Comprehensive Notes on Statistics and Probability Theory

This document provides a detailed overview of fundamental concepts in Statistics and Probability Theory. It covers definitions, essential formulas, and illustrative examples for each topic.

1. Importance of Statistics and Probability Theory

Statistics and Probability Theory are indispensable disciplines that provide the mathematical framework for dealing with uncertainty and making sense of data. They are crucial for:

- **Data Analysis:** Transforming raw data into meaningful information and actionable insights.
- **Decision Making:** Providing a rigorous basis for making informed decisions under conditions of uncertainty in business, science, engineering, and policy.
- **Risk Assessment:** Quantifying and managing risks in finance, insurance, and project management.
- **Scientific Research:** Designing experiments, analyzing results, and drawing valid conclusions.
- **Predictive Modeling:** Building models to forecast future trends and outcomes (e.g., weather, stock prices, disease spread).
- **Quality Control:** Monitoring processes and ensuring product quality.

2. Samples vs. Populations

Understanding the distinction between a sample and a population is fundamental to statistical inference.

- **Population:**
  - **Definition:** The entire group of individuals, objects, or data points about which information is desired or conclusions are to be drawn. It represents the complete

set of possible observations.

- **Characteristics:** Often very large, potentially infinite, and usually impractical or impossible to observe in its entirety.
- **Parameters:** Numerical descriptors of a population. These are typically unknown fixed values that we try to estimate (e.g., population mean μ, population standard deviation σ, population proportion p).
- **Example:** All registered voters in a country; all manufactured light bulbs from a specific production line.

- **Sample:**
  - **Definition:** A subset of the population that is selected for observation and analysis.
  - **Characteristics:** A manageable, finite group from which data are collected.
  - **Statistics:** Numerical descriptors calculated from sample data. These are known values used to estimate unknown population parameters (e.g., sample mean $\bar{x}$, sample standard deviation s, sample proportion $\hat{p}$).
  - **Example:** 1,000 randomly selected registered voters; a batch of 100 light bulbs tested from the production line.
- **Statistical Inference:**
  - The process of using information obtained from a sample to make generalizations, predictions, or decisions about the larger population from which the sample was drawn.

3. Measures and Scales of Measurement

Scales of measurement describe the nature and properties of the numerical values assigned to variables. They determine the types of statistical analyses that are appropriate.

- **Nominal Scale:**
  - **Definition:** Categorical data where values are labels or names that categorize attributes without any inherent order or ranking.
  - **Properties:** Only distinguishes between categories.

- **Examples:** Gender (Male, Female, Non-binary), Eye Color (Blue, Brown, Green), Type of Car (Sedan, SUV, Truck).
- **Applicable Operations/Statistics:** Counting frequencies, percentages, mode.
- **No meaningful mathematical operations (addition, subtraction, etc.).**

- **Ordinal Scale:**
  - **Definition:** Categorical data where values have a meaningful order or ranking, but the intervals between ranks are not necessarily equal or meaningful.
  - **Properties:** Establishes a relative order.
  - **Examples:** Education Level (High School, Bachelor's, Master's, PhD), Likert Scale (Strongly Disagree, Disagree, Neutral, Agree, Strongly Agree), Military Ranks.
  - **Applicable Operations/Statistics:** All nominal operations, plus median, ranks.
  - **Intervals between values are not quantifiable.**
- **Interval Scale:**
  - **Definition:** Numerical data where values have a meaningful order, and the intervals between successive values are equal and meaningful. However, there is no true or absolute zero point (zero does not indicate the absence of the quantity).
  - **Properties:** Allows for addition and subtraction.
  - **Examples:** Temperature in Celsius ($0^\circ C$ does not mean absence of heat), Calendar Years (Year 0 is arbitrary, not the absence of time).
  - **Applicable Operations/Statistics:** All nominal and ordinal operations, plus mean, standard deviation, correlation.
  - **Ratio comparisons are not meaningful (e.g., $20^\circ C$ is not twice as hot as $10^\circ C$).**
- **Ratio Scale:**
  - **Definition:** Numerical data where values have a meaningful order, equal intervals between values, and a true or absolute zero point (zero indicates the complete absence of the quantity being measured).

- **Properties:** Allows for all mathematical operations, including multiplication and division.
- **Examples:** Height, Weight, Age, Income, Number of items sold, Time (duration).
- **Applicable Operations/Statistics:** All statistical operations are appropriate.
- **Ratio comparisons are meaningful (e.g., 20 kg is twice as heavy as 10 kg).**

4. Random Experiments and Probabilities

- **Random Experiment:**
  - **Definition:** A process or action with an uncertain outcome, but where all possible outcomes are known in advance. It can be repeated under identical conditions.
  - **Characteristics:**
    - All possible outcomes are identifiable.
    - The actual outcome cannot be predicted with certainty.
    - Can be repeated multiple times.
  - **Examples:** Flipping a coin, rolling a die, drawing a card from a well-shuffled deck, observing the gender of a newborn baby, measuring the lifespan of a light bulb.
- **Outcome:**
  - **Definition:** A single result of a random experiment.
  - **Example:** When rolling a 6-sided die, '3' is an outcome.
- **Sample Space (Ω or S):**
  - **Definition:** The set of all possible outcomes of a random experiment.
  - **Examples:**
    - Flipping a coin: $\Omega = Head, Tail$
    - Rolling a 6-sided die: $\Omega = 1, 2, 3, 4, 5, 6$
    - Flipping two coins: $\Omega = HH, HT, TH, TT$

- **Event:**
  - **Definition:** A subset of the sample space; a collection of one or more outcomes. An event occurs if any of its constituent outcomes occurs.
  - **Examples:**
    - Getting an even number on a die roll: $E = 2, 4, 6$
    - Getting at least one Head in two coin flips: $E = HT, TH, HH$
- **Probability:**
  - **Definition:** A numerical measure of the likelihood or chance that an event will occur. Probabilities are always between 0 and 1, inclusive.
  - **Axioms of Probability (Kolmogorov Axioms):**
    1. For any event A, $0 \leq P(A) \leq 1$.
    2. The probability of the sample space (the sure event) is 1: $P(\Omega) = 1$.
    3. If $A_1, A_2, \ldots$ are a sequence of mutually exclusive (disjoint) events, then:

$$P(A_1 \cup A_2 \cup \ldots) = P(A_1) + P(A_2) + \ldots$$

- **Basic Probability Rules:**
  - **Complement Rule:** The probability that an event A does not occur, denoted as A^c or A', is:

$$P(A^c) = 1 - P(A)$$

    Example: If $P(\text{rain}) = 0.3$, then $P(\text{no rain}) = 1 - 0.3 = 0.7$.

  - **Addition Rule for Any Two Events A and B:**

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

    Where $P(A \cap B)$ is the probability that both A and B occur.
    Example: If $P(\text{passing Math}) = 0.7$, $P(\text{passing Physics}) = 0.6$, and $P(\text{passing both}) = 0.5$, then $P(\text{passing at least one}) = 0.7 + 0.6 - 0.5 = 0.8$.

  - **Addition Rule for Mutually Exclusive Events:** If A and B are mutually exclusive (they cannot both occur, $A \cap B = \emptyset$), then $P(A \cap B) = 0$, so:

$$P(A \cup B) = P(A) + P(B)$$

Example: Rolling a die, $P(\text{even}) = 0.5$, $P(\text{odd}) = 0.5$. These are mutually exclusive. $P(\text{even or odd}) = 0.5 + 0.5 = 1$.

5. Types of Probability

Probabilities can be interpreted and determined in different ways:

- **Classical Probability (A Priori Probability):**
  - **Definition:** Based on formal reasoning and the assumption that all outcomes in the sample space are equally likely. It can be determined before any experiment is conducted.
  - **Formula:**

$$P(A) = \frac{\text{Number of favorable outcomes for A}}{\text{Total number of equally likely outcomes in the sample space}}$$

  - **Example:** The probability of rolling a '4' on a fair six-sided die is $\frac{1}{6}$, because there is 1 favorable outcome and 6 equally likely total outcomes.
  - **Limitations:** Only applicable when outcomes are truly equally likely.
- **Empirical Probability (A Posteriori Probability / Relative Frequency):**
  - **Definition:** Based on actual observations from repeatedly performing an experiment. It is the ratio of the number of times an event occurs to the total number of trials.
  - **Formula:**

$$P(A) = \frac{\text{Number of times event A occurred}}{\text{Total number of trials}}$$

  - **Example:** If a biased coin is flipped 100 times and lands on heads 60 times, the empirical probability of heads is $\frac{60}{100} = 0.6$. As the number of trials increases, the empirical probability tends to approach the true probability (Law of Large Numbers).
  - **Limitations:** Depends on the number of trials and can vary with different sets of trials.

- **Subjective Probability:**
  - **Definition:** Based on personal judgment, intuition, experience, or expertise, especially when objective data is scarce or impossible to collect.
  - **Example:** A doctor's assessment of the probability that a patient will recover from a rare disease; a sports analyst's prediction of a team winning a championship.
  - **Limitations:** Highly personal and can vary significantly between individuals. Not based on mathematical calculation or observed frequencies.

6. Conditional Probability

- **Definition:** The probability of an event occurring, given that another event has already occurred or is known to have occurred. It reflects how the occurrence of one event modifies the likelihood of another.
- **Notation:** $P(A|B)$ is read as "the probability of event A given event B."
- **Formula:**

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \quad \text{provided } P(B) > 0$$

  Where:
  - $P(A \cap B)$ is the joint probability that both event A and event B occur.
  - $P(B)$ is the marginal probability of event B.
- **Multiplication Rule (derived from conditional probability):**

$$P(A \cap B) = P(A|B) \cdot P(B)$$

  or equivalently:

$$P(A \cap B) = P(B|A) \cdot P(A)$$

- **Example:**
  - Consider a deck of 52 playing cards.
  - Let A be the event "drawing a King."

- Let B be the event “drawing a Face Card (King, Queen, Jack).”
- $P(A) = \frac{4}{52} = \frac{1}{13}$
- $P(B) = \frac{12}{52} = \frac{3}{13}$
- $P(A \cap B)$: The event of drawing a King AND a Face Card is just drawing a King, so $P(A \cap B) = P(A) = \frac{4}{52}$.
- Now, $P(A|B)$ = Probability of drawing a King GIVEN that you drew a Face Card.

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{4/52}{12/52} = \frac{4}{12} = \frac{1}{3}$$

This makes sense: if you know you have a face card, there are 12 face cards, and 4 of them are Kings.

7. Conditional Independence

- **Definition:** Two events, A and B, are **conditionally independent** given a third event C if the occurrence of A does not affect the probability of B once the state of C is known. This means that knowing A provides no additional information about B if C is already known.
- **Formula:**

$$P(A \cap B|C) = P(A|C) \cdot P(B|C)$$

This is equivalent to stating that $P(A|B \cap C) = P(A|C)$ or $P(B|A \cap C) = P(B|C)$.

- **Distinction from (Unconditional) Independence:**
  - **Independent Events:** $P(A \cap B) = P(A) \cdot P(B)$. This means knowing A tells you nothing about B, and vice versa, in the overall sample space.
  - **Conditionally Independent Events:** A and B are independent *within the context* of C. They might not be unconditionally independent. In fact, two events can be conditionally independent but not independent, or independent but not conditionally independent.
- **Example:**

- Let A be the event "a person has a cough."
- Let B be the event "a person has a fever."
- Let C be the event "a person has the flu."
- Without knowing anything, $P(A)$ and $P(B)$ might not be independent (e.g., if you have a cough, you're more likely to have a fever due to common illnesses). So, $P(A \cap B) \neq P(A)P(B)$.
- However, *given that* a person has the flu (C), the presence of a cough might not provide additional information about the presence of a fever beyond what the flu itself suggests. In this context, A and B might be conditionally independent given C.
- So, $P(\text{cough} \cap \text{fever} \mid \text{flu}) = P(\text{cough} \mid \text{flu}) \cdot P(\text{fever} \mid \text{flu})$.

8. Expectation and Variance (General Concepts)

These are two of the most important descriptive statistics for random variables, providing insights into their central tendency and dispersion. Specific formulas for common distributions are provided in Section 13.

Expected Value (Mean), $E[X]$ or μ

- **Definition:** The long-run average value of a random variable if the experiment were to be repeated infinitely many times. It represents the "center" of the probability distribution.
- **Formulas:**
  - **For a Discrete Random Variable X (with possible values x_i and probabilities $P(X = x_i)$):**

    $$E[X] = \sum_i x_i P(X = x_i)$$

    Example (Discrete): If you roll a fair six-sided die, X is the outcome. $E[X] = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \frac{21}{6} = 3.5$
  - **For a Continuous Random Variable X (with probability density function $f(x)$):**

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

Example (Continuous): If X follows an exponential distribution with rate λ, $f(x) = \lambda e^{-\lambda x}$ for $x \geq 0$.
$E[X] = \int_0^\infty x(\lambda e^{-\lambda x}) dx = \frac{1}{\lambda}$

- **Properties of Expectation:**
  - $E[c] = c$ (where c is a constant)
  - $E[aX + b] = aE[X] + b$ (where a, b are constants)
  - $E[X + Y] = E[X] + E[Y]$ (for any random variables X, Y, regardless of independence)
  - $E[XY] = E[X]E[Y]$ (IF X and Y are independent)

Variance, $Var[X]$ or σ^2

- **Definition:** A measure of how spread out the values of a random variable are from its expected value. A higher variance indicates greater variability.
- **Formulas:**
  - **Definition Formula:**

$$Var[X] = E[(X - E[X])^2]$$

  - **Computational Formula (more commonly used):**

$$Var[X] = E[X^2] - (E[X])^2$$

To use this, you first need to calculate $E[X^2]$:

    - **For a Discrete Random Variable X:**

$$E[X^2] = \sum_i x_i^2 P(X = x_i)$$

    - **For a Continuous Random Variable X:**

$$E[X^2] = \int_{-\infty}^{\infty} x^2 f(x) dx$$

Example (Discrete): Rolling a fair six-sided die. We found $E[X] = 3.5$.
$E[X^2] = 1^2 \cdot \frac{1}{6} + 2^2 \cdot \frac{1}{6} + 3^2 \cdot \frac{1}{6} + 4^2 \cdot \frac{1}{6} + 5^2 \cdot \frac{1}{6} + 6^2 \cdot \frac{1}{6}$
$E[X^2] = \frac{1+4+9+16+25+36}{6} = \frac{91}{6}$
$Var[X] = E[X^2] - (E[X])^2 = \frac{91}{6} - (3.5)^2 = \frac{91}{6} - \frac{49}{4} = \frac{182-147}{12} = \frac{35}{12} \approx 2.9167$
Example (Continuous): For an exponential distribution with rate λ:
$E[X^2] = \int_0^\infty x^2(\lambda e^{-\lambda x})dx = \frac{2}{\lambda^2}$
$Var[X] = E[X^2] - (E[X])^2 = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$

- **Properties of Variance:**
  - $Var[c] = 0$ (where c is a constant, as there's no variability)
  - $Var[aX + b] = a^2 Var[X]$ (where a, b are constants)
  - For **independent** random variables X and Y:
    - $Var[X + Y] = Var[X] + Var[Y]$
    - $Var[X - Y] = Var[X] + Var[Y]$
  - For **dependent** random variables X and Y:
    - $Var[X + Y] = Var[X] + Var[Y] + 2Cov(X, Y)$
    - $Var[X - Y] = Var[X] + Var[Y] - 2Cov(X, Y)$
      Where $Cov(X, Y) = E[XY] - E[X]E[Y]$ is the covariance.

Standard Deviation, σ

- **Definition:** The positive square root of the variance. It is expressed in the same units as the random variable itself, making it more interpretable than variance.
- **Formula:**

$$\sigma = \sqrt{Var[X]}$$

Example: For the die roll, $\sigma = \sqrt{\frac{35}{12}} \approx 1.7078$.

9. Independent Random Variables

- **Definition:** Two random variables, X and Y, are **independent** if the probability distribution of one variable is not affected by the values taken by the other variable. In other words, knowing the value of one variable gives no information about the value of the other.

- **Formal Conditions for Independence:**

  - **For Discrete Random Variables:** X and Y are independent if and only if their joint probability mass function (PMF) is the product of their marginal PMFs for all possible values of x and y:

$$P(X = x, Y = y) = P(X = x) \cdot P(Y = y) \quad \text{for all } x, y$$

  - **For Continuous Random Variables:** X and Y are independent if and only if their joint probability density function (PDF) is the product of their marginal PDFs for all possible values of x and y:

$$f(x, y) = f_X(x) \cdot f_Y(y) \quad \text{for all } x, y$$

- **Key Implications/Properties of Independent Random Variables:**

  - **Expectation of Product:** $E[XY] = E[X]E[Y]$

  - **Covariance:** The covariance between independent random variables is zero: $Cov(X, Y) = 0$. (Note: Zero covariance does not necessarily imply independence, but independence *always* implies zero covariance).

  - **Variance of Sum/Difference:**

    - $Var[X + Y] = Var[X] + Var[Y]$

    - $Var[X - Y] = Var[X] + Var[Y]$

  - **Conditional Distributions:** The conditional probability distribution of one variable given the other is equal to its marginal probability distribution:

    - $P(Y = y|X = x) = P(Y = y)$ (discrete)

    - $f_{Y|X}(y|x) = f_Y(y)$ (continuous)

- **Example:**

  - Let X be the outcome of rolling a fair six-sided die.

  - Let Y be the outcome of flipping a fair coin (say, 0 for Tails, 1 for Heads).

- These two experiments are independent. The outcome of the die roll does not influence the coin flip, and vice-versa.
- $P(X = 3, Y = 1) = P(X = 3) \cdot P(Y = 1) = \frac{1}{6} \cdot \frac{1}{2} = \frac{1}{12}$.
- $E[X] = 3.5$, $E[Y] = 0.5$. Then $E[XY] = E[X]E[Y] = 3.5 \cdot 0.5 = 1.75$.

10. Functions of Random Variables

- **Definition:** If X is a random variable, and $g(X)$ is a mathematical function of X, then $Y = g(X)$ is also a random variable. The value of Y depends on the outcome of X.
- **Expected Value of a Function of a Random Variable:**
  - To find the expected value of $Y = g(X)$, you don't necessarily need to find the probability distribution of Y first.
  - **For a Discrete Random Variable X (with possible values x_i and probabilities $P(X = x_i)$):**

    $$E[g(X)] = \sum_i g(x_i)P(X = x_i)$$

    Example (Discrete): Let X be the outcome of a fair die roll. Let $g(X) = X^2$. $E[X^2] = 1^2 \cdot \frac{1}{6} + 2^2 \cdot \frac{1}{6} + 3^2 \cdot \frac{1}{6} + 4^2 \cdot \frac{1}{6} + 5^2 \cdot \frac{1}{6} + 6^2 \cdot \frac{1}{6} = \frac{91}{6}$.
  - **For a Continuous Random Variable X (with PDF $f(x)$):**

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

    Example (Continuous): Let X be the radius of a circle, following a uniform distribution between 0 and 1. $f(x) = 1$ for $0 \le x \le 1$, and 0 otherwise. Let $g(X) = \pi X^2$ be the area of the circle. $E[\text{Area}] = E[\pi X^2] = \int_0^1 \pi x^2 \cdot 1 \, dx = \pi \left[\frac{x^3}{3}\right]_0^1 = \pi \left(\frac{1^3}{3} - \frac{0^3}{3}\right) = \frac{\pi}{3}$.
- **Probability Distribution of a Function of a Random Variable:**
  - Finding the probability mass function (PMF) for a discrete $Y = g(X)$ or the probability density function (PDF) for a continuous $Y = g(X)$ can be more involved.
  - **CDF Method (General):**

1. Find the cumulative distribution function (CDF) of Y: $F_Y(y) = P(Y \le y) = P(g(X) \le y)$.
2. Express the inequality $g(X) \le y$ in terms of X.
3. Use the CDF (or PMF/PDF) of X to calculate this probability.
4. If Y is continuous, differentiate $F_Y(y)$ with respect to y to get $f_Y(y)$.

- **Transformation Method (for continuous, invertible functions):**
  - If $Y = g(X)$ is a continuous, strictly monotonic (increasing or decreasing) function, and $X = h(Y)$ is its inverse, then:

$$f_Y(y) = f_X(h(y)) \left| \frac{dh(y)}{dy} \right|$$

    Where $\left| \frac{dh(y)}{dy} \right|$ is the absolute value of the Jacobian of the transformation.

- **Example (CDF method for continuous):** Let X be exponentially distributed with $\lambda = 1$, so $f_X(x) = e^{-x}$ for $x \ge 0$. Let $Y = X^2$.
  $F_Y(y) = P(Y \le y) = P(X^2 \le y)$. Since $X \ge 0$, this means $P(X \le \sqrt{y})$.
  $F_Y(y) = \int_0^{\sqrt{y}} e^{-x} dx = [-e^{-x}]_0^{\sqrt{y}} = -e^{-\sqrt{y}} - (-e^0) = 1 - e^{-\sqrt{y}}$ for $y \ge 0$.
  Then, $f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy}(1 - e^{-\sqrt{y}}) = -e^{-\sqrt{y}} \cdot \left(-\frac{1}{2\sqrt{y}}\right) = \frac{e^{-\sqrt{y}}}{2\sqrt{y}}$ for $y \ge 0$.

11. Probability Mass Function (PMF) and Cumulative Distribution Function (CDF)

These functions are crucial for describing the probability distribution of a random variable.

Probability Mass Function (PMF) - For Discrete Random Variables

- **Definition:** A function that gives the probability that a discrete random variable is exactly equal to some value.
- **Notation:** $P(X = x)$ or $f_X(x)$.
- **Properties:**

1. $0 \leq P(X = x) \leq 1$ for all possible values of x.
2. $\sum_x P(X = x) = 1$ (the sum of all probabilities for all possible values must equal 1).

- **Example:** For a fair six-sided die roll, X:

$$P(X = x) = \frac{1}{6} \quad \text{for } x \in 1, 2, 3, 4, 5, 6$$

$$P(X = x) = 0 \quad \text{otherwise}$$

Probability Density Function (PDF) - For Continuous Random Variables (Revisited)

- **Definition:** A function whose value at any given sample (or point) in the sample space can be thought of as providing a relative likelihood that the value of the random variable would equal that sample. Unlike PMF, the value of the PDF at a given point is not a probability itself.
- **Notation:** $f(x)$.
- **Properties:**
  1. $f(x) \geq 0$ for all x.
  2. $\int_{-\infty}^{\infty} f(x)dx = 1$ (the total area under the curve must equal 1).
  3. $P(a \leq X \leq b) = \int_a^b f(x)dx$.
- **Example:** For a Uniform distribution on $[0, 10]$:

$$f(x) = \frac{1}{10} \quad \text{for } 0 \leq x \leq 10$$

$$f(x) = 0 \quad \text{otherwise}$$

Cumulative Distribution Function (CDF) - For Both Discrete and Continuous Random Variables

- **Definition:** A function that gives the probability that a random variable X will take a value less than or equal to x. It provides the cumulative probability up to a certain point.
- **Notation:** $F(x)$ or $F_X(x)$.

- **Formulas:**
    - **For a Discrete Random Variable X:**

$$F(x) = P(X \le x) = \sum_{t \le x} P(X = t)$$

Example (Discrete - Die Roll):
$F(3) = P(X \le 3) = P(X = 1) + P(X = 2) + P(X = 3) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{3}{6} = 0.5$

    - **For a Continuous Random Variable X:**

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)dt$$

To find the PDF from the CDF for a continuous random variable, you differentiate the CDF: $f(x) = \frac{d}{dx}F(x)$.
Example (Continuous - Exponential with λ):
$F(x) = 1 - e^{-\lambda x}$ for $x \ge 0$.
$F(5) = P(X \le 5) = 1 - e^{-5\lambda}$.

- **Properties of CDF:**

    1. $0 \le F(x) \le 1$ for all x.
    2. $\lim_{x \to -\infty} F(x) = 0$.
    3. $\lim_{x \to \infty} F(x) = 1$.
    4. $F(x)$ is non-decreasing: if $x_1 \le x_2$, then $F(x_1) \le F(x_2)$.
    5. For continuous random variables, $F(x)$ is continuous. For discrete random variables, $F(x)$ is a step function.
    6. $P(a < X \le b) = F(b) - F(a)$.

12. Common Discrete Probability Distributions

These distributions model the probabilities of outcomes for discrete random variables, often arising from counting events.

12.1. Bernoulli Distribution

- **Definition:** Models a single trial of a random experiment with only two possible outcomes: “success” or “failure.”
- **Parameters:**
    - p: The probability of success ($0 \leq p \leq 1$).
- **Random Variable:** $X = 1$ for success, $X = 0$ for failure.
- **Probability Mass Function (PMF):**

$$P(X = x) = p^x (1 - p)^{1-x} \quad \text{for } x \in 0, 1$$

- **Expected Value ($E[X]$):**

$$E[X] = p$$

- **Variance ($Var[X]$):**

$$Var[X] = p(1 - p)$$

- **Example:** Flipping a coin once where getting Heads is a success. If $p = 0.5$ (fair coin), then $P(X = 1) = 0.5$ and $P(X = 0) = 0.5$. $E[X] = 0.5$, $Var[X] = 0.5(0.5) = 0.25$.

12.2. Binomial Distribution

- **Definition:** Models the number of successes in a fixed number of independent Bernoulli trials.
- **Parameters:**
    - n: Number of trials (fixed integer ≥ 0).
    - p: Probability of success on each trial ($0 \leq p \leq 1$).
- **Random Variable:** X: Number of successes in n trials. $X \in 0, 1, \ldots, n$.
- **Probability Mass Function (PMF):**

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k} \quad \text{for } k \in 0, 1, \ldots, n$$

Where $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ is the binomial coefficient, representing the number of ways to choose k successes from n trials.

- **Expected Value ($E[X]$):**

$$E[X] = np$$

- **Variance ($Var[X]$):**

$$Var[X] = np(1-p)$$

- **Example:** Rolling a fair die 10 times, let X be the number of times a '6' appears. This is Binomial($n = 10, p = 1/6$).
  $P(X = 2) = \binom{10}{2}(\frac{1}{6})^2(\frac{5}{6})^8 \approx 0.2907$.
  $E[X] = 10 \cdot \frac{1}{6} = \frac{10}{6} \approx 1.67$.
  $Var[X] = 10 \cdot \frac{1}{6} \cdot \frac{5}{6} = \frac{50}{36} \approx 1.39$.

12.3. Geometric Distribution

- **Definition:** Models the number of Bernoulli trials needed to get the *first* success. The trials are independent.

- **Parameters:**

  - p: Probability of success on each trial ($0 < p \leq 1$).

- **Random Variable:** X: Number of trials until the first success. $X \in 1, 2, 3, \ldots$.

- **Probability Mass Function (PMF):**

$$P(X = k) = (1-p)^{k-1}p \quad \text{for } k \in 1, 2, 3, \ldots$$

- **Expected Value ($E[X]$):**

$$E[X] = \frac{1}{p}$$

- **Variance ($Var[X]$):**

$$Var[X] = \frac{1-p}{p^2}$$

- **Example:** Flipping a coin until the first Head appears. If $p = 0.5$.
  $P(X = 3)$ (first Head on 3rd flip, i.e., TT H) $= (0.5)^{3-1} \cdot 0.5 = 0.5^3 = 0.125$.
  $E[X] = \frac{1}{0.5} = 2$ flips.

12.4. Negative Binomial (Pascal) Distribution

- **Definition:** Models the number of Bernoulli trials needed to get the r^{th} success. (Generalization of Geometric Distribution, where $r = 1$).
- **Parameters:**
    - r: Number of successes desired (fixed positive integer).
    - p: Probability of success on each trial ($0 < p \leq 1$).
- **Random Variable:** X: Number of trials until the r^{th} success. $X \in r, r+1, r+2, \ldots$.
- **Probability Mass Function (PMF):**

$$P(X = k) = \binom{k-1}{r-1} p^r (1-p)^{k-r} \quad \text{for } k \in r, r+1, r+2, \ldots$$

- **Expected Value ($E[X]$):**

$$E[X] = \frac{r}{p}$$

- **Variance ($Var[X]$):**

$$Var[X] = \frac{r(1-p)}{p^2}$$

- **Example:** Rolling a die until you get the 3rd '6'. Here $r = 3, p = 1/6$. $P(X = 10)$ (it takes 10 rolls to get the 3rd '6') $= \binom{10-1}{3-1}(\frac{1}{6})^3(\frac{5}{6})^{10-3} = \binom{9}{2}(\frac{1}{6})^3(\frac{5}{6})^7 \approx 0.0465$.

12.5. Hypergeometric Distribution

- **Definition:** Models the number of successes in a sample drawn *without replacement* from a finite population where there are a fixed number of successes and failures. Unlike Binomial, trials are *not* independent.
- **Parameters:**
    - N: Total number of items in the population.
    - K: Total number of successes in the population.
    - n: Number of items drawn in the sample.

- **Random Variable:** X: Number of successes in the sample.
  The possible values for X are $\max(0, n - (N - K)) \leq x \leq \min(n, K)$.

- **Probability Mass Function (PMF):**

$$P(X = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$$

- **Expected Value ($E[X]$):**

$$E[X] = n\frac{K}{N}$$

- **Variance ($Var[X]$):**

$$Var[X] = n\frac{K}{N}\frac{N-K}{N}\frac{N-n}{N-1}$$

- **Example:** An urn contains 20 balls: 12 red (successes) and 8 blue (failures). You draw 5 balls without replacement. $N = 20, K = 12, n = 5$.
  $P(X = 3)$ (probability of drawing exactly 3 red balls) $= \frac{\binom{12}{3}\binom{8}{5-3}}{\binom{20}{5}} = \frac{\binom{12}{3}\binom{8}{2}}{\binom{20}{5}} = \frac{220 \cdot 28}{15504} \approx 0.397$.

12.6. Poisson Distribution

- **Definition:** Models the number of events occurring in a fixed interval of time or space, given that these events occur with a known constant mean rate and independently of the time since the last event. Often used for rare events.

- **Parameters:**

  - λ: The average number of events in the given interval (rate parameter). $\lambda > 0$.

- **Random Variable:** X: Number of events in the interval. $X \in 0, 1, 2, \ldots$.

- **Probability Mass Function (PMF):**

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!} \quad \text{for } k \in 0, 1, 2, \ldots$$

  Where e is Euler's number (approx. 2.71828).

- **Expected Value ($E[X]$):**

$$E[X] = \lambda$$

- **Variance ($Var[X]$):**

$$Var[X] = \lambda$$

- **Example:** A call center receives an average of 5 calls per hour. $\lambda = 5$.
  $P(X = 3)$ (probability of receiving exactly 3 calls in an hour) $= \frac{e^{-5}5^3}{3!} = \frac{0.006738 \cdot 125}{6} \approx 0.1404$.
  $E[X] = 5, Var[X] = 5$.

13. Poisson as an Approximation for Binomial

- **Concept:** The Poisson distribution can be used to approximate the Binomial distribution under certain conditions. This approximation is useful when direct calculation with the Binomial PMF becomes computationally intensive due to large n.

- **Conditions for Approximation:** The approximation is generally good when:

  1. The number of trials (n) is large (e.g., $n \geq 20$ or $n \geq 50$).

  2. The probability of success (p) is small (e.g., $p \leq 0.05$ or $p \leq 0.1$).

  3. The product np (which is the mean of the Binomial distribution) is relatively small (e.g., $np < 10$).

- **Approximation Rule:**

  - If $X \sim \text{Binomial}(n, p)$, then X can be approximated by a Poisson distribution with parameter $\lambda = np$.

  - So, $P(X = k)\text{Binomial} \approx P(X = k)\text{Poisson} = \frac{e^{-np}(np)^k}{k!}$

- **Why it works:** When n is large and p is small, the Binomial distribution's trials are almost independent, and the probability of multiple successes in a very small interval becomes negligible, mimicking the assumptions of a Poisson process.

- **Example:** Suppose a manufacturing process has a defect rate of $p = 0.01$. In a batch of $n = 500$ items, what is the probability of exactly 3 defects?

  - Binomial: $X \sim \text{Binomial}(500, 0.01)$. $P(X = 3) = \binom{500}{3}(0.01)^3(0.99)^{497}$. This is tedious.

- Poisson Approximation: Here $n = 500$ (large), $p = 0.01$ (small).
  Calculate $\lambda = np = 500 \cdot 0.01 = 5$.
  Approximate using Poisson($\lambda = 5$): $P(X = 3) \approx \frac{e^{-5}5^3}{3!} \approx 0.1404$.
  (The exact binomial probability is approximately 0.1399, showing a good approximation.)

14. Calculus Fundamentals for Continuous Random Variables

Working with continuous random variables heavily relies on calculus concepts.

14.1. Derivatives

- **Definition:** The derivative of a function measures the instantaneous rate at which the function's value changes with respect to a change in its independent variable. Geometrically, it's the slope of the tangent line to the function's graph at a given point.
- **Notation:** If $y = f(x)$, the derivative is denoted as $f'(x)$, $\frac{dy}{dx}$, or $\frac{d}{dx}f(x)$.
- **Purpose in Probability:**
  - Used to find the Probability Density Function (PDF) from the Cumulative Distribution Function (CDF) for continuous random variables: $f(x) = \frac{d}{dx}F(x)$.
- **Basic Formulas (Examples):**
  - **Power Rule:** $\frac{d}{dx}(x^n) = nx^{n-1}$
    - Example: $\frac{d}{dx}(x^3) = 3x^2$
  - **Constant Rule:** $\frac{d}{dx}(c) = 0$
    - Example: $\frac{d}{dx}(5) = 0$
  - **Constant Multiple Rule:** $\frac{d}{dx}(cf(x)) = c\frac{d}{dx}f(x)$
    - Example: $\frac{d}{dx}(3x^2) = 3 \cdot 2x = 6x$
  - **Sum/Difference Rule:** $\frac{d}{dx}(f(x) \pm g(x)) = \frac{d}{dx}f(x) \pm \frac{d}{dx}g(x)$
    - Example: $\frac{d}{dx}(x^2 + 2x) = 2x + 2$

- **Chain Rule:** $\frac{d}{dx}(f(g(x))) = f'(g(x)) \cdot g'(x)$
  - Example: $\frac{d}{dx}(e^{2x}) = e^{2x} \cdot 2 = 2e^{2x}$
- **Exponential Rule:** $\frac{d}{dx}(e^{ax}) = ae^{ax}$
  - Example: $\frac{d}{dx}(e^{-3x}) = -3e^{-3x}$

14.2. Integrals (Definite and Indefinite)

- **Definition:**
  - **Indefinite Integral (Antiderivative):** The reverse process of differentiation. If $F'(x) = f(x)$, then $\int f(x)dx = F(x) + C$, where C is the constant of integration.
  - **Definite Integral:** Represents the accumulation of a quantity, typically the area under the curve of a function between two specified limits.
- **Notation:**
  - Indefinite: $\int f(x)dx$
  - Definite: $\int_a^b f(x)dx$
- **Purpose in Probability:**
  - Used to find probabilities for continuous random variables: $P(a \leq X \leq b) = \int_a^b f(x)dx$.
  - Used to find the Cumulative Distribution Function (CDF) from the PDF: $F(x) = \int_{-\infty}^x f(t)dt$.
  - Used to calculate expected values and variances for continuous random variables (as shown in Section 8).
  - Used to ensure the total probability for a PDF is 1: $\int_{-\infty}^{\infty} f(x)dx = 1$.
- **Basic Formulas (Examples):**
  - **Power Rule:** $\int x^n dx = \frac{x^{n+1}}{n+1} + C \quad (n \neq -1)$
    - Example: $\int x^2 dx = \frac{x^3}{3} + C$
  - **Logarithmic Rule:** $\int \frac{1}{x} dx = \ln|x| + C$

- **Exponential Rule:** $\int e^{ax}dx = \frac{1}{a}e^{ax} + C$
  - Example: $\int e^{-2x}dx = -\frac{1}{2}e^{-2x} + C$
- **Definite Integral (Fundamental Theorem of Calculus):**

$$\int_a^b f(x)dx = F(b) - F(a)$$

  Where $F(x)$ is any antiderivative of $f(x)$.

  - Example: $\int_0^1 xdx = \left[\frac{x^2}{2}\right]_0^1 = \frac{1^2}{2} - \frac{0^2}{2} = \frac{1}{2}$

14.3. Antiderivatives (Indefinite Integrals)

- **Definition:** An antiderivative of a function $f(x)$ is a function $F(x)$ whose derivative is $f(x)$.
- **Relationship to Integrals:** The term “indefinite integral” is synonymous with “antiderivative.”
- **Importance:** They are the core component for evaluating definite integrals, which are central to probability calculations for continuous distributions.

15. General Properties of Continuous Random Variables

Continuous random variables differ from discrete ones primarily in how probabilities are assigned and calculated.

- **Nature of Values:** A continuous random variable can take on any value within a given interval or set of intervals (e.g., time, height, temperature).
- **Probability at a Single Point:** For any continuous random variable X, the probability of X taking on any *single specific value* is zero.

$$P(X = x) = 0 \quad \text{for any specific } x$$

  This is because there are infinitely many possible values.

- **Probabilities as Areas:** Probabilities for continuous random variables are calculated over intervals, not at specific points. The probability that X falls within an interval $[a, b]$ is the area under its PDF curve from a to b.

$$P(a \leq X \leq b) = P(a < X < b) = P(a \leq X < b) = P(a < X \leq b) = \int_a^b f(x)dx$$

- **Non-negative PDF:** The probability density function $f(x)$ must be non-negative for all x: $f(x) \geq 0$.
- **Total Area is 1:** The total area under the entire PDF curve must be equal to 1: $\int_{-\infty}^{\infty} f(x)dx = 1$.
- **CDF is Continuous:** The Cumulative Distribution Function (CDF) of a continuous random variable is always a continuous function.

16. Common Continuous Probability Distributions

These distributions model the probabilities of outcomes for continuous random variables, often arising from measurements.

16.1. The Uniform Distribution

- **Definition:** Describes a random variable where all outcomes within a given interval $[a, b]$ are equally likely. The probability density is constant over this interval.
- **Parameters:**
  - a: Lower bound of the interval.
  - b: Upper bound of the interval ($a < b$).
- **Random Variable:** X: Any real number in $[a, b]$.
- **Probability Density Function (PDF):**

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

$$f(x) = 0 \quad \text{otherwise}$$

- **Cumulative Distribution Function (CDF):**
  - $F(x) = 0 \quad \text{for } x < a$
  - $F(x) = \frac{x-a}{b-a} \quad \text{for } a \leq x \leq b$

- $F(x) = 1 \quad \text{for } x > b$

- **Expected Value ($E[X]$):**

$$E[X] = \frac{a+b}{2}$$

- **Variance ($Var[X]$):**

$$Var[X] = \frac{(b-a)^2}{12}$$

- **Example:** A bus arrives at a stop every 10 minutes. The waiting time X (in minutes) for a passenger is uniformly distributed between 0 and 10 minutes.
  $a = 0, b = 10.$
  $f(x) = \frac{1}{10-0} = \frac{1}{10}$ for $0 \le x \le 10.$
  $P(X \le 3) = \int_0^3 \frac{1}{10} dx = \frac{3}{10} = 0.3.$
  $E[X] = \frac{0+10}{2} = 5$ minutes.
  $Var[X] = \frac{(10-0)^2}{12} = \frac{100}{12} = \frac{25}{3} \approx 8.33.$

16.2. The Normal Distribution (Gaussian Distribution)

- **Definition:** A symmetric, bell-shaped distribution that is arguably the most important distribution in statistics. Many natural phenomena follow this distribution, and it's central to the Central Limit Theorem.

- **Parameters:**

  - μ: Mean (location parameter), where the peak of the curve is.

  - σ: Standard deviation (scale parameter), which determines the spread of the curve. $\sigma > 0$.

- **Notation:** $X \sim N(\mu, \sigma^2)$ (where σ^2 is the variance).

- **Random Variable:** X: Any real number ($-\infty < x < \infty$).

- **Probability Density Function (PDF):**

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad \text{for } -\infty < x < \infty$$

- **Cumulative Distribution Function (CDF):**

  - $F(x) = P(X \le x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt$

- This integral does not have a closed-form solution and must be evaluated using numerical methods or standard normal tables.

- **Standard Normal Distribution:** A special case where $\mu = 0$ and $\sigma = 1$. The random variable is denoted by Z.

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

Any normal random variable X can be standardized to Z using the formula:

$$Z = \frac{X - \mu}{\sigma}$$

- **Expected Value ($E[X]$):**

$$E[X] = \mu$$

- **Variance ($Var[X]$):**

$$Var[X] = \sigma^2$$

- **Example:** The heights of adult males in a population are normally distributed with a mean of 175 cm and a standard deviation of 7 cm. $X \sim N(175, 7^2)$.

  - To find $P(X \leq 180 \text{ cm})$:
    First, standardize X: $Z = \frac{180-175}{7} = \frac{5}{7} \approx 0.71$.
    Then, look up $P(Z \leq 0.71)$ in a standard normal table or use software. This value is approximately 0.7611.

  - Approximately 68% of values fall within 1 standard deviation of the mean ($\mu \pm \sigma$).

  - Approximately 95% of values fall within 2 standard deviations of the mean ($\mu \pm 2\sigma$).

  - Approximately 99.7% of values fall within 3 standard deviations of the mean ($\mu \pm 3\sigma$).

16.3. The Exponential Distribution

- **Definition:** Models the time until an event occurs in a Poisson process, where events occur continuously and independently at a constant average rate. It is memoryless.

- **Parameters:**

- λ: Rate parameter (average number of events per unit of time/space). $\lambda > 0$.
- Alternatively, $\beta = 1/\lambda$ can be used as the scale parameter (mean time between events).

- **Notation:** $X \sim \text{Exp}(\lambda)$
- **Random Variable:** X: Any non-negative real number ($x \geq 0$).
- **Probability Density Function (PDF):**

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

$$f(x) = 0 \quad \text{otherwise}$$

- **Cumulative Distribution Function (CDF):**

$$F(x) = P(X \leq x) = 1 - e^{-\lambda x} \quad \text{for } x \geq 0$$

$$F(x) = 0 \quad \text{otherwise}$$

From this, $P(X > x) = e^{-\lambda x}$.

- **Expected Value ($E[X]$):**

$$E[X] = \frac{1}{p}$$

- **Variance ($Var[X]$):**

$$Var[X] = \frac{1}{\lambda^2}$$

- **Memoryless Property:** This is a crucial property of the exponential distribution. It states that the probability of an event occurring in the future is independent of how much time has already passed without the event occurring.

$$P(X > t + s \mid X > t) = P(X > s)$$

- **Example:** The lifespan of a certain electronic component follows an exponential distribution with a mean lifespan of 100 hours. So, $\lambda = 1/100 = 0.01$ failures per hour.
  - $P(X > 200)$ (probability it lasts more than 200 hours) $= e^{-0.01 \cdot 200} = e^{-2} \approx 0.1353$.
  - $E[X] = 1/0.01 = 100$ hours.

- $Var[X] = (1/0.01)^2 = 100^2 = 10000$.

17. Joint Probability Distributions

When dealing with more than one random variable, we use joint probability distributions to describe their combined behavior.

17.1. Joint Probability Mass Function (Joint PMF) - For Discrete Random Variables

- **Definition:** For two discrete random variables X and Y, the joint PMF gives the probability that X takes on a specific value x AND Y takes on a specific value y.
- **Notation:** $P(X = x, Y = y)$ or $f_{X,Y}(x, y)$.
- **Properties:**
  1. $0 \leq P(X = x, Y = y) \leq 1$ for all possible pairs (x, y).
  2. $\sum_x \sum_y P(X = x, Y = y) = 1$ (the sum of all joint probabilities must equal 1).
- **Example:** Consider rolling two dice, one red (X) and one blue (Y).

  $$P(X = x, Y = y) = \frac{1}{36} \quad \text{for } x, y \in 1, 2, 3, 4, 5, 6$$

  The probability of rolling a 3 on the red die and a 4 on the blue die is $P(X = 3, Y = 4) = \frac{1}{36}$.

17.2. Joint Probability Density Function (Joint PDF) - For Continuous Random Variables

- **Definition:** For two continuous random variables X and Y, the joint PDF describes the relative likelihood that X takes on a value near x AND Y takes on a value near y.
- **Notation:** $f_{X,Y}(x, y)$.
- **Properties:**
  1. $f_{X,Y}(x, y) \geq 0$ for all x, y.

2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dy\, dx = 1$ (the total volume under the surface must equal 1).

3. $P(a \leq X \leq b, c \leq Y \leq d) = \int_c^d \int_a^b f_{X,Y}(x, y)\, dx\, dy$.

- **Example:** A joint uniform distribution over a square region $[0, 1] \times [0, 1]$:

$$f_{X,Y}(x, y) = 1 \quad \text{for } 0 \leq x \leq 1, 0 \leq y \leq 1$$

$$f_{X,Y}(x, y) = 0 \quad \text{otherwise}$$

The probability $P(X \leq 0.5, Y \leq 0.5)$ is $\int_0^{0.5} \int_0^{0.5} 1\, dx\, dy = 0.5 \cdot 0.5 = 0.25$.

17.3. Marginal Probability Distributions (from Joint)

- **Definition:** The probability distribution of a single random variable in a joint distribution, obtained by summing (for discrete) or integrating (for continuous) over all possible values of the other random variables.

- **Formulas:**

  - **For Discrete:**

$$P(X = x) = \sum_y P(X = x, Y = y)$$

$$P(Y = y) = \sum_x P(X = x, Y = y)$$

  - **For Continuous:**

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

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dx$$

- **Example (Discrete):** Using the two dice example where $P(X = x, Y = y) = \frac{1}{36}$: $P(X = 1) = \sum_{y=1}^{6} P(X = 1, Y = y) = \sum_{y=1}^{6} \frac{1}{36} = 6 \cdot \frac{1}{36} = \frac{1}{6}$.

18. Conditional Probability Distributions

These describe the probability distribution of one random variable given that another random variable has taken on a specific value.

18.1. Conditional Probability Mass Function (Conditional PMF) - For Discrete Random Variables

- **Definition:** The PMF of Y given that X has taken on a specific value x.
- **Notation:** $P(Y = y|X = x)$ or $f_{Y|X}(y|x)$.
- **Formula:**

$$P(Y = y|X = x) = \frac{P(X = x, Y = y)}{P(X = x)} \quad \text{provided } P(X = x) > 0$$

  Similarly for $P(X = x|Y = y)$.

- **Example:** From the two dice example. What is $P(Y = 4|X = 3)$? (Prob. of blue die being 4 given red die is 3)
  $P(X = 3, Y = 4) = \frac{1}{36}$
  $P(X = 3) = \frac{1}{6}$ (marginal probability for one die)
  $P(Y = 4|X = 3) = \frac{1/36}{1/6} = \frac{1}{6}$. This makes sense for independent dice.

18.2. Conditional Probability Density Function (Conditional PDF) - For Continuous Random Variables

- **Definition:** The PDF of Y given that X has taken on a specific value x.
- **Notation:** $f_{Y|X}(y|x)$.
- **Formula:**

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x, y)}{f_X(x)} \quad \text{provided } f_X(x) > 0$$

  Similarly for $f_{X|Y}(x|y)$.

- **Example:** Let $f_{X,Y}(x, y) = 2$ for $0 < y < x < 1$, and 0 otherwise.
  First, find $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dy = \int_0^x 2\, dy = [2y]_0^x = 2x$ for $0 < x < 1$.
  Then, $f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{2}{2x} = \frac{1}{x}$ for $0 < y < x$.
  This means, given $X = x$, Y is uniformly distributed from 0 to x.

19. Conditional Expectation

- **Definition:** The expected value of a random variable, given that another random variable has taken on a specific value. It represents the mean of the conditional distribution.
- **Notation:** $E[Y|X=x]$ or $E[Y|x]$.
- **Formulas:**
  - **For Discrete Random Variables:**

$$E[Y|X=x] = \sum_y yP(Y=y|X=x)$$

  - **For Continuous Random Variables:**

$$E[Y|X=x] = \int_{-\infty}^{\infty} y f_{Y|X}(y|x)\, dy$$

- **Law of Total Expectation (or Law of Iterated Expectations):**

$$E[Y] = E[E[Y|X]]$$

  This states that the overall expected value of Y can be found by taking the expectation of the conditional expectation of Y given X.
  - **For Discrete:** $E[Y] = \sum_x E[Y|X=x]P(X=x)$
  - **For Continuous:** $E[Y] = \int_{-\infty}^{\infty} E[Y|X=x] f_X(x)\, dx$
- **Example (Discrete):** Let X be the number of Heads in two coin flips, and Y be the number of Tails.
  $P(X=0, Y=2) = 1/4$, $P(X=1, Y=1) = 2/4$, $P(X=2, Y=0) = 1/4$
  .
  $E[Y|X=1]$: Given 1 Head, Y must be 1. So $E[Y|X=1] = 1$.
  $E[Y|X=0]$: Given 0 Heads, Y must be 2. So $E[Y|X=0] = 2$.
  $E[Y|X=2]$: Given 2 Heads, Y must be 0. So $E[Y|X=2] = 0$.
  Using Law of Total Expectation:
  $E[Y] = E[Y|X=0]P(X=0) + E[Y|X=1]P(X=1) + E[Y|X=2]P(X=2)$
  $E[Y] = (2)\frac{1}{4} + (1)\frac{2}{4} + (0)\frac{1}{4} = \frac{2}{4} + \frac{2}{4} + 0 = 1$.
  (This matches direct calculation of $E[Y]$ for $Y \sim \text{Binomial}(2, 0.5)$ with parameter $p = 0.5$ for Tails.)

20. Covariance and Correlation

These measures quantify the linear relationship between two random variables.

20.1. Covariance

- **Definition:** A measure of the extent to which two random variables change together. A positive covariance indicates that they tend to increase or decrease together. A negative covariance indicates that one tends to increase while the other decreases. A covariance near zero suggests no linear relationship.
- **Notation:** $Cov(X, Y)$ or σ_{XY}.
- **Formula:**

$$Cov(X, Y) = E[(X - E[X])(Y - E[Y])]$$

  - **Computational Formula:**

$$Cov(X, Y) = E[XY] - E[X]E[Y]$$

    To calculate $E[XY]$:

    - **For Discrete:** $E[XY] = \sum_x \sum_y xyP(X = x, Y = y)$
    - **For Continuous:** $E[XY] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xyf_{X,Y}(x, y)\, dx\, dy$

- **Properties of Covariance:**
  - $Cov(X, X) = Var[X]$
  - $Cov(X, Y) = Cov(Y, X)$
  - $Cov(aX, bY) = abCov(X, Y)$ (for constants a, b)
  - $Cov(X + Y, Z) = Cov(X, Z) + Cov(Y, Z)$
  - If X and Y are independent, then $Cov(X, Y) = 0$. (The converse is not always true, i.e., zero covariance does not imply independence unless the variables are jointly normal).
- **Example (Discrete):** Consider X as the outcome of a fair die roll, and $Y = 0$ if X is even, $Y = 1$ if X is odd.
  $E[X] = 3.5$. $P(Y = 0) = 3/6 = 0.5$, $P(Y = 1) = 3/6 = 0.5$. $E[Y] = 0.5$.

Pairs (x, y) with non-zero probability: $(1,1),(2,0),(3,1),(4,0),(5,1),(6,0)$, each with probability $1/6$.
$E[XY] = (1 \cdot 1)\frac{1}{6} + (2 \cdot 0)\frac{1}{6} + (3 \cdot 1)\frac{1}{6} + (4 \cdot 0)\frac{1}{6} + (5 \cdot 1)\frac{1}{6} + (6 \cdot 0)\frac{1}{6}$
$E[XY] = \frac{1}{6} + 0 + \frac{3}{6} + 0 + \frac{5}{6} + 0 = \frac{9}{6} = 1.5.$
$Cov(X, Y) = E[XY] - E[X]E[Y] = 1.5 - (3.5)(0.5) = 1.5 - 1.75 = -0.25.$
This negative covariance makes sense: as X (die roll) increases, Y (odd/even) tends to alternate, resulting in a slight inverse relationship.

20.2. Correlation Coefficient (Pearson Product-Moment Correlation Coefficient)

- **Definition:** A standardized measure of the linear relationship between two random variables. It is a unitless value between -1 and +1.
  - $+1$: Perfect positive linear relationship.
  - -1: Perfect negative linear relationship.
  - 0: No linear relationship.
- **Notation:** ρ_{XY} or $Corr(X, Y)$.
- **Formula:**

$$\rho_{XY} = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}$$

  Where $\sigma_X = \sqrt{Var[X]}$ and $\sigma_Y = \sqrt{Var[Y]}$ are the standard deviations of X and Y, respectively.

- **Properties of Correlation:**
  1. $-1 \leq \rho_{XY} \leq 1$.
  2. $\rho_{XY} = 0$ if X and Y are linearly uncorrelated (which is true if they are independent, but not vice-versa).
  3. Correlation measures *linear* relationships only. Non-linear relationships might exist even if correlation is zero.
  4. Correlation is symmetric: $\rho_{XY} = \rho_{YX}$.

- **Example (Continued from Covariance example):**
  We found $Cov(X, Y) = -0.25$.
  For X (die roll): $Var[X] = \frac{35}{12} \approx 2.9167$, so $\sigma_X = \sqrt{\frac{35}{12}} \approx 1.7078$.
  For Y (0 for even, 1 for odd): $P(Y = 0) = 0.5, P(Y = 1) = 0.5$. $Var[Y] = p(1 - p) = 0.5(0.5) = 0.25$. So $\sigma_Y = \sqrt{0.25} = 0.5$.
  $\rho_{XY} = \frac{-0.25}{(1.7078)(0.5)} = \frac{-0.25}{0.8539} \approx -0.2928$.
  This indicates a weak negative linear correlation between the die roll outcome and whether it's odd/even.

21. Sampling Distribution

- **Definition:** The probability distribution of a statistic (e.g., sample mean, sample proportion) calculated from samples of the same size drawn from a given population. It describes how the statistic varies from sample to sample.

- **Key Idea:** Instead of looking at the distribution of individual data points, we look at the distribution of the *summary measures* derived from repeated samples.

- **Importance:** It is the bridge between sample statistics and population parameters, crucial for hypothesis testing and constructing confidence intervals.

- **Example:** If you repeatedly take samples of size $n = 30$ from a population and calculate the mean of each sample ($\bar{x}$), the distribution of these sample means would be the sampling distribution of the sample mean.

22. Central Limit Theorem (CLT)

- **Definition:** One of the most fundamental theorems in statistics. It states that, regardless of the shape of the original population distribution, the sampling distribution of the sample mean (or sum) will tend to be approximately normally distributed as the sample size (n) becomes sufficiently large.

- **Conditions/Assumptions:**

  1. The sample must be random and independent.

  2. The sample size n must be sufficiently large. (A common rule of thumb is $n \geq 30$, though it can be smaller for distributions already close to normal, and larger for highly skewed distributions).

- **Implications:**
  - The mean of the sampling distribution of the sample mean ($\mu_{\bar{x}}$) is equal to the population mean (μ): $\mu_{\bar{x}} = \mu$.
  - The standard deviation of the sampling distribution of the sample mean (known as the Standard Error) is given by: $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$, where σ is the population standard deviation.
  - If the original population is normally distributed, the sampling distribution of the mean is normal for *any* sample size n.
- **Example:** Even if a population's income distribution is heavily skewed (most people earn low income, few earn very high), if you take many random samples of, say, 100 people and calculate the average income for each sample, the distribution of these average incomes will be approximately normal.

23. Standard Error

- **Definition:** The standard deviation of the sampling distribution of a statistic. It quantifies the precision of a sample statistic as an estimate of a population parameter. A smaller standard error indicates a more precise estimate.
- **Notation:**
  - For the sample mean: $SE(\bar{x})$ or $\sigma_{\bar{x}}$.
  - For the sample proportion: $SE(\hat{p})$ or $\sigma_{\hat{p}}$.
- **Formulas:**
  - **Standard Error of the Mean (SEM):**
    - If population standard deviation σ is known:

$$SE(\bar{x}) = \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

    - If population standard deviation σ is unknown (and estimated by sample standard deviation s):

$$SE(\bar{x}) = \frac{s}{\sqrt{n}}$$

- **Standard Error of the Proportion:**

$$SE(\hat{p}) = \sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

(For confidence intervals, we often use $\hat{p}$ instead of p: $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$)

- **Importance:** The standard error is a critical component in constructing confidence intervals and performing hypothesis tests, as it tells us how much variability we can expect in our sample statistics if we were to take multiple samples.

24. Confidence Intervals (CI)

- **Definition:** A range of values, derived from sample data, that is likely to contain the true value of an unknown population parameter. It provides a measure of the reliability of an estimate.

- **Confidence Level:** The probability that the confidence interval contains the true population parameter. Commonly used levels are 90%, 95%, 99%.

- **General Form:**

$$\text{Point Estimate} \pm (\text{Critical Value} \times \text{Standard Error})$$

24.1. Confidence Interval for the Population Mean (μ)

24.1.1. When Population Standard Deviation (σ) is Known (Z-interval)

- **Formula:**

$$\bar{x} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Where:

  - $\bar{x}$: Sample mean
  - $Z_{\alpha/2}$: Critical Z-value from the standard normal distribution corresponding to the desired confidence level. α is the significance level ($1 - \text{Confidence Level}$).
  - σ: Population standard deviation

- n: Sample size

- **Example:** A sample of 50 students has a mean test score of 75. The population standard deviation for test scores is known to be 8. Construct a 95% CI for the true mean test score.
  - $\bar{x} = 75, \sigma = 8, n = 50$.
  - For 95% CI, $\alpha = 0.05$, so $\alpha/2 = 0.025$. $Z_{0.025} = 1.96$ (from Z-table).
  - $75 \pm 1.96 \frac{8}{\sqrt{50}} = 75 \pm 1.96 \cdot 1.1314 = 75 \pm 2.2175$
  - CI: $(72.78, 77.22)$. We are 95% confident that the true mean test score is between 72.78 and 77.22.

24.1.2. When Population Standard Deviation (σ) is Unknown (t-interval)

- **Formula:**

$$\bar{x} \pm t_{\alpha/2,df} \frac{s}{\sqrt{n}}$$

  Where:
  - $\bar{x}$: Sample mean
  - $t_{\alpha/2,df}$: Critical t-value from the t-distribution with $df = n - 1$ degrees of freedom.
  - s: Sample standard deviation
  - n: Sample size

- **Assumptions:** Random sample, population normally distributed OR $n \geq 30$, population σ is unknown.

- **Example:** A sample of 25 items has a mean weight of 150 grams and a standard deviation of 10 grams. Construct a 90% CI for the true mean weight.
  - $\bar{x} = 150, s = 10, n = 25$.
  - $df = n - 1 = 24$.
  - For 90% CI, $\alpha = 0.10$, so $\alpha/2 = 0.05$. $t_{0.05,24} = 1.711$ (from t-table).
  - $150 \pm 1.711 \frac{10}{\sqrt{25}} = 150 \pm 1.711 \cdot 2 = 150 \pm 3.422$

- CI: $(146.578, 153.422)$. We are 90% confident that the true mean weight is between 146.58 and 153.42 grams.

24.2. Confidence Interval for the Population Proportion (p)

- **Formula (Large Sample / Normal Approximation):**

$$\hat{p} \pm Z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

  Where:

  - $\hat{p}$: Sample proportion (number of successes / n)
  - $Z_{\alpha/2}$: Critical Z-value.
  - n: Sample size

- **Conditions:** $n\hat{p} \geq 10$ and $n(1-\hat{p}) \geq 10$ (to ensure normal approximation is valid).

- **Example:** In a survey of 200 voters, 120 supported Candidate A. Construct a 95% CI for the true proportion of voters supporting Candidate A.

  - $n = 200, x = 120$. $\hat{p} = \frac{120}{200} = 0.6$.
  - For 95% CI, $Z_{0.025} = 1.96$.
  - $0.6 \pm 1.96\sqrt{\frac{0.6(1-0.6)}{200}} = 0.6 \pm 1.96\sqrt{\frac{0.6 \cdot 0.4}{200}} = 0.6 \pm 1.96\sqrt{\frac{0.24}{200}}$
  - $0.6 \pm 1.96\sqrt{0.0012} = 0.6 \pm 1.96 \cdot 0.03464 = 0.6 \pm 0.0679$
  - CI: $(0.5321, 0.6679)$. We are 95% confident that the true proportion is between 53.21% and 66.79%.

24.3. Confidence Interval for the Population Variance (σ^2)

- **Formula (based on Chi-Squared distribution):**

$$\left(\frac{(n-1)s^2}{\chi^2_{\alpha/2,df}}, \frac{(n-1)s^2}{\chi^2_{1-\alpha/2,df}}\right)$$

  Where:

- s^2: Sample variance
- n: Sample size
- $df = n - 1$: Degrees of freedom
- $\chi^2_{\alpha/2,df}$ and $\chi^2_{1-\alpha/2,df}$: Critical Chi-Squared values from the Chi-Squared distribution.

- **Assumption:** The population must be normally distributed.
- **Example:** A sample of 10 measurements has a variance of $s^2 = 5$. Construct a 90% CI for the population variance.
  - $n = 10$, $s^2 = 5$. $df = 9$.
  - For 90% CI, $\alpha = 0.10$, so $\alpha/2 = 0.05$.
  - From Chi-Squared table with $df = 9$:
    - $\chi^2_{0.05,9} = 16.919$
    - $\chi^2_{1-0.05,9} = \chi^2_{0.95,9} = 3.325$
  - Lower Bound: $\frac{(10-1)\cdot 5}{16.919} = \frac{9\cdot 5}{16.919} = \frac{45}{16.919} \approx 2.66$
  - Upper Bound: $\frac{(10-1)\cdot 5}{3.325} = \frac{9\cdot 5}{3.325} = \frac{45}{3.325} \approx 13.53$
  - CI: $(2.66, 13.53)$. We are 90% confident that the true population variance is between 2.66 and 13.53.

25. Tolerance and Prediction Intervals

Confidence intervals estimate population parameters. Tolerance and prediction intervals, on the other hand, deal with individual observations or a proportion of observations.

25.1. Prediction Interval (PI)

- **Definition:** An interval that provides a range of values within which a *single new observation* from the same population is expected to fall with a certain probability.

- **Key Distinction from CI:** A CI estimates a *parameter* (fixed but unknown), while a PI predicts a *future observation* (random). Prediction intervals are typically wider than confidence intervals because they account for both the uncertainty in estimating the population parameters AND the inherent variability of individual observations.

- **Formula (for a single future observation from a Normal population, σ unknown):**

$$\bar{x} \pm t_{\alpha/2,n-1} s \sqrt{1 + \frac{1}{n}}$$

  Where:

  - $\bar{x}$: Sample mean
  - s: Sample standard deviation
  - n: Sample size
  - $t_{\alpha/2,n-1}$: Critical t-value

- **Example:** Using the example from CI for mean (t-interval): sample of 25 items, mean weight 150 grams, std dev 10 grams. Predict the weight of a single new item with 90% probability.

  - $\bar{x} = 150$, $s = 10$, $n = 25$, $t_{0.05,24} = 1.711$.
  - $150 \pm 1.711 \cdot 10\sqrt{1 + \frac{1}{25}} = 150 \pm 17.11\sqrt{1 + 0.04} = 150 \pm 17.11\sqrt{1.04}$
  - $150 \pm 17.11 \cdot 1.0198 = 150 \pm 17.448$
  - PI: $(132.552, 167.448)$. We can be 90% confident that the next observed item will weigh between 132.55 and 167.45 grams. (Notice it's wider than the CI for the mean).

25.2. Tolerance Interval (TI)

- **Definition:** An interval that, with a certain confidence level, contains at least a specified proportion (e.g., 99%) of the individual observations in the population. It's used to define a range that encompasses a certain percentage of the population values.

- **Key Distinction:** PIs are for a single future observation; TIs are for a *proportion* of the population values.

- **Formula (for Normal population, two-sided, σ unknown):**

$$\bar{x} \pm k \cdot s$$

  Where k is a tolerance factor that depends on n, the confidence level, and the proportion of the population to be covered. The calculation of k is complex and usually requires specialized tables or software.

- **Example:** A manufacturer wants to find a range that contains 99% of all product weights with 95% confidence.
  - This requires a tolerance interval. The calculation for k would involve looking up a tolerance factor from tables or using statistical software (e.g., for $n = 25$, 99% proportion, 95% confidence, k would be approximately 2.9).
  - If $k = 2.9$, then the TI would be $150 \pm 2.9 \cdot 10 = 150 \pm 29$, so $(121, 179)$.
  - This interval is even wider than the prediction interval because it aims to capture a large percentage of the entire population, not just one future observation.

26. Basics of Hypothesis Testing

Hypothesis testing is a formal procedure for making inferences about population parameters based on sample data. It involves making an assumption (hypothesis) about a population parameter and then using sample data to determine whether this assumption is plausible.

- **Goal:** To determine if there is enough statistical evidence to reject a null hypothesis in favor of an alternative hypothesis.
- **Steps of Hypothesis Testing:**
  1. **Formulate Hypotheses:** State the null (H_0) and alternative (H_1) hypotheses.
  2. **Choose Significance Level (α):** Set the probability of making a Type I error. Common values are 0.05 (5%), 0.01 (1%), 0.10 (10%).
  3. **Select Test Statistic:** Choose an appropriate test statistic based on the data type, distribution, and research question (e.g., Z-statistic, t-statistic, Chi-squared statistic).
  4. **Determine Critical Region (or calculate P-value):** Define the range of test statistic values that would lead to rejecting the null hypothesis. Alternatively, calculate the P-value.

5. **Collect Data and Calculate Test Statistic:** Compute the value of the test statistic from the sample data.

6. **Make Decision:** Compare the test statistic to the critical value (or P-value to α) and decide whether to reject or fail to reject H_0.

7. **Draw Conclusion:** State the conclusion in the context of the problem.

- **Null Hypothesis (H_0):**
  - **Definition:** A statement of no effect, no difference, or no relationship. It represents the status quo or a commonly accepted belief.
  - It always contains an equality sign ($=, \leq, \geq$).
  - **Example:** The mean height of students is 170 cm ($\mu = 170$).
- **Alternative Hypothesis (H_1 or H_A):**
  - **Definition:** A statement that contradicts the null hypothesis. It represents what the researcher is trying to find evidence for.
  - It never contains an equality sign ($ \ne, <, >$).
  - **Example:** The mean height of students is not 170 cm ($\mu \neq 170$).

27. Type I and Type II Errors

In hypothesis testing, decisions are made based on sample data, which means there's always a possibility of making an incorrect decision.

- **Type I Error (α error):**
  - **Definition:** Rejecting the null hypothesis (H_0) when it is actually true.
  - **Probability:** The probability of making a Type I error is denoted by α, which is the significance level of the test.
  - **Analogy:** A "false positive" (e.g., convicting an innocent person).
  - **Control:** The researcher directly sets α. Lowering α reduces the chance of a Type I error but increases the chance of a Type II error.
- **Type II Error (β error):**

- **Definition:** Failing to reject the null hypothesis (H_0) when it is actually false.
- **Probability:** The probability of making a Type II error is denoted by β.
- **Analogy:** A “false negative” (e.g., failing to convict a guilty person).
- **Control:** β is influenced by sample size, effect size, and α. Increasing sample size reduces β.

- **Power of the Test:**
  - **Definition:** The probability of correctly rejecting a false null hypothesis.
  - **Formula:** Power $= 1 - \beta$.
  - **Goal:** Researchers generally aim for high power (e.g., 0.80 or 80%).

Decision	H0 is True	H0 is False
Reject H_0	Type I Error (α)	Correct Decision (Power = $1 - \beta$)
Fail to Reject H_0	Correct Decision	Type II Error (β)

28. P-values, Critical Values, and Test Statistics

These are the core components used to make a decision in hypothesis testing.

28.1. Test Statistic

- **Definition:** A single value calculated from sample data that is used to decide whether to reject the null hypothesis. Its calculation depends on the type of test being performed and the parameters being tested.
- **Purpose:** It quantifies how far our sample result deviates from what we would expect if the null hypothesis were true.
- **Examples:**
  - **Z-statistic:** Used for means when population standard deviation is known, or for proportions with large samples.

$$Z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

or

$$Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1 - p_0)/n}}$$

- **t-statistic:** Used for means when population standard deviation is unknown.

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

28.2. P-value (Observed Significance Level)

- **Definition:** The probability of observing a test statistic as extreme as, or more extreme than, the one calculated from the sample data, *assuming the null hypothesis is true*. It is the smallest significance level (α) at which the null hypothesis would be rejected.
- **Interpretation:** A small P-value suggests that the observed data are unlikely under the null hypothesis, thus providing evidence against H_0. A large P-value suggests that the observed data are consistent with H_0.
- **Decision Rule using P-value:**
    - If P-value $\leq \alpha$: Reject H_0. (The result is statistically significant).
    - If P-value $> \alpha$: Fail to reject H_0. (The result is not statistically significant).
- **Example:** If you perform a test and get a P-value of 0.02, and your chosen α is 0.05, then $0.02 \leq 0.05$, so you would reject H_0.

28.3. Critical Value (Rejection Region Approach)

- **Definition:** A threshold value(s) from the sampling distribution of the test statistic that defines the boundary of the rejection region. If the calculated test statistic falls into the critical region, the null hypothesis is rejected.
- **Decision Rule using Critical Value:**
    - If Test Statistic falls into the Critical Region: Reject H_0.
    - If Test Statistic does not fall into the Critical Region: Fail to reject H_0.
- **Example:** For a Z-test with $\alpha = 0.05$ and a two-tailed test, the critical Z-values are ± 1.96. If your calculated Z-statistic is 2.10, it falls outside ± 1.96, so you reject H_0.

29. Tests on Mean and Proportion

These are common hypothesis tests for single population parameters.

29.1. Tests on Population Mean (μ)

29.1.1. Z-Test for Mean (Population σ Known)

- **Hypotheses:**
    - $H_0 : \mu = \mu_0$ (or $\leq, \geq$)
    - $H_1 : \mu \neq \mu_0$ (or $>, <$)
- **Test Statistic:**

$$Z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

- **Assumptions:** Random sample, population normally distributed OR $n \geq 30$, population σ is known.
- **Example:** A company claims that its light bulbs last 1000 hours on average ($\mu = 1000$). A sample of 35 bulbs shows a mean life of 980 hours. If the population standard deviation is 50 hours, test the claim at $\alpha = 0.05$.
    - $H_0 : \mu = 1000$
    - $H_1 : \mu \neq 1000$
    - $Z = \frac{980-1000}{50/\sqrt{35}} = \frac{-20}{8.45} \approx -2.367$
    - For two-tailed test, $\alpha = 0.05$, critical values are ± 1.96.
    - Since $-2.367 < -1.96$, we reject H_0. There is sufficient evidence to claim the mean life is different from 1000 hours.

29.1.2. t-Test for Mean (Population σ Unknown)

- **Hypotheses:**
    - $H_0 : \mu = \mu_0$ (or $\leq, \geq$)

- $H_1 : \mu \neq \mu_0$ (or $>, <$)

- **Test Statistic:**

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

With degrees of freedom $df = n - 1$.

- **Assumptions:** Random sample, population normally distributed OR $n \geq 30$, population σ is unknown.

- **Example:** A new teaching method is claimed to increase test scores. A sample of 20 students using the new method has a mean score of 85 and a standard deviation of 10. The old method had a mean of 80. Test if the new method is better at $\alpha = 0.01$.

  - $H_0 : \mu \leq 80$
  - $H_1 : \mu > 80$ (one-tailed test)
  - $t = \frac{85-80}{10/\sqrt{20}} = \frac{5}{2.236} \approx 2.236$
  - $df = 20 - 1 = 19$. For one-tailed upper test, $\alpha = 0.01$, critical t-value $t_{0.01,19} = 2.539$.
  - Since $2.236 < 2.539$, we fail to reject H_0. There is not enough evidence to claim the new method is better.

29.2. Tests on Population Proportion (p)

- **Hypotheses:**

  - $H_0 : p = p_0$ (or $\leq, \geq$)
  - $H_1 : p \neq p_0$ (or $>, <$)

- **Test Statistic (Z-statistic, using normal approximation):**

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

- **Assumptions:** Random sample, n is large enough such that $np_0 \geq 10$ and $n(1 - p_0) \geq 10$.

- **Example:** A political candidate claims that 50% of voters support them ($p = 0.5$). A poll of 150 voters shows 60 support the candidate. Test the claim at $\alpha = 0.05$.
    - $H_0 : p = 0.5$
    - $H_1 : p \neq 0.5$
    - $\hat{p} = \frac{60}{150} = 0.4$.
    - $Z = \frac{0.4-0.5}{\sqrt{\frac{0.5(1-0.5)}{150}}} = \frac{-0.1}{\sqrt{\frac{0.25}{150}}} = \frac{-0.1}{0.0408} \approx -2.45$
    - For two-tailed test, $\alpha = 0.05$, critical values are ± 1.96.
    - Since $-2.45 < -1.96$, we reject H_0. There is sufficient evidence to claim the proportion of support is not 50%.

30. One-Tailed and Two-Tailed Tests

The type of hypothesis test (one-tailed or two-tailed) depends on the alternative hypothesis.

- **One-Tailed Test:**
    - **Definition:** The alternative hypothesis specifies a direction of difference (e.g., greater than or less than). The rejection region is entirely in one tail of the sampling distribution.
    - H_1 **Forms:**
        - $H_1 : \mu > \mu_0$ (Right-tailed test)
        - $H_1 : \mu < \mu_0$ (Left-tailed test)
    - **Critical Region:** The critical region is entirely on one side. For example, for a right-tailed Z-test with $\alpha = 0.05$, the critical value is $Z_{0.05} = 1.645$. For a left-tailed Z-test with $\alpha = 0.05$, the critical value is $Z_{0.95} = -1.645$.
    - **P-value:** The P-value is the area in the specified tail.
- **Two-Tailed Test:**
    - **Definition:** The alternative hypothesis specifies a difference in *any* direction (e.g., not equal to). The rejection region is split equally into both tails of the sampling distribution.

- H_1 **Form:**
  - $H_1 : \mu \neq \mu_0$
- **Critical Region:** The critical region is split into two equal parts. For example, for a two-tailed Z-test with $\alpha = 0.05$, the critical values are $Z_{\alpha/2} = Z_{0.025} = 1.96$ and $-Z_{\alpha/2} = -Z_{0.025} = -1.96$.
- **P-value:** The P-value is twice the area in one tail (the tail corresponding to the calculated test statistic).

- **Choosing the Test:** The choice of one-tailed or two-tailed test must be made *before* data collection and analysis, based on the research question or the specific claim being investigated.

31. Paired t-Test

- **Definition:** A statistical test used to compare the means of two related groups or measurements. It is appropriate when each observation in one group is paired with an observation in the other group.
- **Common Applications:**
  - "Before-and-after" studies (e.g., weight before and after a diet).
  - Comparing two different treatments applied to the same subjects.
  - Comparing two methods when measurements are taken on the same items.
- **Key Idea:** Instead of analyzing the two groups separately, we calculate the *difference* between the paired observations and then perform a one-sample t-test on these differences.
- **Hypotheses:** Let d_i be the difference between paired observations $(d_i = X_{1i} - X_{2i})$.
  - $H_0 : \mu_d = 0$ (The mean difference is zero; no significant difference between paired measurements)
  - $H_1 : \mu_d \neq 0$ (or $>$, $<$, depending on the desired direction of difference)
- **Test Statistic:**

$$t = \frac{\bar{d} - \mu_{d0}}{s_d/\sqrt{n}}$$

Where:

- $\bar{d}$: Sample mean of the differences.
- μ_{d0}: Hypothesized mean difference (usually 0).
- s_d: Sample standard deviation of the differences.
- n: Number of pairs.
- Degrees of freedom: $df = n - 1$.

- **Assumptions:**
  1. The paired differences are independent.
  2. The population of differences is approximately normally distributed (or n is large enough for CLT).

- **Example:** A new drug is tested for its effect on blood pressure. 12 patients have their blood pressure measured before (X1) and after (X2) taking the drug.
  Calculate the difference for each patient: $d_i = X_{1i} - X_{2i}$.
  Suppose the mean difference $\bar{d} = 5$ mmHg and the standard deviation of differences $s_d = 8$ mmHg.
  Test if the drug significantly reduces blood pressure at $\alpha = 0.05$. (Reduction means before > after, so positive difference).
  - $H_0 : \mu_d \leq 0$
  - $H_1 : \mu_d > 0$ (right-tailed test)
  - $n = 12, df = 11$.
  - $t = \frac{5-0}{8/\sqrt{12}} = \frac{5}{2.309} \approx 2.165$
  - For right-tailed test, $\alpha = 0.05$, critical t-value $t_{0.05,11} = 1.796$.
  - Since $2.165 > 1.796$, we reject H_0. There is sufficient evidence to conclude the drug significantly reduces blood pressure.

32. Simple Linear Regression

Simple linear regression is a statistical method that allows us to summarize and study relationships between two continuous variables: a dependent variable (response) and an independent variable (predictor).

32.1. Regression Equation (Least Squares Regression Line)

- **Definition:** The equation of the straight line that best describes the linear relationship between two variables. It is determined by minimizing the sum of the squared vertical distances from the data points to the line (the "least squares" criterion).
- **Formula for the Estimated Regression Line:**

$$\hat{Y} = b_0 + b_1 X$$

  Where:

  - $\hat{Y}$: The predicted value of the dependent (response) variable.
  - X: The independent (predictor) variable.
  - b_0: The Y-intercept of the regression line. It represents the predicted value of Y when $X = 0$.
  - b_1: The slope of the regression line. It represents the predicted change in Y for a one-unit increase in X.
- **Example:** Predicting a student's final exam score (Y) based on the number of hours they studied (X).
  If the regression equation is $\hat{Y} = 50 + 5X$:
  - A student who studies 0 hours is predicted to score 50.
  - For every additional hour studied, the predicted score increases by 5 points.

32.2. Least Squares Estimates (Calculating b_0 and b_1)

The values of b_0 and b_1 are chosen to minimize the sum of squared residuals.

- **Formulas for b_1 (Slope) and b_0 (Y-intercept):**

$$b_1 = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sum(X_i - \bar{X})^2} = \frac{S_{XY}}{S_{XX}}$$

or using computational formulas:

$$b_1 = \frac{n \sum X_i Y_i - (\sum X_i)(\sum Y_i)}{n \sum X_i^2 - (\sum X_i)^2}$$

$$b_0 = \bar{Y} - b_1 \bar{X}$$

Where:

- n: Number of data pairs.
- X_i, Y_i: Individual data points.
- $\bar{X}, \bar{Y}$: Sample means of X and Y.
- S_{XY}: Sample covariance of X and Y (sum of products of deviations).
- S_{XX}: Sum of squared deviations of X.

- **Example (Continued):** Suppose we have the following data for (Hours Studied, Exam Score):
(2, 60), (4, 70), (6, 80), (8, 90)
  - $\sum X = 20$, $\sum Y = 300$, $n = 4$.
  - $\bar{X} = 5$, $\bar{Y} = 75$.
  - $\sum X^2 = 4 + 16 + 36 + 64 = 120$.
  - $\sum XY = (2 \cdot 60) + (4 \cdot 70) + (6 \cdot 80) + (8 \cdot 90) = 120 + 280 + 480 + 720 = 1600$.
  - $b_1 = \frac{4(1600)-(20)(300)}{4(120)-(20)^2} = \frac{6400-6000}{480-400} = \frac{400}{80} = 5$
  - $b_0 = 75 - 5(5) = 75 - 25 = 50$
  - So, the regression equation is $\hat{Y} = 50 + 5X$.

32.3. Correlation (r)

- **Definition:** (Revisiting from Section 20.2) The Pearson product-moment correlation coefficient, r, measures the strength and direction of the *linear* relationship between two quantitative variables.
- **Formula:**

$$r = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} = \frac{S_{XY}}{\sqrt{S_{XX} S_{YY}}}$$

Where $S_{YY} = \sum(Y_i - \bar{Y})^2$.

- **Properties:**
  - $-1 \leq r \leq 1$.
  - $r = 1$ indicates a perfect positive linear relationship.
  - $r = -1$ indicates a perfect negative linear relationship.
  - $r = 0$ indicates no linear relationship.
- **Relationship with Slope:** The sign of the correlation coefficient (r) will always be the same as the sign of the slope (b_1) of the least squares regression line.
- **Example (Continued):**
  - $S_{XX} = \sum(X_i - \bar{X})^2 = (2-5)^2 + (4-5)^2 + (6-5)^2 + (8-5)^2 = 9 + 1 + 1 + 9 = 20$.
  - $S_{YY} = \sum(Y_i - \bar{Y})^2 = (60-75)^2 + (70-75)^2 + (80-75)^2 + (90-75)^2 = (-15)^2 + (-5)^2 + (5)^2 + (15)^2 = 225 + 25 + 25 + 225 = 500$.
  - $S_{XY} = \sum(X_i - \bar{X})(Y_i - \bar{Y}) = (-3)(-15) + (-1)(-5) + (1)(5) + (3)(15) = 45 + 5 + 5 + 45 = 100$.
  - $r = \frac{100}{\sqrt{20 \cdot 500}} = \frac{100}{\sqrt{10000}} = \frac{100}{100} = 1$.
  - A correlation of $r = 1$ indicates a perfect positive linear relationship, which is expected given the perfectly linear example data.

32.4. Coefficient of Determination (R^2)

- **Definition:** The coefficient of determination, R^2, represents the proportion of the total variation in the dependent variable (Y) that can be explained by the linear relationship with the independent variable (X).
- **Formula:**

$$R^2 = r^2$$

or

$$R^2 = \frac{\text{Explained Variation}}{\text{Total Variation}} = \frac{\sum(\hat{Y}_i - \bar{Y})^2}{\sum(Y_i - \bar{Y})^2} = 1 - \frac{\sum(Y_i - \hat{Y}_i)^2}{\sum(Y_i - \bar{Y})^2} = 1 - \frac{SSE}{SST}$$

Where:

- $SSE = \sum(Y_i - \hat{Y}_i)^2$: Sum of Squares Error (unexplained variation, also called sum of squared residuals).
- $SST = \sum(Y_i - \bar{Y})^2$: Total Sum of Squares (total variation in Y).

- **Interpretation:** R^2 ranges from 0 to 1.
  - $R^2 = 0$: X explains none of the variation in Y.
  - $R^2 = 1$: X explains all of the variation in Y.
- **Example (Continued):**
  - From the previous example, $r = 1$.
  - $R^2 = r^2 = 1^2 = 1$.
  - This means that 100% of the variation in exam scores can be explained by the number of hours studied. This is because the example data points fall perfectly on a straight line. In real-world data, R^2 is typically between 0 and 1, indicating some but not all variation is explained.

32.5. Prediction

- **Definition:** Using the established regression equation to estimate the value of the dependent variable ($\hat{Y}$) for a given new value of the independent variable (X).
- **Process:** Simply substitute the new X value into the estimated regression equation.
- **Caution (Extrapolation):** Avoid making predictions for X values outside the range of the observed data used to build the model. This is called extrapolation and can lead to unreliable predictions.
- **Example (Continued):** Predict the exam score for a student who studies 7 hours.
  - Regression equation: $\hat{Y} = 50 + 5X$
  - For $X = 7$: $\hat{Y} = 50 + 5(7) = 50 + 35 = 85$.
  - A student who studies 7 hours is predicted to score 85.

32.6. Residual Analysis

- **Definition:** A residual is the difference between an observed value of the dependent variable (Y_i) and the predicted value ($\hat{Y}_i$) from the regression model.

$$e_i = Y_i - \hat{Y}_i$$

- **Purpose of Analysis:** Residual analysis is crucial for checking the assumptions of linear regression and assessing the appropriateness of the model. Plotting residuals against predicted values, independent variables, or observation order can reveal patterns that indicate violations of assumptions.

- **Assumptions of Linear Regression (checked using residuals):**

  1. **Linearity:** The relationship between X and Y is linear. (Residuals should show no discernible pattern when plotted against X or $\hat{Y}$).

  2. **Independence of Errors:** Residuals are independent of each other. (No pattern in residuals plotted against time or order).

  3. **Homoscedasticity (Constant Variance):** The variance of the residuals is constant across all levels of X. (The spread of residuals around 0 should be roughly consistent across the range of X or $\hat{Y}$).

  4. **Normality of Errors:** The residuals are normally distributed. (Can be checked with a histogram of residuals or a Q-Q plot).

- **Example (Continued):** Let's calculate residuals for our example data:

X (Hours)	Y (Observed Score)	$\hat{Y} = 50 + 5X$ (Predicted Score)	Residual ($e_i = Y_i - \hat{Y}_i$)
2	60	$50 + 5(2) = 60$	$60 - 60 = 0$
4	70	$50 + 5(4) = 70$	$70 - 70 = 0$
6	80	$50 + 5(6) = 80$	$80 - 80 = 0$
8	90	$50 + 5(8) = 90$	$90 - 90 = 0$

In this perfectly linear example, all residuals are 0. In real-world scenarios, residuals would be non-zero, and their pattern (or lack thereof) would be analyzed. For instance, a "fanning out" pattern in a residual plot would suggest heteroscedasticity (non-constant variance).

33. Random/Stochastic Processes

- **Definition:** A random process (also known as a stochastic process) is a collection of random variables indexed by time or space. Unlike a single random variable that describes a single uncertain outcome, a random process describes the evolution of a random phenomenon over time or space.
- **Key Idea:** A random process produces a sequence of random values. Each value in the sequence is a random variable, and the sequence itself represents a possible "path" or "realization" of the process.
- **Notation:** A random process is often denoted as $X(t), t \in T$ or $X_t, t \in T$, where t is the index (time, usually).
- **Random Variable vs. Random Process:**
  - A **random variable** is a single outcome of an experiment (e.g., the temperature at 10 AM today).
  - A **random process** is a collection of such random variables, describing how that outcome *changes over time* (e.g., the temperature measured every hour throughout the day, or the stock price minute by minute).
- **Examples:**
  - Stock prices over time.
  - Number of customers arriving at a store over time.
  - Temperature readings at different locations on a map.
  - The path of a diffusing particle.

33.1. Poisson Process

- **Definition:** A counting process that models the number of events occurring in a fixed interval of time or space, where these events happen independently and at a constant average rate. It's often used to model rare events.
- **Characteristics/Assumptions:**
  1. **Stationarity:** The rate of events (λ) is constant over time. The number of events in any interval depends only on the length of the interval, not on its starting point.

2. **Independent Increments:** The number of events in non-overlapping intervals are independent.

3. **Ordinary (or Non-simultaneous Events):** It is impossible for two or more events to occur at precisely the same instant.

- **Parameter:**
    - λ: The average rate of events per unit time (or space). It's also known as the intensity or rate parameter.

- **Key Components:**
    - $N(t)$: The number of events that have occurred up to time t. $N(t), t \geq 0$ is the Poisson process.

- **Formulas:**
    1. **Probability of k events in an interval of length t:** The number of events $N(t)$ in an interval of length t follows a Poisson distribution with mean λt:

    $$P(N(t) = k) = \frac{e^{-\lambda t}(\lambda t)^k}{k!} \quad \text{for } k = 0, 1, 2, \ldots$$

    2. **Time between events (Inter-arrival times):** The time between consecutive events in a Poisson process follows an Exponential distribution with rate λ.

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

    The expected time between events is $E[X] = 1/\lambda$.

- **Example:**
    - Customers arrive at a bank at an average rate of 10 customers per hour. This can be modeled as a Poisson process with $\lambda = 10$ per hour.
    - Probability of exactly 3 customers arriving in the next 30 minutes (0.5 hours): Here, $t = 0.5$, $\lambda = 10$. So $\lambda t = 10 \cdot 0.5 = 5$.

    $$P(N(0.5) = 3) = \frac{e^{-5}5^3}{3!} = \frac{0.006738 \cdot 125}{6} \approx 0.1404$$

    - The time a customer waits for the next customer to arrive follows an Exponential distribution with $\lambda = 10$. The expected waiting time is $1/10 = 0.1$ hours or 6 minutes.

33.2. Random Walk

- **Definition:** A mathematical formalization of a path that consists of a sequence of random steps on some mathematical space (e.g., integers, a grid). Each step is chosen randomly.
- **Key Idea:** The position at any given time depends on the previous position plus a random "step."
- **Types:**
  - **Simple Random Walk:** At each step, movement is only allowed to adjacent positions with equal probability.
  - **One-dimensional Random Walk:** Movement along a line (integers).
  - **Multi-dimensional Random Walk:** Movement on a grid or in higher dimensions.
- **Formulas (for a simple symmetric 1D random walk):**
  - Let X_t be the position at time t.
  - $X_0 = 0$ (starting position).
  - At each step i, S_i is a random variable: $S_i = +1$ with probability p (move right), $S_i = -1$ with probability $1 - p$ (move left).
  - Position at time t: $X_t = \sum_{i=1}^{t} S_i$.
  - **Expected Position:** $E[X_t] = t(2p - 1)$. For a symmetric walk ($p = 0.5$), $E[X_t] = 0$.
  - **Variance of Position:** $Var[X_t] = t(1 - (2p - 1)^2)$. For a symmetric walk ($p = 0.5$), $Var[X_t] = t$.
- **Example:**
  - **A Drunkard's Walk:** A drunkard starts at a lamp post (position 0) and takes steps left or right. Each step is 1 meter, and the probability of moving left or right is 0.5.
  - After 10 steps ($t = 10$):
    - The expected position is $E[X_{10}] = 10(2 \cdot 0.5 - 1) = 10(0) = 0$.
    - The variance of the position is $Var[X_{10}] = 10$.

  - The position X_{10} itself will vary. It can range from -10 to +10, with values like 0, 2, -2, etc. (even numbers if starting from 0).

33.3. Markov Chains

- **Definition:** A stochastic process that describes a sequence of possible events in which the probability of each event depends only on the state attained in the previous event.
- **Key Idea:** "The future is independent of the past, given the present." This is known as the **Markov Property**.
- **Components:**
  1. **State:** A possible condition or configuration of the system being modeled. For a Markov chain, the system is always in one of a finite or countably infinite number of states.
  2. **Transitions:** Movement from one state to another.
  3. **Transition Probabilities:** The probabilities of moving from one state to another. These are usually represented in a **transition matrix**.
- **Markov Property:** The assumption that the future state of a system depends only on its current state, and not on any previous states. Formally:

$$P(X_{t+1} = j | X_t = i, X_{t-1} = i_{t-1}, \dots, X_0 = i_0) = P(X_{t+1} = j | X_t = i)$$

- **Notation:** $X_t, t \in T$ where X_t is the state at time t.
- **Transition Matrix (P):** A square matrix where each element P_{ij} represents the probability of transitioning from state i to state j in one step.
  - $P_{ij} = P(X_{t+1} = j | X_t = i)$.
  - Each row of the matrix sums to 1 (since from any state, the process must transition to some state).
  - For an N-state system, the matrix is $N \times N$.

$$P = \begin{pmatrix} P_{11} & P_{12} & \dots & P_{1N} \\ P_{21} & P_{22} & \dots & P_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ P_{N1} & P_{N2} & \dots & P_{NN} \end{pmatrix}$$

The probabilities of being in each state at time t can be represented as a row vector $\pi_t = (\pi_t(1), \pi_t(2), \ldots, \pi_t(N))$.
The state distribution at time $t+1$ is given by $\pi_{t+1} = \pi_t P$.
The state distribution after k steps is $\pi_k = \pi_0 P^k$.

- **Stationary Distribution (Steady-State Distribution):** For some Markov chains (specifically, irreducible and aperiodic Markov chains), after a very long time, the probability of being in any given state becomes constant, regardless of the initial state. This limiting probability distribution is called the stationary distribution, denoted by $\pi = (\pi_1, \pi_2, \ldots, \pi_N)$.
  - It satisfies the equation: $\pi = \pi P$
  - And the sum of probabilities must be 1: $\sum_{i=1}^{N} \pi_i = 1$.
  - This means that if the system starts in the stationary distribution, it will remain in that distribution for all subsequent steps.
- **Expected Value (in the context of Markov Chains):**
  - While individual states don't usually have numerical "values" that are averaged directly like a random variable, we can talk about the expected number of steps to reach a certain state (hitting time) or the expected number of times a state is visited in a certain period.
  - More commonly, "expected value" refers to the expected state of the system if states are mapped to numerical values, or the expected long-run average of some function of the state, calculated using the stationary distribution.
  - If each state i has an associated value v_i, then the long-run expected value of the system would be:

$$E[\text{Value}] = \sum_{i=1}^{N} v_i \pi_i$$

- **Example (Weather Model Continued):**
  - Suppose the weather in a city can be in one of three states: Sunny (S), Cloudy ©, or Rainy ®. The weather tomorrow depends only on the weather today.
  - Transition Probabilities:
    - If Sunny today: 0.8 chance Sunny tomorrow, 0.1 Cloudy, 0.1 Rainy.

- If Cloudy today: 0.2 chance Sunny tomorrow, 0.6 Cloudy, 0.2 Rainy.
- If Rainy today: 0.2 chance Sunny tomorrow, 0.3 Cloudy, 0.5 Rainy.

- Transition Matrix (P):

$$P = \begin{pmatrix} 0.8 & 0.1 & 0.1 \\ 0.2 & 0.6 & 0.2 \\ 0.2 & 0.3 & 0.5 \end{pmatrix}$$

(Rows represent current state, columns represent next state: S, C, R)

- **Stationary Distribution Calculation:** To find the long-term probability of being Sunny, Cloudy, or Rainy, we solve $\pi = \pi P$ and $\sum \pi_i = 1$.
Let $\pi = (\pi_S, \pi_C, \pi_R)$.
$\pi_S = 0.8\pi_S + 0.2\pi_C + 0.2\pi_R$
$\pi_C = 0.1\pi_S + 0.6\pi_C + 0.3\pi_R$
$\pi_R = 0.1\pi_S + 0.2\pi_C + 0.5\pi_R$
And $\pi_S + \pi_C + \pi_R = 1$.
Solving this system of linear equations (e.g., replace one equation with $\pi_S + \pi_C + \pi_R = 1$ and solve the system), you would find the stationary distribution.
For this example, the stationary distribution is approximately $\pi \approx (0.5, 0.3, 0.2)$, meaning in the long run, it's Sunny 50% of the time, Cloudy 30%, and Rainy 20%.

Conclusion

This comprehensive document has been significantly expanded to provide a thorough overview of fundamental concepts in Statistics and Probability Theory. It now includes detailed explanations of random/stochastic processes, distinguishing them from random variables, and delves into specific types such as Poisson processes, random walks, and Markov chains. The section on Markov Chains has been further enriched with explicit definitions and formulas for the Markov property, states, transition matrices, stationary distributions, and the concept of expected value within this context. Each section is complemented with definitions, relevant formulas, and illustrative examples, building a robust foundation for understanding both theoretical and applied aspects of probability and statistics.

You can now copy and paste this Markdown content into your preferred document editor (e.g., Microsoft Word, Google Docs). Most modern editors support direct pasting of Markdown or have import options for it, allowing you to easily format and save it as a Word document or PDF for your notes.