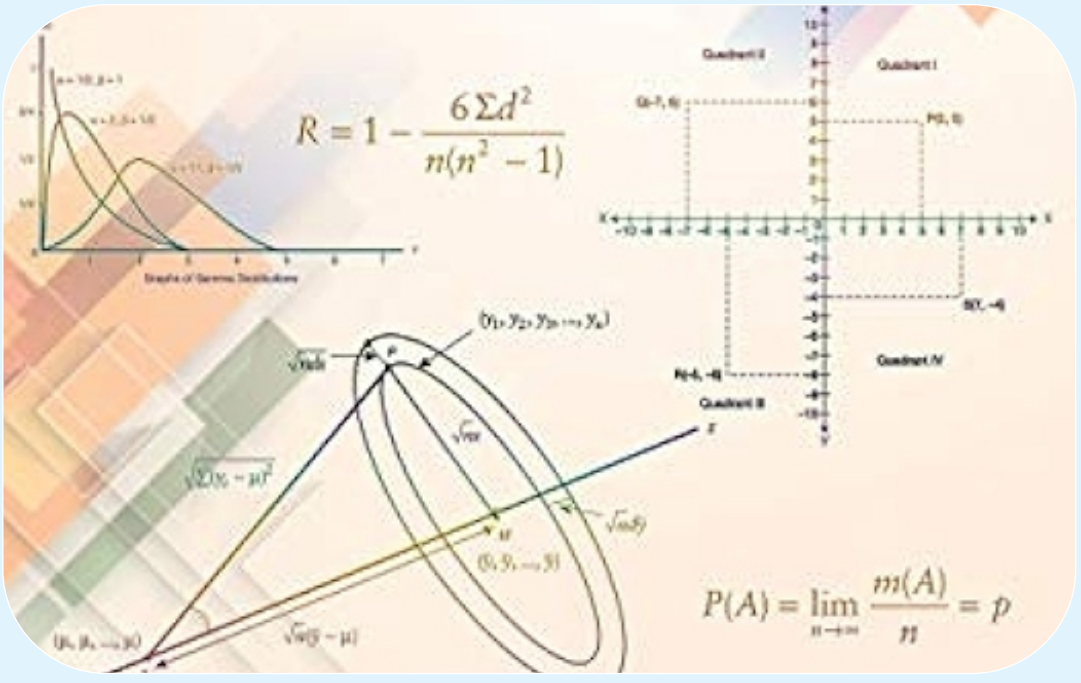
MATS UNIVERSITY

NAAC GRADE A+ ACCREDITED UNIVERSITY

MATS CENTRE FOR DISTANCE & ONLINE EDUCATION

Mathematical Statistics-Elective-I

Master of Science (M.Sc.)
Semester - 1



SELF LEARNING MATERIAL

MSCMODL105
MATHEMATICAL STATISTICS

COURSE DEVELOPMENTEXPERT COMMITTEE

Prof (Dr) K P Yadav	Vice Chancellor, MATS University
Prof (Dr) A J Khan	Professor Mathematics, MATS University
Prof(Dr) D K Das	Professor Mathematics, CCET, Bhilai

COURSE COORDINATOR

Dr Vinita Dewangan	Associate Professor, MATS University

COURSE /BLOCK PREPARATION

Dr. Aachal Mishra	Asst. Professor, MATS University

March 2025 ISBN:978-93-49954-20-5

Printed &Published on behalf of MATS University, Village-Gullu, Aarang, RaipurbyMr. MeghanadhuduKatabathuni, Facilities & Operations, MATS University,Raipur(C.G.)

Printed at: The Digital Press, Krishna Complex, Raipur-492001(Chhattisgarh)

Acknowledgement

The material (pictures and passages) we have used is purely for educational purposes. Every effort has been made to trace the copyright holders of material reproduced in this book. Should any infringement have occurred, the publishers and editors apologize and will be pleased to make the necessary corrections in future editions of this book.

Notes

COURSE INTRODUCTION

Mathematical statistics provides the foundation for data analysis, decision-making, and inference in various fields. This course introduces fundamental statistical concepts, probability theory, probability distributions, and hypothesis testing. Understanding these concepts is essential for statistical modeling, research, and real-world applications in science, business, and engineering.

Module 1: Probability Theory

This module covers the definition and various approaches to probability, including the addition theorem, Boolean equality, conditional probability, multiplication theorem, and independent events. Students will also explore mutual and pairwise independence of events and applications of Bayes' theorem.

Module 2: Random Variables and Probability Functions

Students will learn about random variables, including discrete and continuous types, probability mass and density functions, and distribution functions. The module also introduces bivariate random variables, joint, marginal, and conditional distributions, and mathematical expectation, variance, covariance, and moment-generating functions.

Module 3: Probability Distributions

This module explores important probability distributions, including discrete distributions (Uniform, Bernoulli, Binomial, Poisson, and Geometric) and continuous distributions (Uniform, Exponential, and Normal), along with their properties and applications.

Module 4: Hypothesis Testing

Students will be introduced to statistical hypothesis testing, including concepts such as parameters, statistics, sampling distribution, standard error, and null and alternative hypotheses. The module also covers simple and composite hypotheses.

Module 5: Tests of Significance

This module focuses on the critical region, level of significance, one-tailed and two-tailed tests, and two types of errors. Students will study large sample tests for single mean, single proportion, and differences between two means and two proportions.

MODULE 1

UNIT 1.1

Probability: Definition and various approaches of probability

Objectives

- To understand the concept of likelihood and its different approaches.
- To learn addition and multiplication theorems of likelihood.
- To study Boole's inequality and its applications.
- To analyze conditional likelihood and its significance.
- To explore free and mutually free events.
- To apply Bayes' theorem in real-world problems.

1.1.1: Introduction to Likelihood

It provides a framework for measuring and quantifying the likelihood of events occurring in a given situation. The concept of likelihood is fundamental to many fields including statistics, physics, economics, computer science, and everyday decision-making.

Definition of Likelihood

An event's likelihood is a numerical indicator of how likely it is to happen. The number is always in the range of 0 to 1, inclusive:

- An occurrence is certain when its likelihood is 1.
- The degree of possibility of the event is indicated by a likelihood between 0 and 1.

Basic Terminology

Before diving deeper into likelihood theory, it's essential to understand some fundamental concepts:

1. **Experiment**: Any procedure that can be repeated and has a well-defined set of possible outcomes.

2. **Event (E)**: representing a collection of outcomes. For example, when rolling a die, the event "rolling an even number" would be E = {2, 4, 6}.

3. **Elementary Outcome**: An individual outcome in the sample space.

Mathematical Expression of Likelihood

With n(E) representing number of elements in event E & n(S) representing number of elements in sample space S, the likelihood of an event E in finite sample space with equally likely outcomes is as follows: P(E) = Number of favorable outcomes / Total number of possible outcomes = n(E) / n(S).

Properties of Likelihood

1. **Additivity**: For mutually exclusive events E & F (events that cannot occur simultaneously), P(E ∪ F) = P(E) + P(F)

2. **Complement Rule**: P(E') = 1 - P(E)

The Role of Likelihood in Decision Making

Likelihood is crucial for making informed decisions under uncertainty. individuals and organizations can make more rational choices based on expected values and risk assessments.

1.1.2: Approaches to Likelihood

There are several fundamental approaches to defining and interpreting likelihood, each with its own perspective and applications. These approaches provide different ways to understand and calculate probabilities in various contexts.

Classical Approach

The classical approach, also known as the a priori approach, defines likelihood based on equally likely outcomes.

Definition: equally likely outcomes. Definition: The likelihood of event E is $P(E) = m/n$, which is number of favorable outcomes divided by total number of possible outcomes, if event E includes m of the n equally likely outcomes of an experiment. **Presumptions:** Every possible scenario has an equal chance of happening.

Examples: Coin tosses, dice rolls, card games, and most gambling scenarios where the underlying physical mechanisms produce essentially at random results.

Limitations:

1. It only applies when outcomes are equally likely.
2. It cannot be applied to infinite sample spaces.
3. The concept of "equally likely" is somewhat circular in definition.

Relative Frequency Approach

The relative frequency approach, also known as the a posteriori approach or empirical approach, defines likelihood based on observed data from repeated experiments.

Definition: When an experiment is conducted n times with the same parameters and event E happens m times, the relative frequency m/n gets closer to the likelihood P(E) as n gets closer to infinity.:

$P(E) = \lim_{n\to\infty} m/n$

Applications:

1. Used in statistical studies and data analysis.
2. Useful when theoretical probabilities are difficult to determine.
3. Forms the basis for frequentist statistics.

Limitations:

1. Requires a large number of repetitions for accuracy.
2. Cannot be used for one-time events.
3. Practical constraints may prevent truly identical repetitions.

Subjective Approach

The subjective approach defines likelihood as a measure of personal belief or confidence in the occurrence of an event.

Definition: Likelihood is a numerical measure of a person's degree of belief that an event will occur, based on their knowledge and experience.

Notes

Features:

1. Different individuals may assign different probabilities to the same event.
2. Probabilities can be updated as new information becomes available (Bayesian approach).
3. Used for events that cannot be repeated or where data is limited.

Applications:

1. Decision-making in business and policy.
2. Risk assessment in unique situations.
3. Bayesian statistical inference.

Limitations:

1. Subjective nature may lead to inconsistencies.
2. Difficult to validate objectively.

Axiomatic Approach

The axiomatic approach, developed by Andrey Kolmogorov in the 1930s, provides a mathematical foundation for likelihood theory based on set theory.

Kolmogorov's Axioms:

1. $P(A) > 0$ (non-negativity) for any event A
2. entire sample space has a likelihood of 1 ($P(S) = 1$).
3. total of probabilities of two mutually exclusive events, A_1, A_2,..., indicates the possibility that they will come together: $P(A_1) \cup A_2 \cup ..... = P(A_1) + P(A_2) + ...$
   All other likelihood principles can be derived from this method's strict mathematical base. It allows for the development of likelihood theory as a subfield of measure theory and unifies the many approaches to likelihood.

Comparing the Approaches

Each approach has its strengths and contexts where it is most appropriate:

- The classical approach works well for simple games of chance with known structures.
- The relative frequency approach is useful for empirical studies and statistical analysis.
- The subjective approach helps with decision-making when data is limited or for one-time events.
- The axiomatic approach provides the mathematical foundation that unifies all other approaches.

In practice, these approaches are often complementary rather than competing. The choice of approach depends on the specific problem, available information, and the purpose of the likelihood calculation.

1.1.3: Addition Theorem of Likelihood

A key idea in likelihood theory that explains how to determine the likelihood of an event joining together is the addition theorem. It offers a way to calculate the likelihood that at least one of a number of occurrences will take place.

Applications of the Addition Theorem

1. **Risk Assessment**: Calculating the likelihood of system failure when there are multiple potential failure points.
2. **Medical Diagnosis**: Finding the likelihood that a patient has at least one of several possible conditions based on symptoms.
3. **Financial Planning**: Assessing the likelihood of achieving financial goals through different investment strategies.
4. **Project Management**: Calculating the likelihood of project completion by a deadline when considering various possible delays.

The Complement Method

Using an event's complement can sometimes make calculating its likelihood easier, particularly when working with "at least one" scenarios.

For an event A, the complement method uses:

$P(A) = 1 - P(A')$

Notes

where the complement of A is A'.

This is particularly useful when calculating the likelihood of "at least one success" by first finding the likelihood of "no successes" and then subtracting from 1.

1.1.4: Boole's Inequality and Its Applications

An upper bound for the likelihood of the union of occurrences is provided by Boole's inequality, sometimes referred to as the union bound. It is a fundamental finding in likelihood theory. It bears the name of George Boole, a mathematician.

Statement of Boole's Inequality

Boole's inequality says that for any finite or countably infinite sequence of events $A_1, A_2, \ldots, A_n$, $P(A_1 \cup A_2 \cup \ldots \cup A_n) < P(A_1) + P(A_2) + \ldots + P(A_n)$

In notation for mathematics:

$$P(\bigcup_{i=1}^{n} A_i) \leq \sum_{i=1}^{n} P(A_i)$$

Proof and Intuition

Boole's inequality follows directly from the inclusion-exclusion principle. Terms for every potential intersection are included in the whole inclusion-exclusion calculation for the likelihood of a union:

$$P(\bigcup_{i=1}^{n} A_i) = \sum P(A_i) - \sum P(A_i \cap A_j) + \sum P(A_i \cap A_j \cap A_k) - \ldots + (-1)^{n+1} P(A_1 \cap A_2 \cap \ldots \cap A_n)$$

Since all likelihood values are non-negative, dropping the negative terms yields an upper bound:

$$P(\bigcup_{i=1}^{n} A_i) \leq \sum P(A_i)$$

If &only if the events are mutually exclusive, then this disparity turns into equality.

Key Properties

1. **Sharpness**: The bound is tight (equality holds) when the events are mutually exclusive.
2. **Monotonicity**: Adding more events can only increase the bound.

3. **Conservation of Likelihood Mass**: The bound can exceed 1, which is impossible for an actual likelihood. This happens when there is significant overlap between events.
4. **Relationship to Addition Theorem**: Boole's inequality is a simplification of the addition theorem when the intersection terms are unknown or difficult to calculate.

Applications of Boole's Inequality

1. **Error Likelihood Bounds**: In communication systems, Boole's inequality helps establish upper bounds on the likelihood of error when multiple types of errors can occur.
2. **Multiple Hypothesis Testing**: In statistics, it's used to control the family-wise error rate, providing a bound on the likelihood of making at least one false discovery among multiple hypotheses.
3. **System Reliability**: For complex systems with multiple failure modes, Boole's inequality bounds the overall system failure likelihood.
4. **Algorithm Analysis**: In randomized algorithms, it helps analyze the likelihood of failure when multiple failure conditions exist.
5. **Risk Assessment**: When evaluating the risk of complex scenarios with multiple potential hazards, Boole's inequality provides a conservative estimate of the overall risk.

Bonferroni Correction

A common application of Boole's inequality is the Bonferroni correction in multiple hypothesis testing. If you conduct n free statistical tests at a significance level α, the likelihood of at least one false positive (Type I error) is bounded by $n \cdot \alpha$ according to Boole's inequality.To maintain an overall significance level α for the entire family of tests, each individual test should be conducted at a significance level of α/n. This is known as the Bonferroni correction.

Example of Boole's Inequality

Consider a system with three components, each with the following failure probabilities:

Notes

- Component 1: $P(A_1) = 0.05$
- Component 2: $P(A_2) = 0.03$
- Component 3: $P(A_3) = 0.04$

What is the maximum likelihood that at least one component will fail?
Using Boole's inequality:

$$P(A_1 \cup A_2 \cup A_3) \leq P(A_1) + P(A_2) + P(A_3) \leq 0.05 + 0.03 + 0.04 \leq 0.12$$

So, the likelihood of at least one component failing is at most 0.12 or 12%.

Limitations and Refinements

While Boole's inequality is simple to apply and requires minimal information (just the individual probabilities), it can be quite loose when events have significant overlap. In such cases, more refined bounds like the Bonferroni inequalities or the Hunter-Worsley bound might provide tighter results by incorporating information about pairwise intersections.

1.1.5: Conditional Likelihood and Multiplication Theorem

Conditional likelihood is the likelihood that an event will occur provided that another event has already occurred. This idea is central to likelihood theory and serves as the foundation for Bayesian statistics as well as numerous applications in engineering, science, and decision-making.

Definition of Conditional Likelihood

The conditional likelihood of event B given that event A has taken place, denoted by P(B|A), is defined as follows:

$$P(B|A) = P(A \cap B) / P(A),$$

where P(A) is the likelihood that event A will occur and P(A) is the likelihood that both events A &B will occur.

.

• P(A) > 0 This formula can be seen as the percentage of event A outcomes that are also event B outcomes.

Intuitive Understanding

Notes

A conditional likelihood is a likelihood that has been adjusted in light of fresh data. After discovering that event A has taken place, we limit our sample space to just the results of A.

The likelihood of event B in this new limited sample space is denoted by P(B|A).

The Theorem of Multiplication

The multiplication theorem can be obtained by rearranging Conditional likelihood definition:

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

This theorem allows us to calculate the likelihood of two events occurring at the same time by multiplying the chance of one event by the conditional likelihood of the second event provided beforehand.
The multiplication theorem extends to the following for numerous events:

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

Likelihood Chain Rule

The chain rule is the generic version of the multiplication theorem for n events:

$$P(\bigcap_{i=1}^{n} A_i) = P(A_1) \times P(A_2|A_1) \times P(A_3|A_1 \cap A_2) \times \ldots \times P(A_n|A_1 \cap A_2 \cap \ldots \cap A_{n-1})$$

When determining the combined likelihood of a series of events, this rule is crucial.

Separate Occurrences If the likelihood of one event does not change when the other occurs, then occurrences A & B are free.

A and B are mathematically free if & only if P(B|A) = P(B).

The equivalent expression for independence is $P(A \cap B) = P(A) \times P(B)$.

Mutual independence for multiple events necessitates the independence of each subset of events.
Rule of Multiplication for Free Events Given the independence of events A_1, A_2,..., & A_n, $P(A_1 \cap A_2 \cap \ldots \cap A_n) = P(A_1) \times P(A_2) \times \ldots \times P(A_n)$

Notes

This makes figuring out combined likelihood for individual events easier. The Bayes theorem provides a way to update probabilities in light of new data and is based on the idea of conditional likelihood..

$P(A|B) = [P(B|A) \times P(A)] / P(B)$ for occurrences A and B, where:

• posterior likelihood, or P(A|B), is likelihood that A given B

• likelihood, or likelihood of B given, is P(B|A). A

• prior likelihood is P(A), which is initial likelihood of A.

• The marginal likelihood, or the overall likelihood of B, is denoted by P(B).

In machine learning, medical diagnosis, and numerous other fields where probabilities must be adjusted in light of new information, Bayes' theorem is essential.

Law of Total Likelihood

According to law of total likelihood, for each event A, $P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) +... + P(A|B_n)P(B_n)$ if events $B_1, B_2,..., B_n$ constitute partition of sample space (they are mutually exclusive & their union is entire sample space).

This law assists in determining an event's overall likelihood by taking into account all potential circumstances.

Applications of Conditional Likelihood

1. **Medical Diagnosis**: Interpreting test results based on disease prevalence.
2. **Weather Forecasting**: Predicting tomorrow's weather given today's conditions.
3. **Risk Assessment**: Evaluating the likelihood of accidents under specific circumstances.
4. **Finance**: Estimating future market movements based on current conditions.
5. **Machine Learning**: Algorithms like Naive Bayes classifiers rely on conditional probabilities.

Solved Problems

Problem 1: Classical Likelihood with Playing Cards

Notes

Each of four suits (hearts, diamonds, clubs, & spades) in typical 52-card deck has 13 cards (Ace, 2–10, Jack, Queen, and King). Clubs and spades are black cards, whereas diamonds and hearts are red cards.

What is the likelihood of drawing a face card (Jack, Queen, or King) or a red card if you pull one card at at random from the deck?

Answer: Let R be the occurrence of a red card drawing. Let F be the face card drawing event.
We must determine P(R ∪ F).

Step 1: Determine the likelihood of drawing a red card, or P(R). The deck has 26 red cards (13 diamonds and 13 hearts). P(R) = 26/52 = 1/2

Step 2: Determine the likelihood of drawing a face card, or P(F). The deck has twelve face cards: four Jacks, four Queens, and four Kings. P(F) = 12/52 = 3/13

Step 3: Determine P(R ∩ F), or the likelihood of drawing a face card and a red card. Six cards—the Jack, Queen, King of Hearts, and Diamonds—are both red and face cards. 3/26 = 6/52 = P(R ∩ F)

Use the addition theorem in step four. 1/2 + 3/13 - 3/26 = 13/26 + 6/26 - 3/26 = 16/26 = 8/13 P(R ∪ F) = P(R) + P(F) - P(R ∩ F)

Consequently, the odds of drawing a face card or a red card are 8/13, or roughly 0.615.
Issue 2: Medical Testing's Conditional Likelihood

Five percent of people suffer from a particular ailment. A test for this illness has a 90% sensitivity, which means it can accurately identify 90% of those who have it, and an 80% specificity, which means it can accurately identify 80% of those who do not.
What is the likelihood that an individual truly has the disease if they test positive?
Answer: Let D be the occurrence of the illness. Let T+ be the result of a positive test.
P(D|T+), the likelihood of having the disease in the event of a positive test, must be determined.
Given the following data, P(D) = 0.05 (disease prevalence), P(T+|D) = 0.90 (sensitivity), and P(T-|D') = 0.80 (specificity), P(T+|D') = 0.20

Notes

First, use the Bayes theorem. [P(T+|D) × P(D)] / P(T+) = P(D|T+) Step 2: Apply law of total likelihood to find P(T+). 0.045 + 0.19 = 0.235 = 0.90 × 0.05 + 0.20 × 0.95 = P(T+) = P(T+|D) × P(D) + P(T+|D') × P(D').

Step 3: Apply the Bayes theorem to calculate P(D|T+). 0.90 × 0.05 / 0.235 = 0.045 / 0.235 = 0.1915, or around 19.15%, is the value of P(D|T+).

As a result, the likelihood that an individual has the condition is only 19.15% if they test positive. The significance of taking the base rate (prevalence) into account when interpreting test results is demonstrated by this example.

Issue 3: Multiple Event Addition Theorem
Each of the three parts that make up a software system has the following odds of failing in a 24-hour period:

Components A and B: 0.03 and 0.04, respectively
Component C: 0.02

There is a 0.005 chance that components A and B will both fail. There is a 0.004 chance that components A and C will both fail. There is a 0.006 chance that components B and C will both fail. All three components have a 0.001 chance of failing.
In a 24-hour period, what is the likelihood that at least one component would fail?
Solution: Assume that components A, B, and C fail as a result of the following circumstances. We have to determine P(A ∪ B ∪ C).

Given:
values of P(A) = 0.03, P(B) = 0.04, P(C) = 0.02, P(A ∩ B) = 0.005, P(A ∩ C) = 0.004, & P(B ∩ C) = 0.006

• 0.001 is P(A ∩ B ∩ C).

Three events are examined using inclusion-exclusion principle: The formula for P(A ∪ B ∪ C) is P(A) + P(B) + P(C) - P(A ∩ B) - P(A ∩ C) - P(B ∩ C) + P(A ∩ B ∩ C).

P(A ∪ B ∪ C) = 0.03 + 0.04 + 0.02 - 0.005 - 0.004 - 0.006 + 0.001 = 0.09 - 0.015 + 0.001 = 0.076 or 7.6%

Notes

Consequently, there is a 7.6% chance that at least one component will malfunction within a 24-hour period.

Issue 4: Multiplication Theorem and Free Events
Three times, fair die is rolled.

What is likelihood of receiving a number less than four on third roll, an even number on second, and six first?

Answer: Let A represent the chance of receiving a 6 on the first roll. Let B represent the chance of receiving an even number on the subsequent roll. Let C be the occurrence of a number on the third roll that is less than 4.

Finding $P(A \cap B \cap C)$ is necessary.

We can apply the multiplication rule for separate events as the three rolls are free events: $P(A) \times P(B) \times P(C) = P(A \cap B \cap C)$

Step 1: Calculate P(A), or the likelihood of receiving a 6 on the first roll. $P(A) = 1/6$

Step 2: Determine P(B), or the likelihood that the second roll will provide an even number. On a die, the even numbers are 2, 4, and 6 (three possible outcomes). $P(B) = 3/6 = 1/2$

Step 3: Determine P(C), or the likelihood that the third roll will yield a number smaller than 4.

The numbers 1, 2, and 3 (three outcomes) are fewer than 4. $P(C) = 3/6 = 1/2$
Use the multiplication theorem in step four. $P(A) \times P(B) \times P(C) = (1/6) \times (1/2) \times (1/2) = 1/24$ or around 0.0417

Consequently, there is a 1/24 or around 4.17% chance of receiving a 6 on the first roll, an even number on the second roll, and a number less than 4 on the third roll.

Conditional Likelihood Chain Rule Problem No. 5

Three blue and five red balls are in a bag. Without a replacement, two balls are drawn.

What is the likelihood that both balls will be red?

Notes

Answer: When first ball is red, let R_1 = event. Let R_2 be chance that second ball turns red.

Finding $P(R_1 \cap R_2)$ is necessary.

Applying the theorem of multiplication: $P(R_1) \times P(R_2|R_1) = P(R_1 \cap R_2)$

Step 1: Determine the likelihood that the first ball will be red, or $P(R_1)$. $P(R_1) = 5/8$

Step 2: Given that the first ball is red, find $P(R_2|R_1)$, the likelihood that the second ball is also red. There are four red balls & three blue balls remaining after drawing red ball, for a total of seven balls. $P(R_2|R_1) = 4/7$

Use the multiplication theorem in step three. $P(R_1$ deviates from $R_2) = P(R_1) \times P(R_2|R_1) = (5/8) \times (4/7) = 20/56 = 5/14$, or roughly 0.357

Thus, there is a 5/14, or around 35.7%, chance of drawing two red balls.

Unsolved Problems

Problem 1: Classical Likelihood

Five times, a fair coin is tossed. How likely is it that at least three heads will appear?
Second Issue: The Addition Theorem

Thirty of the fifty students in the class are enrolled in mathematics, twenty-five are enrolled in physics, and ten are enrolled in both. What is the likelihood that a randomly chosen student will be enrolled in either mathematics, physics, or both?

Conditional Likelihood in Problem 3

Eight blue socks and ten red socks are in a drawer. Two socks are chosen at at random and aren't replaced. Given that the first sock drawn was red, what is the likelihood that the second sock will be blue?

Problem 4: Boole's Inequality Application

A security system has four sensors, each with the following probabilities of false alarm:

- Sensor 1: 0.02

- Sensor 2: 0.03
- Sensor 3: 0.01
- Sensor 4:

Notes

UNIT 1.2

Addition theorem, Boolean equality, Conditional probability and multiplication theorem, Independent events

1.2.1: Free and Mutually Free Events

Introduction to Free Events

According to likelihood theory, two events are free if their occurrences have no bearing on each other's probabilities. Put differently, the fact that one event has happened doesn't tell us anything more about whether the other will.

Mathematical Definition of Free Events

A & B are two separate events if and only if:

$P(A) \times P(B) = P(A \cap B)$
Where:

• $P(A \cap B)$ is likelihood that occurrences A and B will occur.

• P(A) is likelihood that event A will occur.

• P(B) is the likelihood that event B will occur.

This formula serves as both the definition and the test for independence of two events.

Alternative Formulation

$P(A|B) = P(A)$ is another way to declare independence if $P(B) > 0$.

conditional likelihood of given that B has occurred is denoted by P(A|B).

Likewise, $P(B|A) = P(B)$ if $P(A) > 0$.

These formulations highlight that the likelihood of one event remains unchanged regardless of whether the other event has occurred.

Mutually Free Events

It is possible to apply the idea of independence to more than two occasions. If each of three or more occurrences occurs freely of any combination of the others, they are said to be mutually free (or jointly free).

Definition of Mutual Independence in Mathematics

If and only if likelihood of intersection for each subset of events A_1, A_2,..., A_n is equal to product of probabilities of individual events, then these events are mutually free.

For instance, all of the following requirements must be met for three events, A, B, and C, to be mutually free:

1. $P(A) \times P(B) = P(A \cap B)$
2. $P(A) \times P(C) = P(A \cap C)$
3. $P(B) \times P(C) = P(B \cap C)$
4. $P(A) \times P(B) \times P(C) = P(A \cap B \cap C)$

   It's crucial to remember that mutual independence is not always implied by pairwise independence, in which each pair of occurrences is free.

Examples of Free Events

1. **Coin Tosses**: The results of separate coin tosses are free events.
2. **Die Rolls**: Each roll of a die is free of previous rolls.
3. **At random Selection from Different Groups**: Selecting a at random male from a population and selecting a at random person with blue eyes are free if gender and eye color are free characteristics in the population.

Examples of Dependent Events (Not Free)

1. **Card Draws without Replacement**: When drawing cards without replacement, each draw depends on the previous draws.
2. **Weather Conditions**: Today's weather and tomorrow's weather are typically dependent events.
3. **Stock Market Movements**: Price movements of related stocks are usually dependent.
4. **Health Outcomes**: Health outcomes for family members may be dependent due to shared genetics and environment.

Importance of Independence in Likelihood

The concept of independence is crucial in likelihood theory and statistics for several reasons:

Notes

1. **Statistical Inference**: Many statistical methods assume independence of observations.
2. **Likelihood Models**: Many likelihood models (like the binomial distribution) are built on the assumption of independence.
3. **Risk Assessment**: In risk analysis, understanding whether risks are free is essential for accurate risk aggregation.

UNIT 1.3
Mutual and pairwise independence of events, Bayes theorem and its applications

1.3.1: Bayes' Theorem and Its Application

Introduction to Bayes' Theorem

A key finding in likelihood theory, Bayes' Theorem explains how to update a hypothesis's likelihood in light of fresh data. It offers a mathematical guideline for updating preexisting hypotheses or forecasts in light of fresh or more data.

Bayes' Theorem in Mathematical Form

According to Bayes' Theorem, $P(A|B) = [P(B|A) \times P(A)]$ for occurrences A and B where $P(B) > 0$.

Where:

• posterior likelihood, or P(A|B), is likelihood that event A will occur given that B has occurred.

• The prior likelihood, or starting likelihood of event A, is denoted by P(A).

A Different Formulation Law of Total Likelihood in Action

law of total likelihood can be used to enlarge the denominator P(B):

$[P(B|A) \times P(A)] = P(A|B)$

The formula is $P(B|A) \times P(A) + P(B|A^c) \times P(A^c)]$.

where A^c is event's complement.

Version of Multiple Hypotheses

Bayes Theorem can be stated as follows when working with several mutually exclusive and exhaustive hypotheses:

$$P(A_i|B) = [P(B|A_i) \times P(A_i)] / \sum_j P(B|A_j) \times P(A_j)$$

Key Components of Bayes' Theorem

1. **Prior Likelihood P(A)**: The initial degree of belief in A before the evidence B is considered.
2. **Likelihood P(B|A)**: How probable the evidence B is, assuming the hypothesis A is true.

Notes

Intuitive Understanding of Bayes' Theorem

Bayes' Theorem can be understood as a method for updating beliefs based on new evidence:

1. Start with a prior belief P(A)
2. Observe new evidence B
3. Consider how likely the evidence would be if A were true P(B|A)
4. Update the belief to obtain the posterior likelihood P(A|B)

Applications of Bayes' Theorem

1. Medical Diagnosis

Bayes' Theorem is used to calculate likelihood of disease given positive test result:

P(Disease|Positive Test) = [P(Positive Test|Disease) × P(Disease)] / P(Positive Test)

This calculation helps understand the true diagnostic value of medical tests, accounting for factors like:

- true positive rate
- true negative rate
- Prevalence of the disease in the population

2. Spam Filtering

Email spam filters often use Bayesian methods to classify messages:

P(Spam|Words) = [P(Words|Spam) × P(Spam)] / P(Words)

The system learns from training data which words are more commonly found in spam versus legitimate emails, and updates its classification accordingly.

3. Machine Learning and AI

Bayesian methods are foundational in many machine learning algorithms:

- Naïve Bayes classifiers
- Bayesian networks

- Bayesian inference in probabilistic models

Notes

4. Risk Assessment and Decision Making

Bayes' Theorem helps update risk assessments as new information becomes available:

- Financial risk models
- Insurance pricing
- Project management risk assessment

5. Forensic Evidence Analysis

In legal settings, Bayes' Theorem can help evaluate the strength of forensic evidence:

$P(Guilty|Evidence) = [P(Evidence|Guilty) \times P(Guilty)] / P(Evidence)$

6. Quality Control

In manufacturing, Bayesian methods help update beliefs about product quality based on sample inspections.

The Bayesian Approach to Likelihood

Bayes' Theorem reflects a broader philosophical approach to likelihood:

- **Frequentist View**: Likelihood represents the long-run frequency of events in repeated trials.
- **Bayesian View**: Likelihood represents a degree of belief that can be updated based on new evidence.

The Bayesian approach treats likelihood as subjective and allows for:

- Incorporation of prior knowledge
- Sequential updating as new data arrives
- Quantification of uncertainty

Common Misconceptions and Challenges

1. **Base Rate Fallacy**: People often neglect the prior likelihood (base rate) when making judgments based on new evidence.

Notes

2. **Appropriate Prior Selection**: Choosing appropriate prior probabilities can be challenging and sometimes controversial.
3. **Computational Complexity**: For complex problems, the calculations required by Bayes' Theorem can be computationally intensive.

Solved Problems

Problem 1: Free Events - Coin and Die

A fair six-sided die is rolled, and fair coin is tossed. How likely is it to obtain an even number on the die and a head on the coin?

Solution:

Let's define our events:

- Event A: Getting a head on the coin
- Event B: Getting an even number on the die

Step 1: Find P(A) For a fair coin, P(A) = P(Head) = 1/2 = 0.5

Step 2: Find P(B) Even numbers on a six-sided die are 2, 4, and 6. P(B) = P(Even number) = 3/6 = 1/2 = 0.5

Step 3: We may use the multiplication formula for free events as the coin toss and die roll are separate occurrences: The formula for $P(A \cap B)$ is $P(A) \times P(B) = 0.5 \times 0.5 = 0.25$. Consequently, there is a 0.25 or 25% chance of getting a head on the coin and an even number on the die..

Problem 2: Testing for Independence

A survey found that among 200 students, 120 play basketball, 80 play football, and 40 play both sports. Are the events "playing basketball" and "playing football" free?

Solution:

Let's define our events:

- Event A: A student plays basketball
- Event B: A student plays football

Step 1: Find the probabilities P(A) = Number of students who play basketball / Total number of students P(A) = 120/200 = 0.6

P(B) = Number of students who play football / Total number of students P(B) = 80/200 = 0.4

P(A ∩ B) = Number of students who play both sports / Total number of students P(A ∩ B) = 40/200 = 0.2

Step 2: Determine whether P(A ∩ B) = P(A) × P(B) = 0.6 × 0.4 = 0.24 to test for independence.
The events "playing basketball" and "playing football" are not free as P(A ∩ B) = 0.2 ≠ 0.24 = P(A) × P(B).

Actually, there is a negative connection between both activities because P(A ∩ B) < P(A) × P(B).

This means that students who participate in one activity are less likely to participate in the other than we would anticipate if the choices were free.

Unsolved Problems

Problem 1: Free Events - Card Drawing

Each of four suits (hearts, diamonds, clubs, & spades) in typical 52-card deck has 13 cards (Ace, 2–10, Jack, Queen, and King). Clubs and spades are black cards, whereas diamonds and hearts are red cards.

The deck is shuffled and two cards are drawn. Given that both cards are face cards (Jack, Queen, or King), determine the likelihood that both are kings.

Problem 2: Testing Independence in a 2×2 Contingency Table

A study surveyed 500 adults about their coffee and tea consumption habits, with the following results:

	Drinks Coffee	**Doesn't Drink Coffee**	**Total**
Drinks Tea	120	180	300
Doesn't Drink Tea	140	60	200
Total	260	240	500

Are the events "drinking coffee" and "drinking tea" free? Justify your answer with calculations.

Problem 3: Bayes' Theorem - Email Spam Filter

Notes

A spam filter has the following characteristics:

- 98% of spam emails are correctly identified as spam (P(Flagged|Spam) = 0.98)
- 5% of non-spam emails are incorrectly flagged as spam (P(Flagged|Not Spam) = 0.05)
- 40% of all emails received are spam (P(Spam) = 0.4)

If an email is flagged as spam by the filter, what is the likelihood that it is actually spam?

Problem 4: Bayes' Theorem - Sequential Testing

A rare genetic condition affects 1 in 10,000 people in a population (P(Disease) = 0.0001). A genetic test for this condition has a sensitivity of 99% (P(Positive|Disease) = 0.99) and a specificity of 99.9% (P(Negative|No Disease) = 0.999).

Problem-Solving Using Likelihood Theorems

Likelihood theorems provide powerful tools for solving complex problems involving uncertainty and randomness. Let's explore these theorems in depth with clear explanations, formulas, solved examples, and practice problems.

Key Likelihood Theorems and Formulas

1. The Rule of Addition in Likelihood

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$ for any two occurrences A & B.

For events that are mutually exclusive: P(A) + P(B) equals $P(A \cup B)$

2. The Rule of Likelihood for Multiplication

For any pair of occurrences A & B: $P(A) \times P(B|A) = P(A \cap B)$

For separate occurrences: $P(A) \times P(B) = P(A \cap B)$

3. Likelihood under Conditions

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

4. The Law of Total Likelihood

If sample space S is divided into events $B_1, B_2,..., B_n$, then $P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) +... + P(A|B_n)P(B_n)$

5. The Bayes Theorem

$[P(A|B) \times P(B)] / P(A) = P(B|A)$

As an alternative, apply the law of total likelihood: $[P(A|B) \times P(B)] =$

P(B|A) The formula is P(A|B) × P(B) + P(A|B') × P(B')

6. Complementary Occasions

P(A') = 1 - P(A)

7. The likelihood of at least one occurrence

P(none of the events) - P(at least one of n events) = 1

Solved Problems

Solved Problem 1: medical testing

When administered to an individual with the disease, a medical test has a 98% chance of producing a positive result; when administered to an individual without the ailment, the likelihood is 3%. Assume that the disease affects 0.5% of the population.

Solution:

Let's define our events:

- D: Person has the disease
- D': Person does not have the disease
- T+: Test result is positive
- T-: Test result is negative

We are provided with:
The test sensitivity is P(T+|D) = 0.98, & false positive rate is P(T+|D') = 0.03.
• disease prevalence, P(D), is 0.005.
• chance of not having disease is P(D') = 0.995.
P(D|T+), or the likelihood that a person has disease if they tested positive, is what we're looking for.
Applying the Bayes Theorem: The formula is [P(T+|D) × P(D)] = P(D|T+) P(T+|D) × P(D) + P(T+|D') × P(D')]
Changing the values: 0.98 × 0.005 / [(0.98 × 0.005) + (0.03 × 0.995)] = P(D|T+) 0.0049 / [0.0049 + 0.02985] is P(D|T+). A = 0.0049 / 0.03475 P(D|T+) P(D|T+) is approximately 14.1%, or 0.141.
As a result, there is a 14.1% chance that a randomly chosen individual who tests positive indeed has the illness.
Problem 2: Card Drawing Solved
A conventional 52-card deck is used, and two cards are drawn consecutively

Notes

without replacement. How likely is it that both cards are aces?

Answer:

Let's specify what our events are:

• A: An ace is first card.

• A: An ace is second card.

Our goal is to locate $P(A_1 \cap A_2)$.

Applying the rule of multiplication to dependent events: $P(A_1 \cap A_2) = P(A_1) \times P(A_2|A_1)$

A normal 52-card deck has four aces. $P(A_1) = 4/52 = 1/13$

Three aces remain out of 51 cards after one ace is drawn. $1/17 = 3/51 = P(A_2|A_1)$

Consequently, $P(A_1 \cap A_2) = (1/13) \times (1/17) = 1/(13\times17) = 1/221 \approx 0.00452$ or almost 0.452%

resolved Issue 3: Likelihood and Three Friends

Charlie, Ben, and Alex, three buddies, show up for a party on their own. Alex has a 0.7 chance of going, Ben has a 0.6 chance, and Charlie has a 0.8 chance. How likely is it that: a) All three will be at the party? b) Does at least one of them show up for the celebration? c) There are precisely two of them at the celebration?

Solution:

Let's define our events:

- A: Alex attends the party, P(A) = 0.7
- B: Ben attends the party, P(B) = 0.6
- C: Charlie attends the party, P(C) = 0.8

We can apply the multiplication rule for free events because arrivals are free. a) The likelihood that all three will show up: 0.7 × 0.6 × 0.8 = 0.336 or 33.6%. b) = P(A ∩ B ∩ C) × P(B) × P(C) The likelihood that one or more people will attend: The complement of the likelihood that no one shows up is this: P(none attend) - P(at least one) = 1. 0.3 × 0.4 × 0.2 = 0.024. P(none attend) = P(A' ∩B' ∩ C') = P(A') × P(B') × P(C') = (1-0.7) × (1-0.6) × (1-0.8)

With that in mind, P(at least one) = 1 - 0.024 = 0.976 or 97.6%. The likelihood that precisely two will be present: Three things can cause this: (A ∩ B' ∩C) ∪ (A' ∩ B ∩C) ∪ (A ∩ B ∩ C')

The formula P(exactly two) is P(A ∩ B ∩ C') + P(A ∩ B' ∩ C) + P(A' ∩ B ∩ C). P (two precise) =(P(A) × P(B) × P(C')) = P(A) × P(B') × P(C) = P(A') × P(B) × P(C)]. P(exactly two) = 0.084 + 0.224 + 0.144 = 0.452 or 45.2% P(exactly two) = [0.7 × 0.6 × 0.2] + [0.7 × 0.4 × 0.8] + [0.3 × 0.6 × 0.8].

Solved Problem 4: Email Spam Filter

A spam filter is designed to identify unwanted emails. In a large sample of emails, it was found that:

- 30% of all emails are spam
- The filter correctly identifies spam emails with a likelihood of 0.95
- The filter incorrectly marks legitimate emails as spam with a likelihood of 0.05

a) What is the likelihood that an email flagged as spam is actually spam? b) What is the likelihood that an email that passes the filter is actually legitimate?

Solution:

Let's define our events:

- S: Email is spam, P(S) = 0.3
- L: Email is legitimate, P(L) = 0.7
- F: Filter flags email as spam
- P: Email passes the filter (not flagged)

We are given:

- P(F|S) = 0.95 (true positive)
- P(F|L) = 0.05 (false positive)
- P(P|S) = 0.05 (false negative)
- P(P|L) = 0.95 (true negative)

Solved Problem 5: Safety Systems

A and B are two separate safety systems on a machine. System A detects malfunctions with a likelihood of 0.95 and system B detects them with a likelihood of 0.90.

Notes

a) How likely is it that at least one of the systems will notice a malfunction? b) How likely is it that both systems will notice a malfunction? c) How likely is it that a single system will identify a malfunction?

Solution:

Let's define our events:

- A: System A detects the malfunction, P(A) = 0.95
- B: System B detects the malfunction, P(B) = 0.90

a) Likelihood that at least one system detects the malfunction: We can use the addition rule: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Since the systems are free: $P(A \cap B) = P(A) \times P(B) = 0.95 \times 0.90 = 0.855$

Therefore: $P(A \cup B) = 0.95 + 0.90 - 0.855 = 0.995$ or 99.5%

Alternatively, we could compute this as the complement of neither system detecting the malfunction: $P(A \cup B) = 1 - P(A' \cap B') = 1 - [(1-0.95) \times (1-0.90)] = 1 - [0.05 \times 0.10] = 1 - 0.005 = 0.995$

b) Likelihood that both systems detect the malfunction: $P(A \cap B) = P(A) \times P(B) = 0.95 \times 0.90 = 0.855$ or 85.5%

c) Likelihood that exactly one system detects the malfunction: P(exactly one) = $P(A \cap B') + P(A' \cap B)$ P(exactly one) = $[P(A) \times P(B')] + [P(A') \times P(B)]$ P(exactly one) = $[0.95 \times (1-0.90)] + [(1-0.95) \times 0.90]$. P(exactly one) = $[0.95 \times 0.10] + [0.05 \times 0.90]$ P(exactly one) = 0.095 + 0.045 = 0.14 or 14%

Applying Likelihood Theorems Strategically

When solving likelihood problems, consider the following approach:

1. Clearly identify the events and their probabilities
2. Determine if events are free, dependent, mutually exclusive, etc.
3. Choose the appropriate theorem (addition rule, multiplication rule, Bayes' theorem, etc.)
4. Consider using complementary probabilities for "at least one" type problems

5. Break complex problems into simpler components

The Role of Likelihood Trees

Likelihood trees are visual tools that can help solve complex likelihood problems, especially those involving sequential events. Each branch in a tree represents a possible outcome, and probabilities are multiplied along paths.

For example, consider the medical testing problem (Solved Problem 1). We could draw a tree with:

- First branch: Disease (0.005) vs. No Disease (0.995)
- Second branches from Disease: Positive test (0.98) vs. Negative test (0.02)
- Second branches from No Disease: Positive test (0.03) vs. Negative test (0.97)

The likelihood of testing positive and having disease would be: P(D and T+) = P(D) × P(T+|D) = 0.005 × 0.98 = 0.0049

Geometric Likelihood

Some problems involve continuous likelihood where outcomes are points in space. Geometric likelihood often uses the principle:

P(E) = Favorable geometric measure / Total geometric measure

The measure could be length, area, volume, etc., depending on the context.

Solving Real-World Problems with Likelihood

Likelihood theory helps us model and make decisions in uncertain situations. In real-world applications, the key challenge is correctly identifying events, assigning appropriate probabilities, and determining the relationships between events.

Quality control, insurance, medical diagnosis, weather forecasting, and risk assessment all rely on likelihood calculations similar to the problems we've explored.

Multiple-Choice Questions (MCQs)

1. **The classical definition of likelihood is based on:**
   a) Experimentation

Notes

b) Equally likely outcomes

c) Subjective judgment

d) At random variations

Answer: b) Equally likely outcomes

2. **likelihood of at least 1 event occurring is found using:**

a) Addition theorem

b) Multiplication theorem

c) Bayes' theorem

d) None of the above

Answer: a) Addition theorem

3. **Boole's inequality states that:**

a) $P(A \cup B) \geq P(A)+P(B)$

b) $P(A \cap B)=P(A)P(B)$

c) $P(A \cup B) \leq P(A)+P(B)$

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

Answer: c) $P(A \cup B) \leq P(A)+P(B)$

4. **The multiplication theorem of likelihood states that:**

a) $P(A \cap B)=P(A)P(B)$

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

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

d) Both (b) and (c)

Answer: d) Both (b) and (c)

5. Two events **A & B are free** if:

a) $P(A|B)=P(A)$

b) $P(B|A)=P(B)$

c) $P(A \cap B)=P(A)P(B)$

d) All of the above

Answer: d) All of the above

6. **Bayes' theorem is used to:**

a) Compute conditional probabilities

b) Find joint probabilities

c) Determine prior and posterior probabilities

d) Both (a) and (c)

Answer: d) Both (a) and (c)

7. **likelihood of complement of an event A is given by:**

   a) $1-P(A)$

   b) $P(A^c)=P(A)$

   c) $P(A)+P(A^c)=0$

   d) None of the above

   Answer: a) $1-P(A)$

8. **If 2 events A & B are mutually exclusive, then:**

   a) $P(A\cap B)=0$

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

   c) $P(A|B)=0$

   d) All of the above

   Answer: d) All of the above

9. **law of total likelihoodstates that:**

   a) $P(B)=P(B|A_1)P(A_1)+P(B|A_2)P(A_2)+...$

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

   c) $P(A\cap B)=P(A)P(B)$

   d) None of the above

   Answer: a) $P(B)=P(B|A_1)P(A_1)+P(B|A_2)P(A_2)+...$

10. **A real-life application of Bayes' theorem is in:**

    a) Spam filtering

    b) Weather prediction

    c) Medical diagnosis

    d) All of the above

    Answer: d) All of the above

Short Answer Questions

1. Define likelihood and explain its different approaches.
2. What is the addition theorem of likelihood? Give an example.
3. Explain Boole's inequality with a practical application.

Notes

4. What is conditional likelihood? Provide an example.
5. Differentiate between free and mutually free events.
6. State and explain Bayes' theorem with an example.
7. Define mutually exclusive events and provide an example.
8. Explain the law of total likelihood with a real-life example.
9. How does Bayes' theorem help in medical diagnosis?
10. What is the importance of likelihood theory in decision-making?

Long Answer Questions

1. Explain the three different approaches to likelihood with examples.
2. Derive and explain the addition theorem of likelihood with an example.
3. Discuss Boole's inequality and its significance in likelihood theory.
4. Explain conditional likelihood and the multiplication theorem with real-life examples.
5. Discuss the concept of free and mutually free events in likelihood.
6. Derive Bayes' theorem and provide a step-by-step example of its application.
7. Solve a real-world problem using Bayes' theorem (e.g., spam filtering or medical diagnosis).
8. Explain how likelihood theory is used in risk assessment and decision-making.
9. Discuss common misinterpretations of likelihood in everyday life.
10. How is likelihood theory applied in artificial intelligence and machine learning?

UNIT 2.1

Random variable and probability functions: Definition and properties of random variables, Discrete and continuous random variables

Objectives

- To understand the concept of random variables and their types.
- To study likelihood mass functions (PMF) and likelihood density functions (PDF).
- To explore distribution functions and their properties.
- To learn about bivariate at random variables and their distributions.
- To define mathematical expectation, variance, and covariance.
- To study moment-generating functions and their applications.

2.1.1: Introduction to random Variables

What is a random Variable?

A random variable is one whose values are determined by the results of a random event. It gives us a method to assign numerical values to results of random experiments, enabling us to use mathematics to evaluate uncertain situations. Random variables serve as a bridge between likelihood theory and statistical analysis. While likelihood theory deals with the likelihood of events occurring, random variables allow us to quantify and analyze these outcomes.

Properties of random Variables

1.R ange: The collection of every numerical result that the random variable could provide

2.**L ikelihood Assignment**: Each value in the range has an associated likelihood, indicating how likely the random variable is to take that value.

Types of random Variables

Notes

Random variables come in two primary varieties:

1. Random variables that are discrete
2. Random variables that are continuous

These types differ in the nature of values they can take and how we calculate probabilities associated with them.

2.1.2: Discrete and Continuous random Variables

Discrete random Variables

Only a countable number of different values, such as integers or a finite set of values, can be assigned to a discrete random variable. There could be a countably infinite or finite number of potential values.

Characteristics:

- Takes distinct, separate values
- Can be counted
- Often represents counts, whole numbers, or categories converted to numbers
- Has gaps between possible values

Continuous random Variables

A continuous random variable can take any value within a range or interval. set of possible values is uncountable and forms a continuum.

Characteristics:

- Can take any value within a range
- Cannot be counted, only measured
- Often represents measurements like time, height, weight, temperature
- Has no gaps between possible values

Examples:

- Height of a person
- Time required to complete a task

- Amount of rainfall in a day
- Weight of a product

Key Differences

1. **Nature of Values**:
    - Discrete: Takes separate, distinct values
    - Continuous: Can take any value within a range
2. **Likelihood Calculation**:
    - Discrete: We can assign a likelihood to each specific value
    - Continuous: The likelihood of any exact single value is zero; we calculate probabilities for ranges of values
3. **Mathematical Representation**:
    - Discrete: Represented by Likelihood Mass Function (PMF)
    - Continuous: Represented by Likelihood Density Function (PDF)
4. **Cumulative Distribution**:
    - Discrete: The CDF has jumps at the possible values
    - Continuous: The CDF is a smooth curve without jumps

Mixed random Variables

Some random variables exhibit both discrete and continuous properties. These are called mixed random variables and have both discrete and continuous components in their distributions.For example, the amount of annual rainfall might be continuous for positive values but have a discrete likelihood mass at zero (for regions that might experience no rainfall in some years).

Notes

UNIT 2.2

Probability mass and density functions, Distribution function. Concepts of bivariate random variable: joint, marginal and conditional distributions. Mathematical expectation: Definition and its properties

2.2.1: Likelihood Mass Function (PMF)

Definition

The likelihood distribution of discrete at random variable is described by Likelihood Mass Function (PMF). PMF provides likelihood that discrete at random variable X will take on exact value x.

Notation

Usually, PMF is written as p(x) or P(X = x), where:

• chance that at random variable X will take value x is equal to p(x) = P(X = x).

Qualities of a Reputable PMF

In order for a function p(x) to be a legitimate PMF, it needs to meet:

1. Non-negativity: for any x, $p(x) \geq 0$. No outcome can have a negative likelihood.

2. Add up to 1: $\sum p(x) = 1$, where the total is the sum of all x's potential values. All conceivable outcomes must have a total likelihood of 1.

3. Domain Restriction: for any value x outside of the range of X, p(x) = 0. The only values with non-zero probabilities are those that can truly happen.

Likelihood Calculation with the PMF

We add the PMF over each value in A to get the likelihood that X will take a value in A: For every x in A, $P(X \in A) = \sum p(x)$.

To determine the likelihood that a dice roll is even, for instance: P(X is even) is equal to 2 + 4 + 6 = 1/6 + 1/6 + 1/6 = 1/2

Value Expected Using PMF

When all potential values of discrete at random variable X are added together, expected value (mean) is $E[X] = \sum x\ p(x)$.

Difference Making Use of PMF

$Var(X) = \sum (x - E[X])$ is variance of discrete at random variable X. sum of all possible values of x is $x^2\ p(x)$.

As an alternative: $E[X^2] - (E[X])^2 = \sum x^2 * p(x) = Var(X) - (E[X])^2$

2.2.2: Likelihood Density Function (PDF)

Definition

likelihood distribution of continuous at random variable is described by Likelihood Density Function (PDF). PDF does not provide probabilities directly, in contrast to PMF for discrete at random variables. Rather, the integral of the PDF over a given range provides the likelihood that a continuous at random variable will fall within that range.

Notation

The standard notation for PDF of continuous at random variable X is f(x).

Characteristics of Legitimate PDF

The following conditions must be met for function f(x) to be a valid PDF:

1. Non-negativity: for any x, $f(x) \geq 0$. Nowhere can the density function be negative.

2. When the integral is taken across the whole domain, the area equals 1: $\int f(x)\, dx = 1$. The PDF curve's entire area under the curve must be 1.

Analysis of the PDF

The likelihood that X = x is not provided by PDF f(x). likelihood of any single point for a continuous at random variable is always zero.

Rather, the "density" of likelihood close to x is represented by f(x). Approximately f(x) times interval width is the likelihood that X falls within small interval surrounding x.

More specifically:

- $P(x \leq X \leq x + \Delta x) = f(x)\ \Delta x$ for a narrow interval $[x, x + \Delta x]$.
- $P(a \leq X \leq b) = \int f(x)\, dx$, where the integral is evaluated from a to b, is precise likelihood that X falls in interval [a, b].

How to Determine Probabilities Making use of PDF

We integrate PDF throughout the range [a, b] to determine likelihood that X takes a value in that range: $P(a \leq X < b) = \int f(x)\, dx$, where the integral is evaluated from a to b.

For a uniform distribution across [0, 1], for instance: The expected value using PDF is $P(0.25 < X \leq 0.75) = \int 1\, dx = 0.75 - 0.25 = 0.5$.

If X is continuous at random variable, its expected value (mean) is: E[X] is the domain-wide integration of $\int x\ f(x)\, dx$.

Variance with PDF

A continuous at random variable X's variance is: $\int (x - E[X]) = Var(X)$

Notes

domain-wide integration of $x^2 * f(x)\, dx$

As an alternative: $E[X^2] - (E[X]) = Var(X) * f(x)\, dx - (E[X]) = \int x^2 f(x)dx$

2.2.3: Cumulative Distribution Function (CDF)

Definition

A at random variable X's Cumulative Distribution Function (or CDF) indicates the likelihood that X will take a value that is less than or equal to x. Both continuous and discontinuous at random variables are covered by CDF.

Notation

For at random variable X, the CDF is commonly represented as F(x):

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

Qualities of a Reputable CDF

To be valid CDF, a function F(x) needs to meet the following requirements:

1. Monotonicity: F(x) is non-decreasing; that is, $F(a) \leq F(b)$ if $a < b$. The cumulative likelihood cannot fall as x rises.
2. Range: for any x, $0 < F(x) \leq 1$. The range of probabilities is 0 to 1.
3. Limits: $\lim F(x) = 0$ when x gets closer to $-\infty$ and $\lim F(x) = 1$ when x gets closer to $+\infty$ There is a zero chance that X will be less than negative infinity, and a one chance that X will be less than positive infinity.
4. Right Continuity: F(x) is right-continuous, meaning that as h gets closer to 0 from the positive side, $\lim F(x + h) = F(x)$.

Discrete At random Variables with CDF

With PMF p(x) for discrete at random variable X, the CDF is $F(x) = \sum p(t)$, where the total of all values $t \leq x$ CDF for continuous at random variables

When integral is assessed from $-\infty$ to x, the CDF for a continuous at random variable X with PDF f(t) is $F(x) = \int f(t)\, dt$.

Partnership Comparing Continuous At randomVariables with PDF & CDF

For a at random variable that is continuous:

1. The derivative of CDF is PDF: $f(x) = d/dx\, F(x)$
2. The CDF is the PDF's essential component: From $-\infty$ to x, $F(x) = \int f(t)\, dt$

How to Determine Probabilities Making use of the CDF

1. The likelihood that $X \leq b$ is $P(X \leq b) = F(b)$.
2. The likelihood that X is greater than b: $P(X > b) = 1-F(b)$

The likelihood that $a < X \leq b$ is as follows: $P(a < X \leq b) = F(b) - F(a)$

Quantiles and Percentiles Using the CDF

A at random variable X's p-th quantile, also known as its 100p-th percentile, is value of xp such that $F(xp) = P(X \leq xp) = p$.

The 50th percentile, for instance, is the median ($p = 0.5$).

Issues Resolved

Problem 1: A Discrete At randomVariable's Likelihood Mass Function (PMF)

Issue Remark: Two rolls of fair six-sided die are made. Assume that at random variable X is sum of two displayed integers. Determine X's PMF.

Answer: X can have any of the following values: 2, 3, 4,...,12.

We must divide total number of possible outcomes by number of ways each sum can occur in order to determine the PMF.

Total number of possible outcomes = $6 \times 6 = 36$ (since each die can show 6 different values)

For each possible value of X, we count the number of ways it can occur:

- X = 2: Only possible with (1,1). Count = 1
- X = 3: Possible with (1,2) or (2,1). Count = 2
- X = 4: Possible with (1,3), (2,2), or (3,1). Count = 3
- X = 5: Possible with (1,4), (2,3), (3,2), or (4,1). Count = 4
- X = 6: Possible with (1,5), (2,4), (3,3), (4,2), or (5,1). Count = 5
- X = 7: Possible with (1,6), (2,5), (3,4), (4,3), (5,2), or (6,1). Count = 6
- X = 8: Possible with (2,6), (3,5), (4,4), (5,3), or (6,2). Count = 5
- X = 9: Possible with (3,6), (4,5), (5,4), or (6,3). Count = 4
- X = 10: Possible with (4,6), (5,5), or (6,4). Count = 3
- X = 11: Possible with (5,6) or (6,5). Count = 2

Notes

- X = 12: Only possible with (6,6). Count = 1

Consequently, p(2) = 1/36 p(3) = 2/36 = 1/18 p(4) = 3/36 = 1/12 p(5) = 4/36 = 1/9 p(6) = 5/36 p(7) = 6/36 = 1/6 p(8) = 5/36 p(9) = 4/36 = 1/9 p(10) = 3/36 = 1/12 p(11) = 2/36 = 1/18 p(12) = 1/36 is PMF of X.

We can confirm that sum of the likelihood equals one: 36/36 = 1 = 1/36 + 2/36 + 3/36 + 4/36 + 5/36 + 6/36 + 5/36 + 4/36 + 3/36 + 2/36 + 1/36

Issue 2: A Discrete At random Variable's Expected Value &Variance

Issue Remark: Determine expected value &variance of X using PMF of total of two dice rolls from Problem 1.

Answer: $E[X] = \sum x * p(x) = 2 * (1/36) + 3 * (2/36) + 4 * (3/36) + 5 * (4/36) + 6 * (5/36) + 7 * (6/36) + 8 * (5/36) + 9 * (4/36) + 10 * (3/36) + 11 * (2/36) + 12 * (1/36) = 2/36 + 6/36 + 12/36 + 20/36 + 30/36 + 42/36 + 40/36 + 36/36 + 30/36 + 22/36 + 12/36 = 252/36 = 7.$

We can use the following to find variance: $E[X^2] - (E[X]) = Var(X)^2$

Let's first compute $E[X^2]$: $E[X^2] = \sum x^2 \ p(x) = 2^2 * (1/36) + 3^2 * (2/36) + 4^2 * (3/36) + 5^2 * (4/36) + 6^2 * (5/36) + 7^2 * (6/36) + 8^2 * (5/36) + 9^2 * (4/36) + 10^2 * (3/36) + 11^2 * (2/36) + 12^2 * (1/36) = 4/36 + 18/36 + 48/36 + 100/36 + 180/36 + 294/36 + 320/36 + 324/36 + 242/36 + 144/36 = 1974/36 = 54.83$

Next, figure out variance: $E[X^2] - (E[X]) = Var(X)\ 54.83\text{-}49 = 5.83^2 = 54.83 - 7^2$

As a result, X's variance is 5.83 and its anticipated value is 7.

Issue 3: Cumulative Distribution Function (CDF) & Likelihood Density Function (PDF)

Statement of the Problem: A certain electrical component's lifetime X (measured in years) has the following PDF: When $x \geq 0$, $f(x) = \lambda e^{-\lambda x}$, and when $\lambda = 0.5$, $f(x) = 0$

(a) Make sure this PDF is legitimate. (b) Find F(x), the CDF. (c) Determine $P(1 < X \leq 3)$ and $P(X > 2)$. (d) Determine the component's anticipated lifespan.

Answer:

(a) To confirm that f(x) is a legitimate PDF, we must make sure that:

1. For every x, $f(x) >= 0$.
2. $\int f(x)\, dx = 1$, in which all values of x are regarded as integrals.

For condition 1, $f(x) = 0$ for $x < 0$ and $f(x) = 0.5e^{-0.5x}$ for $x \geq 0$. $f(x) \geq 0$ for all x since $e^{-0.5x} > 0$ for all x and $0.5 > 0$.

Given that $f(x) = 0$ for $x < 0$, condition 2 is as follows: $\int f(x)\, dx = \int 0.5e^{-0.5x}$

dx from 0 to ∞ is equal to - $e^{-0.5x}$ $|_0^\infty$ = -$e^{-\infty}$ + e^0 =1

Consequently, f(x) is legitimate PDF.

(b) The CDF is: F(x) = $\int_{-\infty}^{x}$ f(t) dt

If x is less than 0: Since f(t) = 0 for t < 0, F(x) = 0.

For x ≥ 0: From 0 to x = -$e^{-0.5t}$, F(x) = ∫ $0.5e^{-0.5t}$ dt = -$e^{-0.5x}$, F(x) = 0 for x < 0 because -$e^{-0.5x}$ –(-e^0) = -$e^{-0.5x}$ + 1 = 1 - $e^{-0.5x}$.

(c) To determine P(X > 2): F(2) = 1 - (1 - $e^{-0.5\times2}$)= 1 - (1 - e^{-1}) = e^{-1} = 0.368. P(X > 2) = 1 - P(X ≤ 2) = P(1 ≤ X ≤ 3) can be found by: P(1 ≤ X ≤ 3) = F(3) - F(1) = (1 - $e^{-0.53}$) - (1 - $e^{-0.51}$) = (1 - $e^{-1.5}$) - (1 - $e^{-0.5}$) = $e^{-0.5}$ - $e^{-1.5}$ = 0.607 - 0.223 = 0.384 (d) E[X] is the anticipated lifetime: From 0 to ∞, E[X] = ∫ x f(x) dx = ∫ x $0.5e^{-0.5x}$ dx

Integration by parts can be used to calculate this: Let dv = $0.5e^{-0.5x}$ dx and u = x. Then v = -$e^{-0.5x}$ and du = dx.

[x * (-$e^{-0.5x}$)] = E[X] [0 - 0] = _0^∞ - ∫ (-$e^{-0.5x}$) dx from 0 to ∞. From 0 to ∞, - ∫ (-$e^{-0.5x}$) dx = ∫ $e^{-0.5x}$ dx = [-$2e^{-0.5x}$]$_0^\infty$ = 2

Consequently, two years is the component's anticipated lifespan.

Issue 4: Locating a PDF Given a problem statement for CDF: The CDF of a at random variable X is as follows: For x < 0, F(x) = 0. For 0 ≤ x < 1, F(x) = x^2, For x ≥ 1, F(x) = 1.

(a) Locate X's PDF. (b) P(0.3 ≤ X ≤ 0.7) is calculated. (c) Determine X's median.

Answer:

(a) The PDF is the CDF's derivative: f(x) = d/dx F(x)

If x is less than 0: f(x) = d/dx (0) = 0.

For 0 ≤ x < 1: d/dx (x^2) = 2x = f(x)

f(x) = d/dx (1) = 0 for x ≥ 1.

Since 0 ≤ x < 1, f(x) = 0; f(x) = 2x; and f(x) = 0 for x ≥ 1,

By determining whether the integral equals 1, we can confirm that this is a legitimate PDF: Since f(x) = 0 outside [0,1] = [x^2]$_0^1$ = 1^2 - 0^2 = 1. (b) ∫ f(x) dx = ∫ 2x dx from 0 to 1 P(0.3≤ X≤ 0.7) can be found by using the formula P(0.3≤ X≤ 0.7) = F(0.7) - F(0.3) = 0.7^2 - 0.3^2 = 0.49 - 0.09 = 0.4. (c). The value m at which F(m) = 0.5 is the median. F(x) = x^2 for 0 ≤ x < 1. Finding m such that m^2 = 0.5 m = √0.5 m ≈ 0.707 is necessary.

As a result, X's median is roughly 0.707.

Joint Likelihood Distribution is the fifth problem.

The issue is that two fair dice are rolled. If X and Y are the same, then X =

Notes

Y. Let X be the greater of the two numbers that appear, & Y be smaller number.

(a) Determine X and Y's joint PMF. (b) Determine X & Y's marginal PMFs. (c) $P(X + Y \leq 5)$ is calculated. (c) Do X & Y stand alone? Describe.

Answer:

(a) The formula for joint PMF is $p(x,y) = P(X = x, Y = y)$.

We must count the number of outcomes that meet $X = x$ and $Y = y$ for each pair of values (x,y), then divide that number by total number of potential outcomes.

Six × 6 = 36 is total number of possible outcomes.

We have $Y \leq X$ since X is larger and Y is smaller. There is just a single die combination that can produce this if X = Y (both dice display the same value). There are two possible dice combinations if X > Y: (X, Y) or (Y, X).

This is joint PMF p(x,y):

For y = 1: p(1,1) = 1/36 p(2,1) = 2/36 = 1/18 p(3,1) = 2/36 = 1/18 p(4,1) = 2/36 = 1/18 p(5,1) = 2/36 = 1/18 p(6,1) = 2/36 = 1/18

For y = 2, p(2,2) = 1/36 p(3,2) = 2/36 = 1/18 p(4,2) = 2/36 = 1/18 p(5,2) = 2/36 = 1/18 p(6,2) = 2/36 = 1/18

p(3,3) = 1/36 p(4,3) = 2/36 = 1/18 p(5,3) = 2/36 = 1/18 p(6,3) = 2/36 = 1/18 for y = 3.

p(4,4) = 1/36 p(5,4) = 2/36 = 1/18 p(6,4) = 2/36 = 1/18 for y = (4).

p(5,5) = 1/36 p(6,5) = 2/36 = 1/18 for y = 5.

If y = 6, then p(6,6) = 1/36

p(x,y) = 0 for every other combination.

(b) $p_X(x) = \sum p(x,y)$ for all y $p_X(1) = p(1,1) = 1/36$ $p_X(2) = p(2,1) + p(2,2) = 2/36 + 2/36 + 3/36 = 1/12$ $p_X(3) = p(3,1) + p(4,2) + p(4,3) + p(4,4) = 2/36 + 2/36 + 2/36 + 1/36 = 5/36$ $p_X(5) = p(5,1) + p(6,2) + p(5,3) + p(5,4) + p(5,5) = 2/36 + 1/36 = 9/36 = 1/4$ $p_X(6) = p(6,1)$

$p_Y(y) = \sum p(x,y)$ for all x, $p_Y(1) = p(1,1) + p(2,1) + p(3,1) + p(4,1) + p(6,1) = 1/36 + 2/36 + 2/36 + 2/36 + 2/36 + 2/36 + 2/36 = 11/36$ $p_Y(2) = p(2,2) + p(3,2) + p(4,2) + p(5,2) + p(6,2) = 1/36 + 2/36 + 2/36 + 2/36 + 2/36 + 2/36 = 9/36 = 1/4$ $p_Y(3) = p(3,3) + p(4,3) + p(5,3) + p(6,3) = 1/36 + 2/36 + 2/36 + 2/36 + 2/36 = 7/36$ $p_Y(4) = p(4,4) + p(5,4) + p(6,4) = 1/36 + 2/36 = 2/36$

2.2.4: Bivariate random Variables

Introduction to Bivariate random Variables

In many practical situations, we need to study two or more random variables simultaneously. For example, in economics, we might be interested in relationship between income &expenditure; in meteorology, we might study relationship between temperature and humidity.

Definition and Notation

Let's use (X, Y) to represent our bivariate random variable. A subset of R2 (the two-dimensional real plane) is the range or set of values that (X, Y) can take on. S(X,Y) represents this set, which is known as support of (X, Y).
We can list all of the potential values for discrete at random variables: \(x_1, y_1), (x_2, y_2),..., (x_n, y_n),...} is the formula for S(X,Y).
An area on the plane could serve as the support for continuous random variables: S(X,Y) = {(x, y): x ∈ A, y ∈ B}, where A &B are subsets of the real line.
Bivariate random Variable Examples
1. Dice Rolling: When two fair dice are rolled, let X be number on first die and Y be number on second. support S(X,Y) is made up of all 36 possible pairs in this case, where both X and Y take values in {1, 2, 3, 4, 5, 6}.
2. Height and Weight: Assume that randomly chosen individual from a population has height X & weight Y. X &Y are both random variables that are continuous.
3. Weather Conditions: Let Y be discrete random variable that indicates whether it rains (Y = 1) or not (Y = 0), and let X be the temperature on a particular day. In this case, Y is discrete and X is continuous.

2.2.5: Joint, Marginal, and Conditional Distributions

Joint Likelihood Distribution

The likelihood behavior of two random variables taken into consideration together is described by joint likelihood distribution of bivariate random variable (X, Y).
For random Variables That Are Discrete
joint likelihood mass function (PMF) for discrete random variables is

expressed as follows: $p(x, y) = P(X = x, Y = y)$.

This is the likelihood that X will take the value x and Y will take the value y at the same time.

The joint PMF's characteristics are as follows: 1. $p(x, y) \geq 0$ for all (x, y)

2. The sum of p(x, y) for all feasible (x, y) values is equal to 1: $\sum\sum p(x, y) = 1$ x y

For discrete at random variables, the joint cumulative distribution function (or CDF) is as follows: $F(x, y) = P(X < x, Y \leq y) = \sum\sum p(s, t)$ $s \leq x t \leq y$

For random Variables That Are Continuous

The joint likelihood density function (PDF) f(x, y) for continuous at random variables fulfills following formula: $P(a \leq X \leq b, c \leq Y \leq d) = \int_a^b \int_c^d f(x, y)$ dy dx

joint PDF's characteristics are as follows: 1. $f(x, y) \geq 0$ for all (x, y)

2. The sum of the integrals is 1: 1 $R^2 = \iint f(x, y)$ dx dy

For continuous at random variables, joint CDF is as follows: $F(x, y) = P(X \leq x, Y \leq y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(s, t)$ dt ds

Distributions of Marginals

Even when two at random variables are being studied together, we may still be curious about how each variable behaves on its own. Marginal distributions are the distributions of the two freeat random variables, X & Y.

For random Variables That Are Discrete

$P(X = x) = \sum p(x, y)$ y is & marginal PMF of X.

$P(Y = y) = \sum p(x, y)$ x is & marginal PMF of Y.

For random Variables That Are Continuous

The formula for $f_1(x) = \int f(x, y)$ dy R is marginal PDF of X.

formula for Y's marginal PDF is $f_2(y) = \int f(x, y)$ dx R.

Distributions Under Conditions

likelihood behavior of one random variable given that other has taken a certain value is described by conditional distributions.

For random Variables That Are Discrete

If $p_1(x) > 0$, then conditional PMF of Y given X = x is as follows: $p(y|x) = P(Y = y \mid X = x) = p(x, y) / p_1(x)$.

If $p_2(y) > 0$, then conditional PMF of X given Y = y is as follows: $p(x|y) = P(X = x \mid Y = y) = p(x, y) / p_2(y)$.

For random Variables That Are Continuous

Given X = x, conditional PDF of Y is $f(y|x) = f(x, y) / f_1(x)$, provided that $f_1(x) > 0$.

Notes

Given Y = y, conditional PDF of X is $f(x|y) = f(x, y) / f_2(y)$, provided that $f_2(y) > 0$.

Random Variables' Independence

If information about one random variable, X &Y, has no effect on the likelihood distribution of the other, then the two variables are free.

In terms of mathematics, X & Y are free if & only if one of the corresponding conditions listed below is true:

For random Variables That Are Discrete

- For all (x, y), $p(x, y) = p_1(x) \times p_2(y)$.
- For every x such that $p_1(x) > 0$, $p(y|x) = p_2(y)$.
- For every y such that $p_2(y) > 0$, $p(x|y) = p_1(x)$.

For random Variables That Are Continuous

- For all (x, y), $f(x, y) = f_1(x) \times f_2(y)$.
- For every x such that $f_1(x) > 0$, $f(y|x) = f_2(y)$.
- For every y such that $f_2(y) > 0$, $f(x|y) = f_1(x)$.

X and Y are free with respect to the CDF if & only if: For all (x, y), $F(x, y) = F_1(x) \times F_2(y)$.

Notes

UNIT 2.3

Variance, Covariance, Moment generating function- Definitions and their properties

2.3.1: Expectation and Variance of a random Variable

Expectation (Mean)

Random variable's "center of mass" or average value is represented by its expectation or mean.

For random Variable That Is Discrete

$E[X] = \mu_x = \sum x \times p(x)$ x is the expectation of discrete random variable X with PMF p(x).

Regarding an Ongoing random Variable

$E[X] = \mu_x = \int x \times f(x)\, dx$ is expected value of continuous random variable X with PDF f(x). R

Expectation Properties

When c is a constant, E[c] = c

2. If c is a constant, then E[cX] = c × E[X]. Since E[X] + E[Y] = E[X + Y], In the event when X and Y are free, E[XY] = E[X] × E[Y]

At random Variable Functions

If X is random variable and g is a function, then g(X) is likewise random variable. The following is the expected value of g(X) for discrete random variable:

$\sum g(x) \times p(x)$ x = E[g(X)]

Regarding an Ongoing random Variable

$\int g(x) \times f(x)\, dx = E[g(X)]$ Standard Deviation and R Variance

A random variable's dispersion or spread around its mean is measured by its variance.

$Var(X) = \sigma^2_x = E[(X - \mu_x)^2] = E[X^2] - (E[X])$ for at random variable X^2

The variance's square root is standard deviation:

$\sigma_x = \sqrt{Var(X)}$

Variance Properties

1. When c is a constant, Var(c) = 0.

2. With c as a constant, $Var(cX) = c^2 \times Var(X)$.

3. Var(X + Y) = Var(X) + Var(Y) if X and Y are equal.

Correlation and Covariance

joint variability of two random variables is measured by covariance. It shows which way the variables' linear relationship is going.

$E[(X - \mu_x)(Y - \mu_\gamma)] = Cov(X, Y) - E[X] \times E[Y] = E[XY]$

The covariance is normalized to a value between -1 and 1 by correlation coefficient:

$Cov(X, Y) / (\sigma_x \times \sigma_\gamma) = \rho(X, Y)$

Covariance and correlation properties:

1. If &only if $Y = aX + b$ with likelihood 1, where $a \neq 0$, then $-1 \leq \rho(X, Y) \leq 1$. 2. $\rho(X, Y) = \pm 1$.

3. $Cov(X, Y) = 0$ and $\rho(X, Y) = 0$ if X and Y are free (but the opposite is not always true).

Bivariate Function Expectations

For two at random variables, X & Y, and function g(X, Y):

For variables that are discrete, $E[g(X, Y)] = \sum\sum g(x, y) \times p(x, y)$ x y

The formula for $E[g(X, Y)]$ for continuous at random variables is $R^2 = \iint g(x, y) \times f(x, y)\, dx\, dy$.

Solved Problems

Problem 1: Joint Likelihood Mass Function

Two fair dice are rolled. Let X stand for the smaller of the two numbers that show up, and Y for the larger one.

a) Determine the (X, Y) joint PMF. c) Determine X and Y's marginal PMFs.

c) Determine $P(X + Y \leq 5)$. d) Do X &Y stand alone?

Answer:

a) $P(X = x, Y = y)$ = the joint PMF $p(x, y)$:

This issue allows X and Y to take values from {1, 2, 3, 4, 5, 6} and {1, 2, 3, 4, 5, 6}, respectively. But we know that $X \leq Y$ since X is least &Y is greatest.

When two dice are rolled, there are 36 equally likely outcomes in the sample space.

If either the first die displays x and the second die displays y, or the first die displays y and the second die displays x, then for $x < y$, event $(X = x, Y = y)$ takes place. Thus, $2/36 = 1/18 = P(X = x, Y = y)$.

If both dice display same number x for $x = y$, the event $(X = x, Y = y)$ takes place. $P(X = x, Y = y) = 1/36$ as a result.

Thus, if $x = y$, $x, y \in \{1, 2, 3, 4, 5, 6\}$, $p(x, y) = 1/36$; if $x < y$, $x, y \in \{1, 2, 3, 4, 5, 6\}$, $p(x, y) = 2/36 = 1/18$; otherwise, $p(x, y) = 0$.

b) $p(1) = P(X = 1)$ is marginal PMF of X. $= p(1, 1) + p(1, 2) = \sum_{k=1}^{6} p(1, k)$

$P(X = 2) + \ldots + p(1, 6) = 1/36 + 5(1/18) = 11/36$ $p_1(2) = 1/36 + 4(1/18) =$

Notes

9/36 = 1/4 $p_1(3) = P(X = 3) = \sum_{k=2}^{6}$ p(2, k) = p(2, 2) + p(2, 3) +... + p(2, 6) = p(3, k) = p(3, 3) + p(3, 4) = $\sum_{k=3}^{6}$ p(3, k)= 1/36 + 3(1/18) = 7/36 $p_1(4)$ +... + p(3, 6) = P(X = 4) = $\sum_{k=4}^{6}$ p(4, k) = p(4, 4) + p(4, 5) + p(4, 6) = 1/36 + 2(1/18) = 5/36 $p_1(5) = P(X = 5) = \sum_{k=5}^{6}$ p(5, k) = p(5, 5) + p(5, 6) = 1/36 + 1/18 = 3/36 = 1/12 $p_1(6)$ = P(X = 6) = p(6, 6) = 1/36

The marginal PMF of Y: $p_2(1)$ = P(Y = 1) = p(1, 1) = 1/36 $p_2(2)$ = P(Y = 2) = p(1, 2) + p(2, 2) = 1/18 + 1/36 = 3/36 = 1/12 $p_2(3)$ = P(Y = 3) = p(1, 3) + p(2, 3) + p(3, 3) = 1/18 + 1/18 + 1/36 = 5/36 $p_2(4)$ = P(Y = 4) = p(1, 4) + p(2, 4) + p(3, 4) + p(4, 4) = 3(1/18) + 1/36 = 7/36 $p_2(5)$ = P(Y = 5) = p(1, 5) + p(2, 5) + p(3, 5) + p(4, 5) + p(5, 5) = 4(1/18) + 1/36 = 9/36 = 1/4 $p_2(6)$ = P(Y = 6) = p(1, 6) + p(2, 6) + p(3, 6) + p(4, 6) + p(5, 6) + p(6, 6) = 5(1/18) + 1/36 = 11/36 c) P(X + Y ≤ 5):

For every pair (x, y), we must add up the likelihood so that x + y ≤ 5:

P(X + Y ≤ 5) = p(1, 1) + p(1, 2) + p(1, 3) + p(1, 4) + p(2, 2) + p(2, 3) = 1/36 + 1/18 + 1/18 + 1/36 + 1/18 = 1/36 + 4(1/18) = 1/36 + 4/18 = 1/36 + 8/36 = 9/36 = 1/4 d) Do X and Y exist freely?

We must determine whether p(x, y) = $p_1(x) \times p_2(y)$ for every (x, y) in order to determine whether X & Y are free.

Let's see if (X = 1, Y = 2) is true. 1/18 $p_1(1) \times p_2(2)$ = (11/36) × (1/12) = 11/432 ≈ 0.0255 p(1, 2)

Given that p(1, 2) ≠ $p_1(1) \times p_2(2)$, we can deduce that X and Y are not connected. This makes intuitive sense since knowing minimum value X limits range of values that can be assigned to the maximum value Y, and vice versa.

Issue 2: Conditional Independence and Likelihood

Assume that X and Y are continuous at random variables with a joint PDF of f(x, y) = 2 for 0 < x ≤ y ≤ 1 and 0 otherwise.

a) Confirm that this PDF is legitimate. c) Determine X and Y's marginal PDFs. c) Locate f(x|y) & f(y|x), conditional PDFs. d) Do X & Y stand alone?

Solution: a) For a PDF to be considered legitimate, all (x, y) must have f(x, y) ≥ 0 and the whole integral must equal 1.

The first condition is obviously satisfied as f(x, y) = 2 > 0 in the designated region and 0 outside of it.

Regarding the second circumstance:

$\iint f(x, y)\, dx\, dy = \int_0^1 \int_0^y 2\, dx\, dy = \int_0^1 [2x]_0^y\, dy = [y^2]_0^1 = 1.$

Given that both requirements are met, f(x, y) is a legitimate PDF.

b) $f_1(x) = \int f(x, y)\, dy = \int_x^1 2\, dy = [2y]_x^1 = 2 - 2x$ is marginal PDF of X. For $0 < x \leq 1$.

PDF of Y's marginal: $f_2(y) = \int f(x, y)\, dx = \int_0^1 [2x] = 2\, dx$, For $0 \leq y < 1$, $f_2(y) = 2y$,. Given the conditional PDF of Y X = x: for $x \leq y \leq 1$, $f(y|x) = f(x, y) / f_1(x) = 2 / (2 - 2x) = 1 / (1 - x)$.

For $0 < x \leq y$ d, conditional PDF of X given Y = y is $f(x|y) = f(x, y) / f_2(y) = 2 / 2y = 1/y$. We must confirm whether $f(x, y) = f_1(x) \times f_2(y)$ for each (x, y) in support in order to check independence.

$f_1(x) \times f_2(y) = (2 - 2x) \times 2y = 4y - 4xy$ for $0 < x \leq y \leq 1$.

X & Y are not free since, for general values of x& y, $f(x, y) = 2 \neq 4y - 4xy$.

Issue 3: Variance and Expected Value

likelihood mass function of a at random variable X is as follows: $p(x) = c \times x^2$ if $x \in \{1, 2, 3, 4\}$, and $p(x) = 0$ otherwise.

a) Determine what c is worth. b) Determine E[X]. d) Determine Var(X). d) Locate E[1/X].

Answer:

a) The sum of all likelihood must equal 1 since p(x) is PMF:

$\sum_x p(x) = p(1) + p(2) + p(3) + p(4) = c(1^2 + 2^2 + 3^2 + 4^2) = c(1 + 4 + 9 + 16) = 30c = 1$

Consequently, c = 1/30.

b) X should have the following value:

$\sum_x x \times p(x) = \sum_x x \times c \times x^2 = c = E[X] \times \sum_x x^3 = (1/30)\ (1/30) \times (1^3 + 2^3 + 3^3 + 4^3) \times (10/3 \approx 3.33\ c) = (1/30) \times 100 = (1 + 8 + 27 + 64)$. We first compute E[X2] in order to determine the variance:

$\sum_x x^2 \times p(x) = \sum_x x^2 \times c = E[X^2] \times x^2 = c \times \sum_x x^4 = (1/30) \times (1^4 + 2^4 + 3^4 + 4^4) = (1/30) \times (1 + 16 + 81 + 256) = (1/30) \times 354 = 354/30 = 11.8$

We can now determine the variance: $E[X^2] - (E[X]) = Var(X)^2 = 11.8 - (10/3)$ The formula is $11.8 - 100/9 \approx 11.8 - 11.11 \approx 0.69$ d. To determine E[1/X], we compute:

Since $c \times \sum_x x = (1/30)$, $E[1/X] = \sum_x (1/x) \times p(x) = \sum_x (1/x) \times c \times x^2\ (1/30) \times 10 = 1/3 \approx 0.33 \times (1 + 2 + 3 + 4)$

Problem 4: Covariance & Correlation

Let X &Y be discrete at random variables with following joint PMF:

p(x, y) Y = 1 Y = 2 Y = 3

Notes

p(x, y)	Y = 1	Y = 2	Y = 3
X = 0	0.1	0.2	0.1
X = 1	0.1	0.3	0.2

Two fair dice are rolled. Let X stand for smaller of two numbers that show up, and Y for the larger one.

a) Determine the (X, Y) joint PMF. c) Determine X &Y's marginal PMFs. c) Determine $P(X + Y \leq 5)$. d) Do X & Y stand alone?

Answer:

a) $P(X = x, Y = y)$ = the joint PMF $p(x, y)$:

This issue allows X and Y to take values from {1, 2, 3, 4, 5, 6} and {1, 2, 3, 4, 5, 6}, respectively. But we know that $X \leq Y$ since X is least & Y is greatest.

When two dice are rolled, there are 36 equally likely outcomes in the sample space.

If either the first die displays x and the second die displays y, or the first die displays y and the second die displays x, then for $x < y$, the event $(X = x, Y = y)$ takes place. Thus, $2/36 = 1/18 = P(X = x, Y = y)$.

If both dice display the same number x for $x = y$, the event $(X = x, Y = y)$ takes place. $P(X = x, Y = y) = 1/36$ as a result.

Thus, if $x = y$, $x, y \in \{1, 2, 3, 4, 5, 6\}$, $p(x, y) = 1/36$; if $x < y$, $x, y \in \{1, 2, 3, 4, 5, 6\}$, $p(x, y) = 2/36 = 1/18$; otherwise, $p(x, y) = 0$.

b) $p(1) = P(X = 1)$ is marginal PMF of X. $= p(1, 1) + p(1, 2) = \sum_{k=1}^{6} p(1, k)$ $P(X = 2) +... + p(1, 6) = 1/36 + 5(1/18) = 11/36$ $p_1(2) = 1/36 + 4(1/18) = 9/36 = 1/4$ $p_1(3) = P(X = 3) = \sum_{k=2}^{6} p(2, k) = p(2, 2) + p(2, 3) +... + p(2, 6) = p(3, k) = p(3, 3) + p(3, 4) = \sum_{k=3}^{6} = 1/36 + 3(1/18) = 7/36$ $p_1(4) +... + p(3, 6) = P(X = 4) = \sum_{k=4}^{6} p(4, k) = p(4, 4) + p(4, 5) + p(4, 6) = 1/36 + 2(1/18) = 5/36$ $p_1(5) = P(X = 5) = \sum_{k=5}^{6} p(5, k) = p(5, 5) + p(5, 6) = 1/36 + 1/18 = 3/36 = 1/12$ $p_1(6) = P(X = 6) = p(6, 6) = 1/36$

The marginal PMF of Y: $p_2(1) = P(Y = 1) = p(1, 1) = 1/36$ $p_2(2) = P(Y = 2) = p(1, 2) + p(2, 2) = 1/18 + 1/36 = 3/36 = 1/12$ $p_2(3) = P(Y = 3) = p(1, 3) + p(2, 3) + p(3, 3) = 1/18 + 1/18 + 1/36 = 5/36$ $p_2(4) = P(Y = 4) = p(1, 4) + p(2, 4) + p(3, 4) + p(4, 4) = 3(1/18) + 1/36 = 7/36$ $p_2(5) = P(Y = 5) = p(1, 5) + p(2, 5) + p(3, 5) + p(4, 5) + p(5, 5) = 4(1/18) + 1/36 = 9/36 = 1/4$ $p_2(6) = P(Y = 6) = p(1, 6) + p(2, 6) + p(3, 6) + p(4, 6) + p(5, 6) + p(6, 6) = 5(1/18) +$

1/36 = 11/36 c) P(X + Y ≤ 5):

For every pair (x, y), we must add up the likelihood so that x + y ≤ 5:

P(X + Y ≤ 5) = p(1, 1) + p(1, 2) + p(1, 3) + p(1, 4) + p(2, 2) + p(2, 3) = 1/36 + 1/18 + 1/18 + 1/36 + 1/18 = 1/36 + 4(1/18) = 1/36 + 4/18 = 1/36 + 8/36 = 9/36 = 1/4 d) Do X and Y exist freely?

We must determine whether $p(x, y) = p_1(x) \times p_2(y)$ for every (x, y) in order to determine whether X &Y are free.

Let's see if (X = 1, Y = 2) is true. 1/18 $p_1(1) \times p_2(2) = (11/36) \times (1/12) = 11/432 \approx 0.0255$ p(1, 2)

Given that $p(1, 2) \neq p_1(1) \times p_2(2)$, we can deduce that X and Y are not connected. This makes intuitive sense since knowing the minimum value X limits the range of values that can be assigned to the maximum value Y, and vice versa.

Issue 2: Conditional Independence and Likelihood

Assume that X &Y are continuous at random variables with a joint PDF of f(x, y) = 2 for 0 < x ≤ y ≤ 1 and 0 otherwise.

a) Confirm that this PDF is legitimate. c) Determine X and Y's marginal PDFs. c) Locate f(x|y) and f(y|x), conditional PDFs. d) Do X and Y stand alone?

Solution:

a) For a PDF to be considered legitimate, all (x, y) must have f(x, y) ≥ 0 and the whole integral must equal 1.

The first condition is obviously satisfied as f(x, y) = 2 > 0 in the designated region and 0 outside of it.

Regarding the second circumstance:

$\iint f(x, y)\, dx\, dy = \int_0^1 \int_0^y 2\, dx\, dy = \int_0^1 [2x]_0^y\, dy = [y^2]_0^1 = 1.$

Given that both requirements are met, f(x, y) is a legitimate PDF.

b) $f_1(x) = \int f(x, y)\, dy = \int_x^1 2\, dy = [2y]$ is marginal PDF of X. For 0 < x ≤ 1, $f_1(x) = 2 - 2x$

PDF of Y's marginal: $f_2(y) = \int f(x, y)\, dx = \int_0^1 [2x] = 2\, dx$ For 0 ≤ y < 1 c, $f_2(y) = 2y$ Given the conditional PDF of Y X = x: for x ≤ y ≤ 1, f(y|x) = f(x, y) / $f_1(x) = 2 / (2 - 2x) = 1 / (1 - x)$.

For 0 < x ≤ y d, conditional PDF of X given Y = y is f(x|y) = f(x, y) / $f_2(y) = 2 / 2y = 1/y$. We must confirm whether $f(x, y) = f_1(x) \times f_2(y)$ for each (x, y)

Notes

in support in order to check independence.

$f_1(x) \times f_2(y) = (2 - 2x) \times 2y = 4y - 4xy$ for $0 < x \leq y \leq 1$.

X& Y are not free since, for general values of x & y, $f(x, y) = 2 \neq 4y - 4xy$.

Issue 3: Variance and Expected Value

likelihood mass function of at random variable X is as follows: $p(x) = c \times x^2$ if $x \in \{1, 2, 3, 4\}$, and $p(x) = 0$ otherwise.

a) Determine what c is worth. b) Determine E[X]. d) Determine Var(X). d) Locate E[1/X].

Answer:

a) The sum of all likelihood must equal 1 since p(x) is a PMF:

$\sum_x p(x) = p(1) + p(2) + p(3) + p(4) = c(1^2 + 2^2 + 3^2 + 4^2) = c(1 + 4 + 9 + 16) = 30c = 1$

Consequently, $c = 1/30$.

b) X should have the following value:

$\sum_x x \times p(x) = \sum_x x \times c \times x^2 = c = E[X] \times \sum_x x^3 = (1/30)\ (1/30) \times (1^3 + 2^3 + 3^3 + 4^3) \times (10/3 \approx 3.33\ c) = (1/30) \times 100 = (1 + 8 + 27 + 64)$ We first compute E[X2] in order to determine the variance:

$\sum_x x^2 \times p(x) = \sum_x x^2 \times c = E[X^2] \times x^2 = c \times \sum_x x^4 = (1/30) \times (1^4 + 2^4 + 3^4 + 4^4) = (1/30) \times (1 + 16 + 81 + 256) = (1/30) \times 354 = 354/30 = 11.8$

We can now determine the variance: $E[X^2] - (E[X]) = Var(X)^2 = 11.8 - (10/3)$ The formula is $11.8 - 100/9 \approx 11.8 - 11.11 \approx 0.69$ d. To determine E[1/X], we compute:

Since $c \times \sum_x x = (1/30)$, $E[1/X] = \sum_x (1/x) \times p(x) = \sum_x (1/x) \times c \times x^2\ (1/30) \times 10 = 1/3 \approx 0.33 \times (1 + 2 + 3 + 4)$

2.3.2: Covariance and its Significance

Introduction to Covariance

Covariance is a statistical measure that quantifies the degree to which two at random variables change together. It indicates both the direction of the linear relationship between variables and its magnitude. When two variables tend to increase or decrease together, their covariance is positive. Conversely, when one variable tends to increase as the other decreases, their covariance is negative. If the variables are free or have no linear relationship, their covariance will be close to zero.

Mathematical Definition of Covariance

covariance between two at random variables, X & Y, is defined as follows:

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

expected value (mean) of X is denoted by E[X], while expected value of Y is denoted by E[Y].

This can be extended to:

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

This turns into the following for discrete at random variables:

$\Sigma \Sigma (x - \mu X)(y - \mu Y)P(X=x, Y=y) = Cov(X, Y)$

When joint likelihood mass function is denoted by P(X=x, Y=y).

We have following for continuous at random variables:

$\iint (x - \mu X)(y - \mu Y)f(x,y)\, dx\, dy = Cov(X, Y)$

where joint likelihood density function is denoted by f(x,y).

Properties of Covariance

1. **Symmetry**: X, Y Cov = Y, X Cov
2. Cov(X, X) = Var(X) is a specific instance of variance.
3. Cov(aX + b, Y) = Cov(X, Y) o Cov(X, aY + b) = a Cov(X, Y) o Cov(X + Z, Y) = Cov(X, Y) + Cov(Z, Y) are examples of bilinearity.
4. Implication of independence: Cov(X, Y) = 0 if X& Y are free. (Note: Zero covariance does not always indicate independence; the opposite is not always true.)
5. **Range**: There is no fixed range for covariance values, making it difficult to interpret the strength of relationships.

Covariance Matrix

For multiple at random variables X_1, X_2, ..., X_n, we can organize their pairwise covariances into a covariance matrix Σ where:

$\Sigma_{ij} = Cov(X_i, X_j)$

This matrix has several important properties:

- It is symmetric
- The diagonal elements are variances of individual variables
- It is positive semi-definite

- For multivariate normal distributions, it completely characterizes the interdependence structure

Correlation vs. Covariance

Covariance does not standardize the strength of a linear relationship between variables, but it does show the direction of such relationship. In order to overcome this constraint, the correlation coefficient scales covariance to a constant range [-1, 1]:

$Cov(X, Y) / (\sigma X \sigma Y) = \rho(X, Y)$

where $\sigma X \& \sigma Y$ are X and Y's respective standard deviations.

Significance of Covariance

Covariance is significant in various fields:

1. **Finance**: It helps in portfolio theory to understand how different assets move together, allowing for diversification and risk management.
2. **Machine Learning**: It's essential in dimensionality reduction techniques like Principal Component Analysis (PCA).
3. **Statistical Inference**: It helps model relationships between variables in regression analysis.
4. **Signal Processing**: It assists in separating signals from noise.
5. **Multivariate Statistics**: It forms the foundation for many multivariate techniques.

Limitations of Covariance

1. **Scale Dependency**: Changing units can change covariance magnitude, making comparisons difficult.
2. **Non-linearity**: It only captures linear relationships between variables.
3. **Outlier Sensitivity**: It can be heavily influenced by outliers.
4. **Interpretation Difficulty**: Without context, the raw covariance value is challenging to interpret.

2.3.3: Moment-Generating Functions and Their Properties

Introduction to Moment-Generating Functions

As the name implies, moment-generating function (MGF), potent mathematical tool in likelihood theory, creates moments of a at random variable. When t is a real parameter, moment-generating function for at random variable X is defined as expected value of e^{tX}: $M_X(t) = E[e^{tX}]$

The MGF's unique ability to ascertain a likelihood distribution and streamline several likelihood computations, particularly those involving sums of free random variables, is what makes it so beautiful.

Mathematical Definition and Derivation

$M_X(t) = \Sigma\, e^{tx_i})$ is the MGF for discrete random variable X with likelihood mass function $P(X = x_i)$. $P(X = x_i)$

$M_X(t) = \int e^{tx} f(x)\, dx$ is the MGF for continuous random variable X with likelihood density function f(x).

Not every distribution has the MGF, but when it does, it exists for t in a neighborhood of zero.

Link to Moments

The derivatives of MGF evaluated at t = 0 provide the moments of distribution, hence term "moment-generating function":

$E[X^n] = M_X^{(n)}(0)$

where the nth derivative of $M_X(t)$ with respect to t is denoted by $M_X^{(n)}(t)$.

The Taylor series can be used to expand e^{tX} in order to demonstrate this relationship: $e^{tX} = 1 + tX + (t^2x^2)/2! + (t^3x^3)/3! + ...$

Taking the expected value:

$M_X(t) = E[1] + tE[X] + (t^2/2!)E[X^2] + (t^3/3!)E[X^3] + ...$

Differentiating once & evaluating at t = 0:

$M_X'(0) = E[X]$

Differentiating twice &evaluating at t = 0:

$M_X''(0) = E[X^2]$

And so on for higher moments.

Properties of Moment-Generating Functions

Notes

1. **Uniqueness**: If two random variables have same MGF, they have same likelihood distribution.
2. **Convergence in Distribution**: If sequence of random variables converges in distribution, their MGFs converge pointwise.

Moment-Generating Functions for Common Distributions

1. Poisson Distribution

For X ~ Poisson(λ):

$M_X(t) = \exp(\lambda(e^t - 1))$

2. Binomial Distribution

For X ~ Binomial(n, p):

$M_X(t) = (pe^t + (1-p))^n$

3. Uniform Distribution

For X ~ Uniform(a, b):

$M_X(t) = (e^{tb} - e^{ta})/(t(b-a))$

Applications of Moment-Generating Functions

1. **Proving the Central Limit Theorem**: MGFs are instrumental in proving this fundamental theorem in likelihood theory.
2. **Distribution Identification**: MGFs can help identify unknown distributions by comparing them with known forms.
3. **Parameter Estimation**: MGFs can be used in method-of-moments estimation.
4. **Cumulant-Generating Functions**: The natural logarithm of the MGF generates cumulants, which have useful statistical properties.

Limitations of Moment-Generating Functions

1. **Existence**: MGFs don't exist for all distributions, particularly those with heavy tails.
2. **Computational Complexity**: Calculating MGFs can be mathematically challenging for complex distributions.

3. **Numerical Stability**: Computing high-order derivatives numerically can lead to instability.

Use of Random Variables and Distributions in Real-World Applications
In many different domains, statistical analysis is based on random variables and probability distributions. Based on each of the fundamental ideas, the following real-world applications exist:
The Types of Random Variables

Accurate weather forecasting is made possible by meteorologists using continuous random variables to model temperature changes and discrete random variables to reflect the number of rainy days in a month.

Quality Control: To count flaws in product batches, manufacturing businesses use discrete random variables. They can optimize production processes and establish acceptable thresholds by using probability distributions.

Financial Risk Assessment: To assist them set fair rates and preserve their financial stability, insurance companies model claim frequencies as discrete random variables and claim amounts as continuous random variables.

Functions for Probability Density (PDF) and Probability Mass (PMF)
Epidemiology: To forecast disease spread patterns, health experts utilize PMFs to simulate the daily number of new infection cases (discrete) and PDFs to depict the distribution of time till recovery (continuous).

Telecommunications: To optimize bandwidth distribution and lessen network congestion, network engineers use PMFs to examine data packet counts and PDFs to estimate transmission durations.
Reliability Engineering: To forecast failure rates and plan preventative maintenance to avert expensive malfunctions, engineers utilize exponential PDFs to simulate the lifespan of electronic components.
Functions of Distribution and Their Characteristics

Investment Strategy: In order to evaluate investment risks and create diversified portfolios, financial analysts employ features such as symmetry and the 68-95-99.7 rule to predict stock returns using normal distribution functions.

Notes

Load testing: To make sure apps can manage high loads and continue to function, software engineers use distribution functions to simulate user behavior and system response times. Environmental Monitoring: By analyzing pollutant concentrations using distribution functions, scientists can spot threshold violations and create efficient environmental protection regulations.

The Distributions of Bivariate Random Variables

Market research: Based on demographic correlations, businesses create customized marketing campaigns by analyzing bivariate distributions between consumer age and spending patterns. Agricultural Yield Optimization: To guide crop selection and irrigation scheduling, farmers utilize bivariate distributions to comprehend the connections between rainfall and crop yields. Traffic Management: In order to install intelligent traffic signal systems that adjust to changing conditions, urban planners research bivariate distributions of traffic volume and time of day. Covariance, Variance, and Expectation in Mathematics

Inventory management: In order to minimize storage expenses and prevent stockouts, retailers estimate the anticipated demand and variance for products.

Portfolio Optimization: In accordance with Modern Portfolio Theory, investment managers construct portfolios that optimize returns while lowering risk by utilizing covariance among various assets.

Healthcare Resource Allocation: To arrange the right staffing levels and improve patient care while keeping expenses under control, hospital administrators compute anticipated patient arrivals and variance.

Functions that Generate Moments and Their Uses medication Development: To ensure safety and effectiveness in clinical trials, pharmaceutical researchers employ moment-generating functions to examine the distribution of medication concentrations in the bloodstream.

Actuarial Science: To calculate reserve requirements and reinsurance needs, insurance analysts simulate aggregate claims using moment-generating functions.

Signal processing: To describe noise patterns in communication networks and create better filters for clearer signal transmission, engineers use moment-generating functions.

Notes

Integrated Applications in the Real World

Manufacturing Predictive Maintenance

IoT sensors are used in modern workplaces to continuously check the condition of their equipment. Engineers can perform the following by representing vibration levels as random variables with certain distributions:

- Use mathematical expectations to determine the predicted time to failure.
- Using variance analysis, establish maintenance intervals.
- Use covariance studies to find relationships between various machine parameters.
- Utilize moment-generating functions to forecast the likelihood of catastrophic failure.
- This all-encompassing strategy increases equipment lifespan and decreases downtime by 30–50%.

Optimizing Treatment and Precision Medicine

- Healthcare professionals use probability distributions to examine patient data in order to:
- Use a particular PDF to model treatment response as a random variable.
- Determine the variation and projected improvements for various dosages of medications.
- Examine the bivariate relationships between treatment results and genetic markers.
- To forecast the likelihood of an undesirable reaction, use moment-generating functions.
- Personalized treatment regimens that increase effectiveness while lowering negative effects are made possible by this statistical method.
- Evaluation of Climate Risk to Agriculture

Notes

Probability theory is used by farmers and agricultural insurance companies to:

- Consider temperature and precipitation as continuous random variables.
- Make bivariate distributions that link crop production with climate conditions.
- Determine the anticipated yields and variations under various climatic conditions.
- Covariance analysis can be used to identify the best crop combinations for lowering risk.
- Despite growing climate variability, these technologies allow for data-driven decisions that enhance food security.

Power utilities use random variable models in smart grid energy management to:

- Display household energy usage using the relevant PDFs.
- Determine the likelihood of peak usage and anticipated demand.
- Calculate the bivariate association between consumption and weather.
- Utilize moment-generating routines to forecast instances of high demand.

Better grid stability, lower carbon emissions, and effective resource allocation are made possible by this statistical methodology.

Multiple-Choice Questions (MCQs)

1. **Random variable is:**
   a) A fixed value
   b) A function that assigns numerical value to each outcome of an experiment
   c) A constant
   d) A likelihood distribution

   Answer: b) A function that assigns numerical value to each outcome of an experiment

2. **A likelihood mass function (PMF)applies to:**

   a) Continuous at random variables

   b) Discrete at random variables

   c) Both discrete &continuous at random variables

   d) None of the above

   Answer: b) Discrete at random variables

3. **likelihood density function (PDF)satisfies condition:**

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

   b) $f(x) \geq 1$

   c) $\sum f(x) = 1$

   d) f(x) is always negative

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

4. **cumulative distribution function (CDF)is defined as:**

   a) $F(x) = P(X = x)$

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

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

   d) Both (b) and (c)

   Answer: d) Both (b) and (c)

5. **If two at random variables X &Y are free, then:**

   a) $P(X \cap Y) = P(X) + P(Y)$

   b) $P(X|Y) = P(X)$

   c) $P(X,Y) = P(X)P(Y)$

   d) Both (b) and (c)

   Answer: d) Both (b) and (c)

6. **expected value of a at random variable X, E(X), is given by:**

   a) $\sum xP(x)$ for discrete variables

   b) $\int xf(x)dx$ for continuous variables

   c) Both (a) and (b)

   d) None of the above

   Answer: c) Both (a) and (b)

7. **Variance of X, denoted as Var(X), measures:**

   a) The central tendency of X

b) The spread of X around its mean

c) likelihood of X

d) cumulative likelihood of X

Answer: b) The spread of X around its mean

8. **moment-generating function (MGF)** is given by:

   a) $MX(t)=E(e^t X)$

   b) $MX(t)=\sum e^{tx} P(x)$ for discrete variables

   c) $MX(t)=\int e^{tx} f(x)dx$ for continuous variables

   d) All of the above

   Answer: d) All of the above

9. If **covariance** between two at random variables X &Y is **zero**, then:

   a) X and Y are free

   b) X & Y are uncorrelated

   c) X & Y are same variable

   d) X and Y are negatively correlated

   Answer: b) X & Y are uncorrelated

10. **property of expectation** states that for any constants a and b:

    a) E(aX+b)=aE(X)+b

    b) E(aX+b)=aE(X)

    c) E(aX+b)=E(X)+b

    d) E(aX+b)=0

    Answer: a) E(aX+b)=aE(X)+b

Short Answer Questions

1. Defineat random variable & give an example.
2. Differentiate between discrete &continuous at random variables.
3. What is likelihood mass function (PMF)? Give an example.
4. Explain the cumulative distribution function (CDF) and its importance.
5. Define joint, marginal, and conditional distributions with examples.
6. What is the mathematical expectation of a at random variable?

7. Define variance and covariance. How are they useful?
8. What is a moment-generating function (MGF)?
9. Explain how MGFs can be used to find moments of at random variable.
10. Why is covariance important in likelihood theory?

Long Answer Questions

1. Explain difference between discrete &continuous at random variables with examples.
2. Discuss likelihood mass function (PMF) and likelihood density function (PDF) with graphs.
3. Derive the properties of cumulative distribution function (CDF) and explain its significance.
4. Define joint likelihood distribution and explain its applications in statistics.
5. Explain the concept of expectation and variance with examples.
6. Describe moment-generating functions (MGFs) and their applications in likelihood.
7. Derive the formula for variance using expectation.
8. Explain how covariance measures relationship between two at random variables.
9. Solve a numerical problem involving joint distributions and conditional probabilities.
10. How are at random variables and likelihood distributions applied in machine learning and AI?

MODULE 3

UNIT 3.1

Discrete distributions: Uniform, Bernoulli

Objectives

- To study discrete likelihood distributions (Uniform, Bernoulli, Binomial, Poisson, and Geometric).
- To analyze continuous likelihood distributions (Uniform, Exponential, and Normal).
- To explore the properties and applications of these distributions.
- To learn how to compute mean, variance, and moment-generating functions for these distributions.

3.1.1: Introduction to Likelihood Distributions

A mathematical function known as likelihood distribution expresses possibility of at random variable taking any of its potential values. Stated otherwise, it provides information on the distribution of the total likelihood of 1 throughout the at random variable's values.

The foundation of likelihood theory and statistics is likelihood distributions. They offer a method for forecasting unpredictable outcomes and modeling at random phenomena. Two primary categories of likelihood distributions exist:

1. Discrete Likelihood Distributions: These explain at random variables, like integers, that can only have a countable number of different values.
2. Continuous Likelihood Distributions: These explain at random variables, like real numbers, that can have any value within a given range.

The likelihood mass function (PMF), represented as $P(X = x)$ or $f(x)$, provides the likelihood distribution for a discrete at random variable X. It indicates likelihood that the at random variable will take exactly value x.

The likelihood density function (PDF), represented as f(x), provides likelihood distribution for a continuous at random variable X. It indicates relative chance that at random variable will have a value close to x.

In both cases, a likelihood distribution must satisfy two conditions:

- The likelihood of any outcome must be non-negative
- The sum (or integral) of probabilities over all possible outcomes must equal 1

3.1.2: Discrete Likelihood Distributions

Uniform Distribution

The discrete uniform distribution assigns equal likelihood to each of finite number of possible outcomes.

P(X = x) = 1/n for x = a, a+1, a+2,..., a+n-1 is Likelihood Mass Function (PMF).
where an is lowest value and n is the number of alternative outcomes.
Average (Predicted Value): (a + a+n-1)/2 = a + (n-1)/2 = E(X)
Difference: Var(X) = $(n^2-1)/12$

Bernoulli Distribution

One experiment with exactly two possible outcomes—"success" (often represented by a letter 1) and "failure" (typically represented by a number 0)—is modeled by the Bernoulli distribution.
For x = 0, 1, Likelihood Mass Function (PMF) is $P(X = x) = p^x \times (1-p)^{(1-x)}$.
where p is the likelihood of success.
E(X) = p is the mean (expected value).
Variance: p(1-p) = Var(X)
As an illustration, think about tossing fair coin once. Let "heads" be represented by X = 1 and "tails" by X = 0. p = 0.5 is likelihood of heads.
• Mean: E(X) = 0.5; PMF: P(X = 1) = 0.5, P(X = 0) = 0.5
Var(X) = 0.5 × 0.5 = 0.25 is the variance.

Notes

UNIT 3.2

Binomial, Poisson and Geometric distributions with their properties

Distribution of Binomials

number of successes in a predetermined number of free Bernoulli trials is modeled by binomial distribution.

$P(X = x) = {}^{n}C_x \times p^x \times (1-p)^{(n-x)}$ for x = 0, 1, 2 , ..., n is Likelihood Mass Function (PMF).

Where:

• n is total number of tries; p is likelihood that each trial will be successful.

• The binomial coefficient is (n pick x) = n!/[x!(n-x)!].

E(X) = np is the mean (expected value).

Variance: np(1-p) = Var(X)

For illustration, think about tossing fair coin five times. Let X be quantity of heads that were acquired. There is a p = 0.5 chance of heads on every flip.

• PMF: For x = 0, 1, 2, 3, 4, 5, $P(X = x) = {}^{5}C_x \times 0.5^x \times 0.5^{(5-x)}$

E(X) = 5 × 0.5 = 2.5 is mean.

Var(X) = 5 × 0.5 × 0.5 = 1.25 is variance.

Poisson Distribution

particular that events happen freely & at constant average rate, Poisson distribution represents number of events that take place within particular time or space interval.

For x = 0, 1, 2,..., Likelihood Mass Function (PMF) is $P(X = x) = (\lambda^x \times e^{(-\lambda)})/x^*$.

Where:

• e is the base of the natural logarithm (about 2.71828); • λ (lambda) is average number of events in the interval.

E(X) = λ is the mean (expected value).

Var(X) = λ is the variance.

Solved Problems

Problem 1: Uniform Distribution

Problem: One roll of fair six-sided die is made. Determine outcome's variance, expected value, and likelihood mass function.

The answer is that there are six possible outcomes for this discrete uniform distribution: 1, 2, 3, 4, 5, and 6.

First, locate the PMF. Each result has an equal chance because the die is fair: For x = 1, 2, 3, 4, 5, and 6, P(X = x) = 1/6.

Step 2: Determine the mean, or expected value. (1 + 6)/2 = 3.5 is E(X).

Step 3: Determine the difference. (6^2-1)/12 = 35/12 ≈ 2.92 is Var(X).

Consequently, for x = 1, 2, 3, 4, 5, 6, the PMF is P(X = x) = 1/6, the variance is roughly 2.92, and the expected value is 3.5.

Bernoulli Distribution is the second issue.

Problem: The odds of a biased coin falling on heads are 70%. One toss of the coin occurs. If the result is heads, let X = 1, and if it is tails, let X = 0. Determine X's variance, anticipated value, and likelihood mass function.

Solution: likelihood of success for this Bernoulli distribution is p = 0.7.

First, locate the PMF. p = 0.7 P(X = 1) = 1 - p = 0.3 P(X = 0) = 0

For x = 0, 1 we can also write: $P(X = x) = 0.7^x \times 0.3^{(1-x)}$.

Step 2: Determine mean, or expected value. E(X) = p = 0.7

Step 3: Determine the difference. 0.7 × 0.3 = 0.21 Var(X) = p(1-p)

Consequently, variance is 0.21, expected value is 0.7, the PMF is P(X = 1) = 0.7, and P(X = 0) = 0.3.

Issue 3: The Binomial Distribution Issue: Ten questions with four alternative answers—only one of which is correct—make up multiple-choice exam. For every question, student makes a guess. Let X represent how many right answers the student receives. Determine X's variance, anticipated value, and likelihood mass function. Additionally, determine the likelihood that the student will receive precisely three right answers as well as the likelihood that they will receive at least eight.

Answer: With n = 10 trials &a likelihood of success of p = 1/4 = 0.25 for each trial, this is an illustration of a binomial distribution.

First, locate the PMF. For x = 0, 1, 2,..., 10, $P(X = x) = {}^{10}C_x \times 0.25^x \times 0.75^{(10-x)}$.

Notes

Step 2: Determine the mean, or expected value. np = 10 × 0.25 = 2.5 = E(X)

Step 3: Determine the difference. np(1-p) = 10 × 0.25 × 0.75 = 1.875 Var(X) = np

Step 4: Determine the likelihood that the pupil will provide precisely three right responses. 120 × 0.015625 × 0.1335 ≈ 0.2503 = $^{10}C_3$ × 0.25^3 × 0.75^7 = P(X = 3).

Step 5: Calculate the likelihood that the student will provide at least eight accurate responses. P(X ≥ 8) = P(X = 8) + P(X = 9) + P(X = 10) = $^{10}C_8$ × 0.25^8 × 0.75^2 + $^{10}C_9$ × 0.25^9 × 0.75^1 + $^{10}C_{10}$ × 0.25^{10} × 0.75^0 = 45 × 0.000001526 × 0.5625 + 10 × 0.0000000954 × 0.75 + 1 × 0.0000000095 × 1 ≈ 0.0000386 + 0.00000072 + 0.0000000095 ≈ 0.0000393

Thus, for x = 0, 1, 2,..., 10, the PMF is P(X = x) = $^{10}C_x$ × 0.25^x × $0.75^{(10-x)}$. The expected value is 2.5, the variance is 1.875, the likelihood that the student receives exactly three right answers is roughly 0.2503, and the likelihood that the student receives at least eight right answers is roughly 0.0000393.

Issue 4: Poisson Distribution Issue: Two and a half consumers visit a service counter every fifteen minutes on average. Determine the likelihood that, in a specific 15-minute period, (a) exactly four customers will arrive, (b) at most one customer will arrive, and (c) more than three customers will arrive, assuming that customer arrivals follow a Poisson distribution.

Solution: A Poisson distribution with λ = 2.5 is shown here.

First, locate the PMF. The formula P(X = x) = (2.5^x × $e^{(-2.5)}$)/x! for x = 0, 1, 2,...

Step 2: Determine the likelihood that precisely four clients will show up. (2.5^4 × $e^{(-2.5)}$)/4! = 39.0625 × 0.082085 / 24 ≈ 0.1336 is the value of P(X = 4).

Step 3: Calculate the likelihood that no more than one consumer will show up. $e^{(-2.5)}$ + 2.5 × $e^{(-2.5)}$ = 0.082085 + 2.5 × 0.082085 = 0.082085 × (1 + 2.5) = 0.082085 × 3.5 ≈ 0.2873 = P(X ≤ 1) = P(X = 0) + P(X = 1) = (2.5^0 × $e^{(-2.5)}$)/0! + (2.5^1 × $e^{(-2.5)}$)/1!

Step 4: Determine the likelihood that more than three clients will show up. $e^{(-2.5)} + 2.5 \times e^{(-2.5)} + (2.5^2 \times e^{(-2.5)})/2! + (2.5^3 \times e^{(-2.5)})/3!] = 1 - [0.082085 + 0.205213 + 0.256516 + 0.213763] = 1 - [P(X > 3) = 1 - P(X \leq 3) = 1 - [P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)] = 1 - 0.757577 \approx 0.2424$

Thus, (a) there is a roughly 0.1336 chance that exactly four customers will attend, (b) there is a roughly 0.2873 chance that at most one customer will arrive, and (c) there is a roughly 0.2424 chance that more than three customers will arrive.

Issue 5: Distribution of Geometry

Problem: There is a 0.8 chance that a basketball player will make a free throw. Until a free throw is made, the player continues to shoot. Let X be required number of tries. Determine X's variance, anticipated value, and likelihood mass function. Additionally, determine the likelihood that the player will require precisely three tries as well as the likelihood that they will require more than two.

Solution: The likelihood of success for this geometric distribution is p = 0.8.

First, locate the PMF. For x = 1, 2, 3,..., $P(X = x) = (1-p)^{(x-1)} \times p = 0.2^{(x-1)} \times 0.8$.

Step 2: Determine the mean, or expected value. 1/p = 1/0.8 = 1.25 is E(X).

Step 3: Determine the difference. $Var(X) = 0.2/0.64 = 0.3125 = (1-p)/p^2$

Step 4: Determine the likelihood that the player will require precisely three tries. 0.2^(3-1) × 0.8 = $0.2^2 \times 0.8 = 0.04 \times 0.8 = 0.032$ is value of P(X = 3).

Step 5: Determine the likelihood that the player will require more than two tries. $P(X > 2) = 1 - P(X \leq 2) = 1 - [P(X = 1) + P(X = 2)] = 1 - [0.8 + 0.2 \times 0.8] = 1 - [0.8 + 0.16] = 1 - 0.96 = 0.04$

Thus, for x = 1, 2, 3,..., the PMF is $P(X = x) = 0.2^{(x-1)} \times 0.8$, the variance is 0.3125, the expected value is 1.25, the likelihood that the player requires precisely three tries is 0.032, and the likelihood that the player requires more than two efforts is 0.04.

Notes

Unresolved Issues

Issue 1: Uniform Distribution

A fair spinner is divided into eight equal sectors, numbered 1 through 8. The spinner is spun once, and the outcome is denoted as X. Determine the following:

a) The probability mass function (PMF) of X.
b) The expected value of X.
c) The variance of X.
d) The probability that the spinner lands on an even number.
e) The probability that the spinner lands on a number greater than five.

Issue 2: Bernoulli Distribution

A quality control inspector examines randomly selected computer chips. Each chip has a 5% probability of being defective. Define X as follows:

- X=1 if the chip is defective
- X=0 if the chip is not defective

Determine the following:

a) The probability mass function (PMF) of X.
b) The expected value of X.
c) The variance of X.
d) The interpretation of the expected value in this context.
e) The expected number of defective chips if the inspector examines 100 chips.

Issue 3: Binomial Distribution

A biased coin has a 60% probability of landing on heads. The coin is flipped 15 times, and X represents the number of heads obtained. Determine the following:

a) The probability mass function (PMF) of X.
b) The expected value of X.
c) The variance of X.

d) The probability of obtaining exactly 10 heads.
e) The probability of obtaining no more than 7 heads.
f) The probability of obtaining between 8 and 12 heads (inclusive).

Issue 4: Poisson Distribution

A website receives an average of 5 comments per hour. Assume the number of comments follows a Poisson distribution. Determine the following:

a) The probability that exactly 7 comments are posted in a given hour.
b) The probability that no comments are posted in an hour.
c) The probability that at least 3 comments are posted in an hour.
d) The probability that the number of comments in an hour falls between 2 and 6 (inclusive).
e) The expected number of comments in a 12-hour period.

Issue 5: Geometric Distribution

A salesman makes cold calls to potential customers, with each call having a 15% chance of resulting in a sale. The salesman continues calling until a sale is made. Let X represent the number of calls needed to close a deal. Determine the following:

a) The probability mass function (PMF) of X.
b) The expected value of X.
c) The variance of X.
d) The probability that exactly 5 calls are needed to close a deal.
e) The probability that a sale is made within the first 3 calls.

Notes

UNIT 3.3

Continuous distributions: Uniform, Exponential and Normal distributions with their properties

3.3.1: Continuous Likelihood Distributions

Continuous likelihood distributions deal with at random variables that might have any value within a given range, whereas discrete likelihood distributions deal with countable outcomes. The likelihood that a at random variable will take on any given precise value is zero in continuous distributions. Rather, we determine likelihood that at random variable will fall inside a specific range.

(PDF) is a mathematical tool used to characterize continuous likelihood distributions.

$\int_a$ f(x) dx = P(a ≤ X ≤ b)

Essential characteristics of any legitimate PDF f(x):

For any x, f(x) ≥ 0 (the function is non-negative).

2. $\int_{-\infty}^{\infty}$ f(x) dx = 1 (the entire likelihood equals 1)

(CDF), represented as F(x), is another crucial function that provides the likelihood that a at random variable X is less than or equal to a value x:

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

The three basic continuous distributions—normal, exponential, and uniform—will be examined now.

Even Distribution

The most basic continuous likelihood distribution is uniform distribution. It characterizes a at random variable that has an equal chance of taking any value between [a, b].

Function of Likelihood Density (PDF)

The uniform distribution's PDF is:

For ≤ x ≤ b, f(x) = 1/(b-a); otherwise, f(x) = 0.

Over the range [a, b], this produces a rectangle of height 1/(b-a).

The CDF, or Cumulative Distribution Function

The uniform distribution's CDF is:

For $x < a$, $F(x) = 0$. The formula $F(x) = (x-a)/(b-a)$ for $a \leq x \leq b$. When $x > b$, $F(x) = 1$.

Properties of the Uniform Distribution

1. **Mean (Expected Value)**: $\mu = (a + b)/2$
2. **Variance**: $\sigma^2 = (b - a)^2/12$
3. **Standard Deviation**: $\sigma = (b - a)/\sqrt{12}$
4. **Median**: Median = $(a + b)/2$
5. **Mode**: The uniform distribution has no unique mode; every value in [a, b] is equally likely.

Applications of the Uniform Distribution

- At random number generators (over a specific range)
- Rounding errors in measurements
- Arrival times when no specific time is more likely than another
- Modeling situations where all outcomes within a range are equally likely

Exponential Distribution

In a process where events happen freely at a constant average rate, the time between occurrences is modeled by the exponential distribution. It is frequently applied in survival analysis, queuing theory, and reliability theory.
Function of Likelihood Density (PDF)
For $x \geq 0$, exponential distribution's PDF is $f(x) = \lambda e^{-\lambda x}$. For $x < 0$, $f(x) = 0$. where the rate parameter, λ (lambda), is positive ($\lambda > 0$).
The CDF, or Cumulative Distribution Function
For $x < 0$, CDF of exponential distribution is $F(x) = 0$. For $x \geq 0$, $F(x) = 1 - e^{-\lambda x}$.

Properties of Exponential Distribution

1. **Mean (Expected Value)**: $\mu = 1/\lambda$
2. **Variance**: $\sigma^2 = 1/\lambda^2$
3. **Standard Deviation**: $\sigma = 1/\lambda$

Notes

4. **Median**: Median = $\ln(2)/\lambda$
5. **Mode**: Mode = 0
6. **Moment Generating Function**: $M(t) = \lambda/(\lambda-t)$ for $t < \lambda$

Applications of Exponential Distribution

- Time between arrivals in Poisson process
- Time until decay of radioactive particles
- Length of phone calls
- Time until equipment failure

Normal Distribution

Perhaps most significant likelihood distribution in statistics is normal distribution, sometimes referred to as the Gaussian distribution. It is symmetrical, bell-shaped, and naturally occurring in a wide range of social and physical events.

Function of Likelihood Density (PDF)

The normal distribution's PDF is:

For $-\infty < x < \infty$, $f(x) = (1/(\sigma\sqrt{(2\pi)})) * e^{-(x-\mu)^2/(2\sigma^2)}$

Where:

μ (mu) represents mean.

• standard deviation is represented by σ (sigma).

σ^2 represents variance.

Normal Distribution Standard

standard normal distribution, with $\mu = 0$ and $\mu = 1$, is a specific case of the normal distribution. Here is its PDF:

For $-\infty < z < \infty$, $f(z) = (1/\sqrt{(2\pi)}) * e^{-z^2/2}$.

following formula can be used to transform any normal at random variable X into a standard normal at random variable Z:

$Z = (X - \mu)/\sigma$

The CDF, or Cumulative Distribution Function

Without a closed-form formula, CDF of normal distribution is typically represented as follows: $\Phi(x) = P(X \leq x) = \int_{-\infty}^{x} (1/(\sigma\sqrt{(2\pi)})) e^{-(t-\mu)^2/(2\sigma^2)} dt$

standard normal CDF is commonly tabulated and is represented by the symbol $\Phi(z)$.

Properties of the Normal Distribution

1. **Mean (Expected Value)**: μ (the parameter in the PDF)
2. **Variance**: σ^2 (the parameter in the PDF)
3. **Standard Deviation**: σ (the parameter in the PDF)
4. **Median**: Median = μ
5. **Mode**: Mode = μ
6. **Moment Generating Function**: $M(t) = \exp(\mu t + \sigma^2 t^2/2)$

Applications of the Normal Distribution

- Heights and weights of populations
- Measurement errors
- IQ scores and other standardized test scores
- Financial returns
- Many natural phenomena

3.3.2: Properties of Distributions

Common Properties of Likelihood Distributions

1. Mean Expected Value

With PDF f(x), the mean, or anticipated value, of continuous at random variable X is:

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

The long-term average of the at random variable is represented by the expected value.

2. Standard Deviation and Variance variance quantifies a distribution's dispersion or spread:

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

An other method of calculating $\int_{-\infty}^{\infty} x^2 \cdot f(x)\, dx - \mu^2 = Var(X) = E[X^2] - (E[X])^2$

variance's square root is the standard deviation:

Notes

$\sigma = \sqrt{Var(X)}$

3. The median The value that splits the distribution in half is called the median. The median m for continuous at random variable with CDF F(x) satisfies F(m) = 0.5.

4. Mode

The value at which the PDF reaches its maximum is known as the mode. f'(x_mode) = 0 &f''(x_mode) < 0 are satisfied by the mode x_mode for continuous at random variable with PDF f(x).

5. Skewness

Skewness measures the asymmetry of a distribution:

Skewness = $E[((X - \mu)/\sigma)^3]$

- Positive skewness: right-tailed distribution (tail extends to the right)
- Negative skewness: left-tailed distribution (tail extends to the left)
- Zero skewness: symmetric distribution (like the normal distribution)

6. Kurtosis

Kurtosis measures the "tailedness" of a distribution:

Kurtosis = $E[((X - \mu)/\sigma)^4]$

- Negative excess kurtosis (< 3): lighter tails than the normal distribution
- Excess kurtosis = 0 (kurtosis = 3): similar tails to the normal distribution

7. Moment Generating Function (MGF)

continuous at random variable X's MGF is: $\int_{-\infty}^{\infty} e^{tx} \bullet f(x)\, dx = M(t) = E[e^{tX}]$

The distribution's moments can be found using the MGF:

$E[X^n] = M^{(n)}(0)$

The nth derivative of M(t) evaluated at t = 0 is denoted by $M^{(n)}(0)$.

8. Percentiles and Quantiles value x_p such that $F(x_p) = p$ is the p-th quantile of a distribution.

Where F is the CDF. Common quantiles include:

Notes

- Median (p = 0.5)
- Quartiles (p = 0.25, 0.5, 0.75)
- Percentiles (p = 0.01, 0.02, ..., 0.99)

Relationships Between Distributions

1. Sum of random Variables

PDF of Z = X + Y is the convolution if X & Y are free random variables with PDFs $f_X(x)$ & $f_Y(y)$:

$f_Z(z) = \int_{-\infty}^{\infty} f_X(z-y) \cdot f_Y(y)\, Dy$

Particular situations:

• Total of the normal, freeat random variables: $X + Y \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$ if $X \sim N(\mu_X, \sigma_X^2)$ and $Y \sim N(\mu_Y, \sigma_Y^2)$.

• The total of free, parameter-sharing exponential random variables: More intricate, resulting in a gamma distribution

2. Random Variable Transformation

If Y = g(X) is a strictly monotonic function and X is random variable with PDF $f_X(x)$, then $f_Y(y) = f_X(g^{-1}(y))$ is the PDF of Y. • $|d/dy\ g^{-1}(y)|$ where the inverse function of g is g^(-1).

3. Statistics on Orders

kth order statistic $X_{(k)}$ has following PDF if $X_1, X_2, ..., X_n$ are free random variables from same distribution with CDF F(x) & PDF f(x): $f_{X_{(k)}}(x) = n!/(((k-1)!\cdot(n-k)!)) \cdot [F(x)]^{(k-1)}\ [1-F(x)]^{(n-k)} \cdot f(x)$

The minimum (k=1) and maximum (k=n) are examples of special situations.

Solved Examples

Solved Example 1: Uniform Distribution

Problem: A bus is scheduled to arrive at a stop between 10:00 AM and 10:30 AM with equal likelihood for any time in this interval. What is the likelihood that, if you arrive at 10:15 AM, you will: a) have to wait longer than ten minutes for the bus? c) Is the bus here already? c) If you arrive at 10:15 AM, how long should you expect to wait?

Solution:

Notes

Let X be arrival time of bus in minutes after 10:00 AM. Then X follows a uniform distribution on [0, 30].

a) You arrive at 10:15 AM, which is 15 minutes after 10:00 AM. You'll wait more than 10 minutes if the bus arrives after 10:25 AM, which is 25 minutes after 10:00 AM.

$P(X > 25) = 1 - P(X \leq 25) = 1 - (25 - 0)/(30 - 0) = 1 - 25/30 = 1 - 5/6 = 1/6$

b) The bus has already arrived if it comes before 10:15 AM, which is before 15 minutes after 10:00 AM.

$P(X \leq 15) = (15 - 0)/(30 - 0) = 15/30 = 1/2$

The likelihood that the bus has already arrived is 1/2 or 0.5.

c) If you arrive at 10:15 AM (15 minutes after 10:00 AM), there are two cases:

- Case 1: The bus has already arrived ($X \leq 15$). In this case, you missed the bus and will have to wait for the next one, but this waiting time is not calculated here.
- Case 2: The bus has not arrived yet ($X > 15$). In this case, you'll wait ($X - 15$) minutes.

The expected waiting time given that you haven't missed the bus is: $E[X - 15|X > 15] = \int_{15}^{30} (x - 15)\cdot(1/15)\, dx = (1/15)\cdot \int_{15}^{30} (x - 15)\, dx = (1/15)\cdot[x^2/2 - 15x]_{15}^{30} = (1/15)\cdot[(30^2/2 - 15\cdot30) - (15^2/2 - 15\cdot15)] = (1/15)\cdot[450 - 450 - 112.5 + 225] = (1/15)\cdot[112.5] = 7.5$

So, if you arrive at 10:15 AM and the bus hasn't arrived yet, you can expect to wait an additional 7.5 minutes on average.

Solved Example 2: Exponential Distribution

Problem: With an average lifespan of 5000 hours, lifespan of a particular electrical component follows an exponential distribution. a) What is this distribution's rate parameter, λ? b) How likely is it that a component will have a lifespan of fewer than 3000 hours? c) What is the likelihood that a component will endure a another 2000 hours if it has already been in use for 4000 hours? d) What is these components' median lifespan?

Solution:

a) For an exponential distribution, the mean $\mu = 1/\lambda$. Given that $\mu = 5000$ hours: $\lambda = 1/5000 = 0.0002$ per hour

b) The likelihood that the lifetime X is less than 3000 hours: $P(X < 3000) = 1 - e^{-\lambda \cdot 3000} = 1 - e^{-0.0002 \cdot 3000} = 1 - e^{-0.6} = 1 - 0.5488 = 0.4512$

So the likelihood is approximately 0.4512 or 45.12%.

c) Due to memoryless property of exponential distribution: $P(X > 4000 + 2000 \mid X > 4000) = P(X > 2000) = e^{\wedge}(-\lambda \cdot 2000) = e^{-0.0002 \cdot 2000} = e^{-0.4} = 0.6703$

So the likelihood is approximately 0.6703 or 67.03%.

d) The median of an exponential distribution is: Median $= \ln(2)/\lambda = \ln(2)/0.0002 = 0.693/0.0002 = 3465$ hours

So the median lifetime is approximately 3465 hours.

Solved Example 3: Normal Distribution

Problem: With mean of 75 kg &standard deviation of 8 kg, weights of adult males in given population are normally distributed. a) What proportion of male adults weigh over 85 kg? b) What proportion of male adults weigh between 70 and 80 kilograms? c) What is the likelihood that five adult males chosen at at random will weigh more than 78 kg on average? d) For this population, what weight represents the 90th percentile?

Answer:

a) Let X be a male adult's weight. We are looking for $P(X > 85)$. Standardize first: $P(X > 85) = P(Z > 1.25) = 1 - P(Z \leq 1.25) = 1 - \Phi(1.25)$ $Z = (X - \mu)/\sigma = (85 - 75)/8 = 10/8 = 1.25$

Using a calculator or the usual normal table: $\Phi(1.25) \approx 0.8944$

$P(X > 85) = 1 - 0.8944 = 0.1056$

So approximately 10.56% of adult males weigh more than 85 kg.

b) We want to find $P(70 < X < 80)$. Standardize the endpoints: $Z_1 = (70 - 75)/8 = -0.625$ $Z_2 = (80 - 75)/8 = 0.625$

$P(70 < X < 80) = P(-0.625 < Z < 0.625) = \Phi(0.625) - \Phi(-0.625)$

Using symmetry of the normal distribution: $\Phi(-0.625) = 1 - \Phi(0.625)$

$P(70 < X < 80) = \Phi(0.625) - (1 - \Phi(0.625)) = 2 \cdot \Phi(0.625) - 1$

Notes

$\Phi(0.625) \approx 0.7339$

$P(70 < X < 80) = 2 \cdot 0.7339 - 1 = 1.4678 - 1 = 0.4678$

So approximately 46.78% of adult males weigh between 70 kg and 80 kg.

c) Let $\bar{X}$ be the mean weight of five adult males chosen at random. $\bar{X}$ has a normal distribution according to Central Limit Theorem, which has:
$\mu_{\bar{X}} = \mu = 75$ kg is the mean.
• $\sigma_{\bar{X}} = \sigma/\sqrt{n} = 8/\sqrt{5} = 8/2.236 = 3.578$ kg is the standard deviation.
Our goal is to determine $P(\bar{X} > 78)$. Standardize: $P(\bar{X} > 78) = P(Z > 0.838) = 1 - P(Z \leq 0.838) = 1 - \Phi(0.838)$ $Z = (\bar{X} - \mu_{\bar{X}})/\sigma_{\bar{X}} = (78 - 75)/3.578 = 3/3.578 = 0.838$
$P(\bar{X} > 78) = 1 - 0.7991 = 0.2009$ $\Phi(0.838) \approx 0.7991$
Thus, the likelihood is roughly 0.2009, or 20.09%.

d) Let x_{90} be the 90th percentile, so $P(X \leq x_{90}) = 0.90$. This means that $\Phi((x_{90} - 75)/8) = 0.90$. The z-score corresponding to the 90th percentile is $\Phi^{-1}(0.90) = 1.282$.

So $(x_{90} - 75)/8 = 1.282$ $x_{90} - 75 = 8 \cdot 1.282 = 10.256$ $x_{90} = 75 + 10.256 = 85.256$ kg

The 90th percentile is approximately 85.26 kg.

Solved Example 4: Distribution Properties

Problem: When $0 < x \leq 1$, PDF of a at random variable X is $f(x) = 3x^2$; otherwise, it is $f(x) = 0$. a) Confirm that this PDF is legitimate. c) Determine X's CDF. b) Determine X's variance and mean. c) Determine X's median. e) Is there a bias in this distribution? In what direction, if at all? Answer:
a) Two requirements must be met for a PDF to be considered valid:
1. For all x (non-negativity), $f(x) \geq 0$.

2. $\int_{-\infty}^{\infty} f(x)\, dx = 1$ (total likelihood = 1)

Condition 1 is satisfied since $f(x) = 3x^2 \geq 0$ for $0 \leq x < 1$ and $f(x) = 0$ elsewhere.

For condition 2: $\int_{-\infty}^{\infty} f(x)\, dx = \int(0 \text{ to } 1)\, 3x^2\, dx = 3 \cdot \int_0^1 x^2\, dx = 3 \cdot [x^3/3]_0^1 = 3 \cdot (1/3) = 1$

Since both conditions are met, this is a valid PDF.

b) $F(x) = P(X < x) = \int_{-\infty}^{x} f(t)\,dt$ is the CDF.

If x is less than 0: F(x) = 0.

For $0 \le x \le 1$: $F(x) = \int_0^x 3t^2\,dt = 3. \cdot [t^3/3]_0^x = 3. \cdot (x^3/3) = x^3$

If x > 1: F(x) = 1.

Thus, the CDF is: For x < 0, F(x) = 0. For $0 < x \le 1$, $F(x) = x^3$ For x > 1 c), F(x) = 1. Average

Predicted Value): $E[X] = \int_{-\infty}^{\infty} x \cdot \int_0^1 x = f(x)\,dx \cdot 3x^2\,dx = 3. \cdot \int_0^1 x^3\,dx = 3$ $[x^4/4]_0^1 = 3. \cdot (1/4) = 3/4 = 0.75$

We first compute E[X2] in order to determine the variance: $E[X2] = \int_{-\infty}^{\infty} x^2 \cdot \int_0^1 x^2 = f(x)\,dx \cdot \int_0^1 x^4\,dx = 3 \cdot 3x^2\,dx = 3$ Between 0 and 1, $[x^5/5] = 3$; $(1/5) = 3/5 = 0.6$

Difference: $E[X^2] = Var(X) - (E[X])^2 = 3/5 - (3/4)\,(3/5) = ^2 = 3/5 - 9/16 - (9/16) = 3/80 = 0.0375$ d (48/80) - (45/80) F(m) = 0.5 is satisfied by the median m.

$F(x) = x^3$ for $0 < x \le 1$ is known from component (b).

Thus, we must resolve: $m^3 = 0.5$ Using the cube root: $m = \sqrt[3]{0.5} = 0.5^{1/3} \approx 0.7937$

The median is approximately 0.7937.

e) To determine if the distribution is skewed, we can compare the mean (0.75) and median (0.7937).

Since mean < median, the distribution is negatively skewed (skewed to the left).

We can also calculate the skewness coefficient, but the comparison of mean and median gives us the direction of skewness.

Solved Example 5: Mixed Distributions

Problem: Electronic components having a mean lifespan of 2000 hours and an exponential distribution are produced throughout a manufacturing process. A box is deemed faulty if more than two components fail within the

Notes

first 500 hours of operation. The components are packed in boxes of ten. a) How likely is it that one part will malfunction in the first 500 hours? b) How likely is it that a box is flawed? c) How many defective boxes should be expected if 100 boxes are shipped? d) How likely is it that one box out of every 100 is flawed?

Answer:

a) Assume that X is a component's lifetime in hours, with a mean of μ = 2000 hours and an exponential distribution. $\lambda = 1/\mu = 1/2000 = 0.0005$ per hour is the rate parameter.

likelihood that a component fails within the first 500 hours is: $P(X \leq 500) = 1 - e^{-\lambda \cdot 500} = 1 - e^{-0.0005 \cdot 500} = 1 - e^{-0.25} = 1 - 0.7788 = 0.2212$

So the likelihood is approximately 0.2212 or 22.12%.

b) Let Y represent how many parts in a box break down in the first 500 hours. With n = 10 (number of components) and p = 0.2212 (likelihood of failure for each component), Y has binomial distribution.

We want $P(Y > 2) = 1 - P(Y \leq 2)$ since a box is faulty if $Y > 2$. $C(10,0) = P(Y \leq 2) = P(Y = 0) + P(Y = 1) + P(Y = 2)$ $[(1-p)^{10} + C(10,1) + p^0 \bullet p^1 \bullet (1-p)^9 + C(10,2)$ $(1) \bullet (p^2 \bullet (1-p)^8 = (1) \bullet (0.7788)^{10} + (10) \bullet (0.2212) \bullet (0.7788)^9 + (45) \bullet (0.2212)^2 \bullet (0.7788)^8$

First term: $(0.7788)^{10} \approx 0.0858$ Second term: $10 \cdot 0.2212 \cdot (0.7788)^9 \approx 10 \cdot 0.2212 \cdot 0.1102 \approx 0.2439$ Third term: $45 \cdot (0.2212)^2 \cdot (0.7788)^8 \approx 45 \cdot 0.0489 \cdot 0.1415 \approx 0.3112$

$P(Y \leq 2) \approx 0.0858 + 0.2439 + 0.3112 = 0.6409$

$P(Y > 2) = 1 - 0.6409 = 0.3591$

So the likelihood that a box is defective is approximately 0.3591 or 35.91%.

c) Let Z be the number of defective boxes out of 100. Z follows binomial distribution with n = 100 (number of boxes) & p = 0.3591 (likelihood of a box being defective).

The expected number of defective boxes is: $E[Z] = n \cdot p = 100 \cdot 0.3591 = 35.91$

So we expect approximately 36 defective boxes out of 100.

d) There is a 1 in 100 chance that at least one box is faulty: $P(Z \geq 1) = 1 - P(Z = 0) = 1 - (1-p)^n = 1 - (1-0.3591)^{100} = 1 - (0.6409)^{100}$

$(0.6409)^{100}$ is extremely small (approximately 0 for practical purposes).

Therefore, $P(Z \geq 1) \approx 1 - 0 = 1$

So the likelihood that at least one box out of 100 is defective is essentially 1 (or 100%).

Unsolved Problems

Problem 1: Uniform Distribution

A random number generator produces numbers uniformly distributed between -3 and 5.

a) Determine the probability density function (PDF) and cumulative distribution function (CDF) of this distribution.
b) Calculate the probability that a randomly generated number is greater than zero.
c) Find the probability that the generated number falls between -1 and 2.
d) Compute the mean and variance of the distribution.
e) Determine the probability that at least one of ten randomly generated numbers is less than -2.

Problem 2: Exponential Distribution

The time interval between customer arrivals at a bank follows an exponential distribution with an average of three minutes.

a) What is the probability that the next customer arrives within two minutes?
b) What is the probability that the next customer arrives in five minutes or more?
c) Given that no customer has arrived in the last four minutes, what is the probability that the bank will wait at least three more minutes for the next arrival?
d) Determine the waiting time between arrivals at the 75th percentile.
e) If the bank opens at 9:00 AM, what is the probability that at least five customers will arrive by 9:15 AM?

Notes

Problem 3: Normal Distribution

The height of adult women in a certain population follows a normal distribution with a mean of 165 cm and a standard deviation of 6 cm.

a) What proportion of women are taller than 175 cm?
b) If a woman is considered "tall" when her height falls within the top 10%, what is the minimum height required to be classified as tall?
c) What is the probability that a randomly selected woman has a height between 160 cm and 170 cm?
d) If four women are randomly selected, what is the probability that their average height exceeds 168 cm?
e) If a sample of 100 women is randomly selected, what is the probability that the sample mean deviates by more than 1 cm from the population mean?

3.3.3: Mean and Variance of Distributions

Two essential metrics that aid in describing likelihood distributions are variance and mean (or expected value). They offer crucial details regarding the dispersion and central tendency of at random variables.

Average (Predicted Value)

A at random variable's "average" value or center of mass of its distribution is represented by its mean or anticipated value.

With likelihood mass function p(x), for discrete at random variables X, $E[X] = \sum(x * p(x))$, where total is computed over all possible values of x.

The integral is taken across the whole domain of X for continuous at random variables X with likelihood density function f(x): $E[X] = \int(x * f(x))dx$.

Difference

A at random variable's variance quantifies how widely it deviates from its mean.
For every arbitrary variable X:

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

This can be expressed differently as follows: $Var(X) = E[X^2] - (E[X])^2$

standard deviation variance's square root is all that standard deviation is:

SD(X) = √Var(X)

Notes

Expected Value Properties

1. Expectation Linearity:

o E[aX + b] = a •E[X] + b

For example, E[X + Y] = E[X] + E[Y]

2. E[X•Y] = E[X]•E[Y] for freeat random variables

Variance Properties

1. Scaling characteristic:

A^2•Var(X) = o Var(aX)

2. For at random variables that are free:

Var(X + Y) is equal to Var(X) + Var(Y).

3. The translation property states that for each constant a, Var(X + a) = Var(X).

Averaging and Varying Typical Discrete Distributions

1. The Bernoulli Distribution

With likelihood p, a Bernoulli at random variable takes on value 1, and with likelihood (1-p), it takes on the value 0.

E[X] = p is the mean.

• Var(X) = p(1-p) is the variance.

2. Distribution of Binomials

number of successes in n free Bernoulli trials, each with a chance of p, is described by the binomial distribution with parameters n & p.

E[X] = np is the mean.

• Variance: np(1-p) = Var(X)

3. Distribution in Geometry

The number of tries required to get first success in series of free Bernoulli trials is described by geometric distribution with parameter p.

E[X] = 1/p is the mean.

• Var(X) = $(1-p)/p^2$ is the variance.

4. The Poisson Distribution

The number of events that occur at constant average rate within a specific interval is described by Poisson distribution with parameter λ.

E[X] = λ is the mean.

• Var(X) = λ is the variance.

5. Distribution of Negative Binomials

The number of attempts required to get r successes is described by the

Notes

negative binomial distribution with parameters r and p.

E[X] = r/p is the mean.

• Var(X) = r(1-p)/p² is the variance.

Common Continuous Distributions: Mean and Variance

1. Equitable Dispersion

It is equally likely that a continuous uniform at random variable on [a,b] will take any value within that range.

E[X] = (a+b)/2 is the mean.

• Var(X) = (b-a)²/12 is the variance.

2. Gaussian Normal Distribution

The well-known bell-shaped curve is present in the normal distribution with parameters μ and σ².

E[X] = μ is the mean.

• Variance: σ² = Var(X)

3. The Exponential Distribution

The time interval between events in a Poisson process is described by the exponential distribution with parameter λ.

E[X] = 1/λ is the mean.

• Variance: 1/λ² = Var(X)

4. Distribution of Gamma

The exponential distribution is generalized by the gamma distribution with parameters α (shape) and β (scale).

E[X] = αβ is the mean.

• Variance: αβ² = Var(X)

5. Distribution of Beta

The interval [0,1] defines the beta distribution with parameters α and β.

E[X] = α/(α+β) is the mean.

Var(X) = αβ/((α+β)²(α+β+1)) is the variance.

3.3.4: Moment-Generating Functions of Distributions

The following is the definition of a at random variable X's moment-generating function (MGF):

$E[e^{tX}] = M_X(t)$

Regarding distinct at random variables: $\sum(e^{tx} * p(x)) = M_X(t)$

For at random variables that are continuous: $\int(e^{tx} * f(x)) = M_X(t)\ dx$

Characteristics of Functions that Generate Moments

Notes

1. Uniqueness: Two at random variables have the same likelihood distribution if they have the same MGF.

2. Moments The formula $E[X^k] = M_X^{(k)}(0)$ can be used to determine the kth moment of X.

where the kth derivative of $M_X(t)$ assessed at t=0 is denoted by $M_X^{(k)}(0)$.

3. For X and Y, freeat random variables: $M_{X+Y}(t) = M_X(t) \bullet M_Y(t)$

4. Linear Transformations: $M_Y(t) = e^{bt} \bullet M_X(at)$ if Y = aX + b
Functions of Common Distributions that Generate Moments

Bernoulli Distribution (p) = $(1-p) + p \bullet e^t = M_X(t)$

2. Binomial Distribution (n,p) = $(1-p + p \bullet e^t)^n$ $M_X(t)$

3. $M_X(t) = p \bullet e^t / (1 - (1-p) \bullet e^t)$ (for $t < -\ln(1-p)$) is the geometric distribution (p).

Fourth, Poisson Distribution (λ) = $\exp(\lambda(e^t - 1))$ $M_X(t)$

5. Even Dispersion on [a,b]

$(e^{tb} - e^{ta}) / (t(b-a)) = M_X(t)$ (for $t \neq 0$) $M_X(t) = \exp(\mu t + (\sigma^2 t^2)/2)$ is the normal distribution (μ, σ^2).

7. For $t < \lambda$, the Exponential Distribution (λ) is $M_X(t) = \lambda/(\lambda - t)$.

8. For $t < 1/\beta$, the Gamma Distribution (α, β) $M_X(t) = (1-\beta t)^{-\alpha}$

Finding Moments with MGFs

To determine X's mean (initial moment): $M_{X'}(0) = E[X]$

To determine X's second moment: $M_{X''}(0) = E[X^2]$

To determine the difference: $Var(X) = M_{X''}(0) - (M_{X'}(0))^2$

Finding Distributions with MGFs

The distribution of sums of free random variables can be found using MGFs. The MGF of their sum can frequently be identified as a member of a known distribution if X_1, X_2,..., X_n are free random variables with the same distribution.

3.3.5: Applications of Likelihood Distributions

Notes

1. Quality Control and Manufacturing

Binomial Distribution: used to simulate the quantity of faulty products in a sample.

As an illustration, a manufacturing process yields products with a 5% fault rate.

- Expected number of defective items: $E[X] = np = 100 \times 0.05 = 5$
- Variance: $Var(X) = np(1-p) = 100 \times 0.05 \times 0.95 = 4.75$

Modeling the quantity of flaws per unit area or volume is done using the Poisson Distribution.

For instance, a surface's defects have an average of 2.5 per square meter and follow a Poisson distribution. $P(X=3) = (e^{-2.5} \times 2.5^3) / 3! \approx 0.2138$ is the likelihood that a given square meter would have precisely three flaws.

2. Finance and Economics

Normal Distribution: Used to model returns on investment, price fluctuations, and other financial variables.

Example: A stock's daily returns have mean of 0.001 (0.1%) & standard deviation of 0.02 (2%), indicating a normal distribution. This is the likelihood that the return on a certain day will be greater than 3%: $P(X > 0.03) = 1 - P(X < 0.03) = 1 - \Phi((0.03-0.001)/0.02) = 1 - \Phi(1.45) \approx 0.0735 = 7.35\%$

Exponential Distribution: Used to model the time between financial events, such as trades or defaults.

Log-normal Distribution: Used to model asset prices, as they cannot be negative.

3. Reliability Engineering

The lifespan of components with a constant failure rate is modeled using the exponential distribution.

For instance, the annual failure rate of an electrical component is $\lambda = 0.05$. For the component to survive for more than five years, likelihood is $P(X > 5) = e^{-\lambda t} = e^{-0.05 \times 5} = e^{-0.25} \approx 0.7788 = 77.88\%$.

Weibull Distribution: Used to model the lifetime of components with increasing or decreasing failure rates.

4. Queuing Theory

The number of arrivals within a certain time period is modeled using the Poisson Distribution.
The time interval between arrivals and service times is modeled using the exponential distribution.
Example: Twelve consumers an hour on average arrive at a service counter based on a Poisson process. In a one-hour period, the likelihood of precisely ten arrivals is $P(X=10) = (e^{-12} \times 12^{10} / 10! \approx 0.1048 = 10.48\%$.

5. Insurance and Risk Assessment

- Normal Distribution: Used to model aggregate claims in large portfolios.
- Pareto Distribution: Used to model the size of large insurance claims.

6. Biostatistics and Medicine

Binomial Distribution: Used in clinical trials to model the number of successes (e.g., recoveries).

Poisson Distribution: Used to model rare events like disease occurrences.

Example: A rare disease has an average of 3.5 new instances every month, according to a Poisson distribution. $P(X \leq 2) = P(X=0) + P(X=1) + P(X=2) = e^{-3.5} + e^{-3.5} \times 3.5 + e^{-3.5} \times 3.5^2 / 2! \approx 0.0302 + 0.1057 + 0.1850 = 0.3209 =$ 32.09% is the chance of having no more than two new cases in a given month.

7. Physics and Engineering

- Normal Distribution: Used to model measurement errors and physical quantities.
- Maxwell-Boltzmann Distribution: Used to model the speed of molecules in a gas.

8. Computer Science and Networks

Geometric Distribution: Used to model the number of attempts until a successful transmission.

Notes

Example: The failure likelihood for each packet transmission is 0.2. E[X] = 1/p = 1/0.8 = 1.25 is the anticipated number of tries required before a successful transmission.

The amount of network events in given time period is modelled using Poisson Distribution.

Issues Resolved

Issue 1: A Custom Discrete Distribution's Mean and Variance

Issue: For discrete at random variable X, the likelihood mass function is as follows: p(2) = 0.3, p(3) = 0.4, p(4) = 0.1, and p(1) = 0.2. Determine X's variance and mean.

Solution:

Step 1: Determine the average. 0.2 + 0.6 + 1.2 + 0.4 = 2.4 = 1 × 0.2 + 2 × 0.3 + 3 × 0.4 + 4 × 0.1 = E[X] = $\sum$(x * p(x))

Step 2: Determine E[X2]. 0.2 + 1.2 + 3.6 + 1.6 = 6.6 = 1^2 × 0.2 + 2^2 × 0.3 + 3^2 × 0.4 + 4^2 × 0.1 = E[X^2] = $\sum$(x^2 * p(x))

Compute the variance in step three. E[X^2] - (E[X]) = Var(X)$(2.4)^2$ = 6.6 - 5.76 = 0.84 ^{2} = 6.6

As a result, X's variance is 0.84 and its mean is 2.4.

Issue 2: Locating Moments with the MGF

Problem: $M_X(t) = (1-2t)^{-3}$ for t < 1/2 is the moment-generating function ofat random variable X. Determine X's third central moment, variance, and mean.

Answer:

Step 1: Determine the MGF's first derivative. $(-3)(1-2t)^{-4}(-2) = 6(1-2t)^{-4} = M_{X'}(t)$

Step 2: Determine the MGF's second derivative. $6(-4)(1-2t)^{-5}(-2) = 48(1-2t)^{-5} = M_{X''}(t)$

Step 3: Determine the MGF's third derivative. $48(-5)(1-2t)^{-6}(-2) = 480(1-2t)^{-6}$ is $M_{X'''}(t)$.

Step 4: To determine the moments, evaluate the derivatives at t = 0. $M_{X'}(0) = 6(1) = E[X]^{-4} = 6$.

$M_{X''}(0) = 48(1)^{-5} = 48\ E[X^2]$

$M_{X'''}(0) = 480(1) = E[X^3]^{-6} = 480$

Determine the variance in step five. $E[X^2] = Var(X) - (E[X])36 - 48 = 12^2 = 48 - 6^2$

Determine the third central moment in step six. $3E[X] - E[X^3] = E[(X-\mu)^3]E[X^2] + 2(E[X])3(6)(48) + 2(6) = 480 - 864 + 432$

Consequently, the third central moment is 48, variance is 12, and mean of X is 6.

Issue 3: Using the Normal Distribution

The issue is that the average height of adult females in a given group is 165 cm, with a standard deviation of 6 cm.

a) What is the likelihood that a woman chosen at at random will be taller than 175 cm? b) How many women with heights between 160 and 170 cm are predicted if 100 are chosen at random?

Answer:

To ascertain $P(X > 175)$:

Standardize the value in step one. $(175 - 165) / 6 = 1.67$ is z.

Step 2: Use the conventional normal table to get the likelihood. $P(Z > 1.67) = 1 - P(Z \leq 1.67) = 1 - 0.9525 = 0.0475 = P(X > 175))$.

Consequently, the likelihood is roughly 4.75%.

b) To determine how many women should have heights between 160 and 170 cm, start by calculating the likelihood that one woman will have a height between 160 and 170 cm. $-0.83 < Z < 0.83) = P(Z < 0.83) - P(Z < -0.83) = 0.7967 - 0.2033 = 0.5934 = P(160 < X < 170) = P((160-165)/6 < Z < (170-165)/6))$.

Step 2: Determine the anticipated number of women out of 100. The anticipated value is 100×0.5934, or $59.34 \approx 59$

Notes

As a result, we anticipate that roughly 59 women will be between 160 and 170 cm tall.

Issue 4: Poisson and Binomial Approximation

Issue: Computer chips produced on a production line have a 2% failure rate. Four hundred chips are examined as a sample.

a) What is the likelihood of discovering precisely ten faulty chips using the binomial distribution? b) Use the Poisson distribution to approximate this likelihood.

Using the binomial distribution as a solution: $P(X = 10) = C(400,10) \times (0.02)^{10} \times (0.98)^{390}$ $X \sim Bin(400, 0.02)$

Calculating this directly: $(400! / (10! \times 390!)) \times (0.02)^{10} \times (0.98)^{390} \approx 0.1122$ $P(X = 10)$

Consequently, the likelihood is roughly 11.22%.

b) We utilize $\lambda = np = 400 \times 0.02 = 8$. $X \sim Poisson(8)$ $P(X = 10) = (e^{-8} \times 8^{10}) / 10! \approx 0.0992$ for a Poisson approximation.

Consequently, the Poisson approximation yields roughly 9.92%.

Since p is tiny (0.02) and n is high (400), the approximation is fairly near to the binomial likelihood.

Issue 5: Function that Generates Moments for the Total of Free random Variables

The issue is that X_1, X_2, and X_3 are free random variables that all have exponential distributions with parameter $\lambda = 2$. Determine the distribution of Y and the moment-generating function of $Y = X_1 + X_2 + X_3$.

Answer:

First Step: Determine the MGF for every single X_i. $M_X(t) = \lambda/(\lambda-t)$ for $t < \lambda$ is the MGF of an exponential distribution with parameter λ.

For $t < 2$, the MGF for $X_i \sim Exp(2)$ is $M_{Xi}(t) = 2/(2-t)$.

Step 2: Determine $Y = X_1 + X_2 + X_3$'s MGF. The MGF of their sum is the product of their individual MGFs because X_1, X_2, and X_3 are free:

For t less than 2, $M_Y(t) = [2/(2-t)] \times M_{X_i}(t) \times M_{X_2}(t) \times M_{X_3}(t)^3 = 8/(2-t)^3$

Notes

Step 3: Determine Y's distribution. One may identify the MGF of a Gamma distribution with parameters $\alpha = 3$ and $\beta = 1/2$ as $M_Y(t) = 8/(2-t)^3$.

$M_X(t) = (1-\beta t)^{-\alpha}$ for $t < 1/\beta$ is the generic form of the MGF for a Gamma distribution with parameters α and β.

This is contrasted with $M_Y(t) = 8/(2-t)^3$:

• Let's revise: One $\times (1-t/2)^{-3} = 1 \times (1-t/2)^{-3} = 8 \times (2-t)^{-3} = 8 \times (2)^{-3} = M_Y(t)$

• This corresponds to the form $(1-\beta t)^{-\alpha}$ where $\beta = 1/2$ and $\alpha = 3$.

Consequently, Y has a Gamma distribution with $\alpha = 3$ and $\beta = 1/2$. $f\_Y(y) = (1/\beta^{\alpha}) \times (y^{\alpha-1} \times e^{-y/\beta}) / \Gamma(\alpha) = 1/(1/2)^3$ is the PDF of Y. If $y > 0$, then $(y^{3-1} \times e^{-y/(1/2)}) / \Gamma(3) = 8 \times (y^2 \times e^{-2y}) / 2 = 4y^2 \times e^{-2y}$.

Unresolved Issues

Issue 1: Discrete Random Variable

For a discrete random variable X, the probability mass function (PMF) is given as:

$p(x)=k(x+2)$,for $x=0,1,2,3,4$

where k is a constant.

a) Determine the value of k.
b) Compute the mean and variance of X.
c) Derive the moment-generating function (MGF) of X.

Issue 2: Continuous Random Variable

The probability density function (PDF) of a continuous random variable X is given by:

$f(x)=3x2$, $0<x\leq1$,0 otherwise.

a) Calculate the mean and variance of X.
b) Find the probability $P(X>0.8)$
c) Derive the moment-generating function (MGF) of X.
d) Compute the first three moments using the MGF.

Notes

Issue 3: Normal Distribution

IQ scores in a given population follow a normal distribution with a mean of 100 and a standard deviation of 15.

a) What proportion of the population has an IQ greater than 130?
b) What is the probability that a randomly selected person has an IQ between 85 and 115?
c) If 25 individuals are randomly selected, what is the probability that their average IQ is greater than 105?
d) Determine the 90th percentile of the IQ distribution.

Issue 4: Binary Communication System

A communications system transmits messages as a series of bits. Due to noise, each bit has a 10% chance of flipping (changing from 0 to 1 or from 1 to 0). Errors occur independently.

a) What is the probability that exactly 3 out of 20 transmitted bits are flipped?
b) In a 20-bit message, what is the probability that at least one bit is flipped?
c) How many flipped bits should be expected in a 100-bit message?
d) For a critical application, no more than 5% of the bits in a message should be flipped. What is the maximum message length that ensures this requirement is met with at least 95% confidence?

Issue 5: Sum of Uniformly Distributed Random Variables

Let $X_1, X_2, \ldots, X_{20}$ be independent random variables, each uniformly distributed on [0,1].

Define:

$Y = X_1 + X_2 + \cdots + X_{20}$.

a) Compute the mean and variance of Y.
b) Using the Central Limit Theorem, identify an appropriate approximation for the distribution of Y.
c) Estimate $P(9<Y<11)$ using this approximation.

d) Derive the moment-generating function (MGF) of a single uniform random variable X_i.

e) Use the result from part (d) to derive the MGF of Y.

Application of Discrete and Continuous Probability Distributions in Real-World Scenarios

Discrete Probability Distributions

1. Uniform Distribution

- **Quality Control**: Used in manufacturing when defects are equally likely across different production batches.
- **Randomized Clinical Trials**: Ensures unbiased group assignment in medical experiments.
- **Cryptography**: Forms the basis of secure random number generators.
- **Gaming**: Governs fair outcomes in lotteries and dice rolls.

2. Bernoulli Distribution

- **A/B Testing**: Used in digital marketing for testing variations of emails or website layouts.
- **Medical Diagnostics**: Models test results as positive/negative.
- **Risk Assessment**: Helps insurers estimate the probability of claims.
- **Quality Assurance**: Used for pass/fail testing in electronics manufacturing.

3. Binomial Distribution

- **Political Polling**: Models the probability of survey respondents supporting a candidate.
- **Manufacturing Defect Analysis**: Estimates the likelihood of a certain number of defective products.
- **Epidemiology**: Used to model the spread of infections within populations.
- **Finance**: Binomial pricing models are used in options trading.

4. Poisson Distribution

Notes

- **Call Centers**: Predicts the number of incoming calls to optimize staffing.
- **Network Traffic**: Helps analyze data packet arrivals and congestion.
- **Emergency Services**: Models hospital emergency room patient arrivals.
- **Retail Inventory**: Estimates demand for irregularly selling items.

5. Geometric Distribution

- **Quality Control**: Predicts the number of inspections needed before a defect is found.
- **Reliability Engineering**: Models system failures after repeated cycles.
- **Cybersecurity**: Estimates the number of attempts before a system is breached.
- **Customer Acquisition**: Determines how many interactions are needed to convert a lead into a customer.

Continuous Probability Distributions

1. Uniform Distribution

- **Signal Processing**: Simulates quantization errors in digital conversion.
- **Computer Simulations**: Used in Monte Carlo methods.
- **Scheduling**: Models arrival times when there are no peak periods.
- **Queueing Theory**: Helps optimize resource allocation in service industries.

2. Exponential Distribution

- **Reliability Engineering**: Models machine failure times.
- **Customer Service**: Estimates waiting times.
- **Telecommunications**: Models call durations.
- **Nuclear Physics**: Describes radioactive decay processes.

3. Normal Distribution

- **Financial Markets**: Models stock returns and risk assessments.
- **Manufacturing Tolerances**: Predicts variations in product dimensions.
- **Educational Testing**: Standardized test scores often follow a normal distribution.
- **Biometrics**: Models human height, weight, and blood pressure distributions.

Computing Statistical Measures in Real-World Applications

1. Expected Value and Mean

- **Inventory Management**: Helps retailers optimize stock levels.
- **Project Planning (PERT)**: Estimates realistic activity durations.
- **Insurance Actuarial Science**: Used to price policies based on expected claims.
- **Portfolio Management**: Helps investors calculate expected returns.

2. Standard Deviation and Variance

- **Quality Control**: Ensures product consistency in manufacturing.
- **Financial Risk Management**: Measures investment volatility.
- **Clinical Trials**: Quantifies variability in treatment effects.
- **Weather Forecasting**: Assesses uncertainty in predictions.

3. Moment-Generating Functions (MGF)

- **Option Pricing**: Used in risk-neutral pricing models.
- **Reliability Engineering**: Models failure rates over time.
- **Econometrics**: Helps estimate economic indicators with uncertainty.

Multiple-Choice Questions (MCQs)

1. **The Bernoulli distribution is used for:**
   a) Multiple trials
   b) A single trial with two possible outcomes

c) Continuous at random variables

d) None of the above

Answer: b) A single trial with two possible outcomes

2. **Binomial distribution models number of:**

   a) Successes in fixed number of trials

   b) Failures in an infinite number of trials

   c) Continuous outcomes

   d) Free events with varying probabilities

   Answer: a) Successes in fixed number of trials

3. **Poisson distribution is used to model:**

   a) number of occurrences in fixed interval of time & space

   b) Continuous data

   c) The likelihood of an event occurring in a single trial

   d) Data that follows a normal distribution

   Answer: a) number of occurrences in fixed interval of time & space

4. **mean of binomial distributionB(n,p) is:**

   a) np(1−p)

   b) np

   c) p(1−p)

   d) n2p

   Answer: b) np

5. **geometric distribution models:**

   a) number of failures before first success

   b) total number of successes in fixed number of trials

   c) The likelihood of success in one trial

   d) The distribution of continuous variables

   Answer: a) number of failures before first success

6. **exponential distribution is used to model:**

   a) The time between events in a Poisson process

   b) number of successes in fixed trials

   c) distribution of binary data

   d) sum of free variables

Answer: a) The time between events in a Poisson process

7. **normal distributionis also called:**

   a) Poisson distribution

   b) Gaussian distribution

   c) Bernoulli distribution

   d) Binomial distribution

   Answer: b) Gaussian distribution

8. **If normal distribution has mean of 0 and standard deviation of 1, it is called a:**

   a) Standard normal distribution

   b) Skewed normal distribution

   c) Poisson distribution

   d) Geometric distribution

   Answer: a) Standard normal distribution

9. **The moment-generating function (MGF) for normal distribution helps find:**

   a) Mean & variance

   b) Likelihood mass function

   c) Cumulative distribution function

   d) None of above

   Answer: a) Mean & variance

10. **The Poisson distribution is an approximation of the binomial distribution when:**

    a) n is large, and p is small

    b) p is large, and n is small

    c) p is close to 0.5

    d) The number of trials is small

    Answer: a) n is large, and p is small

Short Answer Questions

1. Define likelihood distribution with an example.
2. What is a Bernoulli distribution, and where is it used?
3. Explain the Binomial distribution and its parameters.

Notes

4. Define Poisson distribution and state its properties.
5. What is the Geometric distribution, and what does it model?
6. How is the Exponential distribution related to the Poisson process?
7. What are the key properties of the Normal distribution?
8. Why is the Normal distribution important in statistics?
9. How does a moment-generating function (MGF) help in likelihood distributions?
10. Compare discrete and continuous likelihood distributions with examples.

Long Answer Questions

1. Explain the Bernoulli and Binomial distributions with real-world examples.
2. Derive mean &variance of a Binomial distribution.
3. Explain Poisson distribution and derive its likelihood mass function.
4. Discuss the Geometric distribution and find its expectation.
5. Derive mean &variance of Exponential distribution.
6. Explain Normal distribution and prove its properties.
7. How does Poisson distribution approximate Binomial distribution?
8. Explain moment-generating function (MGF) & use it to find moments of the normal distribution.
9. Compare and contrast Binomial, Poisson, and Normal distributions.
10. How are likelihood distributions applied in real-world scenarios, such as quality control, reliability engineering, and risk analysis?

MODULE 4

UNIT 4.1
Testing of hypothesis: Parameter and statistic

Objectives

- To understand the concept of hypothesis testing.
- To define parameters and statistics in hypothesis testing.
- To learn about null and alternative hypotheses.
- To study sampling distributions and standard errors.
- To analyze critical regions, significance levels, and errors in hypothesis testing.
- To apply large sample tests for mean and proportions.

4.1.1: Introduction to Hypothesis Testing

Hypothesis testing is a fundamental procedure in statistical analysis that allows us to make decisions about populations based on sample data. It provides a framework for determining whether experimental results contain enough evidence to reject a null hypothesis.The basic idea behind hypothesis testing is to state a hypothesis about a population parameter, collect sample data, and then use that data to determine whether there is enough evidence to suggest that the hypothesis is incorrect.

The Process of Hypothesis Testing

1.F ormulate the hypotheses (null and alternative)

2.C hoose a significance level (α)

3.C ollect sample data

4.C alculate the test statistic

5.D etermine the p-value or critical region

6.M ake a decision about the null hypothesis

7.I nterpret the results in context

Notes

Types of Hypothesis Testing Errors

There are two kinds of mistakes that might happen when performing a hypothesis test:

• Type I Error (α): False positive, or rejecting a valid null hypothesis

• Type II Error (β): Neglecting to reject a false negative null hypothesis

The test's significance level, α, represents the likelihood of a Type I mistake. The power of the test is 1-β, and β represents the likelihood of a Type II mistake.

One-Tailed vs. Two-Tailed Tests

Hypothesis tests can be either one-tailed or two-tailed:

- **One-tailed test**: The alternative hypothesis specifies a direction (either greater than or less than)
- **Two-tailed test**: The alternative hypothesis specifies a difference in either direction (not equal to)

4.1.2: Parameters and Statistics

Population Parameters

Since examining an entire population is often impractical, parameters are usually unknown and need to be estimated. Common population parameters include:

- μ (mu): Population mean
- σ^2 (sigma squared): Population variance
- σ (sigma): Population standard deviation
- p: Population proportion
- ρ (rho): Population correlation coefficient

Sample Statistics

A sample's numerical properties that are utilized to estimate the associated population parameter are called statistics. Common sample statistics include:

- $\bar{x}$ (x-bar): Sample mean
- s^2: Sample variance

- s: Sample standard deviation
- $\hat{p}$ (p-hat): Sample proportion
- r: Sample correlation coefficient

Relationship between Parameters and Statistics

Point estimators for population parameters are provided by sample statistics. Since we utilize sample statistics to draw conclusions about population parameters, the link between parameters and statistics is essential to hypothesis testing.

For instance:
• The population mean (μ) is estimated using the sample mean ($\bar{x}$).
• The population proportion (p) is estimated using the sample proportion ($\hat{p}$).

4.1.3: Null and Alternative Hypotheses

The Null Hypothesis (H_0)

The null hypothesis, represented by the letter H_0, asserts that there is no relationship, no effect, or no difference in the population. It stands for the current situation or the assertion that has to be verified. Until there is evidence to the contrary, the null hypothesis is taken to be true.

Examples of null hypotheses:

- H_0: $\mu = 100$ (The population mean equals 100)
- H_0: $p = 0.5$ (The population proportion equals 0.5)
- H_0: $\mu_1 - \mu_2 = 0$ (There is no difference between two population means)

The Alternative Hypothesis (H_1 or H_a)

The alternative hypothesis, denoted as H_1 or H_a, is a statement that contradicts the null hypothesis. It represents what we are trying to establish or prove.

Examples of alternative hypotheses:

- H_1: $\mu \neq 100$ (The population mean does not equal 100) - Two-tailed
- H_1: $\mu > 100$ (The population mean is greater than 100) - One-tailed

- H_1: $p < 0.5$ (The population proportion is less than 0.5) - One-tailed

Formulating Hypotheses

When formulating hypotheses, consider the following guidelines:

1. The null hypothesis should always contain an equals sign (=, ≤, or ≥)
2. The alternative hypothesis should never contain an equals sign (≠, >, or <)
3. The hypotheses should be mutually exclusive (they cannot both be true)
4. The hypotheses should be collectively exhaustive (one of them must be true)

Directional vs. Non-directional Hypotheses

- **Non-directional hypothesis**: States that there is a difference but does not specify the direction (H_1: $\mu \neq 100$)
- **Directional hypothesis**: States that there is a difference in a specific direction (H_1: $\mu > 100$ or H_1: $\mu < 100$)

UNIT 4.2
Sampling distribution and standard error of estimate, Null and alternative hypotheses Simple and composite hypotheses

4.2.1: Sampling Distributions and Standard Errors

Sampling Distribution

The likelihood distribution of a statistic derived from a at random sample of the population is known as the sampling distribution. The sampling distribution of the mean is the most often utilized sampling distribution in hypothesis testing.
Important characteristics of the mean's sample distribution:

• The population mean ($\mu_{\bar{x}} = \mu$) and the sample distribution mean are equivalent.
• The sample distribution's standard deviation, or standard error, is equal to $\sigma/\sqrt{n}$.
• The sample distribution is normal if the population is normally distributed.
• According to the Central Limit Theorem, the sampling distribution is roughly normal if the population is not normal but the sample size is high ($n \geq 30$).

StandardError

The standard deviation of a sample distribution is known as the standard error. It gauges a statistic's precision or variability.
Typical standard errors:

1. **(SEM)**:
    - Population known: $\sigma_{\bar{x}} = \sigma/\sqrt{n}$
    - Population unknown: $s_{\bar{x}} = s/\sqrt{n}$
2. **S tandard Error of the Proportion**:
    - $\sigma_{\hat{p}} = \sqrt{[p(1-p)/n]}$
    - When p is unknown, we use $\hat{p}$ to estimate: $s_{\hat{p}} = \sqrt{[\hat{p}(1-\hat{p})/n]}$
3. **S tandard Error of the Difference Between Two Means**:
    - Free samples: $\sigma_{\bar{x}_1-\bar{x}_2} = \sqrt{(\sigma_1^2/n_1 + \sigma_2^2/n_2)}$
    - When population standard deviations are unknown: $s_{\bar{x}_1-\bar{x}_2} = \sqrt{(s_1^2/n_1 + s_2^2/n_2)}$

Notes

4. **Standard Error of the Difference Between Two Proportions**:
    - $\sigma\hat{p}_1\text{-}\hat{p}_2 = \sqrt{[p_1(1\text{-}p_1)/n_1 + p_2(1\text{-}p_2)/n_2]}$
    - When population proportions are unknown: $s\hat{p}_1\text{-}\hat{p}_2 = \sqrt{[\hat{p}_1(1\text{-}\hat{p}_1)/n_1 + \hat{p}_2(1\text{-}\hat{p}_2)/n_2]}$

The Central Limit Theorem (CLT)

According to the Central Limit Theorem, regardless of the initial population distribution's shape, if you collect large enough at random samples from any population, the sampling distribution of the sample mean will be roughly normally distributed.

Important ramifications of the CLT for testing hypotheses:

• We can presume that the sampling distribution is about normal for large samples ($n \geq 30$).
• For small samples ($n < 30$), the population should be regularly distributed in order to employ t-tests; this enables us to use z-tests for big samples even when the population distribution is unknown or not normal.
The t-Distribution
The t-distribution is used in place of the normal distribution when the population standard deviation (σ) is unknown and needs to be calculated using the sample standard deviation (s).
The t-distribution's characteristics:
• Symmetric and bell-shaped, similar to the typical distribution
• More dispersed (heavier tails) than the average distribution
Degrees of freedom (df), which are correlated with sample size, determine the form. The t-distribution gets closer to the conventional normal distribution as df rises.
• The t-distribution is roughly equivalent to the conventional normal distribution when df> 30.

Z-Scores and t-Scores

In hypothesis testing, test statistics are often calculated as z-scores or t-scores:

- **Z-score** (used when σ is known): $z = (\bar{x} - \mu) / (\sigma/\sqrt{n})$
- **t-score** (used when σ is unknown): $t = (\bar{x} - \mu) / (s/\sqrt{n})$

These test statistics measure how many standard errors the sample statistic is from the hypothesized parameter value.

Solved Problems

Solved Problem 1: One-Sample Z-Test for Population Mean

According to a researcher, adult males in a particular area are taller than 175 cm on average. The average height of 100 adult males from the area selected at at random is 177.5 cm. Test the researcher's assertion at a 5% significance level, assuming that the population standard deviation is 8 cm.

Solution:

Step 1: Set up the hypotheses. H_0: $\mu = 175$ cm (The average height is 175 cm) H_1: $\mu > 175$ cm (The average height is greater than 175 cm)

This is a one-tailed test because the researcher's claim is directional.

Step 2: Determine the significance level. $\alpha = 0.05$

Step 3: Calculate the test statistic. $z = (\bar{x} - \mu) / (\sigma/\sqrt{n})$ $z = (177.5 - 175) / (8/\sqrt{100})$ $z = 2.5 / 0.8$ $z = 3.125$

Step 4: Determine the p-value, or critical value. The critical value for a one-tailed test with $\alpha = 0.05$ is $z_\alpha = 1.645$. $P(Z > 3.125) = 0.00089$ is the p-value.

Step 5: Choose a choice. We reject the null hypothesis because $z = 3.125 > 1.645$ (or p-value = $0.00089 < 0.05$).

Step 6: Evaluate the findings. The researcher's assertion that the average height of adult males in the area is more than 175 cm is sufficiently supported by the available data.

Resolved Issue 2: Population One-Sample t-Test Mean

500 grams of product are meant to be filled into containers by a machine. 25 containers are chosen at at random by a quality control inspector, who discovers that each one has an average of 495 grams with a standard deviation of 10 grams. Is there proof that the containers are being underfilled by the machine? Employ a significance level of 1%.

Solution:

Notes

Step 1: Set up the hypotheses. H_0: $\mu = 500$ grams (The machine is filling properly) H_1: $\mu < 500$ grams (The machine is underfilling)

This is a one-tailed test because we're specifically concerned with underfilling.

Step 2: Determine the significance level. $\alpha = 0.01$

Step 3: Determine the test statistic. We choose a t-test because the sample size is limited ($n < 30$) and the population standard deviation is unknown.

$(\bar{x} - \mu) / (s/\sqrt{n}) = t \Rightarrow t = (495\text{-}500) / (10/\text{-}25) \Rightarrow t = \text{-}5 / 2 \Rightarrow t = \text{-}2.5$

Step 4: Determine the p-value, or critical value. Freedom of degrees = n - 1 = 25 - 1 = 24 The critical value for a one-tailed test with df = 24 and $\alpha = 0.01$ is around $t_\alpha = -2.492$. The value of p is equal to $P(t < -2.5) \approx 0.0096$.

Step 5: Choose a choice. We reject the null hypothesis because $t = -2.5 < -2.492$ (or p-value = $0.0096 < 0.01$).

Step 6: Evaluate the findings. There is enough data to draw the conclusion that the containers are being underfilled by the machine.

Resolved Issue 3: Z-Test for Difference in Population Proportions in Two Samples

The goal of the study is to ascertain whether the percentage of smokers in two cities differs. Out of 400 randomly chosen adults in City A, 120 smoke. Ninety of the 350 randomly chosen adults in City B smoke. Determine whether there is a difference in the percentage of smokers between the two cities at a 5% significance level.

Solution:

Step 1: Set up the hypotheses. H_0: $p_1 = p_2$ (There is no difference in the proportion of smokers) H_1: $p_1 \neq p_2$ (There is a difference in the proportion of smokers)

This is a two-tailed test because we're interested in any difference, regardless of direction.

Step 2: Determine the significance level. $\alpha = 0.05$

Step 3: Calculate the sample proportions. $\hat{p}_1 = 120/400 = 0.3$ (City A) $\hat{p}_2 = 90/350 \approx 0.257$ (City B)

Step 4: Determine the pooled proportion, which is applied in the null hypothesis. (120 + 90) / (400 + 350) = 210/750 = 0.28 is the value of $\hat{p} = (x_1 + x_2) / (n_1 + n_2)$.

Determine the test statistic in step five. $z = \sqrt{[\hat{p}(1-\hat{p})(1/n_1 + 1/n_2)]} / (\hat{p}_1 - \hat{p}_2)$ To calculate z, divide (0.3 - 0.257) by $\sqrt{[0.28(1-0.28)(1/400 + 1/350)]}$. [0.2016(0.0025 + 0.00286)] = 0.043 / √ [0.2016 × 0.00536] = 0.043 / √ z = 0.0329 / 0.043 $Z \approx 1.31$

Step 6: Determine the p-value, or critical value. The critical values for a two-tailed test with $\alpha = 0.05$ are $z_{\alpha}/2 = \pm 1.96$. $2 \times P(Z > 1.31) \approx 2 \times 0.0951 \approx 0.19$ is the p-value.

Make a choice in step seven. Given that $|z| = 1.31 < 1.96$ (or p-value = 0.19 > 0.05), the null hypothesis cannot be ruled out.

Step 8: Evaluate the findings. There is not enough data to draw the conclusion that the two cities' smoking rates are different.

Solved Problem 4: One-Sample Z-Test for Population Proportion

A polling organization claims that more than 60% of adults support a new environmental policy. In a at random sample of 1000 adults, 650 expressed support for the policy. Test the polling organization's claim at a 1% significance level.

Solution:

Step 1: Set up the hypotheses. H_0: p = 0.6 (60% of adults support the policy) H_1: p > 0.6 (More than 60% of adults support the policy)

This is a one-tailed test because the claim is directional.

Step 2: Determine the significance level. $\alpha = 0.01$

Step 3: Calculate the sample proportion. $\hat{p} = 650/1000 = 0.65$

Step 4: Determine the test statistic. $z = \sqrt{[p(1-p)/n]} / (\hat{p} - p)$ $z = \sqrt{[0.6(0.4)/1000]} / (0.65 - 0.6)$ 0.05 / √[0.24/1000] is z. z = 0.05 / 0.0155 $Z \approx 3.23$

Step 5: Determine the p-value, or critical value. The critical value for a one-tailed test with $\alpha = 0.01$ is $z_{\alpha} = 2.326$. P(Z > 3.23), the p-value, is 0.0006

Notes

Make a choice in step six. We reject the null hypothesis because $z = 3.23 > 2.326$ (or p-value = $0.0006 < 0.01$).

Step 7: Evaluate the findings. The polling group's assertion that over 60% of adults favor the new environmental policy is well supported by the available data.

Resolved Issue 5: Two-Sample t-Test for Population Mean Difference

A researcher wishes to evaluate the efficacy of two distinct teaching strategies. A mean test score of 78 with a standard variation of 12 is obtained when 30 students employ Method A. A mean test score of 85 with a standard deviation of 15 is obtained when Method B is used to 25 pupils. Examine whether there is a difference in the mean test scores between the two approaches at a 5% significance level, assuming that the populations have identical variances.

Solution:

Step 1: Set up the hypotheses. H_0: $\mu_1 = \mu_2$ (There is no difference in mean test scores) H_1: $\mu_1 \neq \mu_2$ (There is a difference in mean test scores)

This is a two-tailed test because we're interested in any difference, regardless of direction.

Step 2: Determine the significance level. $\alpha = 0.05$

Step 3: Determine the standard deviation of the pooled data. To calculate s_p, divide $((n_1-1)s_1^2$ by $(n_2-1)s_2^2)$ by $(n_1 + n_2 - 2)$ $((30-1)12^2 + (25-1)15^2) / (30 + 25 - 2) = \sqrt{[s_p]} \Rightarrow ((29)(144) + (24)(225)) / 53 = \sqrt{[s_p]} \Rightarrow [(4176 + 5400) / 53] = \sqrt{[s_p]} \Rightarrow s_p = \sqrt{[9576 / 53]} \Rightarrow s_p \approx 13.44 \Rightarrow s_p = \sqrt{180.68}$

Step 4: Calculate the test statistic. $t = (\bar{x}_1 - \bar{x}_2) / (s_p \times \sqrt{(1/n_1 + 1/n_2)})$ $t = (78 - 85) / (13.44 \times \sqrt{(1/30 + 1/25)})$ $t = -7 / (13.44 \times \sqrt{(0.0333 + 0.04)})$ $t = -7 / (13.44 \times \sqrt{0.0733})$ $t = -7 / (13.44 \times 0.2708)$ $t = -7 / 3.64$ $t \approx -1.92$

Step 5: Find the critical value or p-value. Degrees of freedom = $n_1 + n_2 - 2 = 30 + 25 - 2 = 53$ For $\alpha = 0.05$ in a two-tailed test with df = 53, the critical values are approximately $t_{\alpha}/2 = \pm 2.006$. The p-value = $2 \times P(t < -1.92) \approx 2 \times 0.03 \approx 0.06$

Step 6: Decide on something. Given that $|t| = 1.92 < 2.006$ (or p-value = $0.06 > 0.05$), the null hypothesis cannot be ruled out.

Step 7: Evaluate the findings. There is not enough data to draw the conclusion that the two teaching strategies differ in terms of mean test results.

Notes

Unresolved Problems

Problem 1:

A manufacturer claims that its light bulbs have an average lifespan of at least 1000 hours. A random sample of 36 light bulbs shows an average lifespan of 980 hours, with a standard deviation of 120 hours. Test the manufacturer's claim at a 5% significance level.

Problem 2:

The effectiveness of a new medication in reducing cholesterol levels is under investigation. A study of 20 individuals who took the medication recorded an average cholesterol reduction of 25 mg/dL, with a standard deviation of 12 mg/dL. Test whether the medication effectively lowers cholesterol at a 1% significance level.

Problem 3:

A survey suggests that the proportion of adults who exercise regularly has increased from 40% five years ago. In a random sample of 500 individuals, 220 report exercising regularly. Test whether the percentage has increased at a 5% significance level.

Problem 4:

The effects of two different fertilizers on crop yield are being compared. A sample of 40 plots treated with Fertilizer A shows an average yield of 25 bushels per acre, with a standard deviation of 4 bushels. A sample of 45 plots treated with Fertilizer B shows an average yield of 27 bushels per acre, with a standard deviation of 5 bushels. Test whether the mean yields of the two fertilizers differ at a 1% significance level.

Problem 5:

Notes

A researcher aims to determine whether support for a new municipal policy differs between men and women. In a random sample of 300 men, 165 support the policy, while in a random sample of 350 women, 175 express support. Test for a significant difference in support between men and women at a 5% significance level.

4.2.2: Critical Region and Level of Significance

Introduction to Critical Region

The set of test statistic values that result in the null hypothesis being rejected is known as the critical zone (or rejection region) in hypothesis testing. We compute a test statistic from our sample data and compare it with a crucial value while doing a hypothesis test. The null hypothesis is rejected if the test statistic is within the crucial zone; if not, it is not rejected. The vital area is dependent upon:

1. The employed test statistic
2. The test's selected significance threshold (α)
3. Is the test two-tailed or one-tailed?
4. Significance Level (α)

The chance of rejecting the null hypothesis when it is true is represented by the level of significance, which is represented by α (alpha). Another name for this is the likelihood of making a Type I error.

Typical values for α consist of:

- 0.10 (10%)
- 0.05 (5%)
- 0.01 (1%)
- 0.001 (0.1%)

fields where errors can have serious consequences (e.g., medical research), smaller values of α like 0.01 or 0.001 are often used.

Determining the Critical Region

To determine the critical region, we need to:

1. Choose the significance level (α)

2. Determine the test statistic's distribution under the null hypothesis.
3. Identify the crucial value or values that divide the non-rejection region from the rejection area.

For a z-test (when the normal distribution applies):

- For a two-tailed test with significance level α:
    - Critical values: ±z(α/2)
    - Critical region: $z < -z(\alpha/2)$ or $z > z(\alpha/2)$
- For a right-tailed test with significance level α:
    - Critical value: z(α)
    - Critical region: $z > z(\alpha)$
- For a left-tailed test with significance level α:
    - Critical value: -z(α)
    - Critical region: $z < -z(\alpha)$

The z-score with an area of α to its right under the conventional normal curve is denoted by the symbol z(α).

The connection between the p-value and the critical region
The critical region and the p-value are directly correlated:

• Assuming the null hypothesis is correct, the p-value is the likelihood of receiving a test statistic at least as extreme as the one observed.

• If α is less than or equal to the p-value, we reject the null hypothesis.

• The p-value will be less than or equal to α if the test statistic is within the critical zone.

Critical Value Examples for Various Levels of Significance

Regarding the z-distribution, or standard normal distribution:

Two-Tailed Test:

- For α = 0.10: Critical values = ±1.645
- For α = 0.05: Critical values = ±1.96
- For α = 0.01: Critical values = ±2.576

- For α = 0.001: Critical values = ±3.291

One-Tailed Test:

- For α = 0.10: Critical value = ±1.28 (sign depends on direction)
- For α = 0.05: Critical value = ±1.645 (sign depends on direction)
- For α = 0.01: Critical value = ±2.33 (sign depends on direction)
- For α = 0.001: Critical value = ±3.09 (sign depends on direction)

For the t-distribution, critical values depend on both α and the degrees of freedom.

4.2.3: Types of Errors in Hypothesis Testing

There are two kinds of mistakes that might happen when making decisions in hypothesis testing:
Error Type I (False Positive)

Rejecting a correct null hypothesis is a Type I mistake. Stated differently, we get the conclusion that there is an effect or difference when, in fact, none exists.

• Type I error likelihood = α (significance level)

• In a symbolic sense, P(Type I Error) = P(Reject $H_0|H_0$ is true) = α For instance, convicting an innocent person in a criminal trial entails rejecting the null hypothesis of innocence when the individual is in fact innocent.

Error Type II (False Negative)

Failure to reject a faulty null hypothesis is a Type II mistake. To put it another way, we assume that there is no difference or effect when, in fact, there is.

• Type II error likelihood = β (beta)

In a criminal trial, a Type II error is acquitting a guilty person (failing to reject the null hypothesis of innocence when the individual is actually guilty). This can be expressed symbolically as follows: P(Type II Error) = P(Fail to reject $H_0|H_0$ is false) = β.

Test Power

The likelihood of successfully rejecting a false null hypothesis is known as a statistical test's power. It's equivalent to 1 minus β.
• $P(\text{Reject } H_0 | H_0 \text{ is false}) = 1 - \beta = \text{Power}$

A Type II error is unlikely to occur in a test with high power.

Type I and Type II Error-Related Factors

1. Sample Dimensions:

Greater sample numbers lower the likelihood of both kinds of errors.

Power increases (β lowers) as sample size grows.

2. Significance Level (α): o Type I errors are less likely when α is decreased.

Nevertheless, lowering α raises the possibility of Type II error (decreases power) for a fixed sample size.

3. Effect Size: o Greater deviations from the null hypothesis are associated with larger effect sizes. boost power

Power is reduced by smaller effect sizes (increasing β).

4. Variability: o Both kinds of errors are more likely to occur when data variability is higher.

Power rises when variability decreases.

The Connection Between α and β

Both Type I and Type II errors have a trade-off:

• It is easier to miss a real effect when α is decreasing, which makes it harder to reject H_0; conversely, when α is increasing, which makes it easier to reject H_0, β is decreasing, which makes it less likely to miss a real effect.

Increasing the sample size is the only method to concurrently reduce both kinds of mistakes.

Error Types in Decision Table Format

	H_0 is True	**H_0 is False**
Reject H_0	Type I Error (α)	Correct Decision (1-β) (Power)
Fail to Reject H_0	Correct Decision (1-α)	Type II Error (β)

4.7 One-Tailed and Two-Tailed Tests

Notes

Depending on the alternative hypothesis's shape, hypothesis testing can be classified as either one-tailed (directed) or two-tailed (non-directional).

Test with Two Tails

When the alternative hypothesis asserts, without indicating a direction, that the parameter of interest differs from the value given in the null hypothesis, a two-tailed test is employed.

Form of hypotheses:

- H_0: Parameter = specified value
- H_1: Parameter ≠ specified value

Example:

- H_0: $\mu = 100$
- H_1: $\mu \neq 100$

The critical region in a two-tailed test is divided between the distribution's two tails, each of which contains $\alpha/2$ of the area.

When to use:

- When you want to detect a difference in either direction
- When there is no prior expectation about the direction of the effect
- When you're equally interested in deviations above or below the value specified in H_0

One-Tailed Test

When the alternative hypothesis indicates a direction for the difference from the value in the null hypothesis, a one-tailed test is employed.

Right-Tailed Test

Form of hypotheses:

- H_0: Parameter ≤ specified value
- H_1: Parameter > specified value

Example:

- H_0: $\mu \leq 100$

- H_1: $\mu > 100$

In a right-tailed test, the entire critical region (α) is in the right tail of the distribution.

Left-Tailed Test

Form of hypotheses:

- H_0: Parameter $\geq$ specified value
- H_1: Parameter $<$ specified value

Example:

- H_0: $\mu \geq 100$
- H_1: $\mu < 100$

In a left-tailed test, the entire critical region (α) is in the left tail of the distribution.

When to use one-tailed tests:

- When you have a specific direction of interest based on theory or prior research
- When you're only concerned with detecting an effect in one direction
- When detecting an effect in the opposite direction would lead to the same decision as no effect

Comparison of Critical Values

For the same significance level (α), the critical value for a one-tailed test is less extreme than for a two-tailed test:

- For $\alpha = 0.05$:
    - Two-tailed test critical z-value: ± 1.96
    - One-tailed test critical z-value: ± 1.645 (sign depends on direction)

This makes one-tailed tests more powerful for detecting effects in the specified direction, but they have no power to detect effects in the opposite direction.

Notes

Choosing Between One-Tailed and Two-Tailed Tests

Consider using a one-tailed test when:

1. You have a clear directional hypothesis based on theory or prior research
2. You're only interested in detecting an effect in one specific direction
3. An effect in the opposite direction would be treated the same as no effect

Consider using a two-tailed test when:

1. You have no prior expectation about the direction of the effect
2. You want to detect any deviation from the null hypothesis value
3. An effect in either direction would be meaningful and lead to different conclusions

Many researchers prefer two-tailed tests because:

1. They protect against unexpected findings in the opposite direction
2. They are more conservative and generally more accepted in scientific publications
3. They allow for the possibility that your directional hypothesis might be wrong

Solved Problems

Solved Problem 1: Critical Region for a Z-Test

Problem:A researcher is examining whether a novel approach to instruction raises student achievement. The new method's mean test score is 75, the same as the old method's, according to the null hypothesis. The mean score is different from 75, according to the alternative hypothesis. Assume that 36 students will be chosen at at random and that the population standard deviation is 15. Find the two-tailed test's crucial region when $\alpha = 0.05$.

Solution:

1. The hypotheses are:
    - H_0: $\mu = 75$

- $H_1: \mu \neq 75$

2. This is a two-tailed test with $\alpha = 0.05$.
3. For a z-test, the critical values are $\pm z(\alpha/2) = \pm z(0.025) = \pm 1.96$.
4. The test statistic is: $z = (\bar{x} - \mu_0)/(\sigma/\sqrt{n}) = (\bar{x} - 75)/(15/\sqrt{36}) = (\bar{x} - 75)/2.5$
5. The critical region is: $z < -1.96$ or $z > 1.96$
6. In terms of the sample mean:
   - $(\bar{x} - 75)/2.5 < -1.96$ or $(\bar{x} - 75)/2.5 > 1.96$
   - $\bar{x} < 75 - (1.96 \times 2.5)$ or $\bar{x} > 75 + (1.96 \times 2.5)$
   - $\bar{x} < 70.1$ or $\bar{x} > 79.9$

Therefore, the critical region in terms of the sample mean is: $\bar{x} < 70.1$ or $\bar{x} > 79.9$.

Solved Problem 2: Type I and Type II Errors

Problem:A quality control engineer checks to see if the manufactured bearings' mean diameter is 10 mm. H_0: $\mu = 10$ mm is the null hypothesis, whereas H_1: $\mu \neq 10$ mm is the alternative. Clearly state the meaning of Type I and Type II errors in this situation.

Solution:

Type I Error (Rejecting a true H_0):

- This occurs if the engineer concludes that the mean diameter is not 10 mm when it actually is 10 mm.
- This might lead to unnecessary adjustments to the manufacturing process, wasting time and resources.
- The likelihood of this error is α (the significance level chosen for the test).

Type II Error (Failing to reject a false H_0):

- This occurs if the engineer concludes that the mean diameter is 10 mm when it actually is not 10 mm.

- This might lead to continued production of bearings with incorrect diameters, potentially causing problems in applications where these bearings are used.
- The likelihood of this error is β, which depends on the true value of μ, the sample size, and the significance level.

In quality control, both errors have consequences:

- Type I error leads to false alarms and unnecessary adjustments.
- Type II error allows defective products to pass inspection.

The engineer must balance these risks by choosing an appropriate significance level and ensuring adequate sample size for sufficient power.

Solved Problem 3: One-Tailed vs. Two-Tailed Test

Problem:A new medication, according to a pharmaceutical company, lowers blood pressure by at least 10 mmHg on average. The likelihood that the reduction is less than 10 mmHg is something that a researcher want to test against this assertion. Set up the appropriate hypotheses and determine whether a one-tailed or two-tailed test is appropriate. Explain your reasoning.

Solution:

The claim is that the drug reduces blood pressure by at least 10 mmHg.

Let μ = the mean reduction in blood pressure due to the drug.

The claim is $\mu \geq 10$ mmHg.

The researcher wants to test against the possibility that the reduction is less than 10 mmHg.

Appropriate hypotheses:

- H_0: $\mu \geq 10$ mmHg (The drug reduces blood pressure by at least 10 mmHg)
- H_1: $\mu < 10$ mmHg (The drug reduces blood pressure by less than 10 mmHg)

This is a left-tailed test because:

1. The alternative hypothesis specifies a direction (less than 10 mmHg).
2. The critical region will be entirely in the left tail of the distribution.
3. The researcher is only interested in detecting if the drug's effect is less than claimed.

Reasoning:

- The company claims a reduction of at least 10 mmHg, which forms our null hypothesis.
- The researcher's concern is specifically about the drug not meeting this claim (i.e., having a smaller effect than claimed), not about it exceeding the claim.
- Since there's a specific directional concern, a one-tailed test is appropriate.
- Specifically, it's a left-tailed test because the alternative hypothesis involves values less than the null hypothesis value.

Solved Problem 4: Calculating Type II Error Likelihood

Problem: According to the manufacturer, their light bulbs have a typical lifespan of at least 1000 hours. A researcher wishes to use a at random sample of twenty-five bulbs to test this assertion. It is known that the population standard deviation is 200 hours. For a left-tailed test, the researcher will employ a significance level of $\alpha = 0.05$. Determine the likelihood of a Type II error in the event that the actual mean lifetime is 950 hours..

Solution:

1. For a z-test with $\alpha = 0.05$, the critical value is $-z(\alpha) = -z(0.05) = -1.645$.
2. The critical region is: $z < -1.645$
3. In terms of the sample mean:
    - $(\bar{x} - 1000)/(200/\sqrt{25}) < -1.645$
    - $(\bar{x} - 1000)/40 < -1.645$

- $\bar{x} < 1000 - (1.645 \times 40)$
- $\bar{x} < 934.2$

4. Calculate β, the likelihood of Type II error when $\mu = 950$:
   - $\beta = P(\text{Fail to reject } H_0 | \mu = 950)$
   - $\beta = P(\bar{x} \geq 934.2 | \mu = 950)$
5. Standardize this likelihood:
   - $\beta = P((\bar{x} - 950)/(200/\sqrt{25}) \geq (934.2 - 950)/(200/\sqrt{25}))$
   - $\beta = P(z \geq (934.2 - 950)/40)$
   - $\beta = P(z \geq -0.395)$
   - $\beta = 1 - P(z < -0.395)$
   - $\beta = 1 - 0.3464$
   - $\beta = 0.6536$ or approximately 65.36%

Consequently, the likelihood of failing to reject the null hypothesis (committing a Type II error) is roughly 65.36% if the true mean lifetime is 950 hours.

Solved Problem 5: Comparing One-Tailed and Two-Tailed Tests

Problem:A researcher is examining the potential effects of a novel medication on heart rate. The mean change in heart rate is zero beats per minute, according to the null hypothesis. Calculate the critical regions and critical values for: A test with two tails and $\alpha = 0.05$, a test with a right tail and $\alpha = 0.05$, and a test with a left tail and $\alpha = 0.05$

Presume that the test statistic has a normal distribution. What is the comparison of the critical regions?

Solution:

a) Two-tailed test with $\alpha = 0.05$:

- $H_0: \mu = 0$
- $H_1: \mu \neq 0$
- Critical values: $\pm z(\alpha/2) = \pm z(0.025) = \pm 1.96$

- Critical region: $z < -1.96$ or $z > 1.96$

b) Right-tailed test with $\alpha = 0.05$:

- H_0: $\mu \leq 0$
- H_1: $\mu > 0$
- Critical value: $z(\alpha) = z(0.05) = 1.645$
- Critical region: $z > 1.645$

c) Left-tailed test with $\alpha = 0.05$:

- H_0: $\mu \geq 0$
- H_1: $\mu < 0$
- Critical value: $-z(\alpha) = -z(0.05) = -1.645$
- Critical region: $z < -1.645$

Comparison of critical regions:

1. The two-tailed test has critical values that are more extreme (± 1.96) than the one-tailed tests (± 1.645).
2. The two-tailed test divides the significance level between both tails (0.025 in each tail), while the one-tailed tests place the entire significance level (0.05) in one tail.
3. The one-tailed tests have more power to detect effects in the specified direction but no power to detect effects in the opposite direction.
4. If the true effect is in the direction specified by the alternative hypothesis, a one-tailed test is more likely to detect it than a two-tailed test at the same significance level.

Unsolved Problems

Problem 1:

The objective of the study is to determine whether a new fertilizer enhances plant growth. The null hypothesis states that the mean growth with the new fertilizer is 25 cm, the same as the conventional fertilizer. The alternative

hypothesis asserts that the mean growth exceeds 25 cm. A random sample of 16 plants is selected, with a population standard deviation of 4 cm.

a) Identify the critical region for a right-tailed test at a significance level of $\alpha = 0.01$.
b) Compute the test statistic and determine whether to reject the null hypothesis if the sample mean is 27.5 cm.
c) Calculate and interpret the p-value.

Problem 2:

A company claims that its batteries have an average lifespan of at least 30 hours. A consumer group seeks to test this claim using a sample of 40 batteries, given that the population standard deviation is 5 hours.

a) Formulate the null and alternative hypotheses.
b) Determine the critical region for a left-tailed test at $\alpha = 0.05$.
c) Calculate the probability of making a Type II error if the true mean lifespan is 28 hours.
d) Discuss how increasing the sample size to 60 batteries would affect the probability of a Type II error.

Problem 3:

A researcher is investigating the effect of a new teaching method on student performance. The null hypothesis states that the mean test score using the new method is 70, which is the historical average for the traditional method. The researcher is interested in any deviation from this historical average.

a) Determine whether a one-tailed or two-tailed test is appropriate by setting up the correct hypotheses. Justify your choice.
b) Identify the critical values if the researcher uses $\alpha = 0.05$ and the test statistic follows a t-distribution with 24 degrees of freedom.
c) If the researcher later decides that only an improvement in test scores is of interest, how would the hypotheses and critical region change?

Problem 4:

Notes

A quality control engineer is evaluating whether the average package weight on a production line is 500 grams. The null hypothesis is H_0: $\mu = 500$ grams, while the alternative hypothesis is H_1: $\mu \neq 500$ grams. The significance level is set at $\alpha = 0.05$.

a) Explain the meaning of Type I and Type II errors in this context.
b) Calculate the probability of a Type II error if a sample of 25 packages is taken, the true mean weight is 505 grams, and the standard deviation is 10 grams.
c) Discuss how the probability of a Type II error would change if the significance level were increased to $\alpha = 0.10$.

Problem 5:

A medical researcher is examining whether a new treatment lowers cholesterol levels. The null hypothesis states that the mean reduction is 0 mg/dL (no effect), while the alternative hypothesis asserts that the treatment is effective if the mean reduction is greater than 0 mg/dL.

a) Compute the test statistic for a right-tailed test at $\alpha = 0.05$, given a sample size of 30, a sample mean reduction of 8 mg/dL, and a sample standard deviation of 15 mg/dL. Determine whether to reject the null hypothesis.
b) Calculate and interpret the p-value.
c) Explain how the results would differ if the researcher had used a two-tailed test instead.
d) Given a population standard deviation of 15 mg/dL, a true mean reduction of 5 mg/dL, and $\alpha = 0.05$, determine the required sample size to achieve a power of 0.90.

Formula Sheet

Critical Values

For z-tests (standard normal distribution):

Two-tailed test (α):

- The crucial variables are $\pm z(\alpha/2)$.

Test with a right tail (α):

- $z(\alpha)$ is the critical value.

Test with a left tail (α):

- Critical value: $-z(\alpha)$

Common critical z-values:

For $\alpha = 0.10$:

- Two-tailed: ± 1.645
- One-tailed: 1.28 (right) or -1.28 (left)

For $\alpha = 0.05$:

- Two-tailed: ± 1.96
- One-tailed: 1.645 (right) or -1.645 (left)

For $\alpha = 0.01$:

- Two-tailed: ± 2.576
- One-tailed: 2.33 (right) or -2.33 (left)

For $\alpha = 0.001$:

- Two-tailed: ± 3.291
- One-tailed: 3.09 (right) or -3.09 (left)

Test Statistics

Z-test (known population standard deviation): $z = (\bar{x} - \mu_0)/(\sigma/\sqrt{n})$

T-test (unknown population standard deviation): $t = (\bar{x} - \mu_0)/(s/\sqrt{n})$

Where:

- $\bar{x}$ = sample mean
- μ_0 = hypothesized population mean
- σ = population standard deviation
- s = sample standard deviation
- n = sample size

Likelihood of Type II Error (β)

For a right-tailed z-test with alternative $\mu = \mu_1 > \mu_0$: $\beta = P(z < z(\alpha) - (\mu_1 - \mu_0)/(\sigma/\sqrt{n})) = \Phi(z(\alpha) - (\mu_1 - \mu_0)/(\sigma/\sqrt{n}))$

For a left-tailed z-test with alternative $\mu = \mu_1 < \mu_0$: $\beta = P(z > -z(\alpha) - (\mu_1 - \mu_0)/(\sigma/\sqrt{n})) = 1 - \Phi(-z(\alpha) - (\mu_1 - \mu_0)/(\sigma/\sqrt{n}))$

When using the alternative $\mu = \mu_1 \neq \mu_0$ for a two-tailed z-test: $(z(\alpha/2) - |\mu_1 - \mu_0|/(\sigma/\sqrt{n})) = \Pi + \Phi$ The value of $\Phi(-z(\alpha/2) - |\mu_1 - \mu_0|/(\sigma/\sqrt{n}))$

where

- $\Phi(z)$ is the standard normal distribution's cumulative distribution function.
- μ_1 = true value of the population mean
- μ_0 = hypothesized value in the null hypothesis

Power Calculation

Power = $1 - \beta$

Sample Size Determination

To achieve a specific power $(1-\beta)$ for detecting a difference of size $|\mu_1 - \mu_0|$:

For a two-tailed test: $n = [(z(\alpha/2) + z(\beta))^2\sigma^2]/(\mu_1 - \mu_0)^2$

For a one-tailed test: $n = [(z(\alpha) + z(\beta))^2\sigma^2]/(\mu_1 - \mu_0)^2$

Where:

- $z(\alpha)$ = critical value for significance level α
- $z(\beta)$ = critical value corresponding to β (Type II error likelihood)
- σ = population standard deviation
- $\mu_1 - \mu_0$ = effect size (difference to be detected)

4.2.4 Large Sample Tests for Mean and Proportion

The Central Limit Theorem, which asserts that regardless of the population distribution's form, the sampling distribution of the sample mean approaches a normal distribution, can be used when working with large samples (usually $n \geq 30$). Our hypothesis testing processes are made simpler by this potent theorem, which enables us to execute statistical inference using the normal distribution as an approximation.

Notes

When the sample size is large enough, statistical procedures known as large sample tests are employed to draw conclusions regarding population parameters. These assessments are reliable and broadly relevant in a number of disciplines, such as the social sciences, psychology, economics, and medicine.

Key Concepts in Hypothesis Testing

Before diving into specific tests, let's review the fundamental concepts of hypothesis testing:

1. **Null Hypothesis (H_0)**: The default assumption or status quo that we aim to test.
2. **Alternative Hypothesis (H_1 or H_a)**: The claim that challenges the null hypothesis.
3. **Test Statistic**: A value calculated from sample data used to determine whether to reject H_0.
4. **Critical Region**: The set of values for the test statistic that lead to rejecting H_0.

Large Sample Test for Population Mean

The following test statistic is used for testing hypotheses on a population mean μ with a large sample:
$(\bar{x} - \mu_0)/(\sigma/\sqrt{n}) = z$

Where:
$\bar{x}$ represents the sample mean.

σ is the population standard deviation, and μ_0 is the estimated population mean (derived from H_0).
• The sample size is n.

When n is big, we can use the sample standard deviation s in place of σ, which is frequently unknown:
$(\bar{x} - \mu_0)/(s/\sqrt{n}) = z$

Rule of Decision

For a two-tailed test (H_0: $\mu = \mu_0$, H_1: $\mu \neq \mu_0$) at a significance threshold of α:
o Dismiss H_0 if $|z| > z_{\alpha/2}$ or if p-value < α

• For a test with a right tail (H_0: $\mu \leq \mu_0$, H_1: $\mu > \mu_0$), reject H_0 if $z > z_\alpha$ or if the p-value is less than α.

When a test is left-tailed (H_0: $\mu \geq \mu_0$, H_1: $\mu < \mu_0$), reject H_0 if $z < -z_\alpha$ or if the p-value is less than α.
where the critical values from the standard normal distribution are denoted by $z_{\alpha/2}$ and z_α.

Test of Population Proportion with a Large Sample

We utilize $z = (\hat{p} - p_0)/\sqrt{[p_0(1-p_0)/n]}$ to test hypotheses on a population proportion p with a large sample.
Where:
The sample proportion is denoted by $\hat{p}$.

• n is the sample size; • p_0 is the estimated population proportion (derived from H_0).

Conditions for Validity

To ensure the validity of the large sample proportion test:

1. The sample needs to be chosen at random.

2. The sample size needs to be sufficiently big so that both np_0 and $n(1-p_0) > 5$.

Rule of Decision

The choice criteria are the same as for the mean, evaluating the p-value against α or comparing the computed z-statistic with the relevant critical value.

The Connection Between Hypothesis Testing and Confidence Intervals

The formula $\bar{x} \pm z_{\alpha/2} \times (\sigma/\sqrt{n})$ yields a $(1-\alpha)\times100\%$ confidence range for a population mean μ.

Alternatively, if σ is unknown:

$z_{\alpha/2} \times (s/\sqrt{n})$ $\bar{x} \pm$

For a population proportion p, the $(1-\alpha)\times100\%$ confidence interval is $\hat{p} \pm z_{\alpha/2} \times \sqrt{[\hat{p}(1-\hat{p})/n]}$.

Two-tailed hypothesis tests and confidence intervals are directly related:

Notes

H_0 is rejected if the hypothesized value μ_0 (or p_0) lies outside the confidence interval; if it lies inside the confidence interval, H_0 is not rejected.

Errors of Type I and Type II

There are two kinds of mistakes that might happen in hypothesis testing:

1. Type I Error: False positive, or rejecting H_0 when it is true

Likelihood = α (level of significance)

2. Failure to reject H_0 when it is incorrect (false negative) is a Type II error.

Test power = 1 - β (likelihood of successfully rejecting a fake H_0) o Likelihood = β

A test's power can be impacted by the following factors:

• Sample size (n): Power rises with larger samples.

• Significance level (α): raising α raises the possibility of Type I error while also increasing power.

• Effect size: Power increases with more discrepancies between the actual parameter value and the predicted value.

• Variability: Power rises with less variability (smaller σ).

4.2.5: Hypothesis Testing in Real-Life Applications

Applications in Medicine and Healthcare

Hypothesis testing is fundamental in clinical trials and medical research. Researchers use these statistical methods to determine whether new treatments, drugs, or medical procedures are effective.

Application: COVID-19 Vaccine Efficacy

During the COVID-19 pandemic, large-scale clinical trials used hypothesis testing to evaluate vaccine efficacy. For instance:

- H_0: Vaccine efficacy $\leq$ 50% (FDA threshold for approval)
- H_1: Vaccine efficacy > 50%

Researchers calculated: Efficacy = 1 - (Infection rate in vaccinated group)/(Infection rate in placebo group)

Statistical significance in these trials provided evidence for vaccine approval and distribution.

Notes

Applications in Business and Economics

Market Research

Companies use hypothesis testing to make data-driven decisions about products, services, and marketing strategies.

A/B Testing Example

A company testing two different website designs might use:

- H_0: There is no difference in conversion rates between designs A and B
- H_1: There is a difference in conversion rates between designs A and B

After collecting data, they can calculate: $z = (\hat{p}_a - \hat{p}_\beta)/\sqrt{[\hat{p}(1-\hat{p})(1/n_a + 1/n_\beta)]}$

Where $\hat{p} = (x_a + x_\beta)/(n_a + n_\beta)$ is the pooled proportion.

Economic Policy Analysis

Economists apply hypothesis testing to evaluate policy effectiveness:

- Analyzing unemployment rates before and after policy implementation
- Comparing economic growth across different regions with different policies
- Assessing the impact of interest rate changes on inflation

Applications in Quality Control

Manufacturing companies employ statistical quality control to maintain product standards. Hypothesis testing helps monitor production processes and detect deviations.

Example: Production Line Monitoring

Consider a process producing components with a target diameter of 10 mm:

- H_0: The mean diameter = 10 mm (process is in control)

- H_1: The mean diameter ≠ 10 mm (process needs adjustment)

Regular sampling and testing allow for timely intervention when the process drifts out of specification.

Applications in Social Sciences

Hypothesis testing helps researchers in psychology, sociology, and education validate theories and evaluate interventions.

Example: Educational Method Comparison

When comparing traditional teaching methods with a new approach:

- H_0: There is no difference in student performance between methods
- H_1: The new method results in different student performance

Test scores or other performance metrics can be analyzed using appropriate statistical tests to guide educational policy.

Environmental Applications

Scientists use hypothesis testing to monitor climate change, pollution effects, and conservation efforts.

Example: Climate Data Analysis

Testing whether average temperatures have increased:

- H_0: The mean annual temperature has not changed
- H_1: The mean annual temperature has increased

Long-term temperature data can be analyzed to detect significant trends that inform environmental policy.

4.2.6: Examples and Case Studies

Solved Problems

Solved Problem 1: Large Sample Test for Population Mean

Problem: According to a manufacturer, the average lifespan of its light bulbs is at least 1000 hours. After testing 100 bulbs, a consumer advocacy group discovers that the sample mean lifespan is 985 hours, with a sample standard variation of 120 hours. Does the manufacturer's assertion have evidence to refute it at a 5% significance level?

Solution:

Step 1: Define the hypotheses

- H_0: $\mu \geq 1000$ (manufacturer's claim)
- H_1: $\mu < 1000$ (consumer group's suspicion)

The test is left-tailed.

Step 2: Determine that $z = (\bar{x} - \mu_0)/(s/\sqrt{n}) \Rightarrow z = (985 - 1000)/(120/\sqrt{100}) \Rightarrow z = -15/(120/10) \Rightarrow z = -15/12 \Rightarrow z = -1.25$ by computing the test statistic.

Step 3: Determine the critical value For $\alpha = 0.05$ in a left-tailed test, $z_\alpha = -1.645$

Step 4: Make a decision Since $-1.25 > -1.645$, we do not reject H_0.

Step 5: Give the conclusion. There is not enough data to refute the manufacturer's assertion that the average lifespan of their light bulbs is at least 1000 hours at the 5% significance level.

Solved Problem 2: Large Sample Test for Population Proportion

Problem: A political analyst claims that more than 60% of voters support a new policy. In a at random sample of 500 voters, 325 express support for the policy. Test the analyst's claim at a 1% significance level.

Solution:

Step 1: Define the hypotheses

- H_0: $p \leq 0.60$
- H_1: $p > 0.60$

This is a right-tailed test.

Step 2: Check the conditions for using the normal approximation

- $np_0 = 500 \times 0.60 = 300 \geq 5$
- $n(1-p_0) = 500 \times 0.40 = 200 \geq 5$

The conditions are satisfied.

Step 3: Calculate the sample proportion $\hat{p} = 325/500 = 0.65$

Notes

Step 4: Determine $z = (\hat{p} - p_0)/\sqrt{[p_0(1-p_0)/n]}$ $z = (0.65 - 0.60)/\sqrt{[0.60(0.40)/500]}$ as the test statistic. $0.05/0.022$ $z = 2.27$ $z = 0.05/\sqrt{0.00048}$ z

Step 5: Determine the critical value For $\alpha = 0.01$ in a right-tailed test, $z\_\alpha = 2.33$

Step 6: Make a decision Since $2.27 < 2.33$, we do not reject H_0.

Step 7: Explain the conclusion. There is not enough data to back up the analyst's assertion that over 60% of voters like the new policy at the 1% significance level.

Solved Problem 3: Confidence Interval and Hypothesis Testing Relationship

Problem: According to one study, college students read 300 words per minute on average. With a standard deviation of 48 words per minute, the mean reading speed of 64 pupils selected at at random is 312 words per minute. To test the researcher's claim, create a 95% confidence interval for the mean reading speed.

Answer:
First Step: Determine what the 95% confidence interval is. $z_{\alpha/2} = 1.96$ $\bar{x} \pm z_{\alpha/2} \times (s/\sqrt{n})$ $312 \pm 1.96 \times (48/\sqrt{64})$ $312 \pm 1.96 \times (48/8)$ $312 \pm 1.96 \times 6$ 312 ± 11.76 at a 95% confidence level
The range of the 95% CI is [300.24, 323.76].

Step 2: Use the confidence interval to test the hypothesis

- H_0: $\mu = 300$
- H_1: $\mu \neq 300$

At the 5% significance level, we do not reject H_0 because 300 is (very barely) included in the confidence interval.

Step 3: Provide a conclusion The researcher's assertion that the average reading speed is 300 words per minute cannot be refuted due to the lack of proof. .

Solved Problem 4: Comparing Two Population Proportions

Problem: A new teaching method is being evaluated. Of 200 students taught using the traditional method, 140 passed the exam. Of 250 students taught using the new method, 195 passed. At a 5% significance level, is there evidence that the new method has a higher pass rate?

Solution:

Step 1: Define the hypotheses

- H_0: $p_1 \geq p_2$ (traditional method's pass rate is greater than or equal to the new method's)
- H_1: $p_1 < p_2$ (new method has a higher pass rate)

This test is left-tailed.

Step 2: Calculate sample proportions $\hat{p}_1 = 140/200 = 0.70$, $\hat{p}_2 = 195/250 = 0.78$

Step 3: Calculate the pooled proportion (assuming H_0 is true) $\hat{p} = (140 + 195)/(200 + 250) = 335/450 = 0.744$

Step 4: Determine $z = (\hat{p}_1 - \hat{p}_2)/\sqrt{[\hat{p}(1-\hat{p})(1/n_1 + 1/n_2)]}$ as the test statistic. 0.70 - 0.78 / √[0.744(0.256)(1/200 + 1/250)] is z. [0.190464(0.005 + 0.004)] = -0.08/√ [0.190464 × 0.009] z = -0.08/√= -0.08/√0.001714 ⇒Z = -0.08/0.0414 ⇒ z = -1.932

Step 5: Determine the critical value For $\alpha = 0.05$ in a left-tailed test, $z_\alpha = -1.645$

Step 6: Make a decision Since -1.932 < -1.645, we reject H_0.

Step 7: Give the conclusion. There is enough data to draw the conclusion that the new teaching strategy outperforms the conventional one in terms of pass rate at the 5% significance level.

Solved Problem 5: Hypothesis Testing in Real-Life Application

ProblemA pharmaceutical company creates a novel cholesterol-lowering medication. The medication was administered to 45 high-cholesterol individuals for three months during clinical trials. Their mean cholesterol level was 240 mg/dL prior to therapy. The mean level was 218 mg/dL with a standard variation of 25 mg/dL following treatment. Check to see if the medication lowers cholesterol at a 1% significance level.

Notes

Solution:

Step 1: Define the hypotheses

- H_0: $\mu d \geq 0$ (drug does not lower cholesterol)
- H_1: $\mu d < 0$ (drug lowers cholesterol)

Where μd is the mean difference (after - before).

Step 2: Calculate the mean difference and test statistic Mean difference = 218 - 240 = -22 mg/dL

Since we're testing the mean difference: $z = (-22 - 0)/(25/\sqrt{45}) \Rightarrow z = -22/(25/6.71) \Rightarrow z = -22/3.73 \Rightarrow z = -5.90$

Step 3: Determine the critical value For $\alpha = 0.01$ in a left-tailed test, $z_\alpha = -2.33$

Step 4: Make a decision Since $-5.90 < -2.33$, we reject H_0.

Step 5: Give the conclusion. There is enough data to draw the conclusion that the medication lowers cholesterol at the 1% significance level.

Unsolved Problems

Problem 1: Large Sample Test for Population Mean

A corporation claims that its employees work an average of 45 hours per week. However, a labor union suspects that this number is higher. A random sample of 100 employees shows an average weekly work time of 47.2 hours with a standard deviation of 8.5 hours.

a) Formulate the null and alternative hypotheses.
b) Compute the test statistic.
c) Draw a conclusion at a 5% significance level.
d) Determine and interpret the p-value.

Problem 2: Test of Population Proportion with a Large Sample

A quality control manager asserts that no more than 5% of manufactured products are defective. However, in a random sample of 400 products, 30 are found to be defective.

a) Define the null and alternative hypotheses.
b) Calculate the sample proportion.
c) Identify the appropriate statistical test.
d) Make a conclusion at a 1% significance level.
e) Construct a 99% confidence interval to estimate the actual proportion of defective products.

Problem 3: Comparing Two Population Means

A researcher is comparing the effectiveness of two different standardized test preparation methods. A sample of 60 students using Method A achieves an average score of 78.5 with a standard deviation of 8.2, while a sample of 75 students using Method B attains an average score of 82.1 with a standard deviation of 9.5.

a) Establish the null and alternative hypotheses to test if the mean scores differ between the two methods.
b) Compute the test statistic.
c) Make a conclusion at a 5% significance level.
d) Construct a 95% confidence interval for the difference in mean scores.

Problem 4: Application in Marketing

A company wants to determine whether a new advertising campaign has increased daily sales. Before the campaign, the average daily sales were $12,000. After 50 days of the campaign, the average daily sales increased to $13,200, with a standard deviation of $1,800.

a) Set up the null and alternative hypotheses to evaluate the campaign's effectiveness.
b) Compute the test statistic.
c) Determine if there is sufficient evidence to conclude that the campaign increased sales at a 5% significance level.
d) Discuss the possible errors in this hypothesis test and their consequences.

Problem 5: Case Study in Public Health

A public health authority wants to assess whether a new health education program has increased the community's vaccination rate. Before the

Notes

program, the vaccination rate was 65%. After implementation, a random sample of 300 community members reveals that 210 have received the vaccine.

a) Define the null and alternative hypotheses.
b) Identify the appropriate statistical test.
c) Evaluate the conclusion at a 1% significance level.
d) Compute and interpret the p-value.
e) Explain the implications of committing a Type I error in this context.

Additional Considerations in Hypothesis Testing

Practical Significance vs. Statistical Significance

It's important to distinguish between statistical significance and practical significance:

- **Statistical Significance**: shows that it is unlikely that the observed findings happened by accident.
  • Practical Significance: Shows that the effect is significant enough in a real-world setting.

With large samples, even small differences can be statistically significant but may lack practical importance. Researchers should consider the magnitude of the effect and its real-world implications.

Effect Size Measures

Effect size measures quantify the magnitude of the difference between groups or the strength of a relationship.

Common effect size measures include:

1. **Cohen's d** for comparing means: $d = |\mu_1 - \mu_2|/\sigma$

Where σ is the pooled standard deviation.

Interpretations:

- d = 0.2: Small effect
- d = 0.5: Medium effect
- d = 0.8: Large effect

2. **Correlation coefficient (r)** for measuring association: Values range from -1 to 1, with the magnitude indicating the strength of association.
3. **Odds ratio** for comparing proportions: The ratio of the odds of an event occurring in one group to the odds of it occurring in another group.

Multiple Testing Problem

The likelihood of producing at least one Type I error rises while performing numerous hypothesis tests. This is referred to as the multiplicity problem or the numerous testing problem.

Methods to address this issue include:

1. **Bonferroni Correction**: Adjust the significance level by dividing α by the number of tests. $\alpha' = \alpha/m$, where m is the number of tests.
2. **False Discovery Rate (FDR) Control**: Controls the expected proportion of false discoveries among all discoveries.
3. **Family-Wise Error Rate (FWER) Control**: Controls the likelihood of making one or more Type I errors.

Assumptions and Robustness

The Central Limit Theorem, which permits a normal distribution to approximate the sampling distribution, is the foundation of large sample tests. Other presumptions, nevertheless, might still be relevant:

1. **Independence**: Observations should be free of each other.
2. **At random Sampling**: The sample should be randomly selected from the population.
3. **Large Sample Size**: The sample size should be sufficiently large (generally $n \geq 30$).

Tests are considered robust if moderate violations of assumptions still yield reliable results.

Power Analysis and Sample Size Determination

Notes

Power analysis helps determine the sample size needed to detect an effect of a specific size with a given level of confidence.

The power of a test (1 - β) depends on:

- Sample size (n)
- Significance level (α)
- Effect size
- Variability in the population

For a test of a population mean: $n = (z_\alpha + z_\beta)^2 \times \sigma^2/\Delta^2$

Where:

- z_α is the critical value for Type I error
- z_β is the critical value for Type II error
- σ is the population standard deviation
- Δ is the minimum detectable difference

Sequential and Adaptive Testing

In some applications, especially clinical trials, sequential or adaptive testing approaches may be used. These methods allow for interim analyses and potential early stopping of a study based on accumulated data.

Benefits include:

- Ethical considerations (stopping a trial early if treatment shows clear benefit or harm)
- Efficiency in resource allocation
- Flexibility in study design

However, these approaches require careful planning and appropriate statistical adjustments to maintain the integrity of the analysis. Large sample tests for means and proportions form the foundation of many statistical analyses in research and real-world applications. Understanding the principles of hypothesis testing, interpreting results correctly, and recognizing the limitations and assumptions of these methods are essential skills for making data-driven decisions. The examples and case studies

presented in this document illustrate how these statistical techniques can be applied across various fields to answer important questions and guide decision-making processes. By combining theoretical knowledge with practical applications, researchers and professionals can effectively utilize hypothesis testing to extract meaningful insights from data and make informed conclusions about populations based on sample evidence.

Practical Applications of Hypothesis Testing

Hypothesis testing forms the backbone of inferential statistics across numerous fields. Here are practical applications based on the key concepts you've outlined:

Banking and Finance

Credit Scoring Models: Banks develop hypothesis tests to determine if new scoring algorithms significantly improve default prediction rates. The null hypothesis might state that the new algorithm performs no better than the existing one. Investment Strategies: Portfolio managers test hypotheses about whether certain investment approaches yield significantly higher returns. They define parameters (like mean return) and calculate statistics from market data to test claims about performance.

Fraud Detection: Financial institutions analyze transaction patterns using hypothesis testing to identify anomalous behaviors. Critical regions are established where unusual activity triggers additional verification, balancing false positives (legitimate transactions flagged as fraud) against false negatives (missed fraud).

Healthcare

- Pharmaceutical Trials: Drug developers test whether new medications produce significant improvements over placebos or existing treatments. The alternative hypothesis typically suggests the new drug is more effective, while the null hypothesis indicates no difference.
- Medical Screening: Hospitals analyze the sensitivity and specificity of diagnostic tests through hypothesis testing. They carefully monitor Type I errors (false positives) and Type II errors (false negatives), adjusting significance levels based on the severity of missing a diagnosis versus unnecessary treatment.

- Public Health Monitoring: Health departments use proportion tests to determine if disease rates in specific populations differ significantly from baseline levels, helping identify emerging outbreaks early.

Manufacturing

- Quality Control: Manufacturers implement statistical process control using hypothesis testing to monitor if production remains within acceptable parameters. Sampling distributions help them understand expected variation in measurements.
- Product Reliability: Engineers test whether design improvements significantly extend product lifespan. Using standard error calculations, they determine if observed differences in durability are statistically meaningful or due to random variation.
- Supply Chain Optimization: Companies analyze whether new logistics approaches significantly reduce delivery times, using large sample tests when working with historical shipment data.
- Marketing and Retail
- A/B Testing: E-commerce sites test whether different webpage designs significantly impact conversion rates. Null hypotheses typically assume no difference between designs, with critical regions determined by desired significance levels.
- Pricing Strategy: Retailers test hypotheses about optimal price points by analyzing sales data across different store locations. The parameter of interest is usually mean revenue or profit, with statistics calculated from sample data.
- Customer Retention: Subscription businesses test whether new engagement programs significantly improve retention rates, using proportion tests to determine if differences are statistically meaningful.

Agriculture

- Crop Yield Improvement: Farmers test whether new fertilizers or farming techniques significantly increase yields. Sampling distributions help account for natural variation in growing conditions.

- Pest Resistance: Agricultural researchers use hypothesis testing to determine if certain crop varieties show significantly improved resistance to pests, defining parameters like infestation rates.
- Soil Quality Management: Land managers test hypotheses about whether soil amendment practices significantly improve nutrient content, using appropriate significance levels to guide investment decisions.

Environmental Science

- Climate Change Analysis: Researchers test whether observed temperature changes differ significantly from historical patterns. The null hypothesis typically represents natural variation, while the alternative suggests human influence.
- Water Quality Monitoring: Environmental agencies use hypothesis testing to determine if pollutant levels exceed regulatory thresholds, carefully defining significance levels to balance environmental protection against false alarms.
- Conservation Efforts: Wildlife biologists test whether population management strategies significantly increase endangered species numbers, using appropriate statistical methods to account for sampling challenges in wildlife counts.

Technology

- Algorithm Performance: Software engineers use hypothesis testing to determine if new algorithms significantly improve processing speed or accuracy. Critical regions help them decide when improvements are substantial enough to implement.
- User Experience: Product designers test whether interface changes significantly improve user satisfaction or task completion rates, drawing conclusions from sample data to infer population-wide effects.
- Network Reliability: Telecommunications companies test hypotheses about whether infrastructure upgrades significantly reduce outage rates, using large sample tests to analyze performance data.

Education

- Teaching Methods: Educators test whether new instructional approaches significantly improve student outcomes. The parameter of interest is typically mean test scores, with statistics calculated from class samples.
- Admissions Criteria: Universities analyze whether certain admission factors significantly predict student success, using hypothesis testing to evaluate the predictive power of various metrics.
- Resource Allocation: School districts test hypotheses about whether additional funding in specific areas significantly improves educational outcomes, guiding budget priorities based on statistical evidence.

Academic Research

- Psychology Studies: Researchers test whether experimental conditions produce significant differences in human behavior or cognitive processing, carefully defining null and alternative hypotheses to align with research questions.
- Social Science Research: Social scientists use hypothesis testing to determine if demographic factors significantly influence social outcomes, employing appropriate statistical methods based on sampling distributions.
- Scientific Discoveries: Researchers across disciplines test whether observed phenomena differ significantly from theoretical predictions, using significance levels to determine when findings warrant publication and further investigation.

In all these applications, practitioners must carefully:

- Define clear parameters and measurable statistics
- Formulate appropriate null and alternative hypotheses
- Understand the underlying sampling distributions
- Select appropriate significance levels based on the consequences of Type I and Type II errors

Apply the correct statistical tests based on sample size and data characteristics

Interpret results with consideration of both statistical and practical significance

Notes

By mastering these fundamentals of hypothesis testing, professionals across diverse fields can make more informed, evidence-based decisions while properly accounting for uncertainty and random variation.

Multiple-Choice Questions (MCQs)

1. **A hypothesis is a:**
   a) Conclusion based on data
   b) Statement about a population parameter
   c) At random guess
   d) Statistical test

   Answer: b) Statement about a population parameter

2. **The null hypothesis (H_0) represents:**
   a) The claim being tested
   b) The opposite of the research hypothesis
   c) A sample statistic
   d) A confirmed conclusion

   Answer: a) The claim being tested

3. **The alternative hypothesis (H_1) is:**
   a) The hypothesis we seek evidence for
   b) The same as the null hypothesis
   c) Always accepted
   d) A parameter of the population

   Answer: a) The hypothesis we seek evidence for

4. **The likelihood of rejecting a true null hypothesis is known as:**
   a) Type I error
   b) Type II error
   c) Confidence level
   d) Significance level

   Answer: a) Type I error

5. **The likelihood of accepting a false null hypothesis is called:**
   a) Type I error

b) Type II error

c) Level of significance

d) Power of test

Answer: b) Type II error

6. **The level of significance (α) represents:**

   a) likelihood of making Type I error

   b) likelihood of making Type II error

   c) The acceptance region

   d) The sampling error

   Answer: a) likelihood of making Type I error

7. **A one-tailed test is used when:**

   a) We are testing for an extreme deviation in one direction

   b) The population mean is unknown

   c) sample size is large

   d) hypothesis is two-sided

   Answer: a) We are testing for an extreme deviation in one direction

8. **two-tailed test is applied when:**

   a) population standard deviation is unknown

   b) critical region is in both tails of the distribution

   c) sample size is large

   d) likelihood is greater than 1

   Answer: b) critical region is in both tails of the distribution

9. **The Z-test is used when:**

   a) sample size is small

   b) population variance is known

   c) sample variance is unknown

   d) data is not normally distributed

   Answer: b) population variance is known

10. **A large sample test for single mean is conducted using:**

    a) Chi-square test

    b) Z-test

c) t-test

d) F-test

Answer: b) Z-test

Short Answer Questions

1. Define hypothesis testing and explain its purpose.
2. Differentiate between null and alternative hypotheses.
3. What are Type I and Type II errors? Provide an example.
4. Explain one-tailed and two-tailed tests with examples.
5. What is the level of significance, and why is it important?
6. How do sampling distributions affect hypothesis testing?
7. What is a critical region, and how is it determined?
8. Explain the concept of standard error in hypothesis testing.
9. Describe the Z-test and its applications.
10. When should a large sample test for a single proportion be used?

Long Answer Questions

1. Explain the steps involved in hypothesis testing with an example.
2. Derive the standard error formula for mean and proportion.
3. Discuss one-tailed and two-tailed tests with real-life applications.
4. Explain Type I and Type II errors and their impact on decision-making.
5. How is Z-test used for hypothesis testing of means and proportions? Provide examples.
6. Discuss the role of significance level (α) and confidence intervals in hypothesis testing.
7. Explain critical region and p-value with an example.
8. How can hypothesis testing be applied in business and healthcare?

Notes

9. Compare and contrast parametric and non-parametric tests in hypothesis testing.
10. Solve a numerical problem involving large sample test for a mean or proportion.

Unit 5.1
Critical region, Level of significance, One tailed and two tailed tests, Two types of errors

Objectives

- Understand the fundamental concepts of hypothesis testing as a systematic method for making decisions about population characteristics based on sample information.
- Learn to formulate null and alternative hypotheses that represent competing statements about population parameters.
- Comprehend the significance level (α) and its role in determining the probability of committing a Type I error.
- Distinguish between critical regions, critical values, and their importance in hypothesis testing decision-making.
- Differentiate between one-tailed and two-tailed tests and understand appropriate contexts for each.

5.1.1: Fundamentals of Hypothesis Testing

Hypothesis testing is one of the foundations of statistical inference — the formal framework used by statisticians and researchers to make general statements about populations based on information contained in the sample. At its core, hypothesis testing is a systematic method for making decisions about population characteristics based on limited information. This starts by creating a pair of hypotheses that contradict each other on a single characteristic of the population; these are usually on a parameter (mean, proportion, or variance). The competing statements are hypotheses regarding two different views of the population, and the evidence from our sample will help us determine which hypothesis better explains our observed evidence. Hypothesis testing is so beautiful. Using a systematic approach, hypothesis testing applies the principles of mathematics to establish if the difference observed is significant or can be explained by chance resulting from random sampling. We will discuss the key notions that form the basis of hypothesis

Notes

testing to help the reader gain insight on critical regions, significance levels, types of tests and errors in statistical decisions (i.e., false positive and false negative). Some of these may seem a bit technical, they help readers understand how statistical decisions are made and how to assess the reliability and limitations of these decisions.

The Framework Beyond Statistical Hypotheses

Before examining the details of hypothesis testing, we need to understand what statistical hypotheses are. Hypothesis TestingA statistical hypothesis is a formal statement about a population parameter (such as mean, proportion, variance, …etc) or about the distribution of a population. However, in our statistical context, a hypothesis has to be defined in a way so that we can test it using statistical procedures based on mathematics and probability. Hypothesis testing always involves working with two alternative hypotheses representing potential realities. The null hypothesis (H_0) is the default position, the status quo, the statement that there is no effect, no difference, no relationship. In general, it represents the existing belief or the fact to be challenged. In medicinal trials, the null hypothesis may state that "a drug has no effect on recovery time" or that "recovery time on the drug is equal to recovery time without the drug." The statement being directly tested in the statistical procedure is known as the null hypothesis. The alternative hypothesis (denoted as H_1 or H_a) is the position that stands in direct opposition to the null hypothesis. The alternative hypothesis is often the one that the researcher believes is true or hopes to show. In our medication example, the alternative hypothesis would say something like, "the medication decreases recovery time," or "the recovery time for the medication is different from that of the recovery time without it." We cannot directly test the alternative hypothesis; we obtain acceptance through default by rejecting the null hypothesis if the evidence suggests doing so. These competing hypotheses provide a decision framework such that our statistical procedure will either result in our rejection of the null hypothesis (thereby accepting the alternative) or fail to reject the null hypothesis (and maintain our position of the status quo).

Critical Region:

H_0 is said to be rejected if and only if the test statistic takes a value in a so-called critical set (or rejection region) associated with the null hypothesis.

Critical region is the collection of the values of the test statistic that cause rejection of the null. That is to say, if our calculated test statistic from the sample lies in this critical region we reject the null, otherwise we do not reject. Critical region is the region whose determination depends on many factors First, it relies on the theory of the distribution of the test statistic under the null hypothesis. There are various test statistics (z, t, chi-square or F) that follow different distributions, and the shapes of these distributions dictate the location and size of the critical region. Second, the critical region size is affected by the level of significance (α); a larger α means a larger critical region, i.e., more likely to reject the null hypothesis. Third, whether or not the test is one-tailed or two-tailed influences the distribution of the critical region on the tails of the distribution. Critical values are the thresholds separating the rejection region from the acceptance region in hypothesis testing. Those are the location measures of the test statistic; These values are the points on the distribution where the critical region is separated from the non-critical region. Critical values for commonly used distributions such as normal, t, chi-square, and F distributions hand in hand can be looked up in statistical tables or be computed using statistical software. Learn how to calculate and interpret critical values is an important skill for correctly applying hypothesis tests and interpreting the results.

P-Value: One Number To Rule Them All

The comparison of observed to expected frequencies leads to the calculation of p-values and significance levels, where significance level (α (alpha)) is the accepted level of probability for rejecting a null hypothesis that is true (Type I Error) in hypothesis testing. To summarize, α is the probability of falsely claiming an effect or difference exists when it does not. For example, if a researcher desires their overallα set at 0.05, this means they are accepting an error rate of 5% to reject the null when in fact it is true. Consequently, the choice of a significance level has far-reaching implications for hypothesis-testing actions and interpretations made on its outcomes. Values of α often used are 0.10, 0.05, and 0.01, with 0.05 being most typically used across scientific fields. A significance level of 0.05 represents a compromise between the risks of Type I error (which is rejecting, when we should not) and Type II error (which is do not reject, when we should). In more stringent fields or applications, a lower α value may be required, such as 0.01, thus only allowing a 1% risk of false

rejection. In contrast, exploratory research or pilot studies could use a more liberal α such as 0.10 to obtain a 10% risk of false rejection to avoid missing an effect that could be substantial enough to warrant further testing. The critical region depends on the chosen significance level. Increasing α produces a wider critical region, making null hypothesis rejection easier and therefore requiring less evidence, while decreasing α produces a narrower critical region, making rejection harder and thus requiring greater evidence. This reflects the qualitative trade-off in hypothesis testing; as the risk of one kind of mistake (Type I) goes down, the other type (Type II) goes up. Researchers need to evaluate the trade-off between type I and type II errors in their specific research questions, contexts, and design in a context tailored to their work.

Directing Our Focus: One-Tailed vs Two-Tailed Tests

Based on the direction of the alternative hypothesis, the hypothesis test can be classified as one-tailed (directional) or two-tailed (non-directional), which affects the definition of the critical region and how the results are interpreted. This difference indicates whether the researcher wants to detect only a direction of difference (one-tailed) or any difference regardless of direction (two-tailed). The alternative hypothesis in a one-tailed test represents the directionality of the effect, stating that the population parameter is either less than or greater than the value stated in the null hypothesis. For instance, a researcher may suspect that using a new pedagogy raises test scores (H_1: $\mu > \mu_0$) or that using a new medication decreases recovery time (H_1: $\mu < \mu_0$). The critical region is thus completely within one tail of the distribution—the upper (right) tail, for a "greater than" alternative, or the lower (left) tail, for a "less than" alternative. If theory, previous research, or logical constraints suggest that an effect or difference can only exist in one direction, then the one-tailed test is appropriate and provides greater statistical power for detecting the effect compared to the two-tailed test using the same level of significance. A two-tailed test uses an alternative hypothesis that is not directional; it states only that the population parameter is not equal to the null hypothesis value (H_1: $\mu \neq \mu_0$). For example, a researcher may hypothesize that a new drug has an effect on recovery time, but they may not specify whether that effect decreases or increases recovery time. Here, the rejection region or critical region is divided into two tails in each of which, we have placed α2 as our critical

value. On the other hand, two-tailed tests are more conservative and should be used when there is not a strong theoretical reason to predict the effect to be positive or negative or when the researcher wishes to detect any departure from the null hypothesis, regardless of its direction.

The implications of choosing one-tailed versus two-tailed tests are critical both in terms of statistical power and interpretation. However, one-tailed tests have the advantage of providing more power to detect an effect in the specified direction, but they are unable to detect effects in the opposite direction. While two-tailed tests are less powerful at detecting an effect in a particular direction, they protect against the loss of power to detect unanticipated effects in the opposite direction. Deciding which of the two approaches is more appropriate for a particular context, researchers need to think through their research questions, the theoretical underpinnings, and what a potential Type I or Type II error entails.

5.1.2: Recognizing the Two Types of Errors: An Essential part of Decision-Making

Hypothesis testing is a fundamental concept in statistics and researchers must be able to effectively mitigate the risks of making an error when drawing conclusions based on sample data. There are two types of hypothesis testing errors: First, when a false null hypothesis is rejected, and the second when the null hypothesis is not rejected despite true null hypothesis.

Type I error: Finding a difference in the null hypothesis when that null hypothesis is true: A false positive finding. Thus, α — the significance level you choose for the test — is the probability of committing a Type I error. Type I error occurs who make a Type I error: A drug company grows a clinical trial and incorrectly concludes that their drug is effective when it is not (the drug is useless), in this case they used $\alpha = 0.05$. Type I errors can result in the adoption of ineffective treatments (including drugs), inappropriate process changes, or misguided research being published as valid evidence. In medicine, for instance, erroneous positive results may result in prescribing harmful treatments to patients, which could be extremely damaging. On the other hand, a Type II error occurs when the null hypothesis is not rejected when it is actually false, a “false negative” finding. The Type II error probability (β) is a function of several parameters

Notes

(α, sample size, effect size, and population variability). FTA-MultIX (Type 2 error detection) If a medical screening test fails to diagnose a disease that is present, there is happened the Type II error. Type II errors may lead to missed treatment opportunities, failure to adopt beneficial changes, or failing to recognize real effects that warrant further inquiry. The power of a statistical test ($1 - \beta$) describes the likelihood of rejecting a false null hypothesis and as such is the capacity of a test to detect a true effect (when one exists). There is an underlying tension between Type I and Type II errors, in that for a given sample size, decreasing the risk of one type of error generally increases the risk of the other. Lowering α (requiring more evidence to reject the null when that is false) increases β (increasing the chance of missing a real effect), whereas increasing α (requiring less evidence to reject the null when that is false) reduces β (upper bound to detect a real effect when it exists). This interplay highlights the need for careful research design and, where appropriate, sample size determination via power analysis so that these opposing risks are weighed against one another relative to the situation and the implications associated with either decision in a given research context.

Unit – 5.2

Tests of significance: Large sample tests for single mean, Single proportion

Hypothesis Testing Steps: A Systematic Procedure

This is a systematic process developed to obtain objectivity and reliability to the result deduct from sample data. Such systematic thinking lays out a clear roadmap for exploring research inquiries and taking empirically justified actions. This journey starts with the creation of a null and alternative hypothesis. The competing statements must be clearly defined, mutually exclusive, and focused on the research question. The null hypothesis usually represents the existing state of affairs or a proposition of no statistical difference, whereas the alternative hypothesis represents the researcher's hypothesis, or the position that is contrary to the null. In the case of a test to see if a new medication reduces blood pressure, for example, the null might be "the medication has no effect on blood pressure" (H_0: $\mu = \mu_0$) and the alternative might be "the medication reduces blood pressure" (H_1: $\mu < \mu_0$) After formulating the hypotheses, the researcher chooses a suitable significance level (α), this is a probability threshold that the researcher is willing to accept for rejecting the null hypothesis when it is, in fact, true. This should be determined prior to conducting data collection and analysis, based on the implications of Type I and Type II errors in the particular research context. Five percent, one percent, and ten percent are common values for levels of significance, but five percent is used most broadly across disciplines. The next step is to choose the test statistic, a mathematical tool used to assess the evidence against the null hypothesis. When deciding on which test statistic to use, one has to take into account factors such as if the data is independent or paired data, the sample size and the type of parameter being tested (mean, variance, proportion etc.) and assumptions made about the distribution of the population. Examples of test statistics include the z-statistic, t-statistic, chi-square statistic and F-statistic, which are used in specific circumstances. Once the test statistic is selected, the investigator establishes the critical region, that is the collection of values of the test statistic that would cause rejection of the null hypothesis. Whether or not you can reject a null hypothesis is determined by the significance level, the distribution of the test statistic under the null

Notes

hypothesis, and whether the test is one-tailed or two-tailed. A hypothesis test-related term, the critical region sets the decision rule.

Next, the researcher computes the value of the test statistic from the sample data using the relevant formula associated with the selected test statistic. As in hypothesis testing, a decision is then made by comparing the value to the critical region: if the test statistic is in the critical region the null hypothesis is rejected in favor of the alternative. In the last stage, the results are interpreted based on the original research question. This interpretation should balance the limitations of the test, describe what the decision means in practice, and address common misinterpretations, such as confusing a failure to reject H_0 with proving H_0 to be true. Also, a full interpretation takes into account the practical import of the findings as well as their statistical significance and may include effect size measures that quantify the extent of the observed differences or relationships.

Solved Problems:

Problem 1: Testing a Claim About a Population Mean

A manufacturer claims their light bulbs have a mean lifetime of at least 1,000 hours. A quality control engineer tests this claim by randomly sampling 36 bulbs, which have a mean lifetime of 950 hours with a standard deviation of 120 hours. The engineer needs to verify the manufacturer's claim at a 5% significance level.

Step 1: State the Hypotheses

H_0: $\mu \geq 1000$ hours (the manufacturer's claim)

H_1: $\mu < 1000$ hours (the alternative hypothesis)

This is a lower-tailed test since we're testing whether the true mean lifetime is less than the claimed value.

Step 2: Select the Significance Level

- $\alpha = 0.05$

Step 3: Calculate the Test Statistic

Since our sample size is large ($n = 36$), we can use the z-statistic:

$z = (\bar{x} - \mu_0)/(s/\sqrt{n})$

$z = (950 - 1000)/(120/\sqrt{36})$

$z = -50/20$

$z = -2.5$

Step 4: Determine the Critical Value

For a lower-tailed test with $\alpha = 0.05$, the critical value is -1.645.

The critical region is $z < -1.645$.

Step 5: Make a Decision

Since our test statistic $z = -2.5$ is less than the critical value -1.645, it falls in the critical region. Therefore, we reject the null hypothesis.

Step 6: State the Conclusion

There is sufficient evidence at the 5% significance level to reject the manufacturer's claim. The data suggests that the mean lifetime of the light bulbs is less than the claimed 1,000 hours. Based on our sample, the average lifetime is approximately 950 hours, which is statistically significantly lower than the advertised value.

Problem 2: Testing for a Difference from a Known Value

A psychologist wants to determine if students in an experimental educational program have an average IQ different from the national average of 100. A random sample of 25 students from this program yields a mean IQ score of 104 with a standard deviation of 12. The psychologist will test this hypothesis at the 1% significance level.

Step 1: State the Hypotheses

H_0: $\mu = 100$ (The mean IQ equals the national average)

H_1: $\mu \neq 100$ (The mean IQ differs from the national average)

This is a two-tailed test since we're interested in detecting a difference in either direction.

Step 2: Select the Significance Level

- $\alpha = 0.01$

Step 3: Calculate the Test Statistic

Notes

Since the population standard deviation is unknown and we have a small sample size (n = 25), we use the t-statistic:

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

$t = (104 - 100)/(12/\sqrt{25})$

$t = 4/2.4$

$t = 1.667$

Step 4: Determine the Critical Values

For a two-tailed test with $\alpha = 0.01$ and degrees of freedom $df = n - 1 = 24$, the critical values are ± 2.797.

The critical regions are $t < -2.797$ or $t > 2.797$.

Step 5: Make a Decision

Since our test statistic $t = 1.667$ falls between -2.797 and 2.797, it is not in the critical region. Therefore, we fail to reject the null hypothesis.

At the 1% significance level, there is insufficient evidence to conclude that the mean IQ of students in the experimental educational program differs from the national average of 100. While the sample mean (104) is numerically higher than 100, this difference is not statistically significant given our small sample size and strict significance level. The psychologist might consider using a less stringent significance level (e.g., 5%) or gathering a larger sample to detect a potentially meaningful difference.

Problem 3: Testing a Claim About a Population Proportion

A manufacturing company claims that at most 5% of their production is defective. A quality assurance manager randomly selects 200 items and finds 15 defective items. The manager wants to test the company's claim at a 5% significance level.

Step 1: State the Hypotheses

H_0: $p \leq 0.05$ (The company's claim that the defect rate is at most 5%)

H_1: $p > 0.05$ (The defect rate exceeds the company's claim)

This is an upper-tailed test since we're examining whether the true proportion exceeds the claimed maximum.

Step 2: Select the Significance Level

- $\alpha = 0.05$

Step 3: Calculate the Test Statistic

Since we're testing a proportion with a large sample size (n = 200), we use the z-statistic:

Sample proportion: $\hat{p} = 15/200 = 0.075$ (7.5%)

Claimed proportion: $p_0 = 0.05$ (5%)

$z = (\hat{p} - p_0)/\sqrt{[p_0(1-p_0)/n]}$

$z = (0.075 - 0.05)/\sqrt{[0.05(0.95)/200]}$

$z = 0.025/0.0154$

$z = 1.623$

Step 4: Determine the Critical Value

For an upper-tailed test with $\alpha = 0.05$, the critical value is 1.645.

The critical region is $z > 1.645$.

Step 5: Make a Decision

Since our test statistic z = 1.623 is less than the critical value 1.645, it does not fall in the critical region. Therefore, we fail to reject the null hypothesis.

Step 6: State the Conclusion

At the 5% significance level, there is insufficient evidence to contradict the company's claim that at most 5% of their production is defective. Although the observed defect rate (7.5%) is numerically higher than the claimed maximum (5%), this difference is not statistically significant and could be attributed to sampling variation.

However, the test statistic (1.623) is very close to the critical value (1.645), indicating that the result is borderline. The quality assurance manager might want to continue monitoring the production process or collect a larger sample to reach a more definitive conclusion.

Problem 4: Testing the Effectiveness of a New Teaching Method

Notes

An educational researcher has developed a new teaching method that she claims increases students' test scores. The traditional teaching method yields a mean score of 70 points. The researcher tests her new method on a random sample of 20 students and finds a mean score of 75 points with a standard deviation of 10 points. She wants to test if the new method is effective at the 1% significance level.

Step 1: State the Hypotheses

H_0: $\mu \leq 70$ (The new method does not increase scores)

H_1: $\mu > 70$ (The new method increases scores)

This is an upper-tailed test since we want to determine if the new method produces higher scores.

Step 2: Select the Significance Level

- $\alpha = 0.01$

Step 3: Calculate the Test Statistic

Since we have a small sample size (n = 20) and the population standard deviation is unknown, we use the t-statistic:

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

$t = (75 - 70)/(10/\sqrt{20})$

$t = 5/2.236$

$t = 2.236$

Step 4: Determine the Critical Value

For an upper-tailed test with $\alpha = 0.01$ and degrees of freedom $df = n - 1 = 19$, the critical value is 2.539.

The critical region is $t > 2.539$.

Step 5: Make a Decision

Since our test statistic $t = 2.236$ is less than the critical value 2.539, it does not fall in the critical region. Therefore, we fail to reject the null hypothesis.

Step 6: State the Conclusion

At the 1% significance level, there is insufficient evidence to support the researcher's claim that the new teaching method increases test scores. Although the sample mean (75) is higher than the traditional method's mean (70), this difference is not statistically significant at the strict 1% level. It's worth noting that the test statistic (2.236) is relatively close to the critical value (2.539), suggesting that there might be a meaningful effect. The researcher could consider:

1. Testing at a less stringent significance level (e.g., 5%)

2. Increasing the sample size to gain more statistical power

3. Evaluating whether the 5-point increase has practical significance, even if it's not statistically significant at the 1% level

Statistical Hypothesis Tests Solutions

Problem 8.5: Problem Statement: A manufacturer claims that the mean weight of their product is 500 grams. A random sample of 50 products has a mean weight of 495 grams with a standard deviation of 15 grams. Test whether the manufacturer's claim is valid at a 5% level of significance.

Solution:

1. State the hypotheses:

H_0: $\mu = 500$ (The mean weight is 500 grams)

H_1: $\mu \neq 500$ (The mean weight is not 500 grams)

Determine the significance level:

$\alpha = 0.05$

3. Calculate the test statistic:

$\bar{x} = 495$

$\mu_0 = 500$

$s = 15$

$n = 50$

$Z = (\bar{x} - \mu_0) / (s/\sqrt{n}) = (495 - 500) / (15/\sqrt{50}) = -5 / (15/7.071) = -5 / 2.121 = -2.36$

4. Find the critical values (two-tailed test):

   For $\alpha = 0.05$, $Z_{(\alpha/2)} = Z_{(0.025)} = \pm 1.96$

   Critical values are -1.96 and 1.96

5. Make a decision:

   $|Z| = |-2.36| = 2.36 > 1.96$

   Therefore, we reject the null hypothesis.

6. Interpret the results: There is sufficient evidence at the 5% significance level to conclude that the mean weight of the product is not 500 grams as claimed by the manufacturer. The sample data suggests that the actual mean weight is different from the claimed value.

Problem 8.6

Problem Statement: A medical researcher claims that a new treatment reduces the recovery time compared to the standard treatment, which has a mean recovery time of 14 days. A sample of 15 patients treated with the new method has a mean recovery time of 12.5 days with a standard deviation of 2.8 days. Test the researcher's claim at a 5% level of significance.

Solution:

1. State the hypotheses:

   H_0: $\mu = 14$ (The mean recovery time is equal to 14 days)

   H_1: $\mu < 14$ (The mean recovery time is less than 14 days) [Left-tailed test since we're testing if it "reduces" recovery time]

2. Determine the significance level:

   $\alpha = 0.05$

3. Calculate the test statistic:

   $\bar{x} = 12.5$

   $\mu_0 = 14$

   $s = 2.8$

   $n = 15$

Since n < 30, we should use the t-distribution instead of Z-distribution:

$t = (\bar{x} - \mu_0) / (s/\sqrt{n}) = (12.5 - 14) / (2.8/\sqrt{15}) = -1.5 / (2.8/3.873) = -1.5 / 0.723 = -2.07$

Degrees of freedom = n - 1 = 15 - 1 = 14

4. Find the critical value (left-tailed test):

   For $\alpha = 0.05$ with df = 14, $t_{(\alpha)} = t_{(0.05)} = -1.761$

5. Make a decision:

   $t = -2.07 < -1.761$

   Therefore, we reject the null hypothesis.

6. Interpret the results: There is sufficient evidence at the 5% significance level to support the researcher's claim that the new treatment reduces the recovery time compared to the standard treatment. The sample data suggests that the new treatment is effective in reducing recovery time.

Problem 8.7

Problem Statement: A quality control engineer wants to test whether the proportion of defective items in a production process exceeds 3%. In a random sample of 300 items, 12 are found to be defective. Conduct the appropriate hypothesis test at a 1% level of significance.

Solution:

1. State the hypotheses:

   H_0: p = 0.03 (The proportion of defective items is 3%)

   H_1: p > 0.03 (The proportion of defective items exceeds 3%) [Right-tailed test]

2. Determine the significance level:

   $\alpha = 0.01$

3. Check the conditions:

   $np_0 = 300(0.03) = 9 \geq 5$

   $n(1-p_0) = 300(0.97) = 291 \geq 5$

Conditions are satisfied for using the Z-test for proportions.

4. Calculate the test statistic:

$\hat{p} = 12/300 = 0.04$

$p_0 = 0.03$

$n = 300$

$Z = (\hat{p} - p_0) / \sqrt{[p_0(1-p_0)/n]} = (0.04 - 0.03) / \sqrt{[(0.03)(0.97)/300]} = 0.01 / \sqrt{[0.0291/300]} = 0.01 / \sqrt{0.000097} = 0.01 / 0.00985 = 1.02$

5. Find the critical value (right-tailed test):

For $\alpha = 0.01$, $Z_\alpha = Z_{0.01} = 2.33$

6. Make a decision:

$Z = 1.02 < 2.33$

Therefore, we fail to reject the null hypothesis.

7. Interpret the results: There is insufficient evidence at the 1% significance level to conclude that the proportion of defective items exceeds 3%. The observed proportion of 4% defective items is not statistically significantly higher than the 3% threshold at the 1% significance level.

Problem 8.8

Problem Statement: An educational psychologist claims that the mean score of students who receive tutoring is different from the mean score of students who do not receive tutoring, which is 65. A random sample of 30 students who received tutoring has a mean score of 68.5 with a standard deviation of 8.2. Test the psychologist's claim at a 5% level of significance.

Solution:

1. State the hypotheses:

H_0: $\mu = 65$ (The mean score of tutored students is equal to 65)

H_1: $\mu \neq 65$ (The mean score of tutored students is different from 65) [Two-tailed test]

2. Determine the significance level:

$\alpha = 0.05$

3. Calculate the test statistic:

$\bar{x} = 68.5$

$\mu_0 = 65$

$s = 8.2$

$n = 30$

Since n = 30, we can use the Z-test:

$Z = (\bar{x} - \mu_0) / (s/\sqrt{n}) = (68.5 - 65) / (8.2/\sqrt{30}) = 3.5 / (8.2/5.477) = 3.5 / 1.497 = 2.34$

4. Find the critical values (two-tailed test):

For $\alpha = 0.05$, $Z_{\alpha/2} = Z_{0.025} = \pm 1.96$

Critical values are -1.96 and 1.96

5. Make a decision:

$|Z| = |2.34| = 2.34 > 1.96$

Therefore, we reject the null hypothesis.

6. Interpret the results: There is sufficient evidence at the 5% significance level to support the psychologist's claim that the mean score of students who receive tutoring is different from the mean score of students who do not receive tutoring (65). The sample data suggests that tutoring does have an effect on student scores.

Problem 8.9: A company claims that at least 80% of its customers are satisfied with its service. In a random survey of 150 customers, 112 reported being satisfied. Test the company's claim at a 5% level of significance.

Solution:

1. State the hypotheses:

H_0: $p = 0.80$ (The proportion of satisfied customers is 80%)

H_1: $p < 0.80$ (The proportion of satisfied customers is less than 80%)
[Left-tailed test since we're testing if it's less than the claimed "at least 80%"]

2. Determine the significance level:

$\alpha = 0.05$

3. Check the conditions:

$np_0 = 150(0.80) = 120 \geq 5$

$n(1-p_0) = 150(0.20) = 30 \geq 5$

Conditions are satisfied for using the Z-test for proportions.

4. Calculate the test statistic:

$\hat{p} = 112/150 = 0.747$

$p_0 = 0.80$

$n = 150$

$Z = (\hat{p} - p_0) / \sqrt{[p_0(1-p_0)/n]} = (0.747 - 0.80) / \sqrt{[(0.80)(0.20)/150]} = -0.053 / \sqrt{[0.16/150]} = -0.053 / \sqrt{0.00107} = -0.053 / 0.0327 = -1.62$

5. Find the critical value (left-tailed test):

For $\alpha = 0.05$, $Z_{(\alpha)} = Z_{(0.05)} = -1.645$

6. Make a decision:

$Z = -1.62 > -1.645$

Therefore, we fail to reject the null hypothesis.

7. Interpret the results: There is insufficient evidence at the 5% significance level to conclude that the proportion of satisfied customers is less than 80%. The company's claim that at least 80% of its customers are satisfied with its service cannot be rejected based on the available sample data.

Problem 7.5

A company claims the mean salary is $60,000 per year, but a labor union suspects it's less. A sample of 50 employees shows a mean of $58,500 with standard deviation $5,000. Test at 1% significance level.

Solution:

1. Hypotheses:

H_0: $\mu = \$60,000$ (The mean salary is $60,000)

H_1: $\mu < \$60,000$ (The mean salary is less than $60,000)

2. Significance level: $\alpha = 0.01$

3. Test statistic:

$\bar{x} = \$58,500$

$\mu_0 = \$60,000$

$s = \$5,000$

$n = 50$

$Z = (\bar{x} - \mu_0) / (s/\sqrt{n}) = (\$58,500 - \$60,000) / (\$5,000/\sqrt{50}) = -\$1,500 / \$707.11 = -2.121$

4. Critical value (left-tailed test):

For $\alpha = 0.01$, $Z_\alpha = Z_{0.01} = -2.33$

5. Decision:

$Z = -2.121 > -2.33$

Therefore, we fail to reject the null hypothesis.

6. Interpretation: There is insufficient evidence at the 1% significance level to conclude that the mean salary of employees is less than $60,000.

Problem 7.6

The mean height of a plant species is claimed to be 25 cm. A botanist believes a new fertilizer can increase this height. After application, a sample of 40 plants has mean height 26.2 cm with standard deviation 3.8 cm. Test at 5% significance level.

Solution:

1. Hypotheses:

H_0: $\mu = 25$ cm (The fertilizer does not increase height)

H_1: $\mu > 25$ cm (The fertilizer increases height)

2. Significance level: $\alpha = 0.05$

3. Test statistic:

Notes

$\bar{x} = 26.2$ cm

$\mu_0 = 25$ cm

$s = 3.8$ cm

$n = 40$

$Z = (\bar{x} - \mu_0) / (s/\sqrt{n}) = (26.2 - 25) / (3.8/\sqrt{40}) = 1.2 / 0.601 = 1.997$

4. Critical value (right-tailed test):

For $\alpha = 0.05$, $Z_\alpha = Z_{0.05} = 1.645$

5. Decision:

$Z = 1.997 > 1.645$

Therefore, we reject the null hypothesis.

6. Interpretation: There is sufficient evidence at the 5% significance level to conclude that the fertilizer is effective in increasing the mean height of the plants.

Problem 7.7

A factory manager claims the defect rate is at most 3%. In a sample of 500 items, 20 are defective. Test at 5% significance level.

Solution:

1. Hypotheses:

H_0: $p = 0.03$ (The defect rate is 3%)

H_1: $p > 0.03$ (The defect rate is greater than 3%)

2. Significance level: $\alpha = 0.05$

3. Check conditions:

$np_0 = 500 \times 0.03 = 15 \geq 5$

$n(1-p_0) = 500 \times 0.97 = 485 \geq 5$

Conditions are satisfied.

4. Test statistic:

$\hat{p} = 20/500 = 0.04$

$p_0 = 0.03$

$n = 500$

$Z = (\hat{p} - p_0) / \sqrt{[p_0(1-p_0)/n]} = (0.04 - 0.03) / \sqrt{[0.03(0.97)/500]} = 0.01 / 0.00762 = 1.312$

5. Critical value (right-tailed test):

For $\alpha = 0.05$, $Z_\alpha = Z_{0.05} = 1.645$

6. Decision:

$Z = 1.312 < 1.645$

Therefore, we fail to reject the null hypothesis.

7. Interpretation: There is insufficient evidence at the 5% significance level to conclude that the defect rate is greater than 3%.

Problem 7.8

A phone manufacturer claims mean battery life is 15 hours. A consumer organization tests 64 phones and finds mean battery life of 14.6 hours with standard deviation 1.6 hours. Test at 5% significance level.

Solution:

1. Hypotheses:

H_0: $\mu = 15$ hours (The mean battery life is 15 hours)

H_1: $\mu \neq 15$ hours (The mean battery life is not 15 hours)

2. Significance level: $\alpha = 0.05$

3. Test statistic:

$\bar{x} = 14.6$ hours

$\mu_0 = 15$ hours

$s = 1.6$ hours

$n = 64$

$Z = (\bar{x} - \mu_0) / (s/\sqrt{n}) = (14.6 - 15) / (1.6/\sqrt{64}) = -0.4 / 0.2 = -2$

4. Critical values (two-tailed test):

For $\alpha = 0.05$, $Z_{\alpha/2} = Z_{0.025} = \pm 1.96$

Critical values are -1.96 and 1.96

5. Decision:

$|Z| = 2 > 1.96$

Therefore, we reject the null hypothesis.

6. Interpretation: There is sufficient evidence at the 5% significance level to conclude that the manufacturer's claim about the mean battery life is not valid.

Problem 7.9

A polling organization wants to determine if the proportion of voters supporting a candidate differs from 45%. In a sample of 1200 voters, 510 support the candidate. Test at 1% significance level.

Solution:

1. Hypotheses:

H_0: $p = 0.45$ (The proportion is 45%)

H_1: $p \neq 0.45$ (The proportion differs from 45%)

2. Significance level: $\alpha = 0.01$

3. Check conditions:

$np_0 = 1200 \times 0.45 = 540 \geq 5$

$n(1-p_0) = 1200 \times 0.55 = 660 \geq 5$

Conditions are satisfied.

4. Test statistic:

$\hat{p} = 510/1200 = 0.425$

$p_0 = 0.45$

$n = 1200$

$Z = (\hat{p} - p_0) / \sqrt{[p_0(1-p_0)/n]} = (0.425 - 0.45) / \sqrt{[0.45(0.55)/1200]} = -0.025 / 0.01436 = -1.741$

5. Critical values (two-tailed test):

   For $\alpha = 0.01$, $Z_{(\alpha/2)} = Z_{(0.005)} = \pm 2.576$

   Critical values are -2.576 and 2.576

6. Decision:

   $|Z| = 1.741 < 2.576$

   Therefore, we fail to reject the null hypothesis.

6. Interpretation: There is insufficient evidence at the 1% significance level to conclude that the proportion of voters supporting the candidate differs from 45%.

Unit – 5.3
Difference between two means and two proportions

5.3.1: Two Types of Comparative Studies

1. Independent Samples: Two separate groups being compared (like men vs. women)

2. Paired Samples: Same subjects measured twice (like before/after treatment)

Comparing Two Independent Means

When comparing means from two separate populations, we're testing whether the difference $\mu_1 - \mu_2$ is significant.

For equal variances:

Use pooled variance: $s^2p = [(n_1-1)s_1^2 + (n_2-1)s_2^2] / (n_1 + n_2 - 2)$

Test statistic: $t = (\bar{x}_1 - \bar{x}_2) / \sqrt{[s^2p \times (1/n_1 + 1/n_2)]}$

Degrees of freedom: $df = n_1 + n_2 - 2$

For unequal variances (Welch's t-test):

Standard error: $SE = \sqrt{(s_1^2/n_1 + s_2^2/n_2)}$

More complex df calculation using Welch-Satterthwaite approximation

Paired Samples

For before/after or naturally matched pairs:

Analyze the differences between pairs

Test statistic: $t = \bar{d} / (sd/\sqrt{n})$ where $\bar{d}$ is mean difference

Degrees of freedom: $df = n - 1$ (n = number of pairs)

Comparing Two Proportions

When comparing success rates or percentages between two groups:

Test statistic: $z = (\hat{p}_1 - \hat{p}_2) / \sqrt{[\hat{p}(1-\hat{p})(1/n_1 + 1/n_2)]}$

Where $\hat{p} = (x_1 + x_2) / (n_1 + n_2)$ is the pooled proportion

Worked Example Analysis

The first solved problem demonstrates testing the efficacy of two training methods:

Method A: n=35, $\bar{x}$=82, s=8

Method B: n=40, $\bar{x}$=78, s=7

Using pooled variance t-test (assuming equal variances)

t = 2.31 > critical value 1.99

Result: Method A produces significantly higher scores

The second example compares blood pressure medications:

Drug X: n=25, $\bar{x}$=15, s=6

Drug Y: n=30, $\bar{x}$=12, s=3

95% CI for difference: (0.32, 5.68)

Since CI doesn't include zero, Drug X is significantly more effective

Problem 1: Comparing Two Independent Means (Confidence Interval)

Region A: $n_1 = 45$, $\bar{x}_1 = 2450$, $s_1 = 320$ Region B: $n_2 = 50$, $\bar{x}_2 = 2320$, $s_2 = 280$

Step 1: Calculate the standard error. $SE = \sqrt{(s_1^2/n_1 + s_2^2/n_2)}$

$SE = \sqrt{((320^2/45) + (280^2/50))}$

$SE = \sqrt{(2275.56 + 1568)}$

$SE = \sqrt{3843.56}$

SE = 62.00

Step 2: Find the critical value for 90% confidence interval. $\alpha = 0.10$, so $\alpha/2 = 0.05$ Using Welch-Satterthwaite approximation for df:

$df = (s_1^2/n_1 + s_2^2/n_2)^2 / [(s_1^2/n_1)^2/(n_1-1) + (s_2^2/n_2)^2/(n_2-1)]$

$df = (3843.56)^2 / [(2275.56^2/44) + (1568^2/49)]$

$df \approx 89$

For 90% CI with df = 89, $t_{0.05} \approx 1.662$

Notes

Step 3: Calculate the confidence interval. CI = $(\bar{x}_1 - \bar{x}_2) \pm t(\alpha/2) \times SE$ CI = (2450 - 2320) ± 1.662 × 62.00 CI = 130 ± 103.04 CI = (26.96, 233.04)

With 90% confidence, the difference in mean daily caloric intake between Region A and Region B is between 26.96 and 233.04 calories, with students from Region A consuming more calories on average.

Problem 2: Comparing Paired Means

Let me calculate the differences between before and after scores:

Student	1	2	3	4	5	6	7	8	9	10	11	12
Before	72	68	74	77	82	79	65	63	88	76	71	84
After	78	73	77	81	85	82	68	66	91	79	75	87
Diff (d)	6	5	3	4	3	3	3	3	3	3	4	3

Step 1: Calculate mean difference. $\bar{d}$ = (6 + 5 + 3 + 4 + 3 + 3 + 3 + 3 + 3 + 3 + 4 + 3) / 12 $\bar{d}$ = 43 / 12 $\bar{d}$ = 3.58

Step 2: Calculate standard deviation of differences.

$s_d^2 = \Sigma(d - \bar{d})^2 / (n - 1)$

$s_d^2 = [(6-3.58)^2 + (5-3.58)^2 + (3-3.58)^2 + (4-3.58)^2 + (3-3.58)^2 + (3-3.58)^2 + (3-3.58)^2 + (3-3.58)^2 + (3-3.58)^2 + (3-3.58)^2 + (4-3.58)^2 + (3-3.58)^2] / 11$

$s_d^2 = [5.86 + 2.02 + 0.34 + 0.18 + 0.34 + 0.34 + 0.34 + 0.34 + 0.34 + 0.34 + 0.18 + 0.34] / 11 \Rightarrow s_d^2 = 10.96 / 11$

$\Rightarrow s_d^2 = 0.996 \Rightarrow s_d = 0.998$

Step 3: Formulate hypotheses. H_0: $\mu_d = 0$ (no improvement in test scores) H_1: $\mu_d > 0$ (there is improvement in test scores)

Step 4: Calculate the test statistic.

$t = \bar{d} / (s_d/\sqrt{n})$

$t = 3.58 / (0.998/\sqrt{12})$

$t = 3.58 / 0.288$

$t = 12.43$

Step 5: Find critical value. For $\alpha = 0.05$, one-tailed test with df = 11: $t_{0.05,11} = 1.796$

Step 6: Make decision. Since $t = 12.43 > 1.796$, we reject H_0.

There is significant evidence at $\alpha = 0.05$ that the new teaching method improves test scores. The average improvement was 3.58 points.

Problem 3: Comparing Two Proportions (Confidence Interval)

Urban: $n_1 = 500$, $x_1 = 320$, $\hat{p}_1 = 320/500 = 0.64$ Rural: $n_2 = 500$, $x_2 = 280$, $\hat{p}_2 = 280/500 = 0.56$

Step 1: Calculate the standard error.

$SE = \sqrt{[\hat{p}_1(1-\hat{p}_1)/n_1 + \hat{p}_2(1-\hat{p}_2)/n_2]}$

$SE = \sqrt{[0.64(0.36)/500 + 0.56(0.44)/500]}$

$SE = \sqrt{[0.00046 + 0.00049]}$

$SE = \sqrt{0.00095}$

$SE = 0.0308$

Step 2: Find critical value for 95% CI. For 95% confidence, $z_{0.025} = 1.96$

Step 3: Calculate the confidence interval. $CI = (\hat{p}_1 - \hat{p}_2) \pm z(\alpha/2) \times SE$ $CI = (0.64 - 0.56) \pm 1.96 \times 0.0308$ $CI = 0.08 \pm 0.0604$ $CI = (0.0196, 0.1404)$

With 95% confidence, the difference in proportion of supporters between urban and rural residents is between 0.0196 (1.96%) and 0.1404 (14.04%), with urban residents showing more support.

Problem 4: Comparing Two Proportions (Hypothesis Test)

Design A: $n_1 = 1000$, $x_1 = 85$, $\hat{p}_1 = 85/1000 = 0.085$ Design B: $n_2 = 1000$, $x_2 = 110$, $\hat{p}_2 = 110/1000 = 0.11$

Step 1: Formulate hypotheses. H_0: $p_1 = p_2$ (no difference in conversion rates) H_1: $p_2 > p_1$ (Design B has higher conversion rate)

Step 2: Check conditions.

$n_1\hat{p}_1 = 1000(0.085) = 85 \geq 5$

$n_1(1-\hat{p}_1) = 1000(0.915) = 915 \geq 5$

Notes

$n_2\hat{p}_2 = 1000(0.11) = 110 \geq 5$

$n_2(1-\hat{p}_2) = 1000(0.89) = 890 \geq 5$

Step 3: Calculate pooled proportion. $\hat{p} = (x_1 + x_2) / (n_1 + n_2)$ $\hat{p} = (85 + 110) / (1000 + 1000)$ $\hat{p} = 195 / 2000$ $\hat{p} = 0.0975$

Step 4: Calculate standard error.

$SE = \sqrt{[\hat{p}(1-\hat{p})(1/n_1 + 1/n_2)]}$

$SE = \sqrt{[0.0975(0.9025)(1/1000 + 1/1000)]}$

$SE = \sqrt{[0.0975(0.9025)(0.002)]}$

$SE = \sqrt{0.000176}$

$SE = 0.0133$

Step 5: Calculate test statistic. $z = (\hat{p}_2 - \hat{p}_1) / SE$ $z = (0.11 - 0.085) / 0.0133$ $z = 0.025 / 0.0133$ $z = 1.88$

Step 6: Find critical value. For $\alpha = 0.01$, one-tailed test: $z_{0.01} = 2.33$

Step 7: Make decision. Since $z = 1.88 < 2.33$, we fail to reject H_0.

At the 1% significance level, there is insufficient evidence to conclude that Design B has a higher conversion rate than Design A, despite an observed difference of 2.5 percentage points.

Problem 5: Comparing Two Independent Means (Unequal Variances)

Treatment X: $n_1 = 28$, $\bar{x}_1 = 42$, $s_1 = 12$ Treatment Y: $n_2 = 32$, $\bar{x}_2 = 38$, $s_2 = 8$

Step 1: Formulate hypotheses. H_0: $\mu_1 = \mu_2$ (no difference between treatments) H_1: $\mu_1 \neq \mu_2$ (there is a difference between treatments)

Step 2: Calculate standard error.

$SE = \sqrt{(s_1^2/n_1 + s_2^2/n_2)}$

$SE = \sqrt{((12^2/28) + (8^2/32))}$

$SE = \sqrt{(5.14 + 2)}$

$SE = \sqrt{7.14}$

$SE = 2.67$

Step 3: Calculate degrees of freedom using Welch-Satterthwaite approximation.

$df = (s_1^2/n_1 + s_2^2/n_2)^2 / [(s_1^2/n_1)^2/(n_1-1) + (s_2^2/n_2)^2/(n_2-1)]$

$df = (7.14)^2 / [(5.14^2/27) + (2^2/31)]$

$df = 51.0 / [0.98 + 0.13]$

$df = 51.0 / 1.11$

$df \approx 46$

Step 4: Calculate test statistic.

$t = (\bar{x}_1 - \bar{x}_2) / SE$

$t = (42 - 38) / 2.67$

$t = 4 / 2.67$

$t = 1.50$

Step 5: Find critical value. For $\alpha = 0.05$, two-tailed test with $df = 46$: $t_{0.025,46} \approx 2.01$

Step 6: Make decision. Since $|t| = 1.50 < 2.01$, we fail to reject H_0.

At the 5% significance level, there is insufficient evidence to conclude that there is a significant difference between Treatment X and Treatment Y in reducing cholesterol, despite Treatment X showing a 4 mg/dL greater average reduction.

Multiple-Choice Questions (MCQs)

1. The null hypothesis (H_0) in statistical testing is best described as:
    a. The hypothesis the researcher hopes to prove
    b. The statement about population parameters that assumes no effect or difference
    c. The statement claiming a significant difference exists
    d. A hypothesis that can never be directly proven true

    Answer: b) The statement about population parameters that assumes no effect or difference

Notes

2. A researcher conducts a hypothesis test with $\alpha = 0.01$ and calculates a p-value of 0.025. The correct conclusion is:
    a. Reject the null hypothesis
    b. Fail to reject the null hypothesis
    c. Accept the alternative hypothesis with 97.5% confidence
    d. There is a 2.5% chance the null hypothesis is true

    Answer: b) Fail to reject the null hypothesis

3. Which of the following represents a Type II error?
    a. Rejecting a true null hypothesis
    b. Failing to reject a false null hypothesis
    c. Incorrectly accepting the alternative hypothesis
    d. Correctly rejecting the null hypothesis

    Answer: b) Failing to reject a false null hypothesis

4. In hypothesis testing, the power of a statistical test is:
    a. The probability of rejecting the null hypothesis when it is true
    b. The probability of failing to reject the null hypothesis when it is false
    c. The probability of rejecting the null hypothesis when it is false
    d. Equal to the significance level (α)

    Answer: c) The probability of rejecting the null hypothesis when it is false

5. The critical region in hypothesis testing is:
    a. The collection of values where we accept the null hypothesis
    b. The range of values where the test statistic must fall to reject the null hypothesis
    c. The difference between sample and population parameters
    d. The uncertainty associated with the test statistic

    Answer: b) The range of values where the test statistic must fall to reject the null hypothesis

6. Which formula correctly represents the test statistic for testing a single population mean with known population standard deviation?
    a. $t = (\bar{x} - \mu_0)/(s/\sqrt{n})$
    b. $z = (\bar{x} - \mu_0)/(\sigma/\sqrt{n})$
    c. $z = (\hat{p} - p_0)/\sqrt{[p_0(1-p_0)/n]}$
    d. $F = s_1^2/s_2^2$

    Answer: b) $z = (\bar{x} - \mu_0)/(\sigma/\sqrt{n})$

7. For paired samples testing, the degrees of freedom for the t-test is:
    a. $n_1 + n_2 - 2$
    b. n - 1, where n is the number of pairs
    c. $(n_1 - 1) + (n_2 - 1)$
    d. The larger of $(n_1 - 1)$ or $(n_2 - 1)$

    Answer: b) n - 1, where n is the number of pairs

8. Which statement about p-values is correct?
    a. A smaller p-value indicates stronger evidence against the null hypothesis
    b. The p-value is the probability that the null hypothesis is true
    c. If $p > \alpha$, we should accept the alternative hypothesis
    d. The p-value equals the significance level in a well-designed study

    Answer: a) A smaller p-value indicates stronger evidence against the null hypothesis

9. In a two-tailed test with $\alpha = 0.05$, the null hypothesis is rejected if:
    a. The test statistic is greater than the critical value
    b. The test statistic is less than the critical value
    c. The absolute value of the test statistic is greater than the critical value
    d. The p-value is greater than 0.05

    Answer: c) The absolute value of the test statistic is greater than the critical value

10. When comparing two independent population means with unequal variances, the appropriate test is:
    a. Pooled variance t-test

b. Paired samples t-test
c. Welch's t-test (separate variance t-test)
d. One-way ANOVA

Answer: c) Welch's t-test (separate variance t-test)

SHORT QUESTIONS

1. Define statistical hypothesis testing and explain its primary purpose in research.
2. What is the difference between a null hypothesis and an alternative hypothesis? Provide an example of each.
3. Explain what a p-value represents and how it is used in making statistical decisions.
4. Describe the difference between one-tailed and two-tailed tests. When would you use each?
5. What are Type I and Type II errors in hypothesis testing? Give a practical example of each.
6. Explain the concept of significance level (α) and how it relates to the critical region.
7. What factors affect the power of a statistical test and how can researchers increase it?
8. Describe the key differences between independent samples and paired samples tests.
9. What are the assumptions that must be met for a valid t-test?
10. Explain how confidence intervals relate to hypothesis testing results.

LONG QUESTIONS

1. What are the fundamental concepts of hypothesis testing, and how does it serve as a framework for statistical inference within the scientific method?
2. How do critical regions function in hypothesis testing, and in what ways are they mathematically determined based on significance levels and test statistics such as z, t, F, and chi-square?
3. What is the relationship between Type I and Type II errors in hypothesis testing, and how can researchers balance these errors

when designing studies across fields like medicine, law, and business?

4. How do one-tailed and two-tailed testing approaches differ in terms of theoretical justification, statistical power, ethical considerations, and their impact on research conclusions?
5. What is the role of p-values in modern scientific research, and how have issues like the replication crisis and significance threshold controversies influenced their interpretation?
6. How can effect sizes complement p-values in hypothesis testing, and what methods are used to determine practical significance in different disciplines such as psychology, medicine, and economics?
7. How does statistical power influence experimental design, and what is the mathematical relationship between power, effect size, sample size, and significance level in hypothesis testing?
8. What are the major statistical methods for testing differences between means—such as independent samples tests, paired samples tests, and one-way ANOVA—and how should researchers choose among them based on assumptions and data characteristics?
9. How does the frequentist approach to hypothesis testing compare with Bayesian alternatives in terms of philosophical foundations, interpretation, and practical application?
10. What is a comprehensive framework for comparing two populations in applied research, and how should statistical methods for comparing means, proportions, variances, and distributions be selected and interpreted in real-world contexts?

References:

Module 1: Probability and Its Theorems

1. Ross, S. M. (2021). *Introduction to Probability Models* (12th ed.). Academic Press.
2. Blitzstein, J. K., & Hwang, J. (2019). *Introduction to Probability* (2nd ed.). Chapman and Hall/CRC.
3. Grinstead, C. M., & Snell, J. L. (2006). *Introduction to Probability*. American Mathematical Society.
4. Bertsekas, D. P., & Tsitsiklis, J. N. (2008). *Introduction to Probability* (2nd ed.). Athena Scientific.
5. Wasserman, L. (2013). *All of Statistics: A Concise Course in Statistical Inference*. Springer.

Module 2: Random Variables and Probability Functions

1. Casella, G., & Berger, R. L. (2021). *Statistical Inference* (2nd ed.). Cengage Learning.
2. Dekking, F. M., Kraaikamp, C., Lopuhaä, H. P., & Meester, L. E. (2005). *A Modern Introduction to Probability and Statistics: Understanding Why and How*. Springer.
3. DeGroot, M. H., & Schervish, M. J. (2014). *Probability and Statistics* (4th ed.). Pearson.
4. Hogg, R. V., McKean, J. W., & Craig, A. T. (2019). *Introduction to Mathematical Statistics* (8th ed.). Pearson.
5. Rice, J. A. (2006). *Mathematical Statistics and Data Analysis* (3rd ed.). Cengage Learning.

Module 3: Discrete and Continuous Probability Distributions

1. Evans, M., Hastings, N., & Peacock, B. (2000). *Statistical Distributions* (3rd ed.). Wiley.
2. Johnson, N. L., Kotz, S., & Balakrishnan, N. (2005). *Univariate Discrete Distributions* (3rd ed.). Wiley.
3. Forbes, C., Evans, M., Hastings, N., & Peacock, B. (2011). *Statistical Distributions* (4th ed.). Wiley.
4. Balakrishnan, N., & Nevzorov, V. B. (2003). *A Primer on Statistical Distributions*. Wiley.
5. Nadarajah, S., & Kotz, S. (2006). *R Programs for Computing Truncated Distributions*. Journal of Statistical Software, 16(2), 1-8.

Module 4: Testing of Hypothesis

1. Lehmann, E. L., & Romano, J. P. (2005). *Testing Statistical Hypotheses* (3rd ed.). Springer.
2. Schervish, M. J. (2012). *Theory of Statistics*. Springer Science & Business Media.
3. Agresti, A. (2018). *Statistical Methods for the Social Sciences* (5th ed.). Pearson.
4. Kerns, G. J. (2018). *Introduction to Probability and Statistics Using R* (3rd ed.). G. Jay Kerns.
5. Montgomery, D. C., & Runger, G. C. (2018). *Applied Statistics and Probability for Engineers* (7th ed.). Wiley.

Module 5: Grammars and Languages

1. Hopcroft, J. E., Motwani, R., & Ullman, J. D. (2006). *Introduction to Automata Theory, Languages, and Computation* (3rd ed.). Pearson.
2. Sipser, M. (2012). *Introduction to the Theory of Computation* (3rd ed.). Cengage Learning.
3. Aho, A. V., Lam, M. S., Sethi, R., & Ullman, J. D. (2006). *Compilers: Principles, Techniques, and Tools* (2nd ed.). Addison-Wesley.
4. Kozen, D. C. (2007). *Theory of Computation*. Springer.
5. Linz, P. (2017). *An Introduction to Formal Languages and Automata* (6th ed.). Jones & Bartlett Learning.

MATS UNIVERSITY

MATS CENTRE FOR DISTANCE AND ONLINE EDUCATION

UNIVERSITY CAMPUS: Aarang Kharora Highway, Aarang, Raipur, CG, 493 441

RAIPUR CAMPUS: MATS Tower, Pandri, Raipur, CG, 492 002

T : 0771 4078994, 95, 96, 98 **Toll Free ODL MODE :** 81520 79999, 81520 29999

Website: www.matsodl.com