

# Introduction to Probability and Statistics

UG 6th Semester Course (Batch 2022)

By  
Prof. M. Iqbal Bhat  
GDC Beerwah

# Course Overview

## 1 Objectives

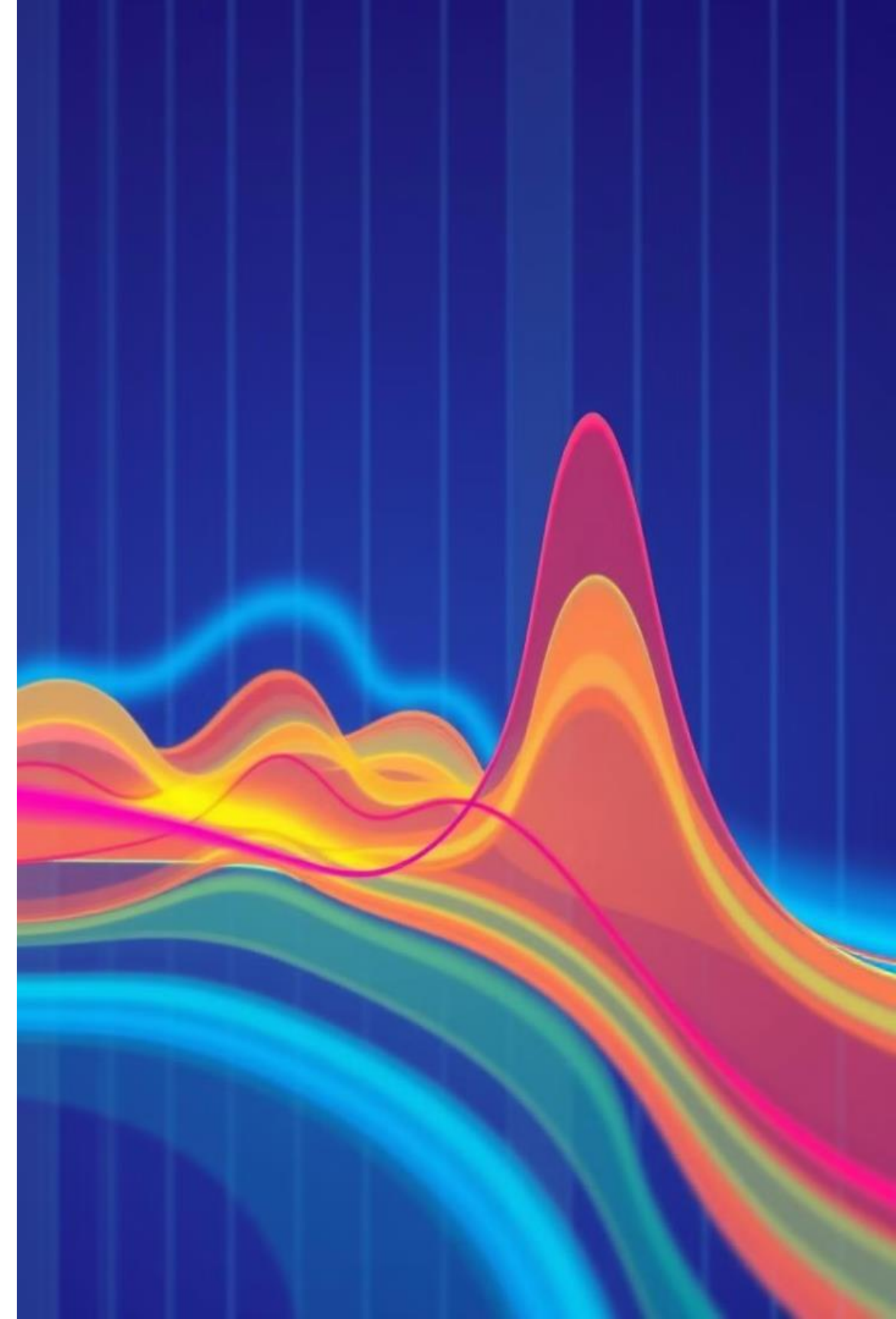
Understand foundational concepts.

## 2 Applications

Apply to real-world scenarios & AI.

## 3 Probability vs Statistics

Understand the relationship





# Ten Amazing Probabilistic Facts in the Universe

Probability shapes our universe in surprising ways. Let's explore ten fascinating examples, from meteor strikes to genetic mutations. These facts highlight the role of chance in everything around us. They also reveal some insights into the amazing processes that shape the world.



# Rare Meteor Strikes

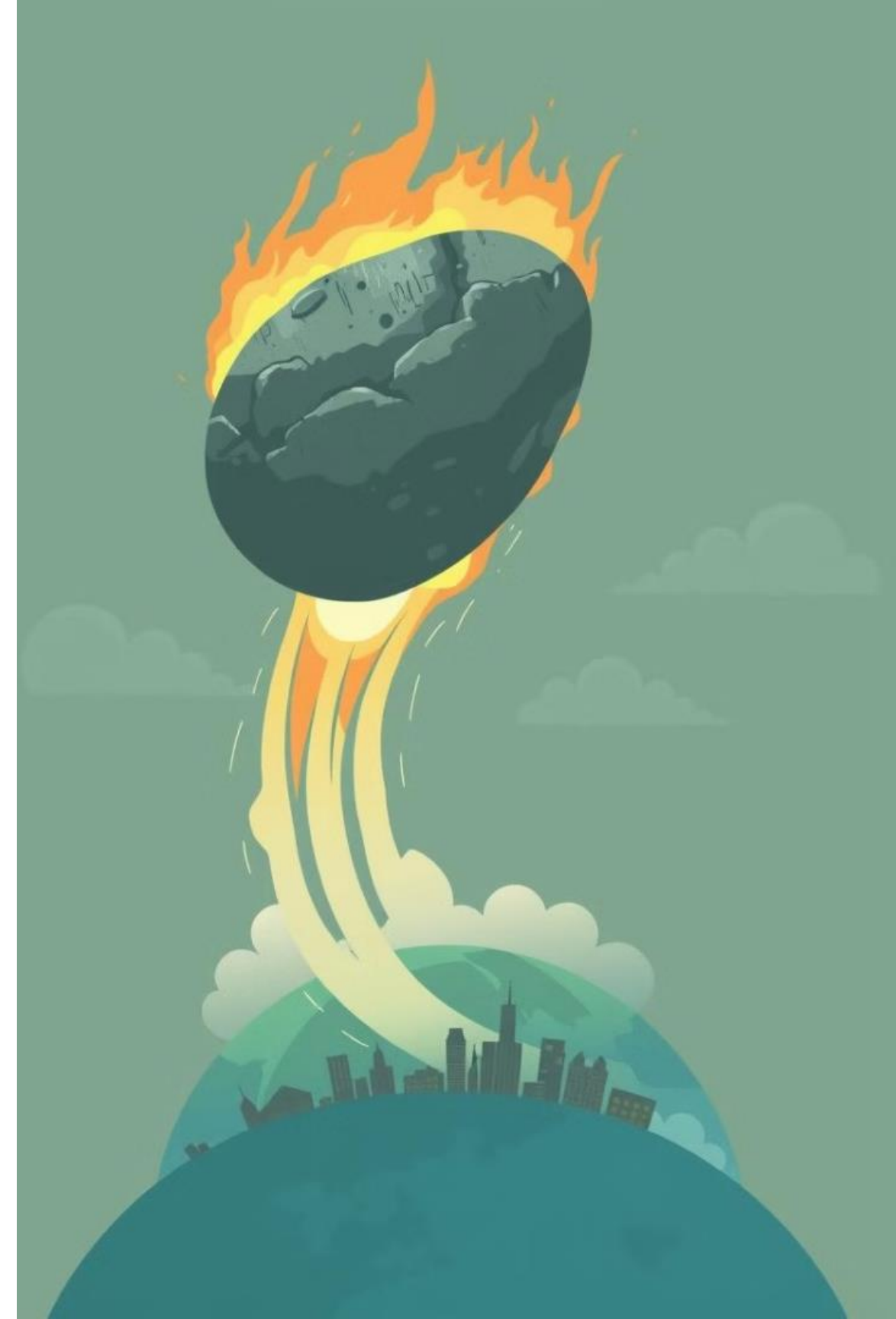
## 1 Low Probability

The annual probability of a 1-kilometer-wide asteroid hitting Earth is roughly 1 in 300,000.

## 2 Catastrophic Impact

Though low, the global impact of such an event would be devastating.

Despite the low chance, the potential for widespread destruction underscores the need for monitoring and defense strategies. Resources can be deployed to protect the planet. Scientists work to predict the path of asteroids.



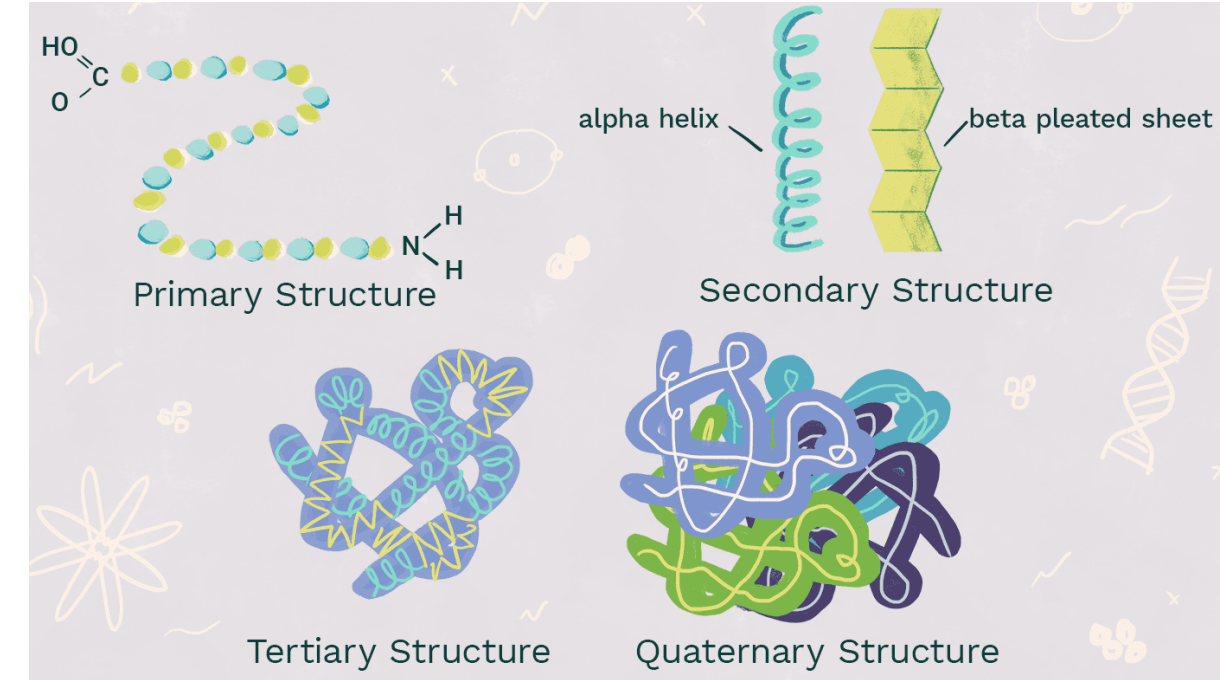
# Formation of Proteins

## Vast Possibilities

The number of possible sequences for a protein of length 300 is astronomically large.

The number of possible sequences for a protein of length 300 is  $20^{300}$ —an astronomically large number (on the order of  $10^{390}$ )

It would be improbable for functional proteins to arise by random chance alone. The process highlights the intricate balance of chemistry, biology, and probability.



## Molecular Machinery

Molecular machinery and evolutionary processes drastically increase the likelihood of correct protein formation.

# Deck of Cards Uniqueness

1

## Staggering Arrangements

The number of ways to arrange a standard 52-card deck is  $52!$  approximately  $8.0658 \times 10^{67}$ .

The probability of any one specific arrangement is  $1 / 52!$

2

## Unprecedented Sequences

Nearly every shuffle results in a sequence of cards that has never been seen before in human history.

The sheer scale emphasizes the concept of combinatorial explosion. With even a small number of items, there are an unbelievable number of arrangements.



# Coin Flip Streaks

## Low Probability

The probability of flipping 10 heads in a row is  $1 / 2^{10} = 1/1024$ .

## Decreasing Odds

This demonstrates how quickly probabilities drop as the required streak lengthens.

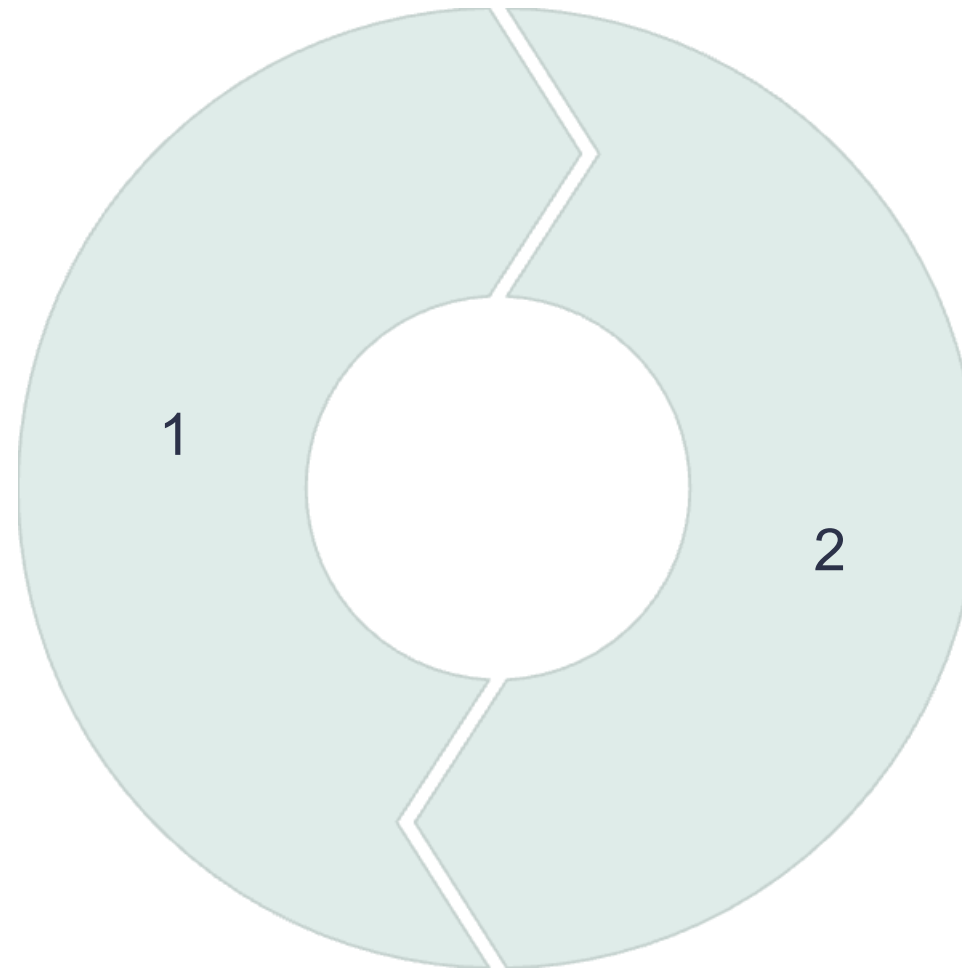
While not extremely rare, such streaks illustrate the nature of random events. Even with equal chances, surprising patterns can emerge. It is not guaranteed to switch.



# Genetic Mutation Rates

## Rare Per Base

The probability that a single base in your DNA mutates in one generation is around 1 in 100 million.



## Billions of Bases

Over billions of bases, mutations do occur but are still relatively rare per site.

The low mutation rate helps preserve the integrity of genetic information. Despite the rarity, these mutations drive evolutionary change over time. They are the engine of change.





# Quantum Uncertainty



Subatomic particles do not have fixed positions.



They are described by probability distributions, also known as wavefunctions.

This is a fundamental aspect of quantum mechanics known as Heisenberg's uncertainty principle. Position and momentum cannot both be known.



# Winning a Major Lottery

1

## Long Odds

For a typical 6/49 lottery, the odds of winning the jackpot can be 1 in ~14 million.

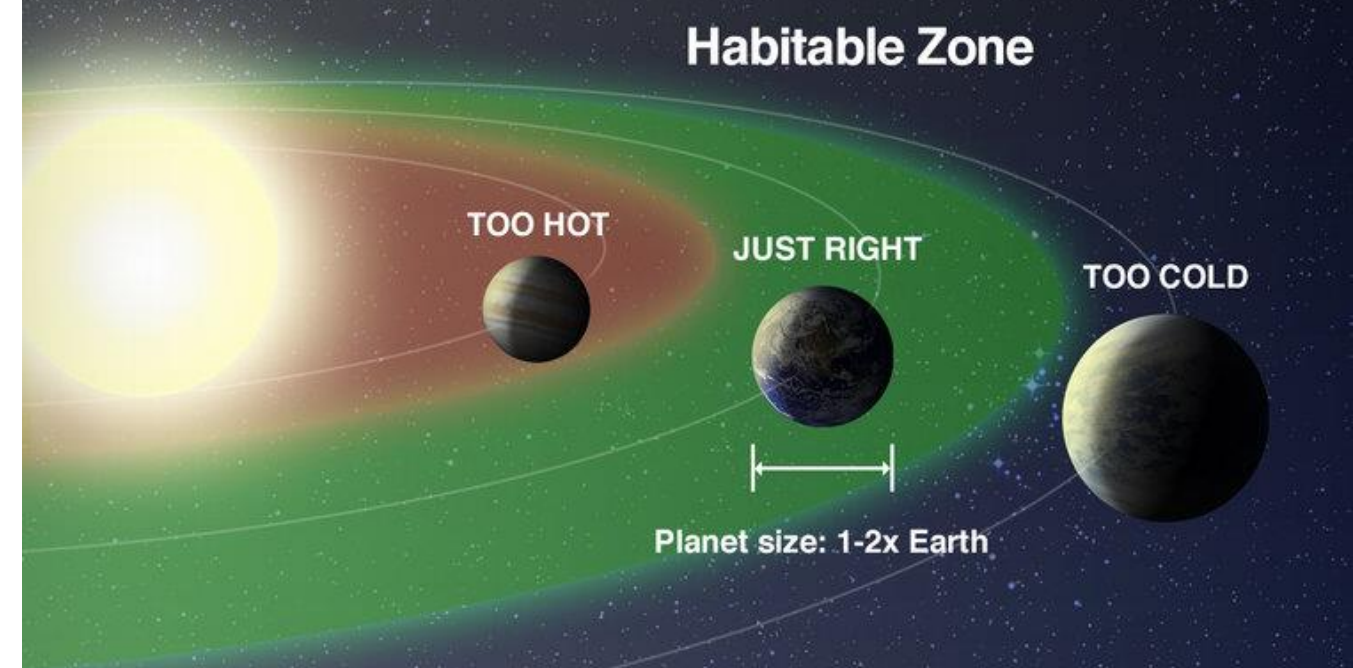
2

## Human Behavior

This illustrates how small probabilities can still motivate significant human behavior.

Despite the odds, many people buy lottery tickets. This highlights the psychology of risk and reward. Even a small chance can be appealing.

# Earth's Goldilocks Zone



1

## Billions of Planets

There are potentially billions of “Earth-like” worlds in the universe.

2

## Habitable Zone

The probability that a planet orbits in the habitable zone is low.

Given the vast number of stars and exoplanets, life could arise. The conditions must be perfect for liquid water to exist, but they do exist.



# Predictive Text in AI



## Conditional Probabilities

Large language models rely on the probability of word sequences. Each next word is chosen based on conditional probabilities learned from text corpora.

AI chatbots leverage vast amounts of data. Probability is used to accurately predict the next word. The applications are endless.

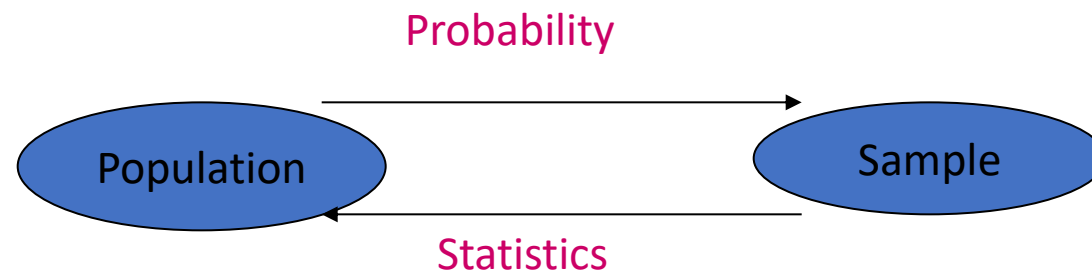
## Learning from Text

Predictive text is based on conditional probabilities learned from huge text corpora.



# Why Learn Probability?

- Nothing in life is certain. In everything we do, we gauge the chances of successful outcomes, from business to medicine to the weather
- A probability provides a quantitative description of the chances or likelihoods associated with various outcomes
- It provides a bridge between descriptive and inferential statistics



# Importance & Need

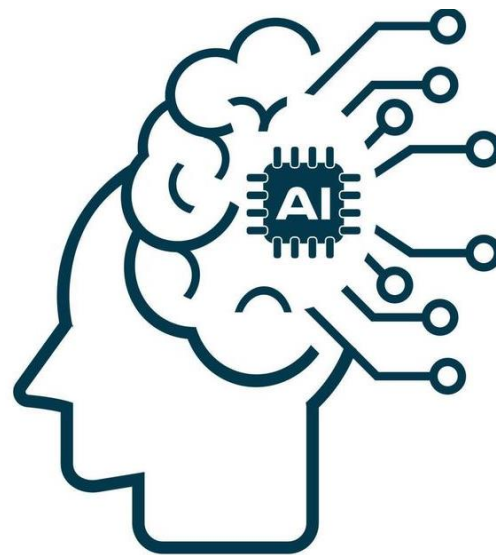
## Decision-Making

Manage uncertainty in real-world problems.



## AI & Machine Learning

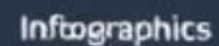
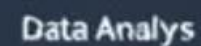
Foundational tool for algorithms.



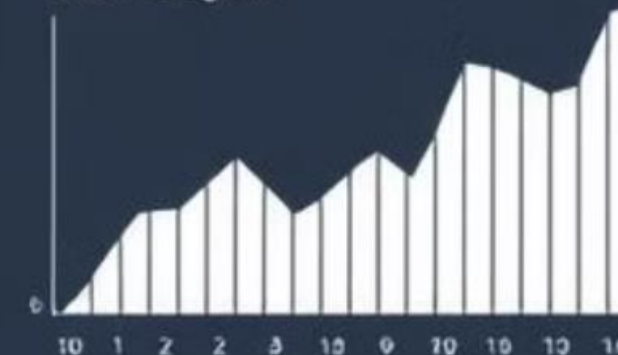
## Applications

Across engineering, finance, medicine.





The pain is a pressure and the correct response is to draw up your eyebrows and relax the chin and lips. You will see that you do not have to hold your breath and you do not have to hold your body in a rigid position.



# Core Concepts

# Probability Theory

Axioms, conditional probability,  
Bayes' theorem.

## Descriptive Statistics

Central tendency, dispersion, data visualization.

# Inferential Statistics

Sampling, estimation, hypothesis testing.



# Why Study in AI?



## Uncertainty

Handle incomplete or noisy data.



## Predictive Modeling

Essential for training ML models.



## Decision Theory

Guide optimal decisions under risk.





# Probability and Statistics in the Real World

# Finance & Risk Management

## Stock Price Modeling

Probability and statistics are used to model stock prices, helping investors understand potential risks and returns. These models often incorporate factors like volatility and market trends to predict future price movements.

## Portfolio Optimization

By applying statistical techniques, financial analysts can optimize investment portfolios to maximize returns while minimizing risk. This involves analyzing correlations between different assets and allocating capital accordingly.

## Credit Risk Assessment

Lenders use statistical models to assess the creditworthiness of borrowers. These models consider factors like credit history, income, and debt levels to determine the probability of default.

Reference: Hull, J. C. (2017). *Options, Futures, and Other Derivatives* (10th ed.). Pearson.

# Insurance & Actuarial Science

## Premium Calculation

Actuaries use probability and statistics to calculate insurance premiums based on risk factors. This ensures that premiums are fair and accurately reflect the likelihood of claims.

## Life Expectancy Tables

Statistical analysis of mortality data is used to create life expectancy tables, which are essential for pricing life insurance policies and managing pension funds.

## Risk Pooling

Insurance companies use risk pooling to spread risk across a large group of policyholders. Statistical models help determine the optimal size and composition of risk pools.

Reference: Bowers, N. L. et al. (1997). *Actuarial Mathematics* (2nd ed.). Society of Actuaries.





# Data Science & Machine Learning



## Bayesian Networks

Bayesian networks are probabilistic graphical models that represent dependencies between variables. They are used for reasoning under uncertainty and making predictions.



## Probabilistic Graphical Models

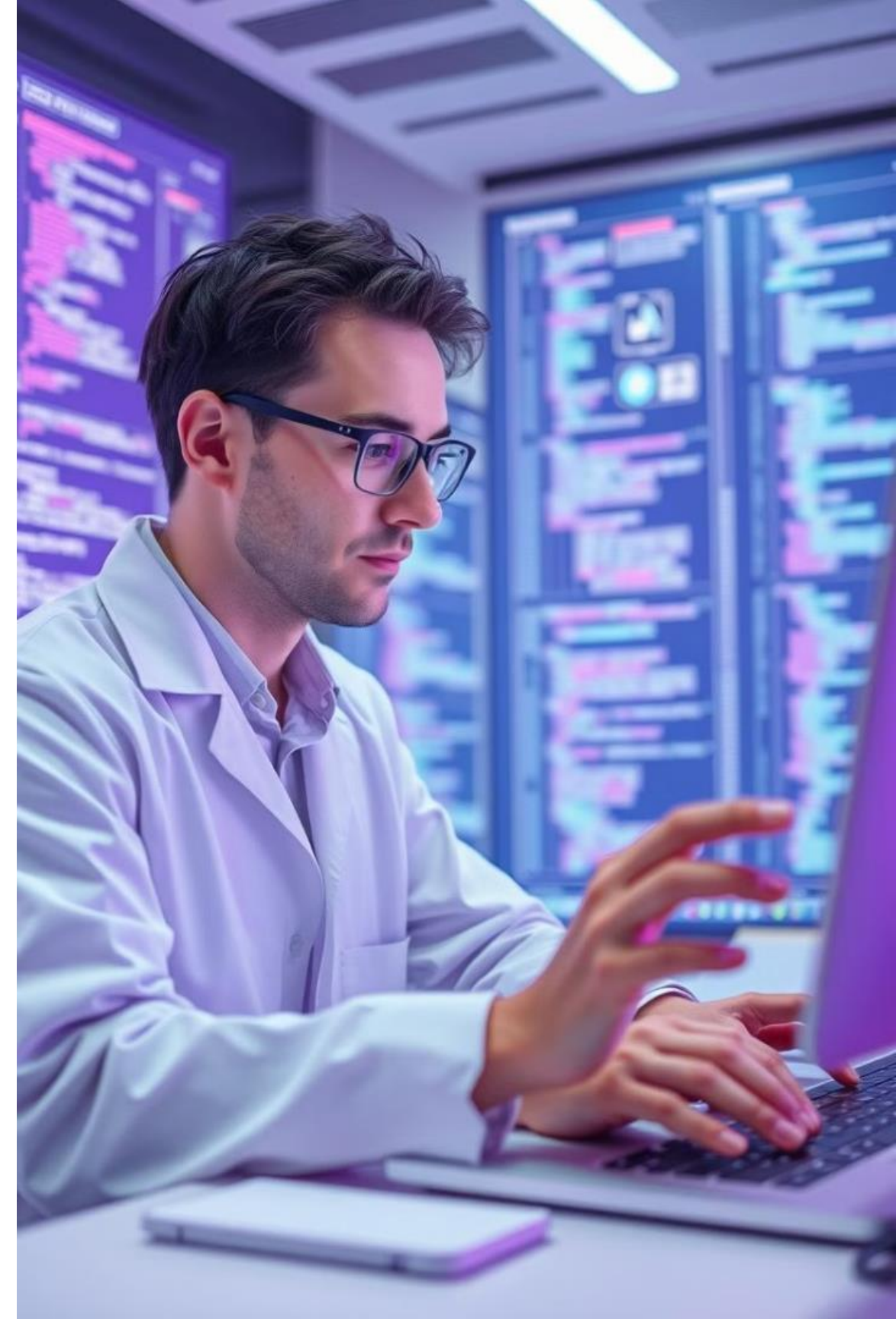
These models provide a framework for representing complex systems and making inferences based on data. They are widely used in machine learning and artificial intelligence.



## Predictive Modeling

Statistical techniques are used to build predictive models that forecast future outcomes based on historical data. These models are used in a variety of applications, from marketing to finance.

Reference: Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective*. MIT Press.







# Healthcare & Epidemiology

1

## Clinical Trials

Statistical methods are essential for analyzing data from clinical trials to determine the efficacy of new drugs and treatments. This ensures that medical interventions are safe and effective.

2

## Disease Spread Models

Epidemiologists use statistical models to understand and predict the spread of infectious diseases. These models help inform public health interventions and control measures.

3

## Survival Analysis

Survival analysis is a statistical technique used to analyze time-to-event data, such as the time until a patient dies or experiences a relapse. This helps researchers understand factors that influence survival rates.

Reference: Rosner, B. (2015). *Fundamentals of Biostatistics* (8th ed.). Cengage Learning.

# Genetics & Bioinformatics

1

## Genetic Mutation Rates

Statistical methods are used to estimate the rates at which genes mutate over time. This helps scientists understand the evolution of species and the development of genetic diseases.

2

## Sequence Alignment

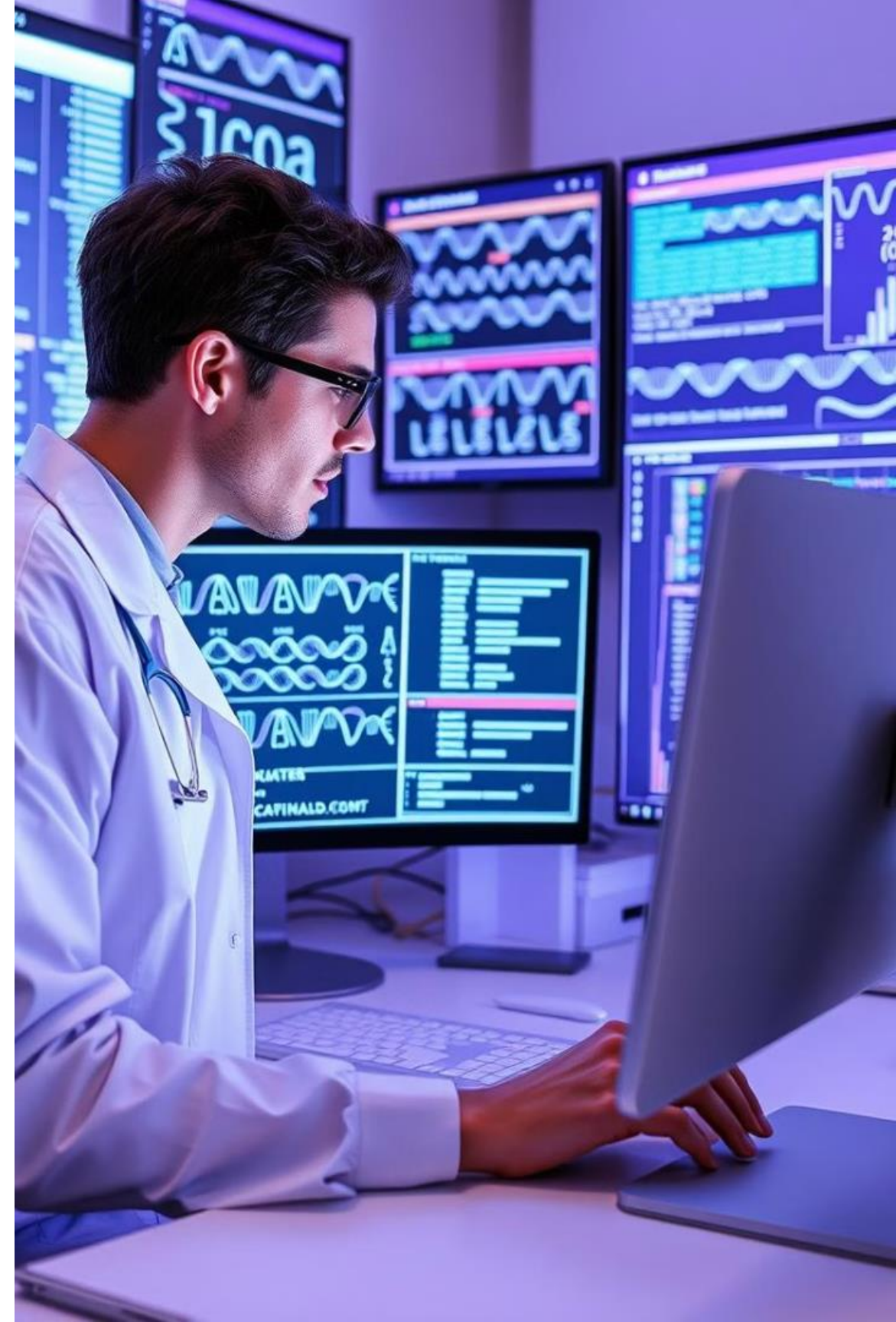
Bioinformaticians use statistical algorithms to align DNA and protein sequences, identifying similarities and differences between them. This helps researchers understand the relationships between different organisms.

3

## Evolutionary Models

Statistical models are used to study the evolutionary relationships between different species. These models help scientists understand how life on Earth has evolved over millions of years.

Reference: Durbin, R. et al. (1998). *Biological Sequence Analysis*. Cambridge University Press.

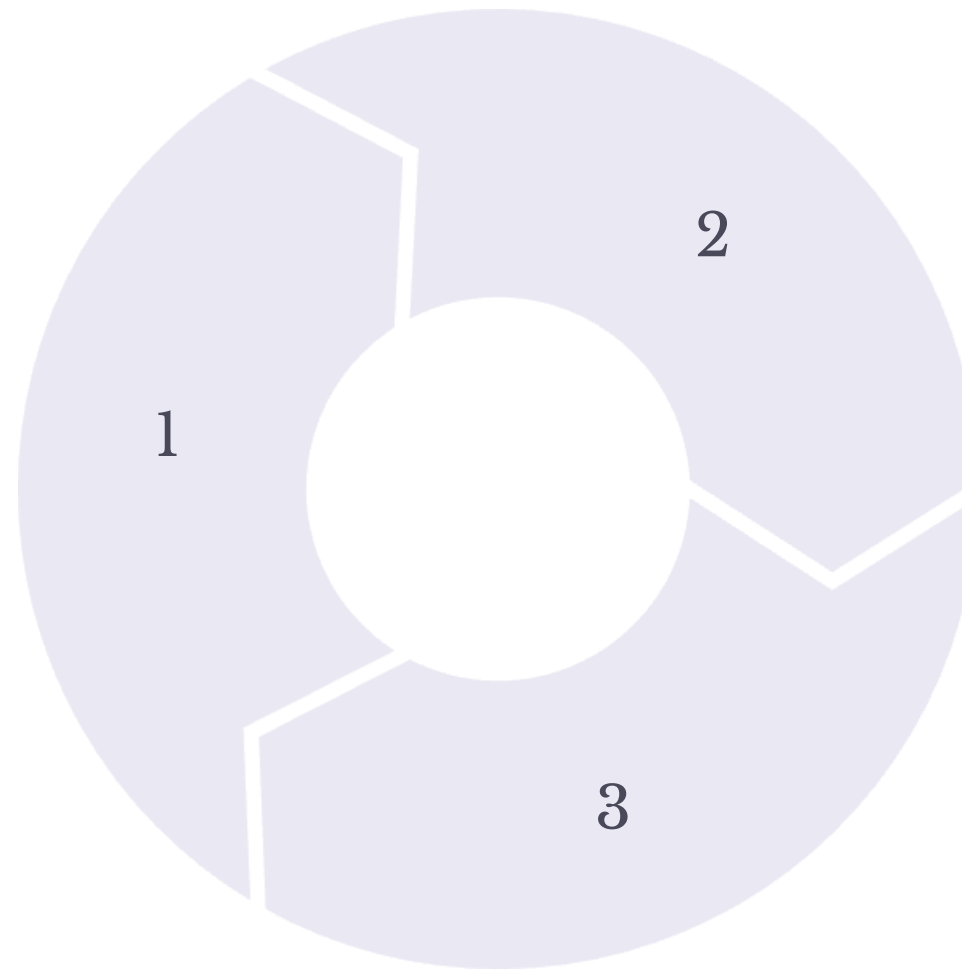




# Telecommunications & Network Reliability

## Error-Correcting Codes

Statistical techniques are used to design error-correcting codes that ensure reliable data transmission over noisy communication channels.



## Bandwidth Allocation

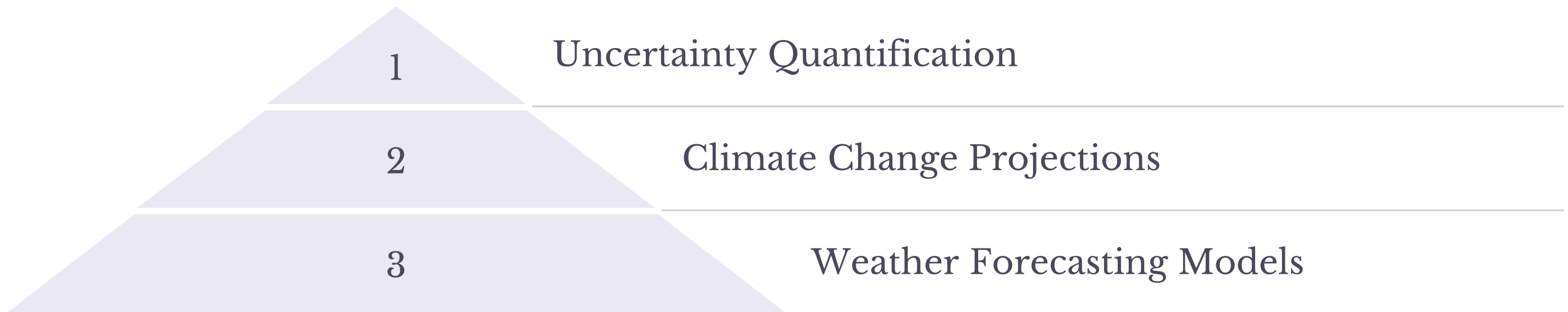
Network engineers use statistical models to allocate bandwidth efficiently, ensuring that users have access to the resources they need.

## Queuing Theory

Queuing theory is used to analyze waiting lines and optimize the performance of telecommunication networks.

Reference: Bertsekas, D., & Gallager, R. (1992). *Data Networks* (2nd ed.). Prentice Hall.

# Meteorology & Climate Science



Meteorologists and climate scientists use statistical models to forecast weather patterns and project future climate change scenarios. These models help inform policy decisions and prepare for the impacts of climate change.

Reference: Wilks, D. S. (2011). *Statistical Methods in the Atmospheric Sciences* (3rd ed.). Academic Press.



# Sports Analytics

1 Betting Odds

---

2 Win/Loss Probability

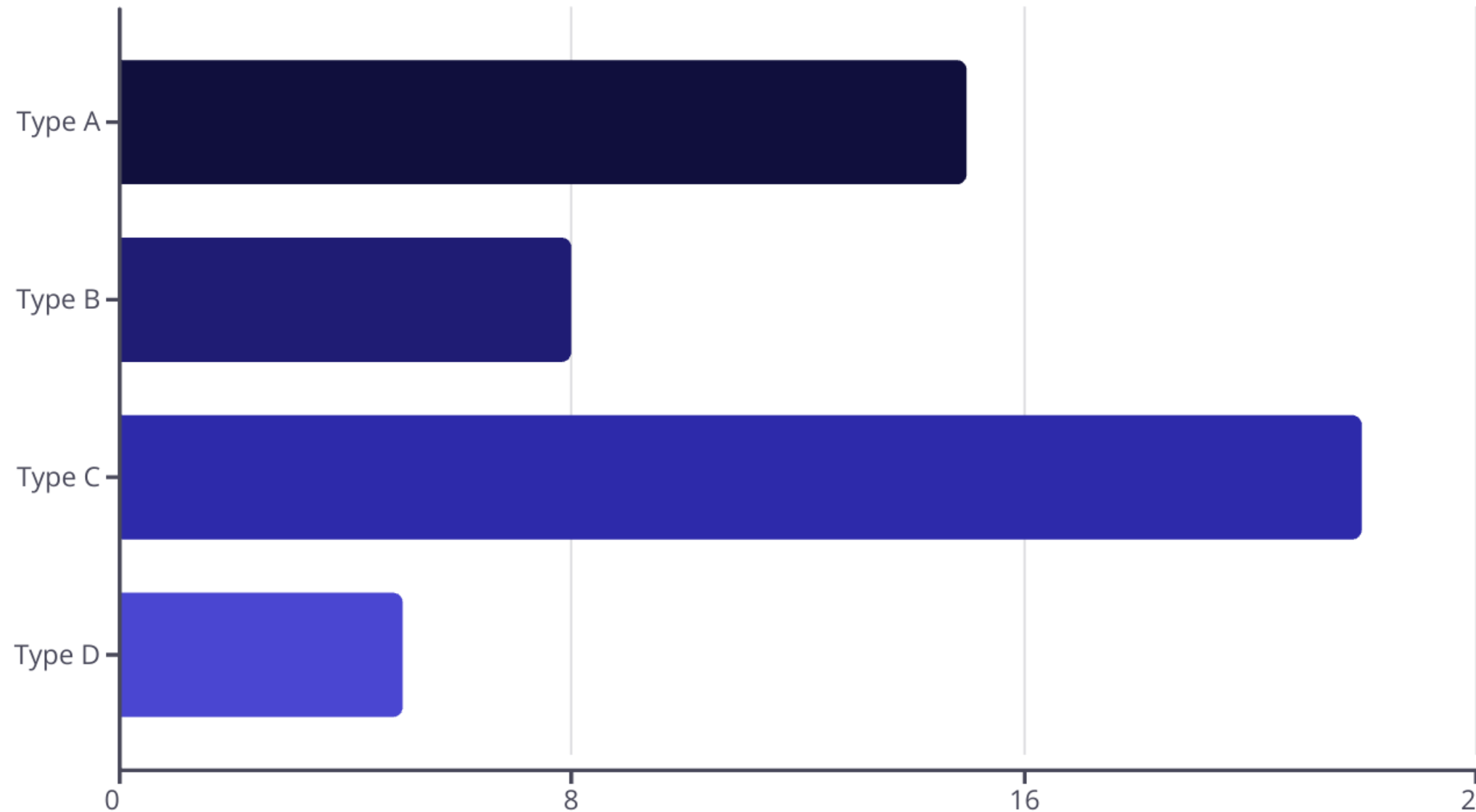
---

3 Player Performance Metrics

Sports analysts use statistical models to evaluate player performance, predict game outcomes, and set betting odds. These models help teams make strategic decisions and improve their chances of winning.

Reference: Winston, W. L. (2009). *Mathletics: How Gamblers, Managers, and Sports Enthusiasts Use Mathematics in Baseball, Basketball, and Football*. Princeton University Press.

# Quality Control & Manufacturing



Manufacturers use statistical process control (SPC) to monitor and improve the quality of their products. This involves collecting data on key process variables and using statistical techniques to identify and correct deviations from desired performance levels.

Reference: Montgomery, D. C. (2020). *Introduction to Statistical Quality Control* (8th ed.). Wiley.

# Probability Vs Statistics

# Probability Theory as a Foundation for Statistics

**Statistics** is a discipline that involves the collection, analysis, interpretation, and presentation of data.

In relation to probability, statistics uses probability theory as its theoretical foundation to model uncertainty and variability in data.

Probability provides a formal framework to quantify uncertainty by assigning numerical values (probabilities) to outcomes.

It describes how likely events are to occur under certain assumptions or models.

These probabilistic models help in understanding random processes and phenomena.

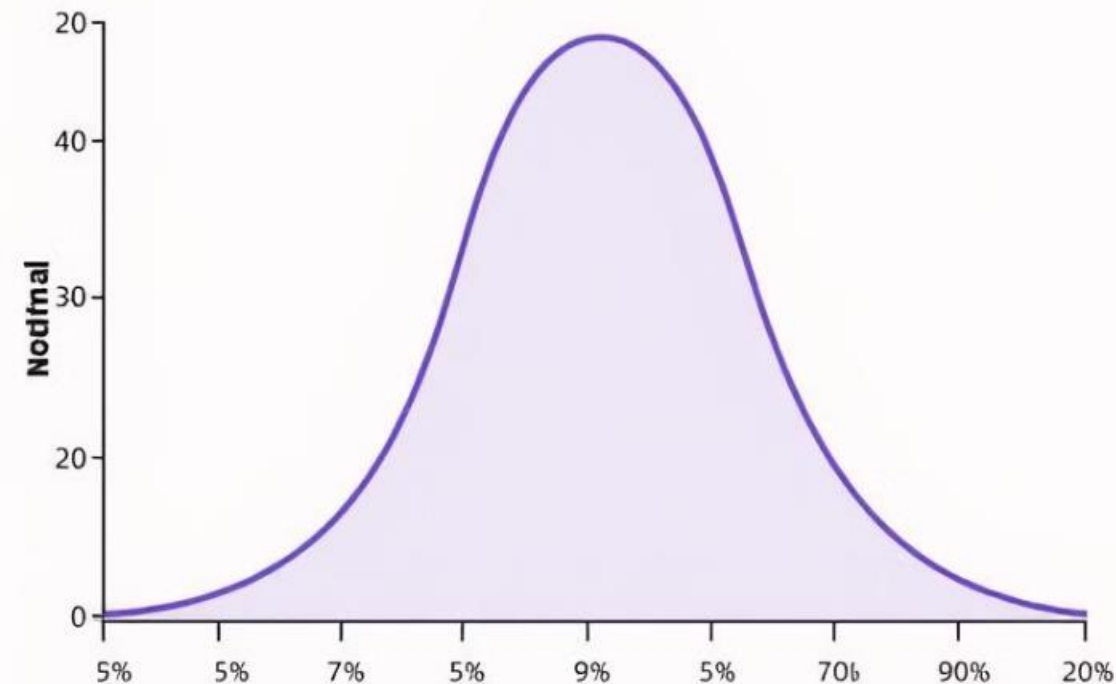
In statistics, we often work with samples drawn from larger populations. Because a sample is subject to random variation, probability theory is essential for making inferences about the population.

This is a symbiotic relationship between probability and statistics that enables researchers and practitioners to rigorously analyze uncertain events and make data-driven decisions.



# Role in Statistical Inference

In statistics, we often work with samples drawn from larger populations. Because a sample is subject to random variation, probability theory is essential for making inferences about the population. Techniques such as hypothesis testing, confidence intervals, and regression analysis all rely on probability distributions (e.g., normal, binomial, Poisson) to assess the likelihood of observed data given certain assumptions.



## 1 Hypothesis Testing

Probability distributions help determine the likelihood of observed data under certain assumptions.

## 2 Confidence Intervals

Probability theory allows calculation of range of values in which the population parameter likely lies.

## 3 Regression Analysis

Probability helps in assessing the significance and reliability of relationships between variables.

# Descriptive vs. Inferential Statistics

Descriptive statistics summarize and describe the features of a dataset using measures like mean, median, and standard deviation. Inferential statistics use probability theory to draw conclusions about a population based on a sample, addressing questions such as "How likely is it that an observed effect is real and not just due to chance?"

## Descriptive Statistics

Summarize and describe the features of a dataset.

- Mean
- Median
- Standard Deviation

## Inferential Statistics

Use probability theory to draw conclusions about a population.

- Likelihood of observed effect
- Drawing conclusion about sample population

A golden scale of justice and a wooden gavel resting on a white block. The scale is in the background, and the gavel is in the foreground, resting on a white rectangular block. The background is dark and out of focus.

# Decision-Making Under Uncertainty

By combining probability and statistics, we can not only describe data but also make decisions in the presence of uncertainty. For example, in clinical trials or quality control, statistical methods help determine whether an observed effect or variation is statistically significant, thereby guiding decisions based on risk assessment.

## Clinical Trials


Statistical methods help determine the significance of observed effects in drug testing.

## Quality Control

Statistical methods help determine variation and ensure product standards.

## Risk Assessment

Guiding decisions by evaluating and quantifying potential risks.



probability

statistics

# Interdependence of Probability and Statistics

While probability deals with the theoretical underpinnings of randomness, statistics applies these concepts to real-world data. Essentially, probability provides the models, and statistics provides the tools to apply these models to draw meaningful conclusions.

Deals with the theoretical underpinnings of randomness.

Probability

1

2

Statistics

Applies probability concepts to real-world data.





# Symbiotic Relationship

This symbiotic relationship between probability and statistics enables researchers and practitioners to rigorously analyze uncertain events and make data-driven decisions. By combining probability and statistics, we can not only describe data but also make decisions in the presence of uncertainty.



Rigorous Analysis: By combining probability and statistics, uncertain events can be analyzed.



Data-Driven Decisions: Using statistical tools and probability models to aid decision making.



# Key Takeaways

Probability theory serves as a critical foundation for statistical analysis by providing the mathematical tools necessary to quantify uncertainty and variability in data. Statistical inference relies on probability distributions to make informed decisions about populations based on sample data. The symbiotic relationship between probability and statistics enables researchers and practitioners to rigorously analyze uncertain events and make data-driven decisions.

1

## Foundation

Probability theory is a critical foundation for statistical analysis.

2

## Inference

Statistical inference relies on probability distributions to make decisions.

3

## Symbiosis

Probability and statistics enable rigorous data analysis and decision-making.



# Course Alignment

## ***COURSE LEARNING OUTCOMES:***

- 1. Learn the language and core concepts of probability theory.*
- 2. Understand basic principles of statistical inference.*
- 3. Become an informed consumer of statistical information and have a good knowledge of what expectation and variance mean and be able to compute them.*

**PREREQUISITE:** Fundamental Mathematics

## **UNIT 1: (15 LECTURES)**

Introduction to Probability: Random experiment, sample space, trial, event. Simple probability, Compound Probability, mutually exclusive events, Addition theorem, independent events, Multiplication theorem, Dependent events, Conditional probability, Bayes' Theorem, Partitions and Total probability law. Exploring Univariate Data: Types of data, Mean, Mode and Median.

## **UNIT 2: (15 LECTURES)**

Standard Deviation and Variance, Range and Finding Outliers. Counting, Random variables, probability mass function, probability density function. distributions, quantiles, mean-variance, Joint distributions, covariance, correlation, independence, and Central limit theorem.

Discrete Distributions, Random Variables, Binomial Distributions, Geometric Distributions Continuous Distributions, Density Curves, The Normal Distribution

## **UNIT 3: (15 LECTURES)**

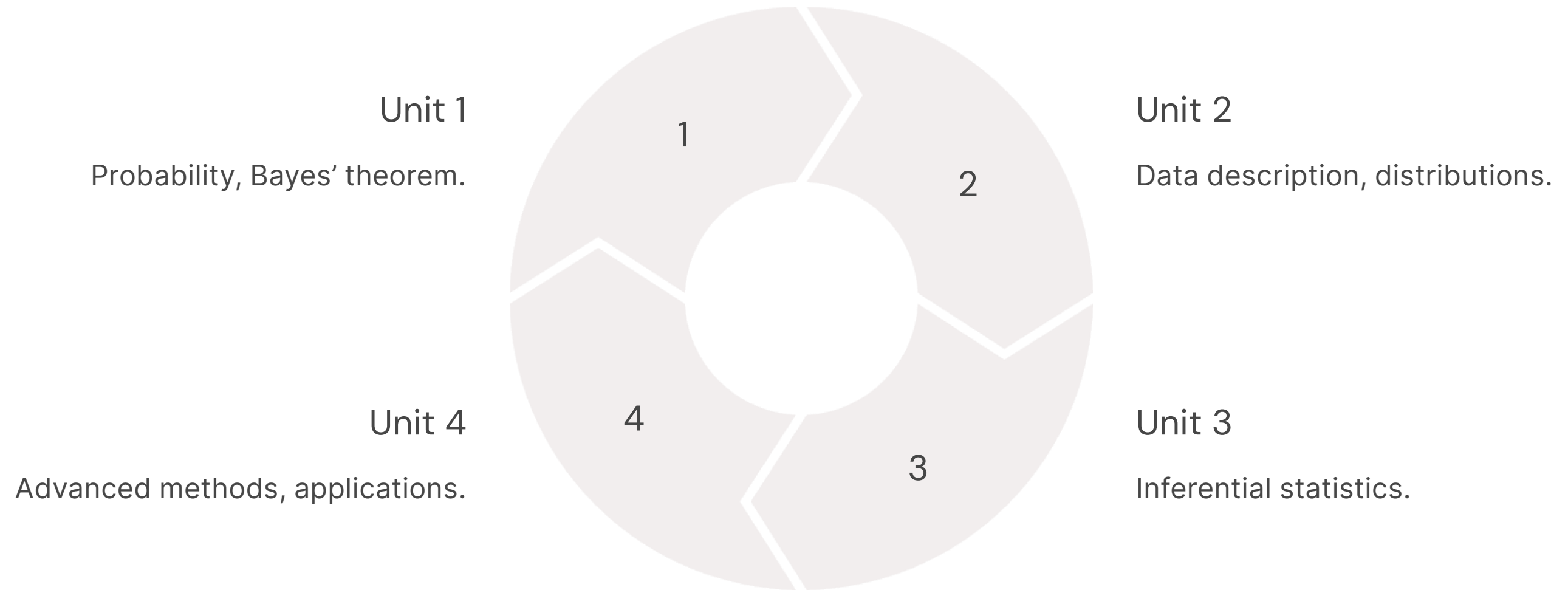
Multivariate Data, Scatter Plots, Correlation, The Least Squares Regression Line, Residuals, Non-Linear Models, Relations in Categorical Data, Margins of Error and Estimates, Confidence Interval for a Proportion, Confidence Interval for the Difference of Two Proportions, Confidence Interval for a Mean, Confidence Interval for the Difference of Two Means.

## **UNIT 4: (15 LECTURES)**

Tests of Significance, Inference for the Mean of a Population, Sample Proportions, Inference for a Population Proportion, Comparing Two Means, Comparing Two Proportions, Goodness of Fit Test, and Two-way Tables. Simple correlation (Pearson's correlation coefficient), Simple linear regression, Prediction, error in prediction, principle of least square.



# Course Alignment





# Next Steps & Reading

1

Definitions

Sample space, events.

---

2

Probability Rules

Simple vs. compound events.

---

3

Bayes' Theorem

Conditional probability.