LES FACULTÉS
DE L'UNIVERSITÉ
CATHOLIQUE DE LILLE

Models examples

PROBABILITY THEORY

Baptiste Mokas

baptiste.mokas@gmail.com

weeki.io/moocs

linktr.ee/baptistemokas

+33 7 69 08 54 19

KNOWLEDGE TREE

Part 1: Foundational Concepts

Part 2: Key elements

Part 3: Advanced Concepts and Simulations

OUR STORY LINE

Probability and Statistics

- STEP -1 _ PROGRAM INTRODUCTION
- STEP 0 _ FOUNDATIONS
- STEP 1 _ THEORY OF SYSTEMS
- STEP 2 _ STOCHASTIC DYNAMICS & PROBABILITY
- STEP 3 _ DATA OBSERVATION
- STEP 4 _ INFERENCE & ESTIMATION THEORY
- STEP 5 _ LINEAR MODEL EXAMPLES
- STEP 6 _ OTHER MODEL EXAMPLES
- STEP 7 _ NON LINEAR MODELS

OUR STORY LINE

Probability and Statistics

STEP -1 _ PROGRAM INTRODUCTION
- -1 - PROGRAM INTRODUCTION

STEP 0 _ FOUNDATIONS
- 0.1 - ELEMENTS OF CALCULUS & TOOLS
- 0.2 - EPISTEMOLOGY & THEORY OF KNOWLEDGE

STEP 1 _ THEORY OF SYSTEMS
- 1.1 - DYNAMICAL SYSTEMS
- 1.2 - COMPLEX ADAPTIVE SYSTEMS

STEP 2 _ STOCHASTIC DYNAMICS & PROBABILITY
- 2.1 - MEASURE THEORY
- 2.2 - PROBABILITY THEORY
- 2.3 - USUAL PROBABILITY DISTRIBUTIONS
- 2.4 - ASYMPTOTIC STATISTICS
- 2.5 - STOCHASTIC PROCESS & TIME SERIES
- 2.6 - INFORMATION GEOMETRY

STEP 3 _ DATA OBSERVATION
- 3.1 - DESCRIPTIVE STATISTICS & DATAVIZUALISATION
- 3.2 - EXPLORATORY DATA ANALYSIS

STEP 4 _ INFERENCE & ESTIMATION THEORY
- 4.1 - PARAMETERS ESTIMATIONS & LEARNING
- 4.2 - EXPERIMENTAL DESIGN & HYPOTHESIS TESTING
- 4.4 - DECISION TREES & MODEL SELECTION
- 4.5 - BAYESIAN INFERENCES

STEP 5 _ LINEAR MODEL EXAMPLES
- 5.1 - SIMPLE LINEAR REGRESSION
- 5.2 - MULTIPLE LINEAR REGRESSION
- 5.3 - OTHER REGRESSIONS MODELS

STEP 6 _ OTHER MODEL EXAMPLES
- 6.1 - USUAL UNIVARIATE TESTING
- 6.2 - USUAL MULTIVARIATE TESTING
- 6.3 - NON PARAMETRIC STATISTICS

STEP 7 _ NON LINEAR MODELS
- 7.1 - PROBABILISTIC GRAPHICAL MODELS
- 7.2 - PERCOLATION THEORY
- 7.3 - SPATIAL STATISTICS
- 7.4 - EXTREM VALUE THEORY
- 7.5 - AGENT BASED MODELING
- 7.6 - NETWORK DYNAMICS

Part 1: Foundational Concepts

1.1 Introduction to Probability

Description:

Probability is a mathematical measure of the likelihood of an event occurring, within a set of possible events. **It quantifies how likely an event is to happen and ranges between 0 and 1,** with 0 meaning the event will not happen and 1 meaning it is certain to happen.

Reference:

Laplace, P. S. (1812). Théorie analytique des probabilités. Courcier.

Link to Biology:

Probability plays a vital role in the study of **genetics and heredity. For instance, Mendelian genetics uses probability to predict the outcomes of genetic crosses.** When Gregor Mendel conducted his pea plant experiments, he used the laws of probability **to predict the outcomes of the crosses.** The probability of inheriting a particular gene or trait is foundational to understanding genetic inheritance patterns.

The History of Probability:

Probability theory, as a formal mathematical discipline, has its roots in the study of **gambling and games of chance. Some key milestones include:**

Ancient Civilizations: Historical records indicate that ancient cultures, like the Chinese and Egyptians, were familiar with **rudimentary games of chance**. Dice games, for example, have been around for millennia.

Renaissance Europe: The modern development of probability theory began in earnest in Europe during the Renaissance. **Notably, an exchange of letters between the French mathematicians Blaise Pascal and Pierre de Fermat in the 1650s laid the groundwork for the discipline.** They discussed the "problem of points", which is about how to split stakes in a game of chance that is interrupted before its end.

18th & 19th Centuries: Key figures like **Jacob Bernoulli, Abraham de Moivre, and Pierre-Simon Laplace** expanded the theory, moving it beyond games of chance to encompass other phenomena.

Interpretations of Probability

Part 1: Foundational Concepts

1.1 Introduction to Probability

Classical Interpretation

If all outcomes are equally likely, **the probability is the ratio of the number of favorable outcomes to the number of possible outcomes.**

Frequentist Interpretation

Probability is the limit of the relative frequency of an event after many trials. Subjective (or Bayesian) Interpretation: Probability measures the degree of belief or confidence in an event occurring based on available evidence.

Link to Biology:

Probability plays a vital role in the study of **genetics and heredity. For instance, Mendelian genetics uses probability to predict the outcomes of genetic crosses.** When Gregor Mendel conducted his pea plant experiments, he used the laws of probability **to predict the outcomes of the crosses.** The probability of inheriting a particular gene or trait is foundational to understanding genetic inheritance patterns.

Interpretations of Probability:

Subjective Interpretation: Here, probability is viewed as **a measure of one's personal belief or confidence in the occurrence of an event.** It's subjective because it can vary from person to person. For instance, two analysts might assign different probabilities to the likelihood of a particular team winning a game based on their individual assessments.

Classical Interpretation: Rooted in the original problems of games of chance, **the classical interpretation states that if all outcomes are equally likely, then the probability of an event is simply the ratio of favorable outcomes to the total number of outcomes**. For example, when rolling a fair six-sided die, the classical probability of getting a 3 is 1/6.

Experiments and Events

Part 1: Foundational Concepts

1.1 Introduction to Probability

Experiments and Events:

In probability theory:

An experiment refers to any procedure or situation that results in **a certain outcome from a list of possible outcomes**. An experiment doesn't necessarily have to be a laboratory procedure. Tossing a coin, rolling a dice, or even just observing the weather on a random day can be considered experiments in this context.

An event is a set of one or more outcomes from an experiment. For instance, when rolling a dice, getting an even number is an event that includes the outcomes 2, 4, and 6.

Equation:
The probability P of an event E occurring is given by:

$$P(E) = \frac{\text{number of favorable outcomes}}{\text{total number of possible outcomes}}$$

For example, when flipping a fair coin, the probability of getting heads $P(H)$ is:

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

Demo:
Consider a standard deck of 52 playing cards. If you draw one card, what is the probability that it is an ace?
Favorable outcomes: There are 4 aces in a deck.
Total possible outcomes: There are 52 cards in a deck.
Using the equation, the probability $P(A)$ of drawing an ace is:

$$P(A) = \frac{4}{52} = \frac{1}{13}$$

Therefore, there's a 1 in 13 chance, or approximately 7.69%, of drawing an ace from a standard deck of cards.

Set Theory

Part 1: Foundational Concepts

1.2 Understanding Sets and Events

Description:
Set theory deals with collections of objects, termed as "sets". It provides a foundation for nearly all of mathematics and has applications in many areas, including logic, computer science, and probability.

History:
The modern study of set theory was initiated by Georg Cantor in the late 19th century.

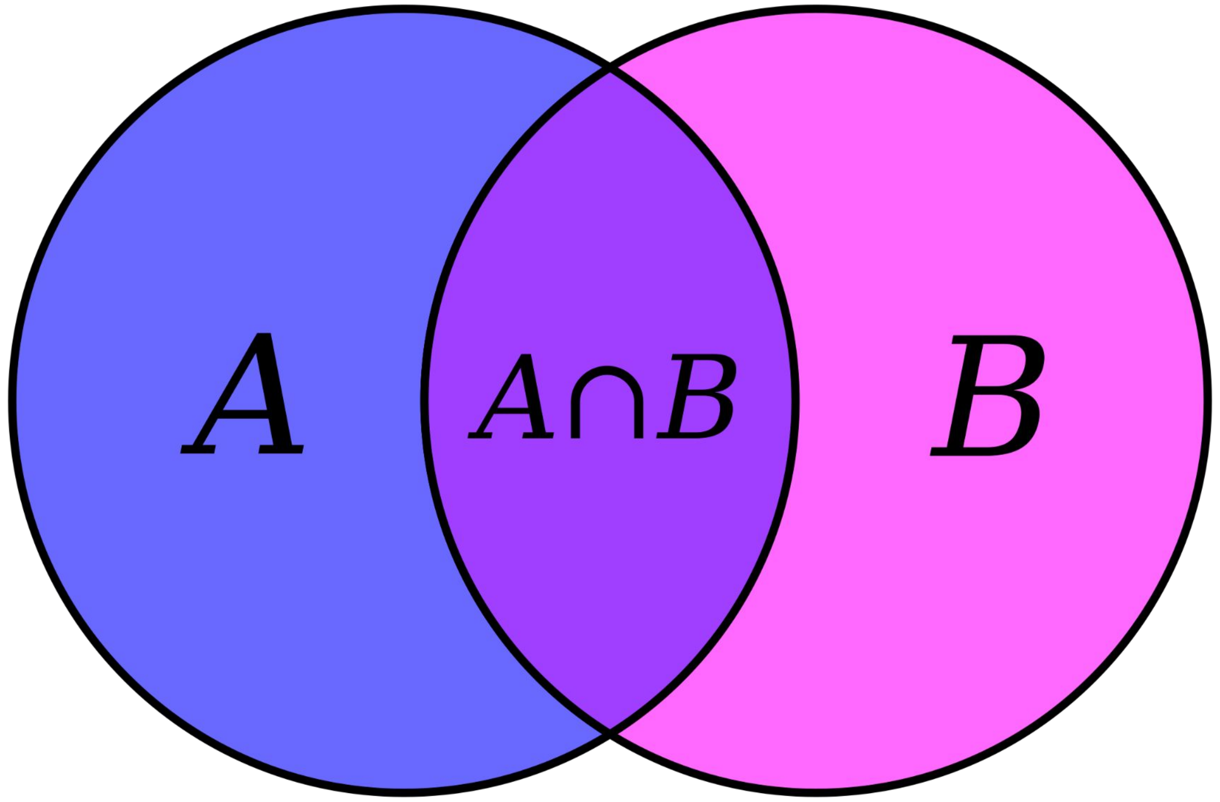
Reference:
Cantor, G. (1874). On a Property of the Collection of All Real Algebraic Numbers. Journal für die reine und angewandte Mathematik.

Link to Biology:
Set theory can be applied in **taxonomy, the classification of organisms.** Species can be grouped into sets based on shared characteristics.

Equation:
There isn't a singular equation for set theory, but a basic notation is:

$$A = \{x, y, z\}$$

where A is a set containing elements x, y, and z.



Part 1: Foundational Concepts

1.2 Understanding Sets and Events

Construction:
Sets can be constructed in multiple ways. The most basic is by listing its members explicitly. However, for larger sets, a rule or property might be defined that members should adhere to.

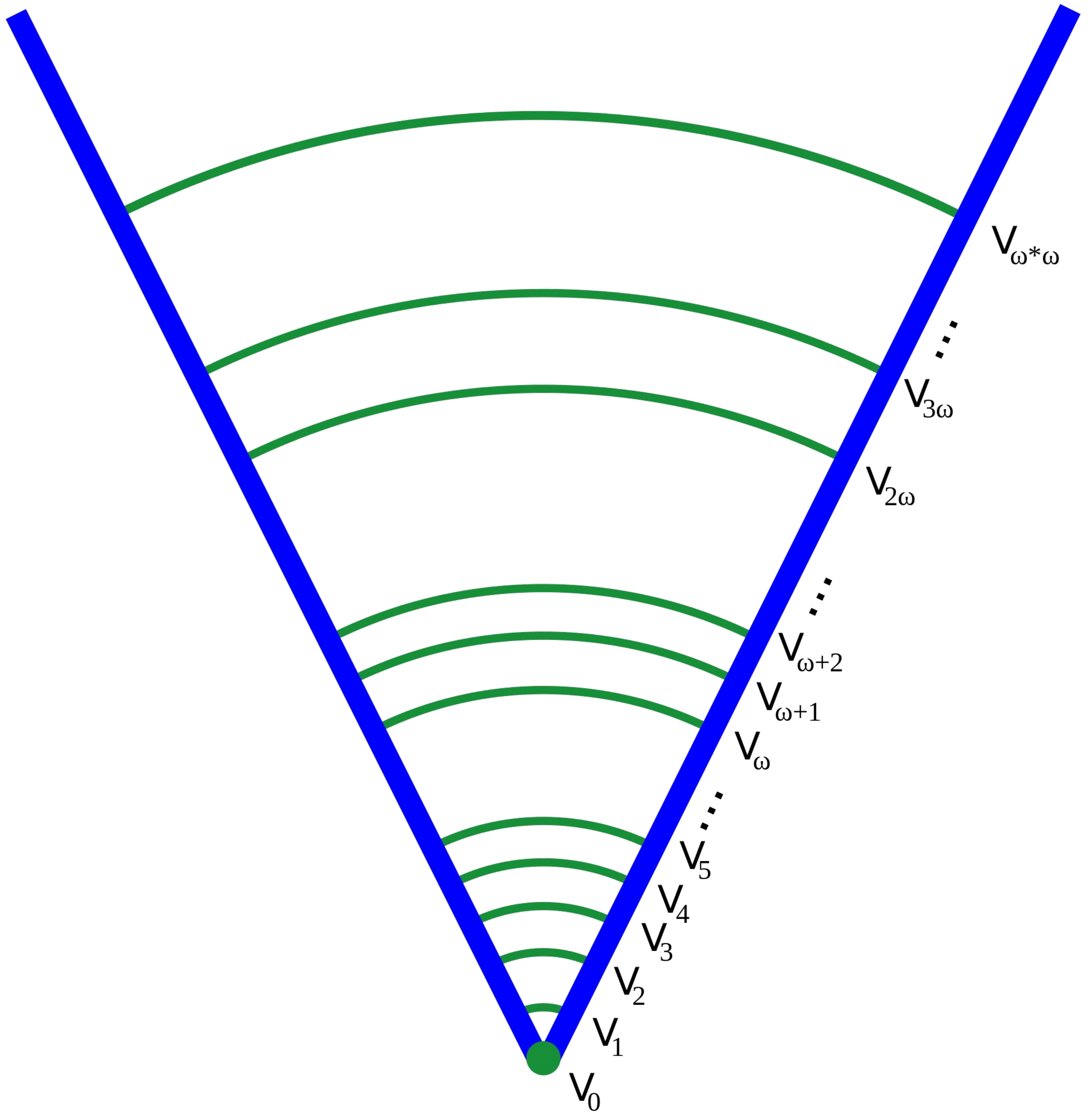
Demo:

1. Explicit Construction:
Consider the set of natural numbers less than 6.

$$N = \{1, 2, 3, 4, 5\}$$

2. Property-based Construction:
We can also define a set by describing a property that its members must have. For instance, consider the set of even natural numbers less than 10. $E = \{x \mid x \text{ is an even natural number and } x < 10\}$ Therefore,

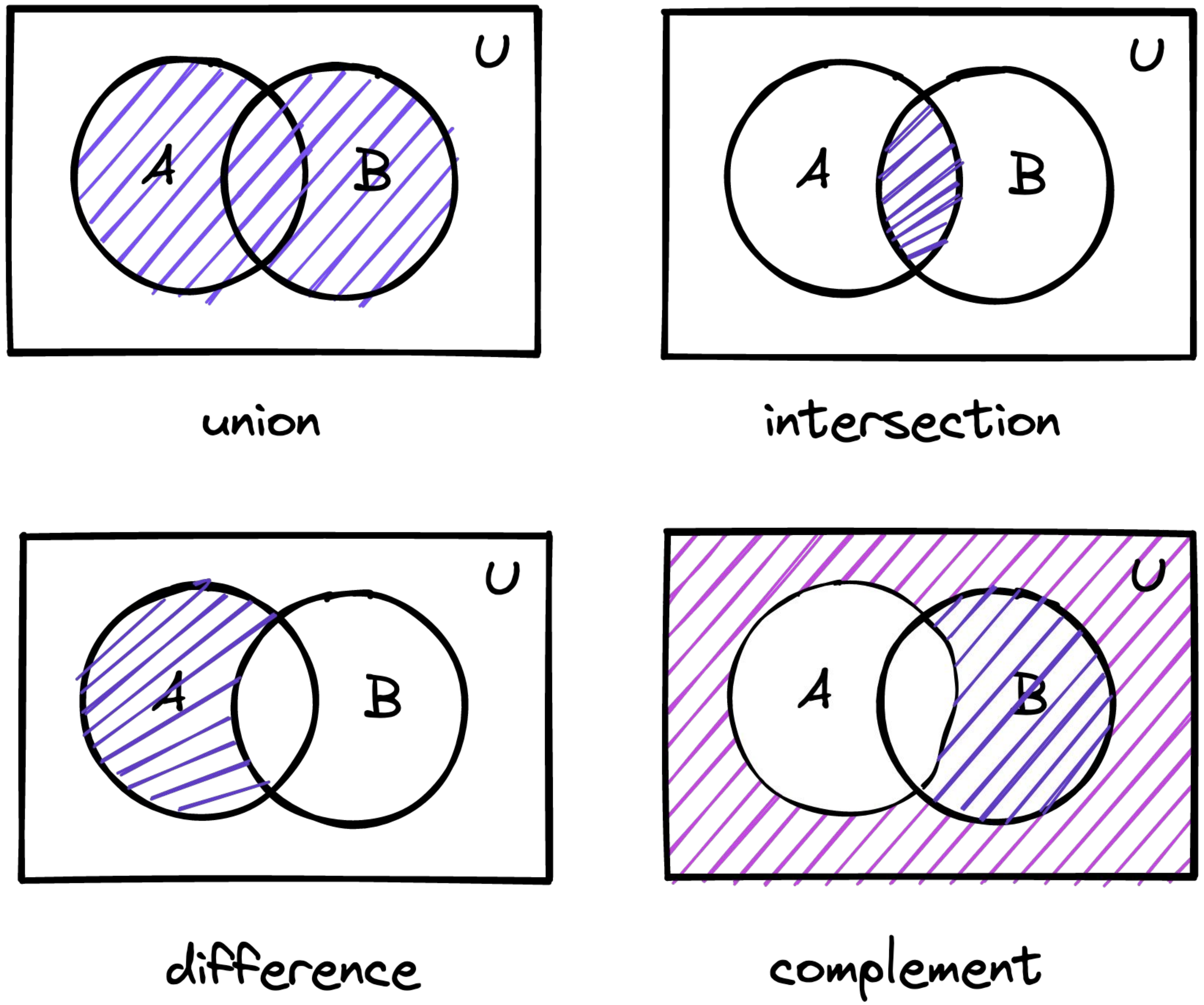
$$E = \{2, 4, 6, 8\}$$



Intersection and Union

Part 1: Foundational Concepts

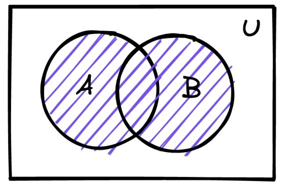
1.2 Understanding Sets and Events



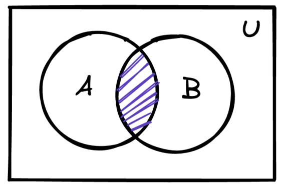
Algebra on Sets

Part 1: Foundational Concepts

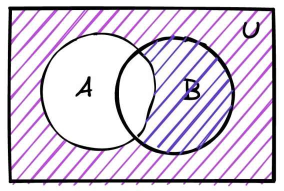
1.2 Understanding Sets and Events



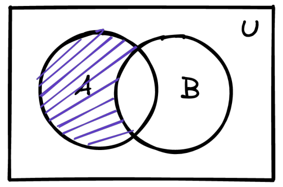
union



intersection



complement



difference

Union (U):
- The union of two sets, A and B, denoted as $A \cup B$, is the set that contains all elements that are in either set A or set B, or in both.
- Mathematically, for any element $x : x \in (A \cup B)$ if and only if $(x \in A) OR (x \in B)$.

Intersection (Ω) :
- The intersection of two sets, A and B, denoted as $A \cap B$, is the set that contains all elements that are common to both sets A and B.
- Mathematically, for any element $x : x \in (A \cap B)$ if and only if $(x \in A)$ AND $(x \in B)$.

Complement ($\neg$ or '):
- The complement of a set A, denoted as $\neg A$ or A', is the set that contains all elements not in set A but are in the universal set.
- Mathematically, for any element $x : x \in \neg A$ if and only if (x is not in A) and (x is in the universal set).

Set Difference (- or):
- The set difference of set A and set B, denoted as A - B or A $\setminus B$, is the set that contains all elements that are in A but not in B.
- Mathematically, for any element $x : x \in (A - B)$ if and only if $(x \in A)$ and (x is not in B).

Cartesian Product (x) :
- The Cartesian product of two sets A and B, denoted as $A \times B$, is a set of ordered pairs where the first element of each pair is from set A and the second element is from set B.
- Mathematically, for any ordered pair $(a, b) : (a, b) \in (A \times B)$ if and only if $(a \in A)$ and $(b \in B)$. Power Set (P):
- The power set of a set A, denoted as $P(A)$, is the set of all possible subsets of A, including the empty set and A itself.

Definition

Part 1: Foundational Concepts

1.3 Kolmogorov Formalism

Sample Space (Ω): The sample space is the set of all possible outcomes of a random experiment. This set is often denoted as Ω. Each element of Ω is called an "elementary event" or "outcome."

Events (Random Events): Events are subsets of the sample space Ω. They represent sets of possible outcomes of the random experiment. Events can be simple (a single outcome) or compound (multiple outcomes).

Probability Function (P): The probability function, denoted as P, assigns a real number between 0 and 1 to each event. This function specifies the probability that an event will occur. It must satisfy certain properties:
a. Non-negativity: $P(E) \geq 0$ for all events E.
b. Normalization: $\mathrm{P}(\Omega) = 1$, meaning the sum of probabilities over all possible events equals 1.
c. Additivity: For mutually exclusive events E1, E2, ..., $P(E1 \cup E2 \cup \ldots) = P(E1) + P(E2) + \ldots$

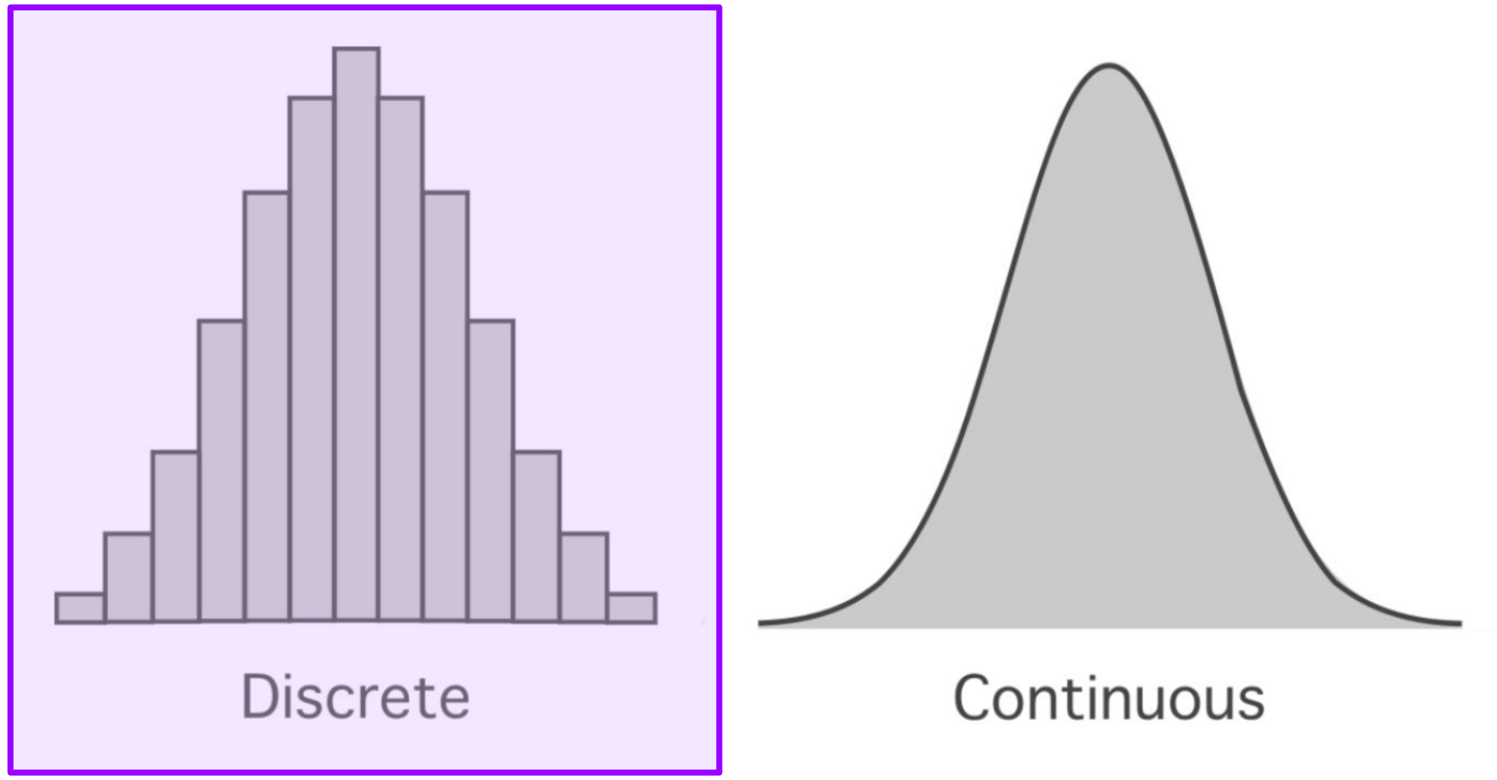
Independence: Two events, E1 and E2, are considered independent if the probability of their joint occurrence is equal to the product of their individual probabilities: $P(E1 \cap E2) = P(E1) * \mathrm{P(E2)}$.
Conditional Probability: This refers to the probability of one event occurring given that another event has occurred. It is denoted as $P(A \mid B)$ and is defined as $P(A \cap B)/P(B)$, where A and B are events.

Random Variables: Random variables are used to associate real numbers with events. They allow us to analyze and model random processes. The probability distribution of a random variable describes how its values are spread over different outcomes.

The Kolmogorov formalism provides a rigorous foundation for the study of probability, allowing for the mathematical modeling of uncertain events and random processes.

Discrete Case: Finite or Countable

Discrete Case: Finite or Countable Sample Space (Ω) :
- In the discrete case, the sample space Ω represents a finite or countable set of possible outcomes.
- Each elementary event in Ω corresponds to a distinct, individual outcome of the random experiment.
- For example, if you're flipping a fair coin, the sample space Ω would consist of two outcomes: $\Omega = \{$ Heads, Tails $\}$.
- Probability Function (P): In the discrete case, the probability function P assigns probabilities to individual outcomes and events. For a finite sample space, you can assign a probability to each elementary event directly. For a countable sample space, you can assign probabilities using a probability mass function (PMF).

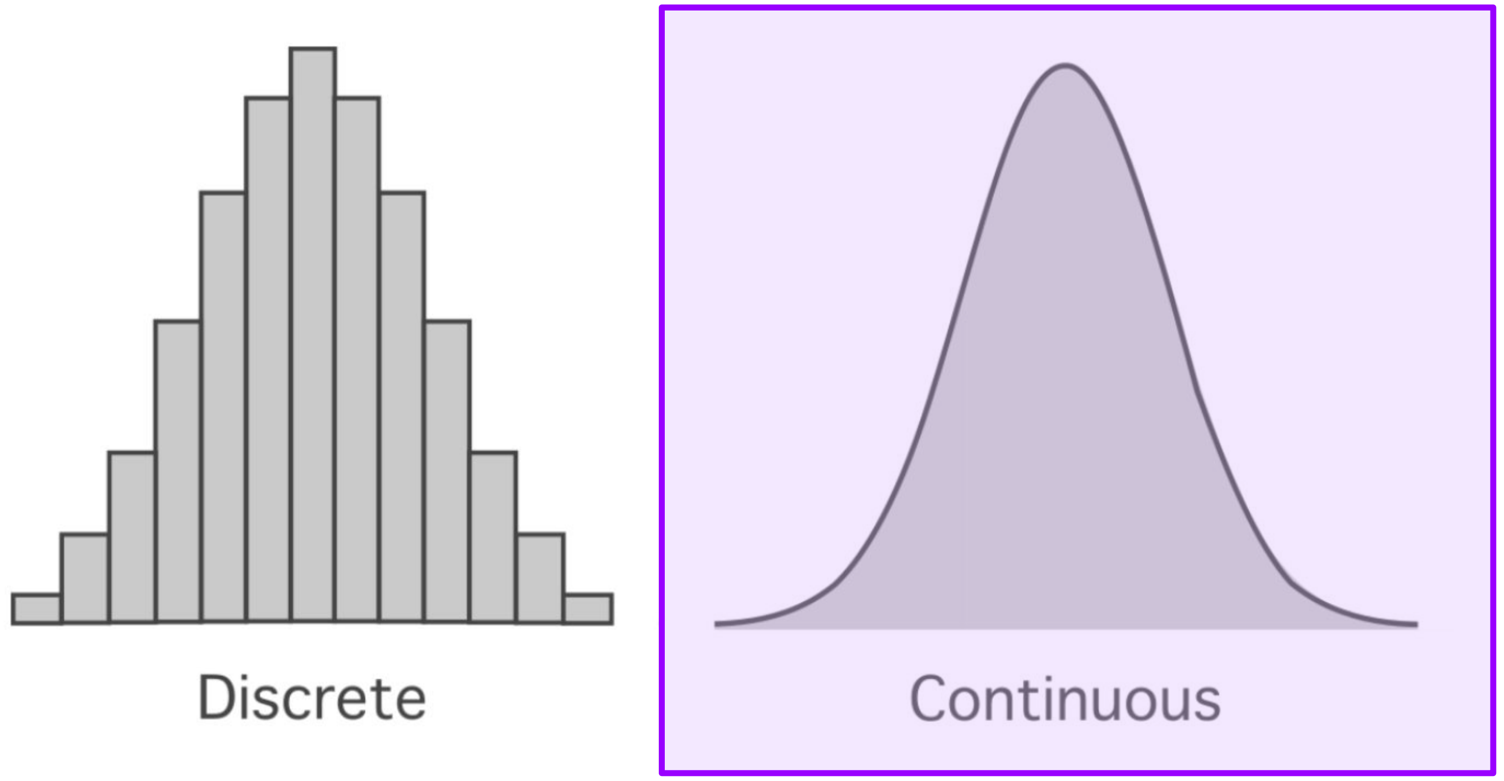


Continuous Case: S = R or R*

Part 1: Foundational Concepts

1.3 Kolmogorov Formalism

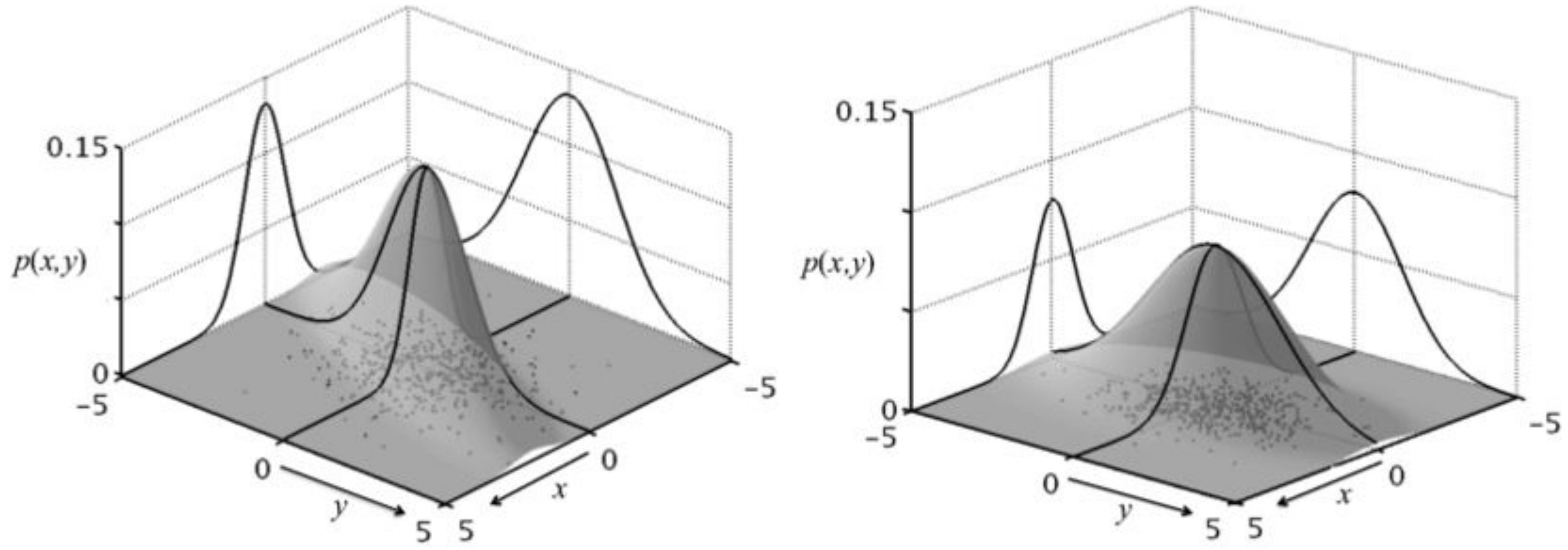
Continuous Case: Sample Space (Ω) as Real Numbers $(S = R)$ or Positive Real Numbers $(S =$ R):*
- In the continuous case, the sample space Ω corresponds to real numbers or positive real numbers, depending on the problem context.
- The sample space in the continuous case represents an uncountable set of possible outcomes. For instance, when measuring a physical quantity, such as the height of individuals, the sample space could be the set of real numbers.
- Probability Density Function (PDF): In the continuous case, probability is described using a probability density function (PDF). The PDF, denoted as $f(x)$, specifies the likelihood of a random variable taking on a particular value. The probability of an event is calculated as the integral of the PDF over the event's range.



Part 1: Foundational Concepts

1.3 Kolmogorov Formalism

Product Spaces Case: Combining Multiple Experiments (Product Space Ω):
- In the product spaces case, you consider the combination of multiple random experiments. The sample space Ω for the combined experiment is a Cartesian product of the individual sample spaces from each experiment.
- For example, if you're rolling two fair dice, the sample space Ω for this combined experiment is a Cartesian product of the sample spaces of each die: $\Omega = \{1, 2, 3, 4, 5, 6\} \times \{1, 2, 3, 4, 5, 6\}$, representing all possible outcomes of both dice rolls.
- To compute probabilities in product spaces, you use joint probability distributions, conditional probabilities, and independence concepts to understand the combined behavior of multiple random variables.



Definition of Conditional Probability

Part 2: Key elements

2.1 Conditional Probability Basics

Conditional Probability (P(A | B)):

- Conditional probability, denoted as $\mathrm{P}(\mathrm{A} \mid \mathrm{B})$, represents the probability of event A occurring given that event B has already occurred.
- It quantifies the likelihood of event A under the condition or assumption that event B is true.
- Mathematically, conditional probability is calculated as: $P(A \mid B) = P(A \cap B)/P(B)$, where $P(A \cap B)$ is the probability of both events A and B occurring, and $P(B)$ is the probability of event B occurring.

Intersection of Events $(A \cap B)$:

- The intersection of events A and B, denoted as $A \cap B$, represents the set of outcomes that are common to both events.
- It is the event where both A and B occur simultaneously.
- The conditional probability $P(A \mid B)$ is calculated using the probability of the intersection of events A and B, $P(A \cap B)$, as part of the formula mentioned above.

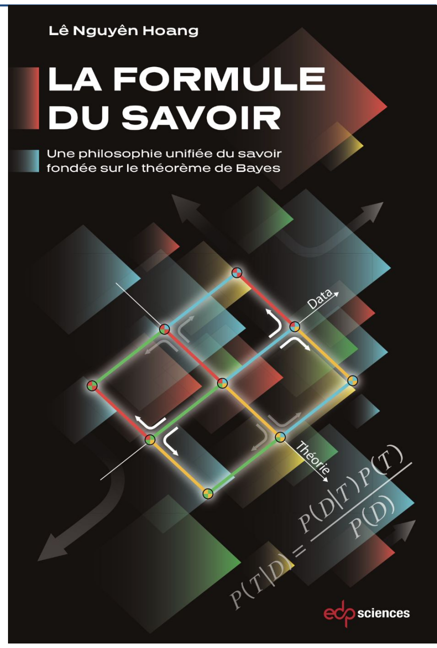
Independence of Events:

- Two events, A and B, are considered independent if the occurrence (or non-occurrence) of one event does not affect the probability of the other event.
- Mathematically, events A and B are independent if $P(A \mid B) = P(A)$ and $P(B \mid A) = P(B)$.
- In the case of independent events, the conditional probabilities are the same as the unconditional probabilities.

Multiplication Rule for Independent Events:

- When two events A and B are independent, the probability of both events occurring is given by the multiplication rule: $P(A \cap B) = P(A) * P(B)$.
- This rule simplifies the calculation of conditional probabilities for independent events.

Bayes' Theorem



Part 2: Key elements

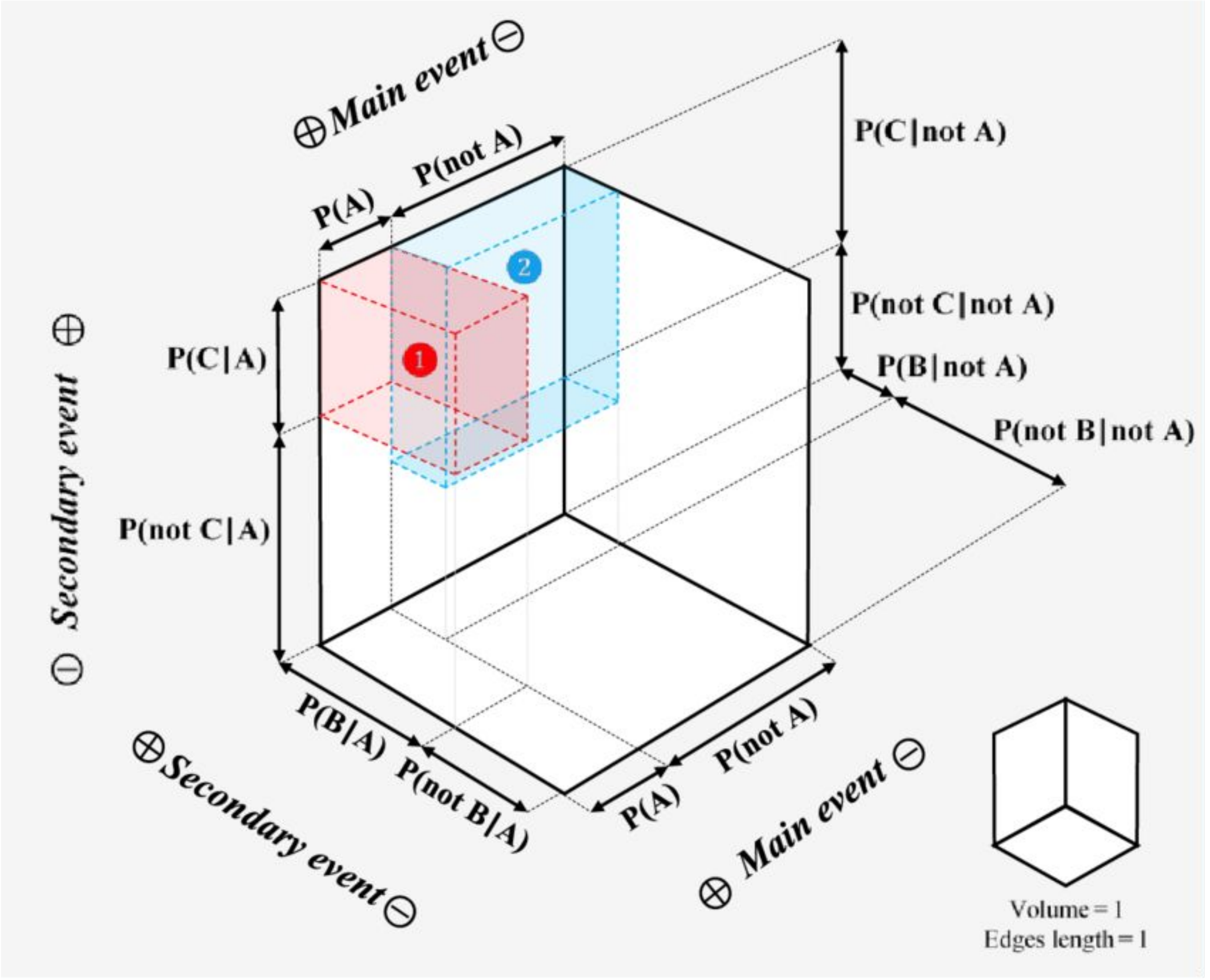
2.1 Conditional Probability Basics

Bayes' Theorem provides a way to update our beliefs or probabilities regarding an event A based on new evidence or information B. It is expressed mathematically as:

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

Where:
- $P(A \mid B)$ is the updated probability of event A given evidence B.
- $P(B \mid A)$ is the probability of observing evidence B given that event A has occurred.
- $P(A)$ is the prior probability of event A, before considering the evidence.
- $P(B)$ is the total probability of observing evidence B.

P(🌧️|🌅) = P(🌅|🌧️) × P(🌧️) / P(🌅)



Part 2: Key elements

2.1 Conditional Probability Basics

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

Prior Probability (P(A)): This is our initial belief or probability regarding event A before taking the new evidence into account. It represents what we know about A based on previous information.
Likelihood (P(B|A)): This term represents the probability of observing evidence B given that event A is true. It quantifies how likely the evidence is when the event of interest has occurred.
Evidence Probability (P(B)): This is the total probability of observing evidence B. It can be thought of as the normalization factor, ensuring that the updated probability is a valid probability distribution. It is calculated as the sum of the joint probabilities of A and B occurring and not occurring:

$$P(B) = P(B \mid A) \cdot P(A) + P(B \mid \neg A) \cdot P(\neg A)$$

- $P(B \mid A)$ is the likelihood of observing B given that A has occurred.
- $P(B \mid \neg A)$ is the likelihood of observing B given that A has not occurred (the complement of A).
- $P(\neg A)$ is the prior probability that A has not occurred (complement of A).

Updated Probability (P(A|B)): This is the probability of event A given the new evidence B. It represents our revised belief about A after considering the evidence. Bayes' Theorem updates our prior probability with the likelihood and the evidence probability.

Definition of a Random Variable

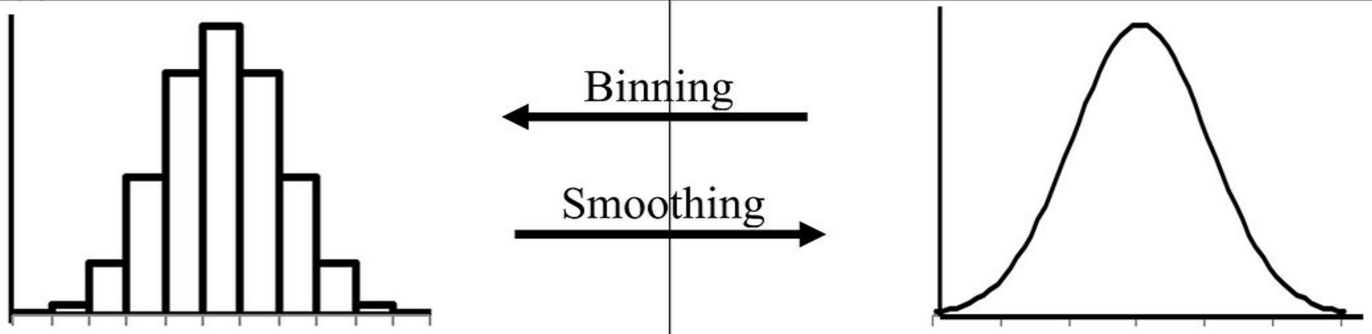
Part 2: Key elements

2.2 Random Variables and Their Laws

Random Variable (X): A random variable is defined as a function that maps the outcomes of a random experiment to real numbers. Mathematically, $X : \Omega \to R$, where Ω is the sample space, and R represents the set of real numbers.

Probability Mass Function (PMF) for Discrete Random Variable: For a discrete random variable X, the PMF is defined as $\mathrm{P(X = x)}$, which gives the probability that X takes on a specific value x. Mathematically, $\Sigma P(X = x) = 1$, where the summation is over all possible values of X.

Probability Density Function (PDF) for Continuous Random Variable: For a continuous random variable X, the PDF is denoted as $f(x)$, and it describes the probability density at a specific value x. Mathematically, $\int f(x)dx$ over the entire range of X equals 1 .

Discrete	Continuous
	
Probability Mass Function	Probability Density Function
Count,Sum,Proportion	Integration
P(X = x) = f(x)	P(X=x) = $\int f(x).dx$
CMF,PMF = Sum, Difference	CDF,PDF = Integrate, Differentiate

Covariance and correlation

Part 2: Key elements

2.2 Random Variables and Their Laws

Covariance (σXY) for a Sample:

The sample covariance between two random variables X and Y, based on n data points (x_i, $\mathrm{y_i}$), is calculated as follows:

$$\mathrm{Cov}(X,Y) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

Where:

- $\mathrm{Cov}(X,Y)$ is the covariance between X and Y.
- n is the number of data points.
- x_i and y_i are individual data points.
- $\bar{x}$ and $\bar{y}$ are the sample means of X and Y, respectively.

Covariance and correlation

Part 2: Key elements

2.2 Random Variables and Their Laws

Population Covariance (σXY) :

$$\operatorname{Cov}(X, Y)=\frac{1}{N} \sum_{i=1}^{N}\left(x_i-\backslash \mathrm{muX}\right)\left(y_i-\backslash \mathrm{muY}\right)$$

Where:
- $\operatorname{Cov}(X, Y)$ is the population covariance between X and Y.
- N is the population size.
- x_i and y_i are individual data points.
- $\backslash$ muX is the population mean of X.
- $\backslash$muY is the population mean of Y.

This formula calculates the population covariance, which measures how X and Y change together in the entire population. It's a measure of their joint variability. A positive population covariance indicates that when X is above its mean, Y tends to be above its mean as well, and vice versa. A negative population covariance indicates an inverse relationship. A value of O means there is no linear relationship between X and Y in the population.

Covariance and correlation

Part 2: Key elements

2.2 Random Variables and Their Laws

The correlation coefficient, often represented by ρ (rho) or r, quantifies the strength and direction of the linear relationship between two random variables X and Y. It's a normalized measure of the covariance. Here is the mathematical expression for the correlation coefficient:

Sample Correlation (r) for a Sample:

The sample correlation coefficient between two random variables X and Y, based on n data points $(\mathrm{x_i}, \mathrm{y_i})$, is calculated as follows:

$$r = \frac{S_{XY}}{S_X \cdot S_Y}$$

Where:

- r is the sample correlation coefficient.
- S_{XY} is the sample covariance between X and Y.
- S_X is the sample standard deviation of X.
- S_Y is the sample standard deviation of Y.

Covariance and correlation

The correlation coefficient, often represented by ρ (rho) or r, quantifies the strength and direction of the linear relationship between two random variables X and Y. It's a normalized measure of the covariance. Here is the mathematical expression for the correlation coefficient:

Population Correlation (ρ) for a Population:

The population correlation coefficient between two random variables X and Y is calculated similarly, using the population standard deviations and population covariance:

$$\rho = \frac{\sigma_{XY}}{\sigma_X \cdot \sigma_Y}$$

Where:
- ρ is the population correlation coefficient.
- σ_{XY} is the population covariance between X and Y.
- σ_X is the population standard deviation of X.
- σ_Y is the population standard deviation of Y.

The correlation coefficient ranges from -1 to 1 :
- A value of 1 indicates a perfect positive linear relationship, meaning that as one variable increases, the other increases proportionally.
- A value of -1 indicates a perfect negative linear relationship, meaning that as one variable increases, the other decreases proportionally.
- A value of O indicates no linear relationship; the variables are not linearly related.

The correlation coefficient is a valuable measure in statistics to assess the strength and direction of the linear relationship between two variables while normalizing for differences in scale.

Law of a Random Variable

Part 2: Key elements

2.2 Random Variables and Their Laws

Probability Density Function (PDF) for Continuous Random Variable: For a continuous random variable X, the PDF is denoted as $f(x)$, and it describes the probability density at a specific value x. Mathematically, $\int f(x)dx$ over the entire range of X equals 1 .

The PDF is defined in such a way that the area under the function's curve over a given interval represents the probability that the random variable falls within that interval. Specifically, for a continuous random variable X, the PDF is defined as follows:
$f(x) \geq 0$ for all values of x (the function is always non-negative).
The integral of the PDF over the entire range of possible values is equal to 1 :

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

This equation ensures that the total probability of all possible values of the random variable X is equal to 1 , meaning that X must take on some value in its entire range. The PDF is a fundamental concept in understanding the distribution of continuous random variables and plays a crucial role in various statistical and probabilistic calculations.

Part 2: Key elements

2.2 Random Variables and Their Laws

Probability Mass Function (PMF) for Discrete Random Variable: For a discrete random variable X, the PMF is defined as $\mathrm{P}(\mathrm{X} = \mathrm{x})$, which gives the probability that X takes on a specific value x. Mathematically, $\Sigma P(X = x) = 1$, where the summation is over all possible values of X.

A "Probability Mass Function" (PMF) is a mathematical function that describes the probability distribution of values for a discrete random variable. The PMF indicates the probability that the discrete random variable takes on a specific value.
For a discrete random variable X, the PMF is defined as follows:
$P(X = x) \geq 0$ for all possible values of x (the probability is always non-negative).
The sum of the probabilities for all possible values of X is equal to 1 :

$$\sum_x P(X = x) = 1$$

This equation ensures that the total probability of all possible values of the random variable X sums to 1 , indicating that X must take on one of its possible values.
The PMF is a fundamental concept in understanding the distribution of discrete random variables and is essential for various statistical and probabilistic calculations. It provides the basis for calculating expected values, variances, and other statistical measures for discrete random variables.

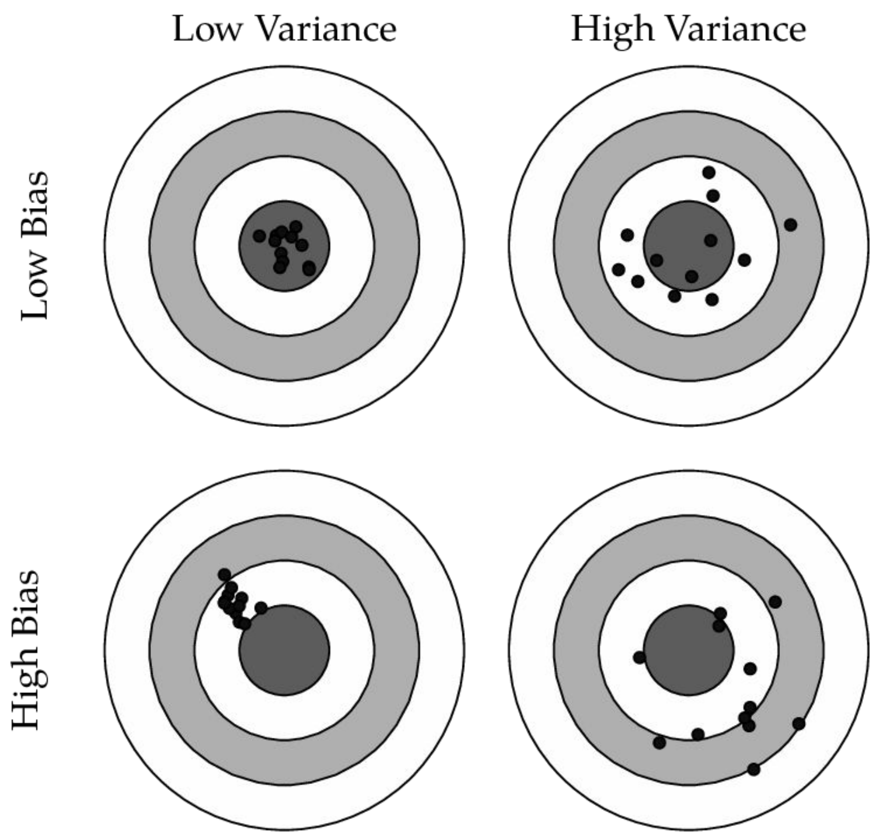
Expectation and Variance

Part 2: Key elements

2.2 Random Variables and Their Laws

Expected Value (Mean, $\mathbf{E}(\mathbf{X})$): The expected value of a random variable X is calculated as $E(X) = \Sigma[x * P(X = x)]$ for discrete X, or $E(X) = \int[X * f(x)]dx$ for continuous X.

Variance ($\mathrm{Var}(X)$) : The variance of X is calculated as $\mathrm{Var}(X) = E\left[(X - \mu)^2\right]$, where μ is the expected value. For discrete X, it can be calculated as $\mathrm{Var}(X) = \Sigma\left[(x - \mu)^{2*} P(X = x)\right]$, and for continuous $X, \mathrm{Var}(X) = \int\left[(x - \mu)^{2*} f(x)\right] dx$.



Law of a Tuple of Random Variables

Part 2: Key elements

2.2 Random Variables and Their Laws

Joint Probability Distribution: The joint probability distribution provides information about the probabilities of combinations of values for the set of random variables. For discrete random variables, it's defined using a joint probability mass function (joint PMF), and for continuous random variables, it's defined using a joint probability density function (joint PDF).

Marginal Distributions: A marginal distribution focuses on a single random variable within the tuple. It provides the probability distribution of that specific variable while ignoring the others. The marginal distribution can be obtained by summing (for discrete) or integrating (for continuous) the joint distribution over all other variables.

Independence: Random variables in a tuple are considered independent if the joint probability of all variables can be expressed as the product of their individual probabilities. Mathematically,
$P(X_1 = x_1, X_2 = x_2, \ldots, Xn = xn) = P(X_1 = x_1) * P(X_2 = x_2) * .. * P(Xn = xn)$.

Covariance and Correlation: The covariance and correlation coefficients are used to measure the degree to which two random variables in the tuple change together. They provide insights into the linear relationship between variables.

Functions of Random Variables: You can analyze functions of random variables within the tuple to understand how these transformations affect the overall joint distribution.
Applications:

The Law of a Tuple of Random Variables is essential in various fields, including statistics, economics, engineering, and machine learning. It's used for modeling complex systems with multiple variables and understanding how they interact, affect each other, and collectively influence outcomes.

Tuple of Random Variables vs Random vectors

Part 2: Key elements

2.2 Random Variables and Their Laws

A "Tuple of Random Variables" and a "Random Vector" both refer to collections of random variables. While they are similar in concept, there is a subtle difference in how they are often used and represented:

Tuple of Random Variables:
A tuple of random variables is a set of individual random variables that may or may not be related to each other. These random variables are typically listed as separate entities, each with its own probability distribution. For example, $\mathrm{X}_1, \mathrm{X}_2, \mathrm{X}_3, \ldots, \mathrm{Xn}$ represent individual random variables.

Random Vector:
A random vector is a mathematical construct that represents multiple random variables as components of a vector. It treats these random variables as elements of a single vector. The random vector is often denoted as a bold letter, such as $X = [X_1, X_2, X_3, \ldots, Xn]$, where X is a vector, and X_1, X_2, X_3, etc., are its components.

Key Differences:
Representation: A tuple of random variables is a simple listing of individual random variables, while a random vector treats these variables as components of a vector. A random vector provides a more compact representation for multiple random variables. Mathematical Operations: Random vectors allow for vector-based mathematical operations, such as addition, multiplication, and vector norms. This is particularly useful in multivariate statistics and linear algebra.

Notation: Random vectors are often denoted using vector notation, which is a convenient way to express multiple random variables together. In contrast, a tuple of random variables uses separate symbols for each variable.

Part 2: Key elements

2.3 Random Variables and Their Distributions

Probability Mass Function (PMF) for Discrete Random Variable: For a discrete random variable X, the PMF is defined as $\mathrm{P}(\mathrm{X}=\mathrm{x})$, which gives the probability that X takes on a specific value x. Mathematically, $\Sigma P(X=x)=1$, where the summation is over all possible values of X.

A "Probability Mass Function" (PMF) is a mathematical function that describes the probability distribution of values for a discrete random variable. The PMF indicates the probability that the discrete random variable takes on a specific value.
For a discrete random variable X, the PMF is defined as follows:
$P(X=x) \geq 0$ for all possible values of x (the probability is always non-negative).
The sum of the probabilities for all possible values of X is equal to 1 :

$$\sum_x P(X=x)=1$$

This equation ensures that the total probability of all possible values of the random variable X sums to 1 , indicating that X must take on one of its possible values.
The PMF is a fundamental concept in understanding the distribution of discrete random variables and is essential for various statistical and probabilistic calculations. It provides the basis for calculating expected values, variances, and other statistical measures for discrete random variables.

Part 2: Key elements

2.3 Random Variables and Their Distributions

Probability Density Function (PDF) for Continuous Random Variable: For a continuous random variable X, the PDF is denoted as $f(x)$, and it describes the probability density at a specific value x. Mathematically, $\int f(x) dx$ over the entire range of X equals 1 .

The PDF is defined in such a way that the area under the function's curve over a given interval represents the probability that the random variable falls within that interval. Specifically, for a continuous random variable X, the PDF is defined as follows:
$f(x) \geq 0$ for all values of x (the function is always non-negative).
The integral of the PDF over the entire range of possible values is equal to 1 :

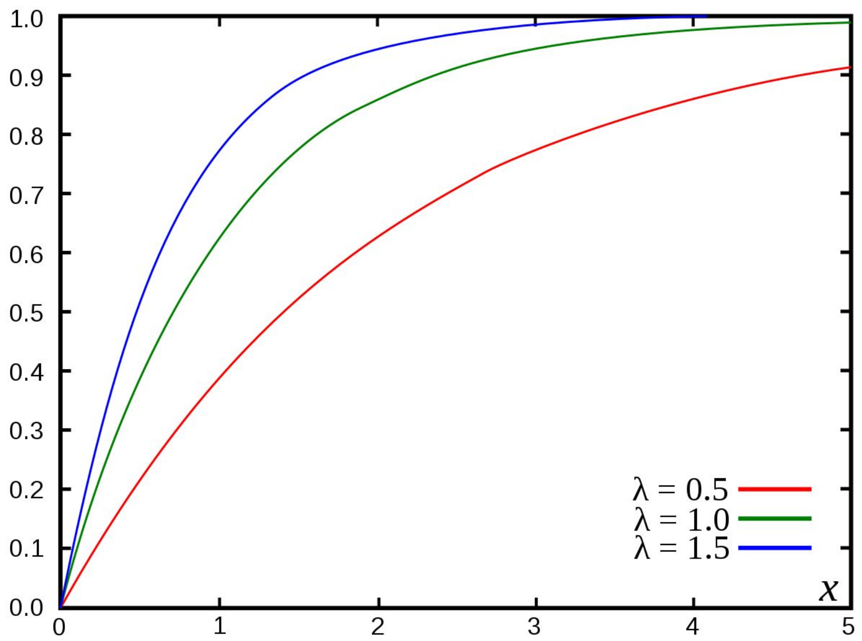
$$\int_{-\infty}^{\infty} f(x) dx = 1$$

This equation ensures that the total probability of all possible values of the random variable X is equal to 1 , meaning that X must take on some value in its entire range. The PDF is a fundamental concept in understanding the distribution of continuous random variables and plays a crucial role in various statistical and probabilistic calculations.

The Cumulative Distribution Function

Part 2: Key elements

2.3 Random Variables and Their Distributions



The Cumulative Distribution Function (CDF) of a random variable X, denoted as $F(x)$, is defined as follows:

$$F(x) = P(X \leq x)$$

This means that the CDF gives the probability that the random variable X is less than or equal to a given value x.

Key Features:

Properties of the CDF:

- The CDF is an increasing function, which means that for any $x_1 < x_2, F(x_1) \leq F(x_2)$.
- The CDF is bounded by 0 and 1 , as the probability of any event falls within this range:

$$0 \leq F(x) \leq 1$$

Interpretation:

- The CDF provides a comprehensive view of how the probabilities of X being less than or equal to various values are distributed across its entire range.

Probability Calculation:

- To find the probability that X falls within a specific interval $[a, b]$, you can use the CDF:

$$P(a \leq X \leq b) = F(b) - F(a).$$

Complementary Probability:

- You can calculate the probability that X is greater than a certain value by using the complementary probability: $P(X > x) = 1 - F(x)$.

Limiting Values:

- As x approaches negative infinity, $F(x)$ approaches 0 .
- As x approaches positive infinity, $F(x)$ approaches 1 .

The CDF is a fundamental tool in probability theory for characterizing the distribution of random variables. It is used to compute various statistical properties, including percentiles and quantiles, and is crucial for understanding the behavior of random variables in various applications, such as risk assessment and hypothesis testing.

Part 2: Key elements

2.3 Random Variables and Their Distributions

A bivariate distribution is a statistical distribution involving two random variables, X and Y, and describes the joint probability distribution of both variables. This distribution is defined by a joint probability density function (PDF) or a joint probability mass function (PMF) for discrete and continuous cases:

Joint Probability Density Function (PDF) for Continuous Variables $(\ f(x, y))$:
- Defines the probability that both X and Y fall within a specific region in their respective domains.

Joint Probability Mass Function (PMF) for Discrete Variables $(P(X = x, Y = y)\ )$:
- Gives the probability that X takes a particular value x and Y takes a particular value y.
These equations encapsulate the bivariate distribution and provide insight into how two random variables are related and the probabilities associated with their combinations.

Marginal Distributions

Part 2: Key elements

2.3 Random Variables and Their Distributions

The marginal distribution often denoted as $P(X)$ for a single random variable X in the context of a bivariate distribution, is the probability distribution of that variable on its own. It represents the probabilities of X taking different values, ignoring the presence of the other variable. Here's the mathematical explanation:
Mathematical Explanation:
For a bivariate distribution with random variables X and Y, the marginal distribution of X is calculated as follows:
Discrete Random Variables:
- To find the marginal distribution $P(X=x)$ for a specific value x, sum the joint probabilities for all possible values of Y, while keeping X fixed:

$$P(X=x)=\sum_{y} P(X=x, Y=y)$$

- This equation calculates the probability that X takes on the specific value x while ignoring the possible values of Y.
Continuous Random Variables:
- To find the marginal probability density function (PDF) $f_X(x)$ for a specific value x, integrate the joint PDF $f(x, y)$ with respect to Y over its entire range:

$$f_X(x)=\int_{-\infty}^{\infty} f(x, y) dy$$

- This equation calculates the probability density for X taking on the specific value x while ignoring the possible values of Y.

The marginal distribution of a random variable provides insights into its behavior and allows you to analyze it independently of the other variable in the bivariate distribution. It's an essential concept in probability and statistics, often used in various applications, such as calculating expected values, variances, and making univariate inferences.

Conditional Distributions

Part 2: Key elements

2.3 Random Variables and Their Distributions

Conditional Probability Mass Function (PMF):

For a discrete random variable X with its PMF denoted as $P(X)$, the conditional $PMFP(X \mid Y =$ y) for a given value of another random variable $Y = y$ is calculated as follows:

$$P(X = x \mid Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)}$$

- $P(X = x \mid Y = y)$: The conditional probability that X takes on a specific value X, given that Y is equal to y.
- $P(X = x, Y = y)$: The joint probability that both X and Y are equal to x and y, respectively.
- $P(Y = y)$: The probability that Y is equal to y.

In words, the conditional PMF for X given $Y = y$ is the joint probability of both X and Y being specific values x and y, divided by the probability that Y is equal to y.

Conditional Probability Density Function (PDF):

For continuous random variables X with its PDF denoted as $f(x)$, the conditional PDF $f(x \mid Y =$ y) for a given value of another random variable Y = y is calculated as follows:

$$f(x \mid Y = y) = \frac{f(x, y)}{f_Y(y)}$$

$f(x \mid Y = y)$: The conditional probability density function for X taking on a value x, given that Y is equal to y
.

$f(x, y)$: The joint probability density function for X and Y.

$f_Y(y)$: The probability density function for Y taking on the value y.

In words, the conditional PDF for X given Y = y is the joint PDF of both X and Y, divided by the PDF of Y being equal to y.

These conditional functions allow you to calculate the probability or density of one random variable, X, given certain conditions or values of another related random variable, Y. They are fundamental for making statistical inferences and modeling relationships between random variables in a dependent system.

Multivariate Distributions

Part 2: Key elements

2.3 Random Variables and Their Distributions

A multivariate distribution deals with multiple random variables, often organized as a vector.

Multivariate Distribution for Random Variables $X_1, X_2, \ldots, Xn$:

Joint Probability Density Function (PDF):
For n continuous random variables $X_1, X_2, \ldots, Xn$, the multivariate distribution is described by the joint probability density function (PDF), denoted as $f(x_1, x_2, \ldots, xn)$. This function defines the probability of the variables jointly taking on specific values $(x_1, x_2, \ldots, xn)$. Joint Probability Mass Function (PMF):
For n discrete random variables $X_1, X_2, \ldots, Xn$, the multivariate distribution is defined by the joint probability mass function (PMF), denoted as $P(X_1 = x_1, X_2 = x_2, \ldots, Xn = xn)$. It provides the probability that the variables simultaneously take particular values $(x_1, x_2, \ldots, xn)$.

Covariance Matrix (For Continuous Variables):
The covariance matrix, often denoted as Σ, captures the relationships between pairs of random variables in the multivariate distribution. The (i, j)-th entry of the matrix represents the covariance between Xi and Xj. A covariance matrix is used in multivariate Gaussian distributions.
Correlation Matrix (For Continuous Variables):

The correlation matrix, often denoted as R, is derived from the covariance matrix and represents the pairwise correlations between random variables. Each entry (i, j) represents the correlation between Xi and Xj, indicating the degree and direction of their linear relationship.

Marginal Distributions:
You can obtain the marginal distributions for individual random variables by integrating (for continuous variables) or summing (for discrete variables) the joint distribution over the other variables. For example, the marginal PDF for X_1 is obtained by integrating $f(x_1, x_2, \ldots, \mathrm{xn})$ with respect to $X_2, X_3, \ldots, Xn$.

The specific equations for these functions will depend on the nature of the random variables (discrete or continuous) and the specific multivariate distribution being considered (e.g., multivariate normal distribution for continuous variables or joint multinomial distribution for discrete variables).

Definition and Examples

Part 2: Key elements

2.4 Random Vectors

A random vector is a mathematical concept used to represent multiple random variables simultaneously. Random vectors are often encountered in multivariate statistics and probability theory. Below are the mathematical notations and key concepts related to random vectors:

Random Vector:

A random vector typically represents a collection of n random variables, which can be denoted as $X = [X_1, X_2, \ldots, Xn]$, forming a vector. Each component X_i can be either a continuous or discrete random variable. The vector X represents a single outcome that includes values for all these random variables.

Joint Probability Density Function (PDF) or Probability Mass Function (PMF):

The joint PDF or PMF for a random vector $X = [X_1, X_2, \ldots, Xn]$ describes the probability distribution of all n random variables simultaneously.

For a continuous random vector (PDF):

$$f(\mathbf{x})$$

For a discrete random vector (PMF):

$$P(\mathbf{x})$$

Here, $\mathbf{x} = [x_1, x_2, \ldots, xn]$ represents a specific vector of values for the random variables in X.

Definition and Examples

Part 2: Key elements

2.4 Random Vectors

Expectation and Variance for Random Vectors:
The expected value (mean) and variance of a random vector are defined as follows:
Expected Value for a Random Vector:

$$E(\mathbf{X}) = [\mu_1, \mu_2, \ldots, \mu_n]$$

Here, μ_i represents the expected value of the i-th random variable, i.e., $\mu_i = E(X_i)$.

Covariance Matrix for a Random Vector (Continuous):
The covariance matrix Σ characterizes the pairwise covariances between random variables in the vector. The (i, j)-th element of the matrix represents the covariance between X_i and X_j.

Covariance Matrix for a Random Vector (Discrete):
The covariance matrix Σ is similarly defined for discrete random vectors but involves the covariances between individual values of X_i and X_j.

Correlation Matrix for a Random Vector (Continuous):
The correlation matrix R is derived from the covariance matrix and represents the pairwise correlations between random variables in the vector.
Marginal Distributions:

You can obtain the marginal distributions for individual random variables by integrating (for continuous variables) or summing (for discrete variables) the joint distribution over the other variables.

Multivariate Distributions:
Random vectors are commonly associated with multivariate distributions, such as the multivariate normal distribution for continuous variables or the joint multinomial distribution for discrete variables.

Random vectors are essential for modeling complex systems with multiple variables, understanding their joint behavior, and performing multivariate statistical analyses. The specific mathematical expressions may vary depending on the nature of the random variables and the distribution being considered.

The Expectation of a Random Variable

Part 2: Key elements

2.5 Expectation and Variance

The expectation of a random variable, often denoted as $E(X)$, represents the mean or average value of that random variable. Here's the mathematical expression for the expectation of a random variable:
Expectation $(E(X))$:
For a discrete random variable X, the expectation is calculated as:

$$E(X) = \sum_x x \cdot P(X = x)$$

Where:
- $E(X)$ is the expectation of the random variable X.
- x represents individual values that X can take.
- $P(X = x)$ is the probability that X equals the specific value x.

For a continuous random variable X with a probability density function (PDF) denoted as f(x), the expectation is calculated as an integral:

$$E(X) = \int_{-\infty}^{\infty} x \cdot f(x) dx$$

Where:
- $E(X)$ is the expectation of the continuous random variable X.
- x represents the values X can take.
- $f(x)$ is the PDF of the continuous random variable X.
- The integral is taken over the entire range of possible values for X.

The expectation provides a measure of the central tendency of a random variable and is akin to its average value when the random variable is repeatedly observed or measured. It is a fundamental concept in probability and statistics.

Properties of Expectations

Part 2: Key elements

2.5 Expectation and Variance

Expectations of random variables have several important properties, which make them a powerful tool in probability theory and statistics. Here are some of the key properties:

Linearity: The expectation of a sum of random variables is equal to the sum of their expectations. This property holds for both discrete and continuous random variables. Formally, for random variables X and Y and constants a and b :

- $E(aX + bY) = aE(X) + bE(Y)$

Constant: The expectation of a constant is simply the constant itself. For any constant c:

- $E(c) = c$

Independence: If two random variables X and Y are independent, the expectation of their product is equal to the product of their expectations:

- $E(XY) = E(X)E(Y)$

Monotonicity: If X is a non-decreasing function of Y (i.e., X $\leq$ Y), then $E(X) \leq E(Y)$.

Expectation of a Function of a Random Variable: If $g(X)$ is a function of the random variable X, then:

- $E[g(X)] = \sum_x g(x)P(X = x)$ for discrete random variables.
- $E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx$ for continuous random variables.

Properties of Expectations

Part 2: Key elements

2.5 Expectation and Variance

Law of the Unconscious Statistician (LOTUS) This property allows you to compute the expectation of a function of a random variable using the random variable's probability distribution. If you have a function h(X), then:
- $E[h(X)] = \sum_x h(x)P(X = x)$ for discrete random variables.
- $E[h(X)] = \int_{-\infty}^{\infty} h(x)f(x)dx$ for continuous random variables.

Iterated Expectations: This property, also known as the law of total expectation or tower property, is used when dealing with conditional expectations. It states that the expectation of a random variable can be computed as the expectation of its conditional expectation:
- $E(X) = E[E(X \mid Y)]$

Variance Property: The variance of a random variable X can be expressed as the expectation of its squared deviation from the mean:
- $\mathrm{Var}(X) = E\left[(X - E(X))^2\right]$

Conditional Expectation

Part 2: Key elements

2.5 Expectation and Variance

Conditional Expectation $(E(X \mid Y = y))$:
The conditional expectation of a random variable X given that another random variable Y takes on a specific value y is calculated as follows:
For a discrete random variable X :

$$E(X \mid Y = y) = \sum_x x \cdot P(X = x \mid Y = y)$$

For a continuous random variable X :

$$E(X \mid Y = y) = \int_{-\infty}^{\infty} x \cdot f(X \mid Y = y) dx$$

Where:
- $E(X \mid Y = y)$ is the conditional expectation of the random variable X given that $\mathrm{Y} = \mathrm{y}$. - x represents individual values that X can take.
- $P(X = x \mid Y = y)$ is the conditional probability that X equals the specific value x, given that Y equals y for discrete random variables.
- $f(X \mid Y = y)$ is the conditional probability density function (PDF) for continuous random variables.
- The sum (for discrete) or the integral (for continuous) is taken over the entire range of possible values for X.

The mean and the median

The mean and the median are two common measures of central tendency used to describe a set of data. They provide insights into the "typical" or "central" value of a dataset. Here's a brief explanation of each:

Mean (Average):

The mean, often denoted as μ (mu) or $\bar{x}$ (x-bar), is the arithmetic average of a set of values. To calculate the mean, sum up all the values in the dataset and then divide by the number of values.
The mean is sensitive to extreme values (outliers) in the dataset.
It is widely used in statistics and is a measure of central tendency.

Median:

The median is the middle value in a dataset when the values are arranged in order. If there is an even number of values, the median is the average of the two middle values.
The median is not affected by extreme values (outliers) in the dataset.
It is often used when the data is not normally distributed or when you want to find a typical value that represents the "middle" of the data.
Here's a simple mathematical representation for each:
Mean $(\mu$ or $\bar{x})$ for a Dataset with $\mathbf{n}$ Values:

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Median for a Dataset with $\mathbf{n}$ Values:
Arrange the data in order.
If n is odd, the median is the value at position $\frac{n+1}{2}$.
If n is even, the median is the average of the values at positions $\frac{n}{2}$ and $\frac{n}{2}+1$.
Both the mean and median have their strengths and are used in different situations. The mean is appropriate when dealing with normally distributed data or when you want to find the average value. The median is useful when you need a robust measure that is not influenced by outliers or when the data is skewed or not normally distributed.

Variance and Covariance

Part 2: Key elements

2.5 Expectation and Variance

Variance (σ^2 or Var(X)) :

The variance of a random variable X measures the spread or dispersion of its values. It is calculated as follows:

For a Discrete Random Variable X:

$$\operatorname{Var}(X) = E\left[(X-\mu)^2\right] = \sum_x (x-\mu)^2 \cdot P(X=x)$$

Where:
- $\operatorname{Var}(X)$ is the variance of the random variable X.
- $E\left[(X-\mu)^2\right]$ represents the expected value of the squared deviations from the mean.
- x represents individual values that X can take.
- μ is the mean (expected value) of X.
- $P(X=x)$ is the probability that X equals the specific value X.

For a Continuous Random Variable X :

$$\operatorname{Var}(X) = E\left[(X-\mu)^2\right] = \int_{-\infty}^{\infty} (x-\mu)^2 \cdot f(x)dx$$

Where:
- $\operatorname{Var}(X)$ is the variance of the continuous random variable X.
- $E\left[(X-\mu)^2\right]$ represents the expected value of the squared deviations from the mean.
- x represents the values that X can take.
- μ is the mean (expected value) of X.
- $f(x)$ is the probability density function (PDF) of X.
- The integral is taken over the entire range of possible values for X.

Variance and Covariance

Part 2: Key elements

2.5 Expectation and Variance

Covariance (σXY or $\text{Cov}(X, Y)$) :

The covariance between two random variables X and Y measures their joint variability or relationship. It is calculated as follows:

For a Discrete Random Variable X and Y :

$$\text{Cov}(X,Y) = E[(X - \setminus\text{muX}) \cdot (Y - \setminus\text{muY})] = \sum_x \sum_y (x - \setminus\text{muX})(y - \setminus\text{muY}) \cdot P(X = x, Y = y)$$

Where:
- $\text{Cov}(X,Y)$ is the covariance between X and Y.
- $E[(X - \setminus\text{muX}) \cdot (Y - \setminus\text{muY})]$ represents the expected value of the product of the deviations of X and Y from their means.
- x and y represent the individual values that X and Y can take.
- $\setminus\text{muX}$ and $\setminus\text{muY}$ are the means (expected values) of X and Y.
- $P(X = x, Y = y)$ is the joint probability that X equals X and Y equals y.

For Continuous Random Variables X and Y :

$$\text{Cov}(X,Y) = E[(X - \setminus\text{muX}) \cdot (Y - \setminus\text{muY})] = \iint (x - \setminus\text{muX})(y - \setminus\text{muY}) \cdot f(x,y) dx dy$$

Where:
- $\text{Cov}(X,Y)$ is the covariance between X and Y.
- $E[(X - \setminus\text{muX}) \cdot (Y - \setminus\text{muY})]$ represents the expected value of the product of the deviations of X and Y from their means.
- x and y represent the values that X and Y can take.
- $\setminus\text{muX}$ and $\setminus\text{muY}$ are the means (expected values) of X and Y.
- $f(x,y)$ is the joint probability density function (PDF) of X and Y.
- The double integral is taken over the entire range of possible values for X and Y.

Variance measures the spread or dispersion of a single random variable, while covariance quantifies the joint variability of two random variables, indicating their linear relationship.

The Law of Large Numbers

Part 3: Advanced Concepts and Simulations

3.1 Large Random Samples and Laws

The Law of Large Numbers (LLN) is a fundamental theorem in probability and statistics that describes the behavior of sample averages as the sample size increases. There are two main versions of the LLN: the Weak Law of Large Numbers (WLLN) and the Strong Law of Large Numbers (SLLN). I'll provide the mathematical expressions for both:

Weak Law of Large Numbers (WLLN):

The WLLN states that, as the sample size (n) becomes large, the sample mean $\bar{X}$ converges in probability to the true population mean μ :

$$\lim_{n \to \infty} P(|\bar{X} - \mu| < \varepsilon) = 1$$

In this expression:
- n is the sample size.
- $\bar{X}$ represents the sample mean.
- μ is the population mean.
- ε is a small positive number representing a margin of error.

In simple terms, as you collect more and more data points, the sample mean $\bar{X}$ becomes increasingly close to the true population mean μ with a high probability.

Part 3: Advanced Concepts and Simulations

3.1 Large Random Samples and Laws

Strong Law of Large Numbers (SLLN):
The SLLN is a stronger version of the LLN. It states that, as the sample size (n) becomes large, the sample mean $\bar{X}$ almost surely converges to the true population mean (\mu):
$\lim_{n\to\infty} \bar{X} = \mu \quad \text{with probability } 1$

In this expression:
n is the sample size.
- $\bar{X}$ represents the sample mean.
μ is the population mean.

In simple terms, the SLLN asserts that as you collect more data, the sample mean $\bar{X}$ will converge to the true population mean μ with probability 1 , which means it is almost certain to happen.

Both versions of the LLN are fundamental in statistics and have important applications in various fields, such as estimating population parameters, conducting hypothesis tests, and making statistical inferences. They provide a theoretical foundation for the use of sample statistics to make inferences about populations.

Statement of the Strong Law of Large Numbers (SLLN):
For a sequence of independent and identically distributed random variables $X_1, X_2, X_3, \ldots$, with the same mean μ and finite variance σ^2, the SLLN asserts that: $\lim_{n \to \infty} \bar{X}_n = \mu \quad$ with probability 1

In this expression:
- n represents the sample size (i.e., the number of observations).
- $\bar{X}_n$ is the sample mean of the first n observations.
- μ is the population mean.

Key Characteristics and Interpretation:
Independence and Identical Distribution: The SLLN assumes that the random variables $X_1, X_2, X_3, \ldots$ are independent and identically distributed (i.i.d.), which means that each random variable comes from the same probability distribution and is not influenced by the others.

Almost Sure Convergence: The SLLN asserts that as the sample size (n) increases to infinity, the sample mean $\bar{X}_n$ converges to the true population mean μ almost surely. "Almost surely" means that this convergence occurs with probability 1 , which implies that it is certain to happen as the sample size becomes sufficiently large.

Applications: The SLLN has important applications in statistics, as it justifies the use of sample means to estimate population means and provides a strong theoretical foundation for statistical inference. It ensures that with a sufficiently large sample size, the sample mean will be a very close estimate of the population mean.

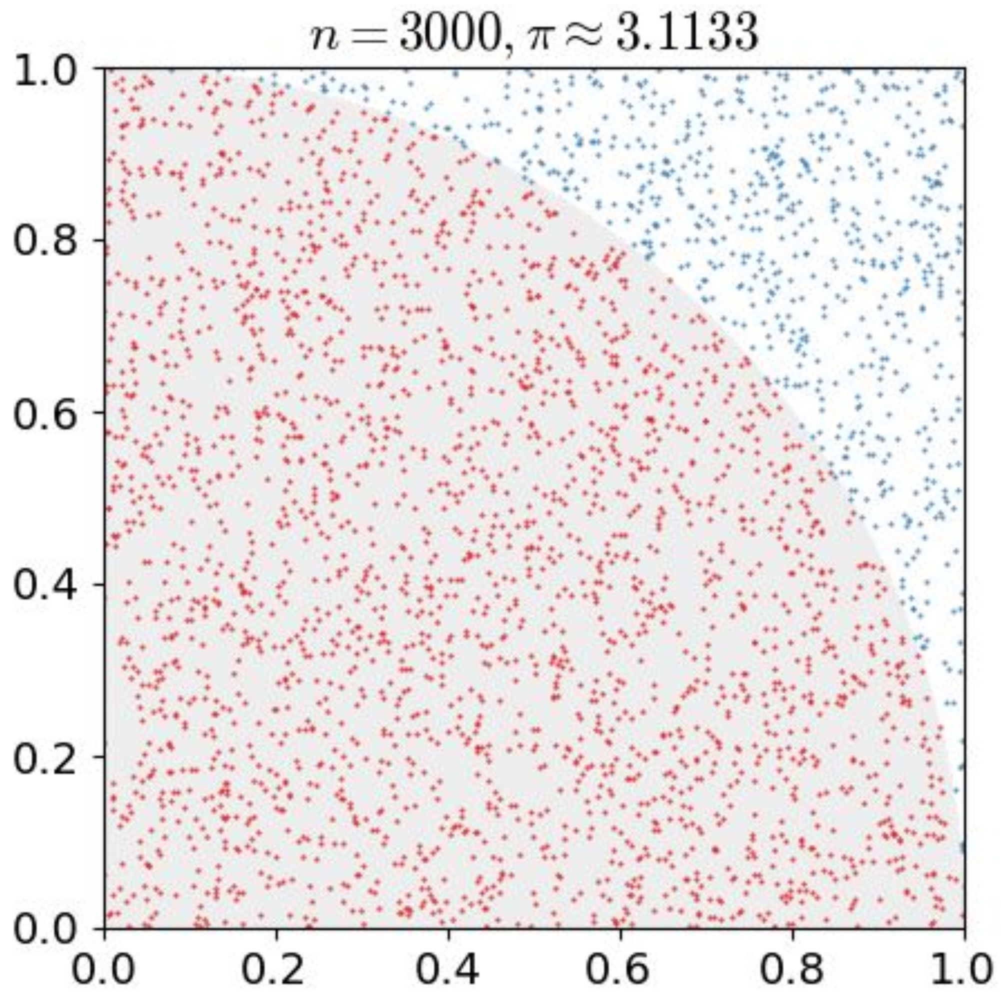
Differences from the Weak Law of Large Numbers (WLLN): The SLLN is a stronger result than the Weak Law of Large Numbers (WLLN). While the WLLN states that the sample mean converges in probability to the population mean, the SLLN guarantees almost sure convergence, which is a stronger and more reassuring result.

The SLLN is a fundamental theorem that underlies many statistical methods and justifies the practice of using sample statistics to make inferences about population parameters, assuming the conditions of the theorem are met. It provides a high level of confidence in the reliability of sample means for estimating population means in large samples.

What Is Simulation?

Part 3: Advanced Concepts and Simulations

3.2 Simulation in Probability



Monte Carlo Simulation Framework:

Problem Description: Define the problem or system you want to analyze or model. This typically involves specifying the relevant variables, parameters, and relationships.

Random Sampling: For each variable in the model, you need to sample random values based on their respective probability distributions. You can use a variety of techniques, including inverse transform sampling or specialized libraries for specific distributions.

Model Evaluation: Evaluate the problem or system using the sampled values of the variables. This may involve running a simulation, solving a complex equation, or performing any other relevant computation.

Collecting Results: Collect the results from each iteration of the simulation. This can include the output of the model, such as the value of an option, the temperature of a system, or the probability of an event.

Repeat: Steps 2-4 are repeated a large number of times (often thousands or millions of times). This creates a distribution of possible outcomes.

Analysis: Analyze the collected results to estimate the desired quantity or make inferences about the system or problem. This can include computing statistics such as the mean, variance, or quantiles of the results.

Estimating Expected Values:
To estimate an expected value (e.g., the mean) using a Monte Carlo simulation, you can use the following formula:
Estimate of $E(X) = \frac{1}{N} \sum_{i=1}^{N} X_i$

Where:
$E(X)$ is the expected value you want to estimate.
N is the number of iterations or samples in the simulation.
X_i represents the result of the simulation in the i-th iteration.
The law of large numbers ensures that as the number of iterations (N) increases, the estimate becomes increasingly accurate and converges to the true expected value.

Part 3: Advanced Concepts and Simulations

3.2 Simulation in Probability

■ Uniform Distribution:

A random variable X follows a continuous uniform distribution on the interval $[a, b]$ if its probability density function (PDF) is given by:

$$f(x) = \frac{1}{b-a}, \quad a \leq x \leq b$$

To simulate a random variable X that follows a uniform distribution on [a, b], you can generate random numbers from a uniform distribution on $[0, 1]$ and transform them using:

$$X = a + (b-a) \cdot U$$

Where:
- U is a random number uniformly distributed on $[0, 1]$.
- X will be uniformly distributed on $[a, b]$.

■ Normal (Gaussian) Distribution:

A random variable X follows a normal distribution with mean μ and standard deviation σ if its PDF is given by:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

To simulate a random variable X that follows a normal distribution with mean μ and standard deviation σ, you can use a standard normal random variable Z (with mean 0 and standard deviation 1) and transform it as:

$$X = \mu + \sigma \cdot Z$$

Where:
- Z is a random number that follows a standard normal distribution (mean 0 , standard deviation 1).
- X will follow a normal distribution with mean μ and standard deviation σ.

■ Exponential Distribution:

A random variable X follows an exponential distribution with rate parameter λ if its PDF is given by:

$$f(x) = \lambda e^{-\lambda x}, \quad x \geq 0$$

To simulate a random variable X that follows an exponential distribution with rate parameter λ, you can generate a random number from a uniform distribution on $[0, 1]$ and transform it as:

$$X = -\frac{1}{\lambda}\ln(U)$$

Where:

- U is a random number uniformly distributed on $[0, 1]$.
- X will follow an exponential distribution with rate parameter λ.

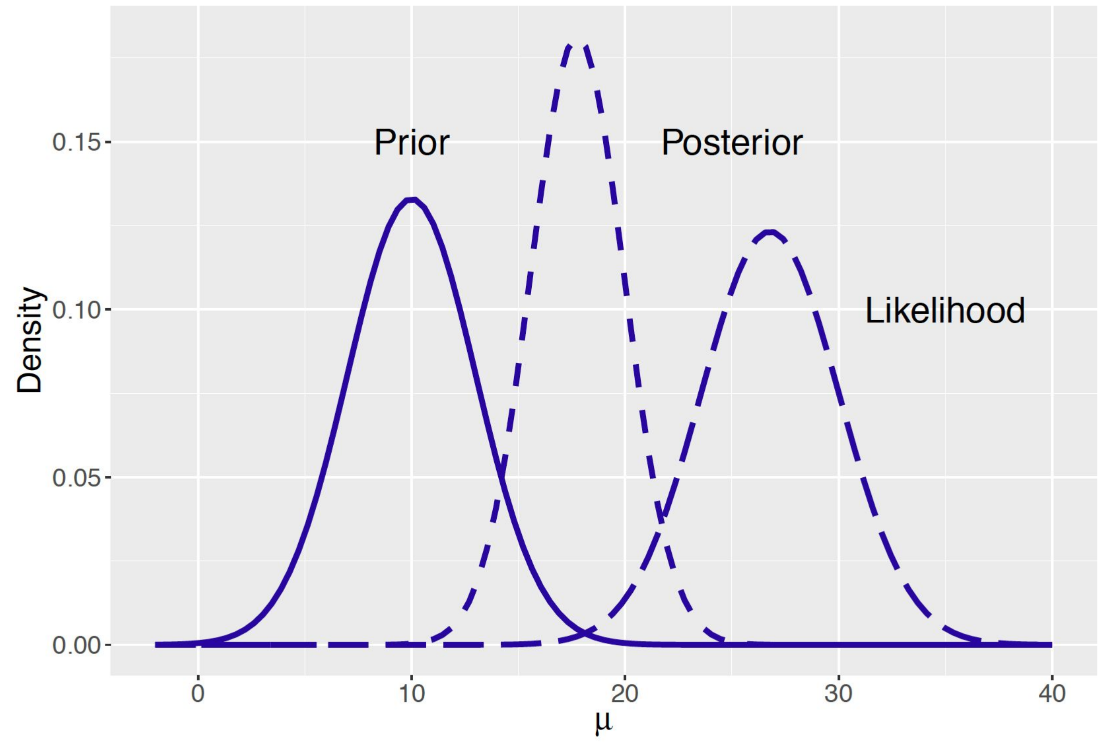
These are just a few examples of simulating specific probability distributions using Monte Carlo methods. More complex distributions require more involved techniques, and you can use software libraries and tools designed for such simulations.

Markov Chain Monte Carlo

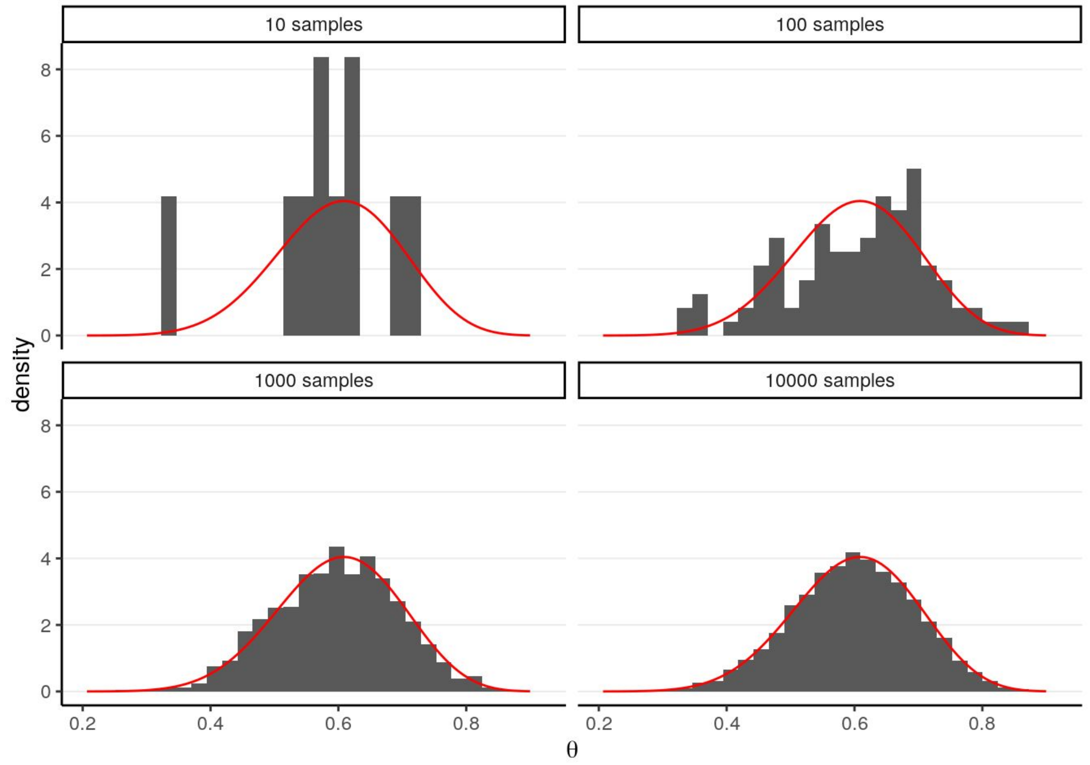
Part 3: Advanced Concepts and Simulations

3.2 Simulation in Probability

Markov Chain Monte Carlo (MCMC) is a powerful statistical technique for sampling from complex probability distributions and estimating various statistical properties. It is often used in Bayesian statistics, machine learning, and other fields where obtaining samples from a target distribution is challenging. The mathematical basis of MCMC involves Markov chains and the detailed balance condition.



Markov Chain:
A Markov chain is a stochastic process where the probability of transitioning from one state to another depends only on the current state and is independent of the past states. Formally, a discrete-time Markov chain is defined by a set of states, a transition probability matrix, and an initial state distribution.
- States: $S=\{s_1, s_2, \ldots, s_n\}$
- Transition Probability Matrix: $P=[p_{ij}]$, where p_{ij} is the probability of transitioning from state s_i to state s_j.
- Initial State Distribution: $\pi=[\pi_i]$, where π_i is the probability of starting in state s_i.



Detailed Balance Condition:
The detailed balance condition is a fundamental concept in MCMC. It ensures that a Markov chain converges to a stationary distribution. For MCMC to work, the target distribution $\pi(x)$ should satisfy the detailed balance condition with respect to a proposal distribution $q(x' \mid x)$:

$$\pi(x) p\left(x \rightarrow x'\right)=\pi\left(x'\right) p\left(x' \rightarrow x\right)$$

Where:
- $\pi(x)$ is the target distribution (e.g., the posterior distribution in Bayesian inference).
- $p(x \rightarrow x')$ is the transition probability from state x to state x' (according to the Markov chain dynamics).
- $p(x' \rightarrow x)$ is the transition probability from state x' to state x.
- $q(x' \mid x)$ is the proposal distribution, representing the probability of proposing a transition from state x to state x'.

Part 3: Advanced Concepts and Simulations

3.2 Simulation in Probability

Metropolis-Hastings Algorithm:

The Metropolis-Hastings (MH) algorithm is a widely used MCMC method. It generates samples from a target distribution by following these steps:

1. Start at an initial state $x^{(0)}$.

2. For each iteration t :

- Propose a new state x' from the proposal distribution $q(x' \mid x^{(t)})$.
- Calculate the acceptance probability $A(x^{(t)}, x')$ as the ratio of the target distribution and the proposal distribution.
- Accept the proposal with probability $A(x^{(t)}, x')$, and set $x^{(t+1)} = x'$. Otherwise, stay at the current state: $x^{(t+1)} = x^{(t)}$.

3. Repeat step 2 for a large number of iterations.

Convergence:
The Metropolis-Hastings algorithm converges to the target distribution when it satisfies the detailed balance condition. In practice, it may take some time for the chain to reach equilibrium, but after sufficient iterations, the samples obtained should closely approximate the desired distribution.
MCMC methods, including the Metropolis-Hastings algorithm, provide a versatile way to sample from complex probability distributions and estimate various properties of those distributions, making them valuable tools in Bayesian statistics, statistical physics, and machine learning.

Part 3: Advanced Concepts and Simulations

3.2 Simulation in Probability

The bootstrap method is a resampling technique used to estimate the sampling distribution of a statistic without making parametric assumptions. Here's a concise mathematical approach:

Data Collection: Start with an original data sample of size n denoted as $X = \{x_1, x_2, \ldots, x_n\}$.

Resampling: Generate a large number of bootstrap samples (typically thousands) by drawing random samples with replacement from the original data. Each bootstrap sample X_i^* is of the same size n as the original sample.

Statistic Estimation: Calculate the statistic of interest for each bootstrap sample. Let $T(X_i^*)$ represent the statistic calculated from the i-th bootstrap sample.

Bootstrap Sampling Distribution: The set of statistics $\{T(X_1^*), T(X_2^*), \ldots, T(X_k^*)\}$ forms the bootstrap sampling distribution of the statistic. This distribution represents the variability of the statistic without assuming any specific population distribution.

Statistical Analysis: You can analyze the bootstrap sampling distribution to estimate properties of the statistic, such as confidence intervals, standard errors, or hypothesis test p-values.

In summary, the bootstrap method is a powerful technique for estimating the sampling distribution of a statistic by repeatedly resampling from the original data. It provides a nonparametric and data-driven approach to assess the variability and uncertainty of statistical estimates.

KEYWORDS (NEW)

- Fonction affine
- Loi faible des grands nombres
- Théorème de la limite centrale tronquée
- Théorie des files d'attente
- Théorie des jeux
- Fonction
- Union (Disjonction)
- Vraisemblance
- Loi forte des grands nombres
- Théorème de convergence en distribution
- Système complet d'événements
- Variable dépendante
- Probabilité
- Variable aléatoire
- Loi binomiale
- Explosion combinatoire
- Théorème de convergence en probabilité
- Théorème de convergence en loi
- Variable indépendante
- Kurtosis
- Variable endogène
- Loi conjointe
- Factorisation canonique
- Théorème de convergence presque certaine
- Théorème de De Moivre-Laplace
- Loi des grands nombres
- Aléatoire
- Variable exogène
- Loi marginale
- Factorisation de la distribution
- Théorème de convergence presque sûre
- Fonction de distribution conjointe
- Théorème central limite
- Variable explicative
- Distribution
- Fonction de répartition
- Théorème de la limite centrale multivariée
- Fonction de distribution cumulative
- Théorème de Bernoulli
- Variable expliquée
- Événement
- Fonction de vraisemblance
- Fonction de densité de probabilité (PDF)
- Fonction de masse de probabilité
- Théorème de Cochran
- Vecteur gaussien
- Notion d'ordre
- Intersection (Conjonction)
- Fréquences relatives cumulées, effectifs cumulés
- Formule des probabilités composées
- Théorème de Cramér
- Fréquence
- Convergence
- Convergence en probabilité
- Convergence essentiellement uniforme
- Formule des probabilités totales
- Inégalité de Markov
- Convergence en loi
- Convergence presque sûre
- Définition classique des probabilités
- Inégalité de Bienaymé-Tchebychev
- Log-vraisemblance
- Convergence faible
- Permutation sans répétition
- Espérance mathématique
- Événements mutuellement exclusifs
- Définition axiomatique de la probabilité
- Ensemble négligeable
- Convergence forte
- Combinatoire analytique
- Événement complémentaire
- Arrangement avec répétition
- Événements incompatibles (disjoints)
- Espace d'événements
- Espace probabilisé
- Événement simple
- Événement composé
- Combinaison avec répétitions (avec remises)
- Événements indépendants
- Événement contraire
- Notion de répétition
- Factoriel d'un entier n
- Combinaison sans répétitions (sans remises)
- Arrangement sans répétition
- Expérience aléatoire
- Permutation avec répétition
- Axiomes de Kolmogorov
- Propriétés des combinaisons, binôme de Newton

Use Case with python

Use Case: Analyzing Risk with Probability Distributions

In the context of the course on Foundations of Probability, Discrete and Continuous Probability Distributions, and Advanced Topics in Probability, let's explore a use case related to analyzing risk using probability distributions. This use case involves the application of probability concepts to assess and manage risk in financial investments.

Description:

In this use case, we will focus on modeling the risk associated with a financial investment using probability distributions. We will consider a hypothetical investment in a stock and assess the potential returns and risks associated with it by modeling the stock's price movement.

Key Components:

1. Foundations of Probability: Understanding the basic concepts of probability, such as sample space, events, and probability axioms, is essential for modeling and assessing risk.

2. Discrete and Continuous Probability Distributions: Knowledge of discrete and continuous probability distributions, including their probability mass functions (PMFs) and probability density functions (PDFs), is crucial for modeling financial data.

3. Expected Value and Variance: Calculating the expected value and variance of a random variable helps assess the average return and risk associated with an investment.

4. Advanced Topics in Probability: Concepts such as joint and marginal distributions, functions of random variables, and convergence concepts are useful for analyzing the relationship between multiple financial instruments.

Python Code Example (Analyzing Investment Risk):

```python
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import norm

# Define parameters of the stock's return distribution
mean_return = 0.08  # Mean return
volatility = 0.15  # Annual volatility (standard deviation)

# Generate a range of possible annual returns
returns = np.linspace(-0.3, 0.3, 1000)

# Calculate the probability density function (PDF) using the normal distribution
pdf_values = norm.pdf(returns, loc=mean_return, scale=volatility)

# Plot the PDF of the stock's returns
plt.figure(figsize=(10, 6))
plt.plot(returns, pdf_values, label='PDF', color='blue')
plt.xlabel('Annual Return')
plt.ylabel('Probability Density')
plt.title('Probability Density Function (PDF) of Stock Returns')
plt.legend()
plt.grid(True)

# Calculate and print the expected return and risk (standard deviation)
expected_return = mean_return
risk = volatility
print(f'Expected Return: {expected_return:.2%}')
print(f'Risk (Standard Deviation): {risk:.2%}')

plt.show()
```

In this code, we model the probability distribution of a stock's annual returns using a normal distribution with a specified mean return and volatility. We calculate and display the probability density function (PDF) of returns and then calculate the expected return and risk (standard deviation) as measures of investment risk.

This use case demonstrates how probability concepts can be applied to assess investment risk, allowing investors and analysts to make informed decisions based on the expected return and risk associated with a financial instrument.

REFERENCES V2

- Kolmogorov, A. N. (1950). Foundations of the Theory of Probability. Chelsea Publishing Company.
- Hacking, I. (2006). The Emergence of Probability: A Philosophical Study of Early Ideas about Probability, Induction and Statistical Inference. Cambridge University Press.
- Ross, S. M. (2014). A First Course in Probability. Pearson.
- Johnson, N. L., Kotz, S., & Balakrishnan, N. (2005). Discrete Multivariate Distributions. Wiley.
- Feller, W. (1968). An Introduction to Probability Theory and Its Applications, Vol. 1. Wiley.
- DeGroot, M. H., & Schervish, M. J. (2018). Probability and Statistics. Pearson.
- Casella, G., & Berger, R. L. (2001). Statistical Inference. Duxbury Press.
- Evans, M., Hastings, N., & Peacock, B. (2000). Statistical Distributions. Wiley.
- Grimmett, G., & Stirzaker, D. (2001). Probability and Random Processes. Oxford University Press.
- Papoulis, A., & Pillai, S. U. (2002). Probability, Random Variables, and Stochastic Processes. McGraw-Hill.
- Durrett, R. (2019). Probability: Theory and Examples. Cambridge University Press.
- Lehmann, E. L., & Casella, G. (1998). Theory of Point Estimation. Springer.
- Resnick, S. (1999). A Probability Path. Birkhäuser.
- Pitman, J. (1993). Probability. Springer.
- David, H. A., & Nagaraja, H. N. (2003). Order Statistics. Wiley.
- Billingsley, P. (2013). Convergence of Probability Measures. Wiley.
- Dudley, R. M. (2002). Real Analysis and Probability. Cambridge University Press.
- Jaynes, E. T. (2003). Probability Theory: The Logic of Science. Cambridge University Press.
- Varadhan, S. R. S. (2001). Probability Theory. American Mathematical Society.
- Athreya, K. B., & Lahiri, S. N. (2006). Measure Theory and Probability Theory. Springer.

COURSE DESCRIPTION

Course Overview: This course provides a comprehensive exploration of probability theory and random variables. It covers fundamental concepts, key elements, and advanced topics, making it valuable for students, researchers, and professionals in mathematics, statistics, engineering, and data science.

Foundational Concepts: In the introductory section, we delve into the historical development of probability, explore different interpretations of probability, and understand the fundamental concepts of experiments and events. The foundation is built on these historical and conceptual pillars.

Understanding Sets and Events: We dive into set theory, where we explore set operations like intersection and union. You'll master algebraic manipulations on sets, understand the correct order of set operation execution, and grasp the principles of distributivity and De Morgan's Laws. These concepts are applied practically to a card selection game.

Kolmogorov Formalism: This section introduces Kolmogorov's formalism, which provides a precise definition and axioms of probability. You'll explore probability in discrete and continuous scenarios and extend your understanding to product spaces, encountered in complex probabilistic systems.

Key Elements: We delve into key elements of probability, including conditional probability, Bayes' Theorem, and independent events. The Gambler's Ruin Problem provides a real-world application of these concepts. You'll also gain insights into random variables, their laws, expectation, variance, and the law governing tuples of random variables.

Random Variables and Distributions: This section deepens your understanding of random variables, focusing on discrete distributions, covariance, correlation, and continuous distributions. You'll learn about cumulative distribution functions and the intricacies of bivariate distributions, marginal distributions, and conditional distributions.

Random Vectors: Explore the concept of random vectors, their definitions, and properties. Dive into joint, marginal, and conditional distributions, and understand how to compute expectation, variance, and covariance of random vectors. Linear transformations of random vectors and Gaussian random vectors are also covered.

Expectation and Variance: Learn about the expectation of a random variable, its properties, conditional expectation, and the distinctions between mean and median. Delve into variance, covariance, and moments of a real random variable, including variance and standard deviation.

Advanced Concepts and Simulations: This section delves into advanced topics, including the Law of Large Numbers, the Central Limit Theorem, and the Correction for Continuity. We explore the concept of simulation in probability, its applications, and various simulation techniques such as Importance Sampling, Markov Chain Monte Carlo, and the Bootstrap.

Advanced Theorems and Applications: Discover advanced theorems and applications, including generating functions, characteristic functions, Gaussian random variables, convergence of sequences of random variables, the Borel-Cantelli Lemma, inequalities, and different types of convergence.

This course equips you with a robust foundation in probability, advanced concepts, and simulation techniques, making it indispensable for those working with uncertainty and statistical analysis in various fields.