.7%.

2. What happens if the test's accuracy increases (whereby it is now 99% true positive rate instead of 90%) – disease prevalence remains 1%?

[0.1667 or 16.67%]

Summary: Even with a more accurate test (99% sensitivity), the probability that a patient has the disease after testing positive only increases slightly to 16.7%. This shows that test accuracy alone does not completely resolve the problem of false positives when the disease is rare.





PROBABILITY DISTRIBUTIONS

Probability distributions play a key role in statistics, allowing healthcare professionals to model and analyze various types of data. This section will outline the differences between discrete and continuous distributions.

DISCRETE VS. CONTINUOUS DISTRIBUTIONS

Discrete Distributions

Discrete distributions deal with countable outcomes such as the number of patients, events, or occurrences.

Continuous Distributions

Continuous distributions apply to measurable outcomes, representing data that can take any value within a range.

Importance in Healthcare

Understanding the differences between these distributions helps in accurately analyzing healthcare statistics and outcomes.

COMMON DISTRIBUTIONS IN HEALTHCARE (BINOMIAL, NORMAL, POISSON)

Binomial Distribution

The binomial distribution models the number of successes in a fixed number of trials, particularly useful in clinical settings.

Normal Distribution

Normal distribution is essential for analyzing continuous data such as blood pressure, weight, and other health metrics.

Poisson Distribution

The Poisson distribution models the number of events occurring in a fixed interval, such as patient arrivals or disease incidence.

CONCLUSION

Importance of Probability

Understanding probability is crucial for healthcare professionals, influencing their decision-making process in patient care.

Risk Assessment

Knowledge of probability helps in accurate risk assessment, leading to better patient outcomes and management strategies.

Statistical Evidence

Statistical evidence derived from probability allows healthcare professionals to provide data-driven care and treatment plans.

WE CONTINUE THE JOURNEY WITH STATISTICAL TEST