Foundations of Data Science

Applied Statistics and Probability with Python

Lecture Note
(Draft)

Dr. Md Rezaul Karim
Professor
Department of Statistics and Data Science
Jahangirnagar University, Savar, Bangladesh

`‘Are those who know equal to those who do not know?’ Only they will remember [who are] people of understanding (Surah Al-Zumar (39:9), Al-Quran).`

Table of Contents

Chapter 4

Introduction to Probability

4.1 Introduction

In the realm of data science, understanding probability is essential for making informed decisions based on uncertain and incomplete information. Probability theory provides the mathematical foundation for analyzing data, modeling uncertainty, and deriving insights from complex datasets. As data scientists, we frequently encounter situations where outcomes are not deterministic but rather subject to variability and chance. Probability offers tools and frameworks to quantify this uncertainty and to make predictions that guide decision-making.

At its core, probability is concerned with measuring the likelihood of various outcomes in uncertain situations. Whether it's predicting customer behavior, assessing risk, or evaluating the effectiveness of an algorithm, probability helps us to model and interpret the inherent randomness in data. By applying probability theory, we can develop robust statistical models, conduct rigorous hypothesis testing, and perform meaningful data analysis.

In this chapter, we will lay the groundwork for understanding probability within the context of data science. We will start with the basic concepts that form the foundation of probability theory, including experiments, sample spaces, and events. We will then explore different methods of assigning probabilities, such as classical, empirical, and subjective approaches, and examine how these methods apply to real-world data problems.

As we delve deeper, we will cover key topics such as joint and marginal probabilities, conditional probability, and posterior probabilities. Each of these concepts is crucial for analyzing relationships between variables, updating beliefs based on new data, and making predictions about future events.

By the end of this chapter, you will gain a solid understanding of probability and its applications in data science. This knowledge will equip you with the tools needed to tackle complex data challenges and to make data-driven decisions with confidence.

4.2 Basic Concepts

4.2.1 Experiment

In the context of probability, an **experiment** refers to any process or action that generates a set of outcomes. The experiment is conducted under specified conditions, and the outcomes of interest are observed and recorded. Each outcome of an experiment is uncertain, but over repeated trials, patterns emerge that allow us to assign probabilities to different outcomes.

Example of an Experiment

To test the unbiasedness of a coin used in a cricket match to decide whether a team bats or bowls first, we would design a simple experiment. The aim is to determine if the coin has an equal probability of landing on heads or tails, indicating that it is fair.

- **Objective:** The experiment's goal is to test whether the coin is fair, meaning it has an equal likelihood of landing on heads or tails.
- **Possible Outcomes:** The set of possible outcomes when the coin is flipped consists of two events: **heads** or **tails**.
- **Procedure:**
  - Flip the coin a large number of times, such as 100 flips.
  - Record each outcome, noting whether the coin lands on heads or tails.
- **Analysis:**
  - Calculate the frequency of heads and tails from the recorded outcomes.
  - Determine the propostion of heads and tails (relative frequency) and compare it to the expected probability of 0.5.
- **Conclusion:** If the empirical probability of heads is significantly different from 0.5, this suggests that the coin may be biased. Conversely, if the probability is close to 0.5, there is no evidence to suggest that the coin is unfair.

4.2.2 Random Experiment

A **random experiment** is a specific type of experiment where the outcome is subject to chance and cannot be predicted with certainty. The outcomes are uncertain and vary each time the experiment is performed.

Characteristics of a Random Experiment

- **Uncertainty**: The outcome cannot be determined beforehand and is influenced by chance.
- **Sample Space**: The set of all possible outcomes of the experiment.
- **Reproducibility**: The experiment can be repeated under the same conditions, but the outcome remains uncertain.

Examples

- **Rolling a Die**: A random experiment with outcomes {1, 2, 3, 4, 5, 6}, where each outcome is unpredictable.
- **Flipping a Coin**: A random experiment with two possible outcomes: heads or tails.
- **Drawing a Card**: A random experiment from a deck of 52 cards, where each card is equally likely to be drawn.

4.2.3 Sample Space and Events

In data science, the concept of a sample space is essential for understanding the possible outcomes of a random experiment or a data-generating process. The sample space is the set of all possible outcomes or values that a random variable can take.

> **Sample Space:** The set of all possible outcomes of a random experiment is denoted as the sample space, typically represented by S or Ω.

Consider the experiment of rolling a fair six-sided die. The possible outcomes of this experiment are the numbers that appear on the top face of the die after a roll. Then the sample space S for this experiment is the set of all possible outcomes. That is,

$$S = \{1, 2, 3, 4, 5, 6\}$$

Here are some examples of sample spaces in different contexts within data science:

- The sample space for tossing a fair coin is

$$S = \{\text{Heads}, \text{Tails}\}.$$

- Sample Space for tossing two coins:

$$S = \{(\text{Heads, Heads}), (\text{Heads, Tails}), (\text{Tails, Heads}), (\text{Tails, Tails})\}.$$

- The sample space S for rolling two six-sided dice consists of all possible ordered pairs (x_1, x_2), where x_1 represents the outcome of the first die and x_2 represents the outcome of the second die. Since each die has 6 faces, the sample space contains 36 possible outcomes:

$$\begin{aligned} S = \big\{ &(1,1),(1,2),(1,3),(1,4),(1,5),(1,6), \\ &(2,1),(2,2),(2,3),(2,4),(2,5),(2,6), \\ &(3,1),(3,2),(3,3),(3,4),(3,5),(3,6), \\ &(4,1),(4,2),(4,3),(4,4),(4,5),(4,6), \\ &(5,1),(5,2),(5,3),(5,4),(5,5),(5,6), \\ &(6,1),(6,2),(6,3),(6,4),(6,5),(6,6) \big\} \end{aligned}$$

  This sample space shows all the possible outcomes when two dice are rolled simultaneously.

Events are subsets of the sample space, representing specific outcomes or combinations of outcomes.

> **Event:** A subset of the sample space.

For example, the event A of rolling an even number is:

$$A = \{2, 4, 6\}$$

4.3 Probability

The **probability** of an event A, denoted by $P(A)$, is a measure of the likelihood that A will occur, which is a number between 0 and 1.

> **Probability:** A numerical measure of the likelihood that a particular event will occur.

If the die is fair, each outcome is equally likely. The probability of the event $A = \{2, 4, 6\}$ (rolling an even number) is the number of favorable outcomes divided by the total number of possible outcomes in $S = \{1, 2, 3, 4, 5, 6\}$. That is,

$$P(A) = \frac{|A|}{|S|} = \frac{3}{6} = \frac{1}{2}$$

Properties of the Probability

A set of probability values for an experiment with a sample space

$$S = \{O_1, O_2, \ldots, O_n\}$$

consists of some probabilities $p_1, p_2, \ldots, p_n$ that satisfy

$$0 \leq p_1 \leq 1, \quad 0 \leq p_2 \leq 1, \quad \ldots, \quad 0 \leq p_n \leq 1$$

and

$$p_1 + p_2 + \cdots + p_n = 1.$$

The probability of outcome O_i occurring is said to be p_i, and this is written

$$P(O_i) = p_i.$$

Given a sample space S, the probability of an event A satisfies:

(i). $0 \leq P(A) \leq 1$

(ii). The sum of the probabilities of all mutually exclusive events

$$A_1, A_2, \ldots, A_n$$

that cover the sample space S must be equal to one. That is,

$$\sum_i P(A_i) = 1$$

Problem 4.1. *An experiment has five outcomes, I, II, III, IV, and V. If* $P(I) = 0.08$*,* $P(II) = 0.20$*, and* $P(III) = 0.33$*, (a) what are the possible values for the probability of outcome V? (b) If outcomes IV and V are equally likely, what are their probability values?*

Solution

(a).
An experiment has five outcomes: I, II, III, IV, and V. Given the probabilities for outcomes I, II, and III are:

$$P(\text{I}) = 0.08, \quad P(\text{II}) = 0.20, \quad P(\text{III}) = 0.33$$

We need to determine the possible values for the probability of outcome V. First, we find the sum of the given probabilities:

$$P(\text{I}) + P(\text{II}) + P(\text{III}) = 0.08 + 0.20 + 0.33 = 0.61$$

Since the sum of the probabilities of all outcomes must equal 1, the sum of the probabilities of outcomes IV and V is:

$$P(\text{IV}) + P(\text{V}) = 1 - 0.61 = 0.39$$

Thus, the possible values for the probability of outcome V, denoted as $P(\text{V})$, depend on the probability of outcome IV, denoted as $P(\text{IV})$:

$$P(\text{V}) = 0.39 - P(\text{IV})$$

Since, $0 \leq P(\text{V}) \leq 0.39$, then the possible values for the probability of outcome V are

$$0 \leq P(\text{V}) \leq 0.39$$

(b).

If outcomes IV and V are equally likely, then:

$$P(\text{IV}) = P(\text{V})$$

Let $P(\text{IV}) = P(\text{V}) = x$. Then:

$$2x = 0.39 \quad \Rightarrow \quad x = \frac{0.39}{2} = 0.195$$

Therefore, the probabilities for outcomes IV and V are:

$$P(\text{IV}) = P(\text{V}) = 0.195$$

Problem 4.2. *An experiment has three outcomes, I, II, and III. If outcome I is twice as likely as outcome II, and outcome II is three times as likely as outcome III, what are the probability values of the three outcomes?*

Solution

An experiment has three outcomes: I, II, and III. Let the probabilities of these outcomes be $P(\text{I})$, $P(\text{II})$, and $P(\text{III})$, respectively.

Let $P(\text{III}) = x$. Then:

- $P(\text{II}) = 3x$ (since outcome II is three times as likely as outcome III).
- $P(\text{I}) = 2 \cdot P(\text{II}) = 2 \cdot 3x = 6x$ (since outcome I is twice as likely as outcome II).

Since the sum of the probabilities of all outcomes must equal 1, we have:

$$P(\text{I}) + P(\text{II}) + P(\text{III}) = 1$$

Substituting the values, we get:

$$6x + 3x + x = 1$$

or,

$$10x = 1$$

$$\therefore x = \frac{1}{10} = 0.1$$

Thus, the probabilities of the three outcomes are:

$$P(\text{III}) = x = 0.1$$

$$P(\text{II}) = 3x = 3 \times 0.1 = 0.3$$

$$P(\text{I}) = 6x = 6 \times 0.1 = 0.6$$

Therefore, the probability values of the three outcomes are:

$$P(\text{I}) = 0.6, \quad P(\text{II}) = 0.3, \quad P(\text{III}) = 0.1$$

4.3.1 Union of Events

The union of two events A and B, denoted $A \cup B$, represents the event that either A, B, or both occur. Formally:

$$A \cup B = \{\omega \mid \omega \in A \text{ or } \omega \in B\}$$

Example: Consider rolling a standard six-sided die. Let:

- A be the event "rolling an even number" (i.e., $A = \{2, 4, 6)\}$)
- B be the event "rolling a number greater than 3" (i.e., $B = \{4, 5, 6)\}$)

The union $A \cup B$ represents rolling a number that is either even or greater than 3 (or both). The possible outcomes for

$$A \cup B = \{2, 4, 5, 6\}.$$

The probability of $A \cup B$ is given by:

$$P(A \cup B) = \frac{\text{Number of favorable outcomes}}{\text{Total number of outcomes}} = \frac{4}{6} = \frac{2}{3}$$

4.3.2 Intersection of Events

The intersection of two events A and B, denoted $A \cap B$, represents the event that both A and B occur simultaneously. Formally:

$$A \cap B = \{\omega \mid \omega \in A \text{ and } \omega \in B\}$$

Example: Using the same die roll, let:

- A be the event "rolling an even number" (i.e., $A = \{2, 4, 6)\}$)
- B be the event "rolling a number greater than 3" (i.e., $B = \{4, 5, 6)\}$)

The intersection $A \cap B$ represents rolling a number that is both even and greater than 3.

The possible outcomes for

$$A \cap B = \{4, 6\}.$$

The probability of $A \cap B$ is given by:

$$P(A \cap B) = \frac{\text{Number of favorable outcomes}}{\text{Total number of outcomes}} = \frac{2}{6} = \frac{1}{3}$$

4.3.3 Complementary Event

A **complementary event** of an event A, denoted by A^c or $\overline{A}$, consists of all outcomes in the sample space S that are not in A. The probability of the complementary event A^c is given by:

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

Example: If A is the event of getting a head when flipping a coin, then the complementary event A^c is the event of getting a tail. If $P(A) = 0.5$, then $P(A^c) = 1 - 0.5 = 0.5$.

Odds: The *odds* in favor of an event A are defined as the ratio of the probability that the event A occurs to the probability that the event A does not occur (i.e., the complement of A). Mathematically, the odds in favor of A are given by:

$$\text{Odds in favor of } A = \frac{P(A)}{P(A^c)}$$

where $P(A^c)$ is the probability of the complement of A.

Problem 4.3. *Suppose p is the probability of the success.*

(a) If the odds is 1, what is p?

(b) If the odds is 2, what is p?

(c) If $p = 0.25$, what is the odds?

Solution

(a). The odds in favor of success are given by:

$$\text{Odds} = \frac{p}{1-p}$$

If the odds are 1, then:

$$1 = \frac{p}{1-p}$$

Solving for p:

$$1 - p = p \quad \Rightarrow \quad 1 = 2p \quad \Rightarrow \quad p = \frac{1}{2}$$

So, $p = 0.5$.

(b). If the odds are 2, then:

$$2 = \frac{p}{1-p}$$

Solving for p:

$$2(1-p) = p \quad \Rightarrow \quad 2 - 2p = p \quad \Rightarrow \quad 2 = 3p \quad \Rightarrow \quad p = \frac{2}{3}$$

So, $p = \frac{2}{3}$.

(c). If $p = 0.25$, then the odds are:

$$\text{Odds} = \frac{p}{1-p} = \frac{0.25}{1-0.25} = \frac{0.25}{0.75} = \frac{1}{3}$$

So, the odds are $\frac{1}{3}$.

4.3.4 Equally Likely Events

In probability theory, equally likely events are events that have the same probability of occurring. If all outcomes in the sample space are equally likely, then the probability of any specific event can be calculated by dividing the number of favorable outcomes by the total number of possible outcomes.

> **Equally Likely Events**: Events that have the same probability of occurring.

Example

Consider the experiment of rolling a fair six-sided die. The sample space is:

$$S = \{1, 2, 3, 4, 5, 6\}$$

Since the die is fair, each of the six outcomes is equally likely. The probability of each outcome is:

$$P(\{1\}) = P(\{2\}) = P(\{3\}) = P(\{4\}) = P(\{5\}) = P(\{6\}) = \frac{1}{6}.$$

4.3.5 Mutually Exclusive Events

In probability theory, mutually exclusive events are events that cannot happen at the same time. In other words, if one event occurs, the other cannot occur at the same time.

> **Mutually Exclusive Events:** Two events A and B are said to be mutually exclusive (or disjoint) if they cannot occur at the same time. Formally, A and B are mutually exclusive if:
>
> $$A \cap B = \emptyset$$
>
> where $\cap$ denotes the intersection of events, and $\emptyset$ represents the empty set, indicating that there are no outcomes common to both A and B.

Example:

Consider rolling a standard six-sided die. Let:

- A be the event "rolling a 2"
- B be the event "rolling a 5"

The events A and B are mutually exclusive because you cannot roll a 2 and a 5 at the same time.

Additivity for Mutually Exclusive Events

If A and B are mutually exclusive events, then the probability of their union is the sum of their individual probabilities:

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

Generalization

For any finite or countable collection of mutually exclusive events $A_1, A_2, \ldots, A_n$:

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

4.3.6 Probability Axioms

The probability of an event A_i; $i = 1, 2, \ldots, n$, denoted as $P(A_i)$, satisfies the following axioms:

(i). $0 \leq P(A_i) \leq 1$,

(ii). $\sum\limits_{i=1}^{n} P(A_i) = 1$,

(iii). For any sequence of mutually exclusive events $\{A_i\}$, we have

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

For mutually exclusive events $A, B \in S$, we have

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

This is an ***addition rule for two mutually exclusive events***

If A and B are not mutually exclusive, then the addition rule is

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B).$$

If two events are mutually exclusive, then the probability of both occurring is denoted as $P(A \cap B)$ and

$$P(A \text{ and } B) = P(A \cap B) = 0.$$

Problem 4.4. *A single 6-sided die is rolled. What is the probability of rolling a 2 or a 5?*

Solution

- $\Pr(2) = \frac{1}{6}$ and $\Pr(5) = \frac{1}{6}$
- Therefore,

$$\begin{aligned} \Pr(2 \text{ or } 5) &= \Pr(2 \cup 5) = \Pr(2) + \Pr(5) \\ &= \frac{1}{6} + \frac{1}{6} \\ &= \frac{2}{6} \\ &= \frac{1}{3} \end{aligned}$$

Problem 4.5. *In a Math class of 30 students, 17 are boys and 13 are girls. On a unit test, 4 boys and 5 girls made an A grade. If a student is chosen at random from the class, what is the probability of choosing a girl or an A-grade student?*

Solution

- $\Pr(\text{girl}) = \frac{13}{30}$, $\Pr(A\text{-grade student}) = \frac{9}{30}$ and $\Pr(\text{girl} \cap A\text{-grade student}) = \frac{5}{30}$
- Therefore,

$$\begin{aligned}\Pr(\text{girl or } A\text{-grade student}) &= \Pr(\text{girl}) + \Pr(A\text{-grade student}) \\ &\quad - \Pr(\text{girl} \cap A\text{-grade student}) \\ &= \frac{13}{30} + \frac{9}{30} - \frac{5}{30} \\ &= \frac{17}{30}\end{aligned}$$

4.4 Approaches to Assigning Probabilities

In probability theory, there are several methods for assigning probabilities to events. These approaches include:

1. **Classical Approach**
2. **Empirical Approach**
3. **Subjective Approach**

4.4.1 Classical Approach

The classical approach to assigning probabilities is based on the assumption that all outcomes in the sample space are **equally likely**. If an experiment has n equally likely outcomes and an event A consists of m of these outcomes, the probability of event A is given by:

$$P(A) = \frac{\text{Number of favorable outcomes}}{\text{Total number of possible outcomes}}.$$

Problem 4.6. *Consider a fair six-sided die. Each side of the die is equally likely to land face up. We want to calculate the probability of rolling an even number.*

Solution

- **Sample Space**: The sample space S is:

$$S = \{1, 2, 3, 4, 5, 6\}$$

- **Favorable Outcomes**: The favorable outcomes for rolling an even number are:

$$\text{Even numbers} = \{2, 4, 6\}$$

- **Number of Outcomes**:
  - Total number of possible outcomes: $|S| = 6$
  - Number of favorable outcomes: Number of even numbers $= 3$

- **Probability Calculation**: The probability P of rolling an even number is:

$$P(\text{Even number}) = \frac{\text{Number of favorable outcomes}}{\text{Total number of possible outcomes}} = \frac{3}{6} = \frac{1}{2}$$

So, the probability of rolling an even number on a fair six-sided die is $\frac{1}{2}$, or 50%.

4.4.2 Empirical Approach

The empirical approach, also known as the frequentist approach, assigns probabilities based on observed frequencies from experimental data. This approach is valuable for estimating probabilities from data. If an event A occurs k times in N trials, the probability of event A is estimated by:

$$P(A) = \frac{\text{Number of times the event } A \text{ occurred}}{\text{Total number of observations}} = \frac{m}{n}.$$

The empirical approach to probability relies on the principle known as the **law of large numbers**. This principle suggests that as the number of observations increases, the estimate of the probability becomes more accurate. Therefore, by gathering more data, one can obtain a more precise estimation of the probability.

Law of large numbers: As the number of trials or observations increases, the empirical probability of an event will get closer to its actual probability.

Example: To estimate the probability of getting heads when flipping a coin, follow these steps:

1. **Conduct the Experiment**: Flip the coin a certain number of times and record the outcomes.

2. **Collect Data**: Suppose you flip the coin 100 times.
3. **Count Occurrences**: Out of 100 flips, you get heads 52 times.
4. **Calculate Probability**: The empirical probability $P(\text{Heads})$ is calculated as:

$$P(\text{Heads}) = \frac{\text{Number of heads}}{\text{Total number of flips}} = \frac{52}{100} = 0.52$$

Based on the empirical data, the estimated probability of getting heads on a coin flip is 0.52, or 52%.

4.4.3 Subjective Approach

The subjective approach assigns probabilities based on **personal judgment** or **belief about the likelihood of an event**. This method does not rely on mathematical calculations or empirical data but rather on an individual's intuition or experience. For an event A, the subjective probability is denoted as:

$$P(A) = \text{Subjective belief about } A$$

In this approach, probabilities are not necessarily based on frequency or equal likelihood but on personal estimation.

Example 1: Consider an entrepreneur deciding whether to launch a new product. Since there is no historical data or empirical studies available for this specific product, the entrepreneur uses their expertise and market knowledge to estimate the likelihood of success.

The entrepreneur assesses various factors:

- Knowledge of current market trends.
- Feedback from potential customers.
- Expertise of the team involved.
- Analysis of the competitive landscape.

Based on these factors, the entrepreneur might estimate the probability of the product's success to be 70%. This subjective estimate is derived from their personal judgment and experience rather than from data analysis.

$$P(\text{Success}) = 0.70$$

This subjective probability is based on the entrepreneur uses their expertise and market knowledge, rather than on empirical data or mathematical models.

Example 2: Consider a football game between Team A and Team B. Based on your personal judgment and knowledge of the teams, you estimate the probability of Team A winning the game.

Let's denote:

- $P(A)$ as the probability of Team A winning.
- $P(B)$ as the probability of Team B winning.

Based on your assessment, you estimate that:

$$P(A) = 0.70$$

This means you believe there is a 70% chance that Team A will win the game. This probability is derived from your subjective evaluation of the teams' recent performances, player conditions, and other relevant factors.

4.5 Joint and Marginal Probabilities

Joint probability is the probability of two (or more) events happening simultaneously. For two events A and B, the joint probability is denoted by $P(A \cap B)$.

Marginal probability is the probability of the occurrence of the single event. It is obtained by summing (or integrating) the joint probabilities over all possible values of the other variable(s).

Example: Consider a study on student performance with the following events:

- A: The event that a student studied for the exam.
- B: The event that a student passed the exam.

The probability table for these events is as follows:

	Studied (A)	Not Studied (A^c)	Total
Passed (B)	0.4	0.1	0.5
Not Passed (B^c)	0.2	0.3	0.5
Total	0.6	0.4	1.0

The **joint probability** is the probability that a student **studied** and **passed** the exam:

$$P(A \cap B) = 0.4$$

Marginal Probability of Studying (A) that a student studied for the exam, regardless of whether they passed or not, is obtained by summing the joint probabilities involving A:

$$P(A) = P(A \cap B) + P(A \cap B^c) = 0.4 + 0.2 = 0.6$$

Marginal Probability of Passing (B): The probability that a student passed the exam, regardless of whether they studied or not, is obtained by summing the joint probabilities involving B:

$$P(B) = P(A \cap B) + P(A^c \cap B) = 0.4 + 0.1 = 0.5$$

Problem 4.7. *Suppose two dice are thrown together. What is the probability that at least one 6 is obtained on the two dice?*

Solution

The total number of outcomes when two dice are thrown is $6 \times 6 = 36$. The number of outcomes with at least one 6 is 11. That is,

$$(6,1),(6,2),(6,3),(6,4),(6,5),(6,6),(1,6),(2,6),(3,6),(4,6),(5,6)$$

.

$$P(\text{at least one 6}) = \frac{11}{36}$$

4.6 Conditional Probability

Conditional probability is essential in data science because it helps model and understand how the probability of an event changes based on the occurrence of another event. It underpins Bayesian inference, supports feature engineering, enhances risk assessment, informs decision-making, aids in anomaly detection, and is pivotal in natural language processing tasks. This ability to adjust probabilities with new information is crucial for accurate predictions and data-driven insights.

Conditional Probability: The conditional probability of an event A given that event B has occurred is denoted by $P(A|B)$ and is defined as:

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

One simple example of conditional probability concerns the situation in which two events A and B are mutually exclusive. Since mutually exclusive events have no common outcomes, the occurrence of event B makes the occurrence of event A impossible. Thus, intuitively, the probability of event A given that B has occurred should be zero. This is confirmed by the formula:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{0}{P(B)} = 0.$$

Another example involves a scenario where event B is a subset of event A, denoted $B \subseteq A$. In this case, if event B occurs, event A must also occur. Thus, the probability of event A given that B has occurred should be one. This is supported by the formula:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1.$$

Problem 4.8. *If somebody rolls a fair die without showing you but announces that the result is even, then what is the probability of scoring a 6?*

Solution

The sample space for a fair die roll is $\{1, 2, 3, 4, 5, 6\}$. The event that the result is even is $\{2, 4, 6\}$.

$$P(\text{Even}) = \frac{3}{6} = \frac{1}{2}$$

The event of scoring a 6 given that the result is even is $\{6\}$.

$$P(6|\text{Even}) = \frac{P(6 \cap \text{Even})}{P(\text{Even})} = \frac{P(6)}{P(\text{Even})} = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3}$$

Problem 4.9. *Suppose that somebody rolls the two dice without showing you but announces that at least one 6 has been scored. Suppose first one is a red die and the second one is a blue die. What is the probability that the red die scored a 6?*

Solution

The total number of outcomes with at least one 6 is 11 as mentioned before. The number of outcomes where the red die scores a 6 is 6. That is,

$$(6, 1), (6, 2), (6, 3), (6, 4), (6, 5), (6, 6)$$

.

$$P(\text{Red die 6}|\text{at least one 6}) = \frac{6}{11}$$

Problem 4.10. *Suppose two dice are thrown together. The first one is a red die and the second one is a blue die. What is the probability that the red die scores a 6 given that exactly one 6 of the two outcomes has been scored?*

Solution

The total number of outcomes when exactly one 6 is scored is 10 (i.e., excluding $(6, 6)$ from the previous list). The number of outcomes where the red die scores a 6 and the blue die does not is 5 (i.e., $(6, 1), (6, 2), (6, 3), (6, 4), (6, 5)$).

$$P(\text{Red die 6}|\text{exactly one 6}) = \frac{5}{10} = \frac{1}{2}$$

4.6.1 Probabilities Computation form Contingency Table

A **contingency table** displays the frequency distribution of two categorical variables. Each cell in the table represents the count of observations where the two variables take specific values. We use contingency tables to compute marginal and joint probabilities.

Consider two categorical variables: **Variable A** with categories A_1 and A_2 and **Variable B** with categories B_1 and B_1 The contingency table 4.1 is structured as follows:

Table 4.1: Contingency Table

	A_1	A_2	**Total**
B_1	a	b	$a+b$
B_2	c	d	$c+d$
Total	$a+c$	$b+d$	$n=a+b+c+d$

The **joint probability table** shows the probability of each combination of categories occurring. It is obtained by dividing each cell count in the contingency table by the overall total number of observations n.

To compute the joint probabilities:

$$\text{Joint Probability} = \frac{\text{Count in cell}}{n}$$

The joint probability table is:

	A_1	A_2	**Total**
B_1	$\frac{a}{n}$	$\frac{b}{n}$	$\frac{a+b}{n}$
B_2	$\frac{c}{n}$	$\frac{d}{n}$	$\frac{c+d}{n}$
Total	$\frac{a+c}{n}$	$\frac{b+d}{n}$	1

Table 4.2: Joint Probability Table

Problem 4.11. *Consider the situation of the promotion status of male and female officers of a major metropolitan police force in the eastern United States. The force consists of 1200 officers, 960 men and 240 women. Over the past two years 324 officers on the public force received promotions. After reviewing the promotion record, a committee of female officers raised a discrimination case on the basis that 288 male officers had received promotions, but only 36 female officers had received promotions.*

	Men	*Women*	*Total*
Promoted	288	36	324
Not Promoted	672	204	876
Total	960	240	1200

(i). *Develop a joint probability table for these data. What are the marginal probabilities? Suppose a male officer is selected randomly, what is the chance that the officer will be promoted?*

(ii). *Suppose a female officer is selected randomly, what is the chance that the officer will not be promoted? Suppose an officer is selected randomly who got promotion, what is the chance that the officer will be male?*

(iii). *Suppose an officer is selected randomly who did not get promotion, what is the chance that the officer will be female?*

Solution

(i) Joint Probability Table and Marginal Probabilities

To develop the joint probability table, we divide each cell count by the total number of officers, which is 1200.

Joint Probability Table :

	Men	Women	Total
Promoted	$\frac{288}{1200}$	$\frac{36}{1200}$	$\frac{324}{1200}$
Not Promoted	$\frac{672}{1200}$	$\frac{204}{1200}$	$\frac{876}{1200}$
Total	$\frac{960}{1200}$	$\frac{240}{1200}$	1

Simplifying the fractions, we get:

	Men	Women	Total
Promoted	0.24	0.03	0.27
Not Promoted	0.56	0.17	0.73
Total	0.80	0.20	1

Marginal Probabilities:

- Probability of promotion: $\frac{324}{1200} = 0.27$
- Probability of not promotion: $\frac{876}{1200} = 0.73$

- Probability of being male: $\frac{960}{1200} = 0.80$
- Probability of being female: $\frac{240}{1200} = 0.20$

Probability that a randomly selected male officer is promoted:

$$P(\text{Promoted} \mid \text{Male}) = \frac{288}{960} = 0.30$$

(ii) Probabilities for Female Officers and Promotion

Probability that a randomly selected female officer is not promoted:

$$P(\text{Not Promoted} \mid \text{Female}) = \frac{204/1200}{240/1200} = 0.85$$

Probability that a randomly selected officer who got promoted is male:

$$P(\text{Male} \mid \text{Promoted}) = \frac{288/1200}{324/1200} \approx 0.889$$

(iii) Probability for Officers Not Promoted

Probability that a randomly selected officer who did not get promoted is female:

$$P(\text{Female} \mid \text{Not Promoted}) = \frac{204/1200}{876/1200} \approx 0.233$$

4.6.2 Independent Events

Two events A and B are said to be **independent** if the occurrence of one event does not affect the probability of the other event occurring. In mathematical terms, this is expressed as:

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

For independent events, the following also holds true:

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

and

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

These equations indicate that knowing the occurrence of one event does not change the probability of the other event.

Independent Events: Two events A and B are said to be **independent** if:

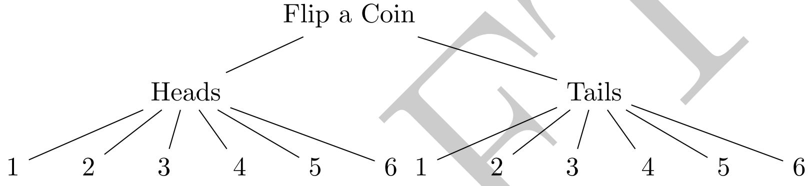
$$P(A \mid B) = P(A), \qquad P(B \mid A) = P(B), \qquad P(A \cap B) = P(A) \cdot P(B).$$

Any one of these three conditions implies the other two.

Example 1: Rolling a Die and Flipping a Coin

Consider rolling a fair six-sided die and flipping a fair coin. Let:

- A be the event that the die shows a 3.
- B be the event that the coin lands on heads.



The outcome of rolling the die does not affect the outcome of flipping the coin, and vice versa. Therefore, events A and B are independent. We can verify this as follows:

$$P(A) = \frac{1}{6}, \quad P(B) = \frac{1}{2}$$

$$P(A \cap B) = P(\text{die shows 3 and coin lands on heads}) = \frac{1}{12}$$

$$P(A \cap B) = P(A) \cdot P(B) = \frac{1}{6} \cdot \frac{1}{2} = \frac{1}{12}$$

Thus, the events are independent.

Example 2: Drawing Cards from a Deck Without Replacement

Consider drawing two cards from a standard deck of 52 cards without replacement. Let:

- A be the event that the first card drawn is a heart.
- B be the event that the second card drawn is a heart.

In this case, the events are not independent because the outcome of the first draw affects the probability of the second draw. If the first card is a heart, there are now only 12 hearts left in a deck of 51 cards, so:

$$P(A) = \frac{13}{52}, \quad P(B \mid A) = \frac{12}{51}$$

The probability of B if A occurs is different from $P(B)$ without conditioning on A, which is:

$$P(B) = \frac{13}{52}$$

Thus, A and B are not independent.

Problem 4.12. *A system has four computers. Computer 1 works with a probability of 0.88; computer 2 works with a probability of 0.78; computer 3 works with a probability of 0.92; computer 4 works with a probability of 0.85. Suppose that the operations of the computers are independent of each other.*

(a). Suppose that the system works only when all four computers are working. What is the probability that the system works?

(b). Suppose that the system works only if at least one computer is working. What is the probability that the system works?

(c). Suppose that the system works only if at least three computers are working. What is the probability that the system works?

Solution

(a). To find the probability in this scenario, we multiply the probabilities of all four computers working, as they are independent.

$$P(\text{system works}) = 0.88 \times 0.78 \times 0.92 \times 0.85 = 0.537$$

(b). To find the probability in this scenario, we find the complement of the probability that none of the computers are working. Then, the probability that at least one computer is working is the complement of the probability that none of the computers are working.

$$\begin{aligned} P(\text{system works}) &= 1 - P(\text{no computers working}) \\ &= 1 - ((1 - 0.88) \times (1 - 0.78) \times (1 - 0.92) \times (1 - 0.85)) \\ &= 0.9997 \end{aligned}$$

(c).

$$
\begin{aligned}
P(\text{system works}) = P&(\text{all computers working}) \\
&+ P(\text{computers 1,2,3 working, computer 4 not working}) \\
&+ P(\text{computers 1,2,4 working, computer 3 not working}) \\
&+ P(\text{computers 1,3,4 working, computer 2 not working}) \\
&+ P(\text{computers 2,3,4 working, computer 1 not working}) \\
&= 0.537 + (0.88 \times 0.78 \times 0.92 \times (1 - 0.85)) \\
&+ (0.88 \times 0.78 \times (1 - 0.92) \times 0.85) \\
&+ (0.88 \times (1 - 0.78) \times 0.92 \times 0.85) \\
&+ ((1 - 0.88) \times 0.78 \times 0.92 \times 0.85) \\
&= 0.903
\end{aligned}
$$

$\square$

Problem 4.13. *Suppose that somebody secretly rolls two fair six-sided dice, and what is the probability that the face-up value of the first one is 2, given the information that their sum is no greater than 5?*

Solution

To find the probability that the face-up value of the first die is 2 given that the sum of the two dice is no greater than 5, we use the concept of conditional probability.

Let A be the event that the face-up value of the first die is 2, and B be the event that the sum of the two dice is no greater than 5. We want to find $P(A \mid B)$.

The conditional probability $P(A \mid B)$ is given by:

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

First, we determine $P(B)$. The possible outcomes for the sum of the two dice being no greater than 5 are:

$$
\begin{aligned}
&(1,1),(1,2),(1,3),(1,4), \\
&(2,1),(2,2),(2,3), \\
&(3,1),(3,2), \\
&(4,1)
\end{aligned}
$$

There are 10 such outcomes, and since there are 36 possible outcomes when rolling two dice, the probability $P(B)$ is:

$$P(B) = \frac{10}{36} = \frac{5}{18}$$

Next, we determine $P(A \cap B)$, which is the probability that the first die is 2 and the sum of the dice is no greater than 5. The possible outcomes for this are:

$$(2, 1), (2, 2), (2, 3)$$

There are 3 such outcomes, so the probability $P(A \cap B)$ is:

$$P(A \cap B) = \frac{3}{36} = \frac{1}{12}$$

Now we can calculate $P(A \mid B)$:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{1}{12}}{\frac{5}{18}} = \frac{1}{12} \times \frac{18}{5} = \frac{18}{60} = \frac{3}{10}$$

Thus, the probability that the face-up value of the first die is 2 given that their sum is no greater than 5 is:

$$\boxed{\frac{3}{10}}$$

□

4.7 Posterior Probabilities

4.7.1 Law of Total Probability

Consider a sample space S partitioned into mutually exclusive events $A_1, A_2, \ldots, A_n$. This means:

$$S = A_1 \cup A_2 \cup \cdots \cup A_n$$

Let B be another event in the sample space given in Figure 4.1. The initial question of interest is how to use the probabilities $P(A_i)$ and $P(B \mid A_i)$ to calculate $P(B)$, the probability of the event B. This can be achieved by noting that

$$B = (A_1 \cap B) \cup (A_2 \cap B) \cup \cdots \cup (A_n \cap B)$$

where the events $A_i \cap B$ are mutually exclusive, so that

$$P(B) = P(A_1 \cap B) + P(A_2 \cap B) + \cdots + P(A_n \cap B)$$

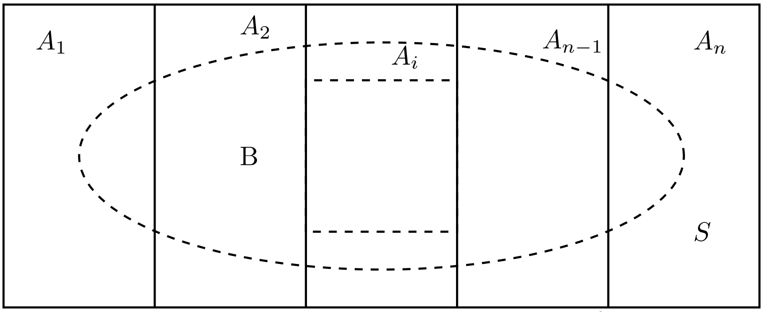


Figure 4.1: A partition $A_1, \ldots, A_n$ and an event B.

Using the definition of conditional probability, this becomes

$$P(B) = P(A_1)P(B \mid A_1) + P(A_2)P(B \mid A_2) + \cdots + P(A_n)P(B \mid A_n)$$

This result, known as the **Law of Total Probability**, has the interpretation that if it is known that one and only one of a series of events A_i can occur, then the probability of another event B can be obtained as the weighted average of the conditional probabilities $P(B \mid A_i)$, with weights equal to the probabilities $P(A_i)$.

Law of Total Probability: If $A_1, \ldots, A_n$ is a partition of a sample space, then the probability of an event B can be obtained from the probabilities $P(A_i)$ and $P(B \mid A_i)$ using the formula

$$P(B) = P(A_1)P(B \mid A_1) + P(A_2)P(B \mid A_2) + \cdots + P(A_n)P(B \mid A_n)$$

The law of total probability states that if you have a partition of the sample space into mutually exclusive events, the probability of an event can be found by summing the probabilities of the event occurring within each partition, weighted by the probability of each partition.

Example

Suppose we have a sample space divided into three mutually exclusive events A_1, A_2, and A_3 with the following probabilities and conditional probabilities:

$$P(A_1) = 0.2, \quad P(A_2) = 0.5, \quad P(A_3) = 0.3$$

$$P(B \mid A_1) = 0.4, \quad P(B \mid A_2) = 0.6, \quad P(B \mid A_3) = 0.3$$

To find $P(B)$, use the Law of Total Probability:

$$P(B) = P(A_1) \cdot P(B \mid A_1) + P(A_2) \cdot P(B \mid A_2) + P(A_3) \cdot P(B \mid A_3)$$

$$P(B) = 0.2 \cdot 0.4 + 0.5 \cdot 0.6 + 0.3 \cdot 0.3$$

$$P(B) = 0.08 + 0.30 + 0.09 = 0.47$$

Problem 4.14. *A company sells a certain type of car that it assembles in one of four possible locations. The probabilities of a car being assembled at each plant are as follows:*

- *Plant I: 20% ($P(\textit{Plant I}) = 0.20$)*
- *Plant II: 24% ($P(\textit{Plant II}) = 0.24$)*
- *Plant III: 25% ($P(\textit{Plant III}) = 0.25$)*
- *Plant IV: 31% ($P(\textit{Plant IV}) = 0.31$)*

Each new car sold carries a one-year bumper-to-bumper warranty. The company has collected data showing the following conditional probabilities of making a warranty claim:

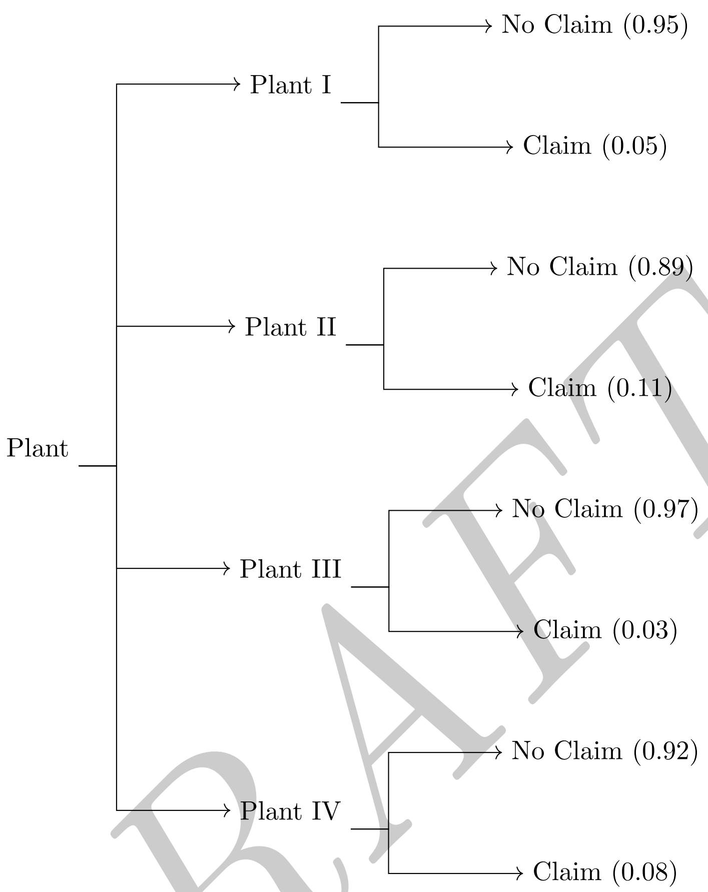
- $P(\textit{claim} \mid \textit{Plant I}) = 0.05$
- $P(\textit{claim} \mid \textit{Plant II}) = 0.11$
- $P(\textit{claim} \mid \textit{Plant III}) = 0.03$
- $P(\textit{claim} \mid \textit{Plant IV}) = 0.08$

The probability of interest is the probability that a claim on the warranty of the car will be required. If B is the event that a claim is made, we want to find $P(B)$.

Solution

We can use the Law of Total Probability to find $P(B)$. According to the Law of Total Probability:

$$\begin{aligned} P(B) =& P(B \mid \text{Plant I}) \cdot P(\text{Plant I}) + P(B \mid \text{Plant II}) \cdot P(\text{Plant II}) \\ &+ P(B \mid \text{Plant III}) \cdot P(\text{Plant III}) + P(B \mid \text{Plant IV}) \cdot P(\text{Plant IV}) \end{aligned}$$



Substitute the given values:

$$
\begin{aligned}
P(B) &= (0.05 \cdot 0.20) + (0.11 \cdot 0.24) + (0.03 \cdot 0.25) + (0.08 \cdot 0.31) \\
&= 0.01 + 0.0264 + 0.0075 + 0.0248 \\
&= 0.0687
\end{aligned}
$$

Thus, the probability that a claim on the warranty will be required is 0.0687.

Problem 4.15. *Consider a clinical study investigating the likelihood of developing high blood pressure (BP). Participants are classified by age and smoking status:*

Age Group Probabilities:

- *Young* (<50 *years*): $P(\textit{Young}) = 0.60$
- *Old* (≥ 50 *years*): $P(\textit{Old}) = 0.40$

Conditional Probabilities of Having High BP:

- *For the Young age group:*
  - *Smokers:* $P(\textit{High BP} \mid \textit{Young, Smoker}) = 0.10$
  - *Non-smokers:* $P(\textit{High BP} \mid \textit{Young, Non-smoker}) = 0.05$
- *For the Old age group:*
  - *Smokers:* $P(\textit{High BP} \mid \textit{Old, Smoker}) = 0.40$
  - *Non-smokers:* $P(\textit{High BP} \mid \textit{Old, Non-smoker}) = 0.25$

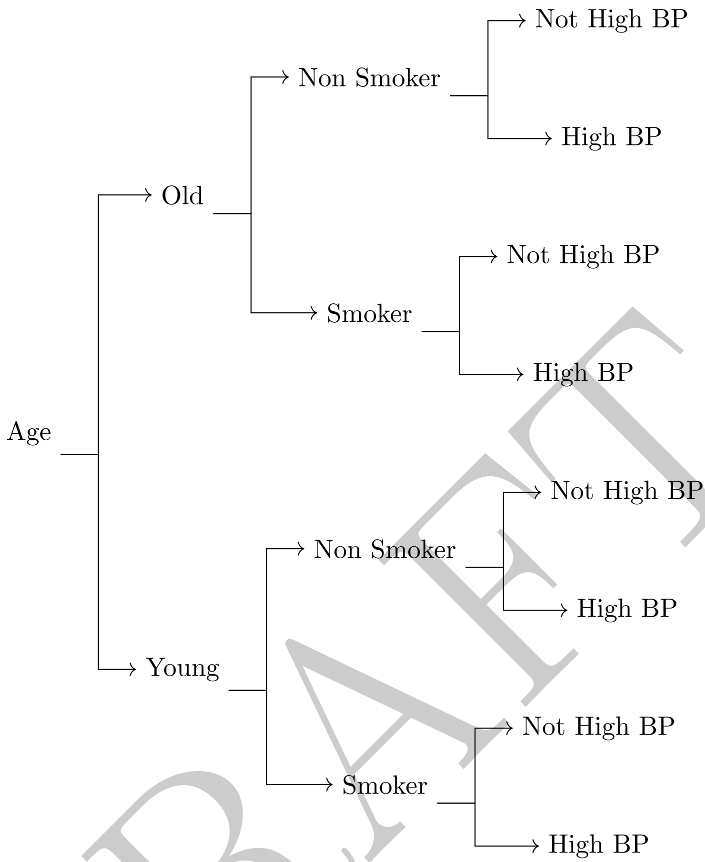
Smoking Probabilities within Each Age Group:

- *Young:*
  - *Smokers:* $P(\textit{Smoker} \mid \textit{Young}) = 0.30$
  - *Non-smokers:* $P(\textit{Non-smoker} \mid \textit{Young}) = 0.70$
- *Old:*
  - *Smokers:* $P(\textit{Smoker} \mid \textit{Old}) = 0.40$
  - *Non-smokers:* $P(\textit{Non-smoker} \mid \textit{Old}) = 0.60$

We want to find the overall probability of having high blood pressure, $P(\textit{High BP})$*.*

Solution

To find the overall probability of having high blood pressure $P(\text{High BP})$, use the law of total probability:



$$
\begin{aligned}
P(\text{High BP}) = {} & P(\text{High BP} \mid \text{Young, Smoker}) \cdot P(\text{Smoker} \mid \text{Young}) \cdot P(\text{Young}) \\
& + P(\text{High BP} \mid \text{Young, Non-smoker}) \cdot P(\text{Non-smoker} \mid \text{Young}) \cdot P(\text{Young}) \\
& + P(\text{High BP} \mid \text{Old, Smoker}) \cdot P(\text{Smoker} \mid \text{Old}) \cdot P(\text{Old}) \\
& + P(\text{High BP} \mid \text{Old, Non-smoker}) \cdot P(\text{Non-smoker} \mid \text{Old}) \cdot P(\text{Old}) \\
& = (0.10 \cdot 0.30 \cdot 0.60) + (0.05 \cdot 0.70 \cdot 0.60) + (0.40 \cdot 0.40 \cdot 0.40) \\
& \quad + (0.25 \cdot 0.60 \cdot 0.40) \\
& = 0.018 + 0.021 + 0.064 + 0.060 \\
& = 0.163
\end{aligned}
$$

Thus, the probability of developing high blood pressure is $P(\text{High BP}) = 0.163$ or 16.3%.

Interpretation: The overall probability of having high blood pressure, considering all age groups and smoking statuses, is 16.3%. This result helps understand the prevalence of high blood pressure in the study population, taking into account various risk factors.

□

4.7.2 Bayes' Theorem

Bayes' Theorem relates the conditional and marginal probabilities of random events. It is used to update the probability of a hypothesis based on observed evidence. The theorem can be stated mathematically as follows:

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

where:

- $P(A \mid B)$ is the posterior probability of event A given that event B has occurred.
- $P(B \mid A)$ is the likelihood of event B given that event A has occurred.
- $P(A)$ is the prior probability of event A before observing event B.

If $A_1, A_2, \ldots, A_n$ is a partition of a sample space, then the marginal probability $P(B)$ is

$$P(B) = \sum_i P(B \mid A_i) \cdot P(A_i).$$

In this case, the conditional probabilities $P(B \mid A_i)$ is

$$P(A_i \mid B) = \frac{P(A_i) \cdot P(B \mid A_i)}{\sum_{j=1}^{n} P(A_j) \cdot P(B \mid A_j)}$$

which is known as **Bayes' theorem**.

Bayes' Theorem for Posterior Probabilities

If $A_1, A_2, \ldots, A_n$ is a partition of a sample space, then the posterior probabilities of the events A_i conditional on an event B can be obtained from the prior probabilities $P(A_i)$ and the conditional probabilities $P(B \mid A_i)$ using the formula:

$$P(A_i \mid B) = \frac{P(A_i) \cdot P(B \mid A_i)}{\sum_{j=1}^{n} P(A_j) \cdot P(B \mid A_j)}$$

where:

- $P(A_i)$ is the prior probability of event A_i,
- $P(B \mid A_i)$ is the conditional probability of event B given A_i,

- The denominator is the total probability of B, computed by summing over all possible partitions A_j.

Bayes' Theorem is particularly useful in scenarios where the probability of an event is updated as more evidence becomes available. It plays a crucial role in fields such as machine learning, data analysis, and decision making under uncertainty.

Problem 4.16. *When a customer buys a car, the (prior) probabilities of it having been assembled in a particular plant are:*

- $P(\textit{Plant I}) = 0.20$
- $P(\textit{Plant II}) = 0.24$
- $P(\textit{Plant III}) = 0.25$
- $P(\textit{Plant IV}) = 0.31$

If a claim is made on the warranty of the car, how does this change these probabilities?

Solution

From Bayes' theorem, the posterior probabilities are calculated as follows:

$$P(\text{Plant I} \mid \text{Claim}) = \frac{P(\text{Plant I}) \cdot P(\text{Claim} \mid \text{Plant I})}{P(\text{Claim})}$$

Substitute the given values:

$$P(\text{Plant I} \mid \text{Claim}) = \frac{0.20 \times 0.05}{0.0687} = 0.146$$

$$P(\text{Plant II} \mid \text{Claim}) = \frac{P(\text{Plant II}) \cdot P(\text{Claim} \mid \text{Plant II})}{P(\text{Claim})}$$

Substitute the given values:

$$P(\text{Plant II} \mid \text{Claim}) = \frac{0.24 \times 0.11}{0.0687} = 0.384$$

$$P(\text{Plant III} \mid \text{Claim}) = \frac{P(\text{Plant III}) \cdot P(\text{Claim} \mid \text{Plant III})}{P(\text{Claim})}$$

Substitute the given values:

$$P(\text{Plant III} \mid \text{Claim}) = \frac{0.25 \times 0.03}{0.0687} = 0.109$$

$$P(\text{Plant IV} \mid \text{Claim}) = \frac{P(\text{Plant IV}) \cdot P(\text{Claim} \mid \text{Plant IV})}{P(\text{Claim})}$$

Substitute the given values:

$$P(\text{Plant IV} \mid \text{Claim}) = \frac{0.31 \times 0.08}{0.0687} = 0.361$$

Interpretation

The posterior probabilities are as follows:

- $P(\text{Plant I} \mid \text{Claim}) = 0.146$
- $P(\text{Plant II} \mid \text{Claim}) = 0.384$
- $P(\text{Plant III} \mid \text{Claim}) = 0.109$
- $P(\text{Plant IV} \mid \text{Claim}) = 0.361$

Notice that Plant II has the largest claim rate (0.11), and its posterior probability (0.384) is much larger than its prior probability (0.24). This is expected since the fact that a claim is made increases the likelihood that the car has been assembled in a plant with a high claim rate. Similarly, Plant III has the smallest claim rate (0.03), and its posterior probability (0.109) is much smaller than its prior probability (0.25), as expected.

Problem 4.17. *Suppose it is known that 1% of the population suffers from a particular disease. A blood test has a 97% chance of identifying the disease for diseased individuals, but also has a 6% chance of falsely indicating that a healthy person has the disease.*

(a) *What is the probability that a person will have a positive blood test?*

(b) *If your blood test is positive, what is the chance that you have the disease?*

(c) *If your blood test is negative, what is the chance that you do not have the disease?*

Solution

(a) Probability of a Positive Blood Test

Let D be the event that a person has the disease, and D^c be the event that a person does not have the disease. Let T^+ be the event of a positive test result, and T^- be the event of a negative test result.

$$\begin{aligned} P(D) &= 0.01 \\ P(D^c) &= 0.99 \\ P(T^+|D) &= 0.97 \\ P(T^+|D^c) &= 0.06 \end{aligned}$$

The total probability of a positive test result is given by:

$$\begin{aligned} P(T^+) &= P(T^+|D)P(D) + P(T^+|D^c)P(D^c) \\ &= (0.97 \times 0.01) + (0.06 \times 0.99) \\ &= 0.0097 + 0.0594 \\ &= 0.0691 \end{aligned}$$

So, the probability that a person will have a positive blood test is 0.0691.

(b) Probability of Having the Disease Given a Positive Test

We use Bayes' theorem:

$$\begin{aligned} P(D|T^+) &= \frac{P(T^+|D)P(D)}{P(T^+)} \\ &= \frac{0.97 \times 0.01}{0.0691} \\ &= \frac{0.0097}{0.0691} \\ &\approx 0.1403 \end{aligned}$$

So, if your blood test is positive, the chance that you have the disease is approximately 0.1403 or 14.03%.

(c) Probability of Not Having the Disease Given a Negative Test

We first find the probability of a negative test:

$$\begin{aligned} P(T^-) &= P(T^-|D)P(D) + P(T^-|D^c)P(D^c) \\ &= (1 - P(T^+|D))P(D) + (1 - P(T^+|D^c))P(D^c) \\ &= (1 - 0.97) \times 0.01 + (1 - 0.06) \times 0.99 \\ &= 0.03 \times 0.01 + 0.94 \times 0.99 \\ &= 0.0003 + 0.9406 \\ &= 0.9409 \end{aligned}$$

Now, using Bayes' theorem for $P(D^c|T^-)$:

$$\begin{aligned} P(D^c|T^-) &= \frac{P(T^-|D^c)P(D^c)}{P(T^-)} \\ &= \frac{0.94 \times 0.99}{0.9409} \\ &= \frac{0.9406}{0.9409} \\ &\approx 0.9997 \end{aligned}$$

So, if your blood test is negative, the chance that you do not have the disease is approximately 0.9997 or 99.97%.

4.8 Concluding Remarks

In this chapter, we covered the essential principles of probability that form the backbone of data science. We discussed experiments, sample spaces, joint and marginal probabilities, and conditional probabilities, providing a solid foundation for analyzing uncertainty and making data-driven decisions. We also explored various methods for assigning probabilities, including classical, empirical, and subjective approaches. The insights gained from understanding joint probabilities, marginal probabilities, and Bayes' Theorem will be invaluable for refining models and interpreting data.

As we move forward, the next chapter will delve into random variables and their properties. Random variables are crucial for quantifying and modeling uncertainty in a more structured way. We will explore different types of random variables, their distributions, and key properties, further building on the probability concepts introduced here. Mastering these topics will enhance your ability to handle complex data challenges and apply statistical techniques effectively. Understanding random variables is essential for advanced data analysis and developing predictive models.

4.9 Chapter Exercises

1. Consider an experiment where a fair six-sided die is rolled. Define the following events:

    - A: The event that the outcome is an even number.
    - B: The event that the outcome is greater than 4.

    Calculate the following probabilities:

    (a) $P(A)$

(b) $P(B)$
(c) $P(A \cap B)$
(d) $P(A \cup B)$
(e) $P(A^c)$

2. In a bag of 10 balls, 4 are red and 6 are blue. Two balls are drawn at random without replacement. Define the following events:

   - A: Drawing a red ball on the first draw.
   - B: Drawing a red ball on the second draw.

   Calculate the following probabilities:

   (a) $P(A)$
   (b) $P(B \mid A)$
   (c) $P(A \cap B)$
   (d) $P(A \cup B)$

3. You are given a deck of 52 playing cards. Define the following events:

   - A: Drawing a card that is a heart.
   - B: Drawing a card that is a queen.

   Calculate the following probabilities:

   (a) $P(A)$
   (b) $P(B)$
   (c) $P(A \cap B)$
   (d) $P(A \cup B)$
   (e) $P(A^c)$

4. A survey finds that 60% of people prefer coffee over tea, and 30% prefer both coffee and tea. What is the probability that a randomly chosen person prefers at least one of the two drinks? Define the following events:

   - A: Preferring coffee.
   - B: Preferring tea.

   Calculate:

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

5. In a company, 70% of employees are full-time workers, and 40% of employees are both full-time and have a college degree. If an employee is selected at random, find the probability that:

- The employee is a full-time worker.
- The employee has a college degree given that they are a full-time worker.
- The employee is either a full-time worker or has a college degree.

Define the following events:

- A: Being a full-time worker.
- B: Having a college degree.

Calculate:

(a) $P(A)$
(b) $P(B \mid A)$
(c) $P(A \cup B)$

6. In a clinical study, researchers are interested in the probability of a patient developing a particular health condition based on the type of treatment received. There are three types of treatments: A, B, and C. The probabilities of receiving each treatment are as follows:

   - Treatment A: 30% ($P(A) = 0.30$)
   - Treatment B: 50% ($P(B) = 0.50$)
   - Treatment C: 20% ($P(C) = 0.20$)

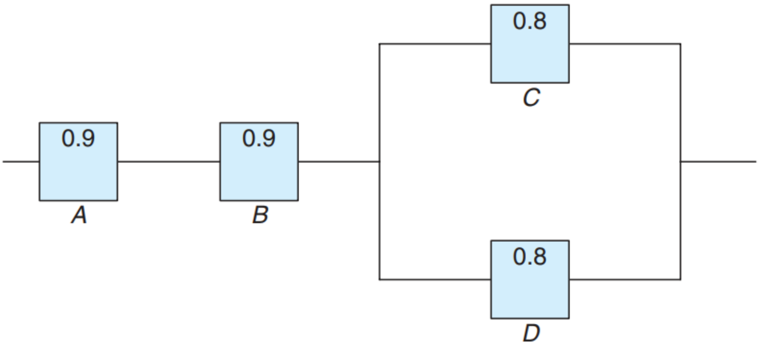
   The probability of developing the health condition given the type of treatment is known to be:

   - $P(\text{Condition} \mid A) = 0.10$
   - $P(\text{Condition} \mid B) = 0.25$
   - $P(\text{Condition} \mid C) = 0.15$

   Find the overall probability of a patient developing the health condition, denoted as $P(\text{Condition})$.

7. Suppose that somebody secretly rolls two fair six-sided dice, and what is the probability that the face-up value of the first one is 3, given the information that their sum is no greater than 5?

8. An electrical system consists of four components as illustrated in the following figure.



The system works if components A and B work and either of the components C or D works. The reliability (probability of working) of each component is also shown in the above figure. Find the probability that

(a) the entire system works.

(b) the component C does not work, given that the entire system works. Assume that the four components work independently.

(c) the component D does not work, given that the entire system works.

9. An agricultural research establishment grows vegetables and grades each one as either good or bad for its taste, good or bad for its size, and good or bad for its appearance. Overall 78% of the vegetables have a good taste. However, only 69% of the vegetables have both a good taste and a good size. Also, 5% of the vegetable have both a good taste and a good appearance, but a bad size. Finally, 84% of the vegetables have either a good size or a good appearance.

   (a). If a vegetable has a good taste, what is the probability that it also has a good size?

   (b). If a vegetable has a bad size and a bad appearance, what is the probability that it has a good taste?

.

10. A company produces electronic components, and it has two types of machines, A and B, that manufacture these components. Machine A produces 60% of the components, while Machine B produces 40%. Historical data shows that 2% of the components produced by Machine A are defective, while 5% of the components produced by Machine B are defective.

    A component is selected at random and found to be defective. What is the probability that this defective component was produced by Machine A?

Chapter 5

Random Variable and Its Properties

5.1 Introduction

In the realm of data science, understanding and manipulating uncertainty is a fundamental skill. At the core of this capability lies the concept of a random variable. A random variable is a quantitative variable whose values are determined by the outcome of a random phenomenon. It serves as a bridge connecting the abstract world of probability theory to the concrete domain of data analysis.

Random variables can be classified into two main types: discrete and continuous. Discrete random variables take on a countable number of distinct values, often representing things like the number of occurrences of an event. Continuous random variables, on the other hand, can take on an infinite number of possible values within a given range, making them essential for representing measurements and other quantities that vary smoothly.

This chapter delves into the foundational aspects of random variables, exploring their properties and the critical role they play in statistical modeling and data analysis. We will discuss probability distributions, expected values, variances, and other By the end of this chapter, readers will gain a robust understanding of how random variables function and how they can be applied to solve real-world problems in data science.

5.2 Random Variable

A **random variable** is a mathematical concept used in probability theory and statistics, representing a variable whose possible values depend on the outcomes of a random phenomenon. It serves as a fundamental tool for defining probability distributions and calculating probabilities associated with events arising from uncertain or stochastic processes. In data science, random variables are fundamental because they allow us to model and reason about uncertainty and variability in data.

A random variable is a numerical outcome of a random phenomenon. It is a function that assigns a real number to each outcome in a sample space of a random experiment. Formally, a random variable X is defined as a function:

$$X : S \to \mathbb{R}$$

where S is the sample space of the experiment, and $\mathbb{R}$ is the set of real numbers.

Random Variable: A variable whose possible values are determined by outcomes of a random experiment or process, with each value associated with a probability.

Random variables can be classified into two types: discrete and continuous.

Example: Testing Electronic Components

Consider a **random experiment** where three electronic components are tested for defects. The sample space, giving a detailed description of each possible outcome, can be written as follows:

$$S = \{\text{NNN}, \text{NDN}, \text{NND}, \text{DNN}, \text{NDD}, \text{DND}, \text{DDN}, \text{DDD}\}$$

where,

- N stands for a non-defective component.
- D stands for a defective component.

Defining the Random Variable

Let X be the random variable representing the **number of defective components** in the sample. The possible values of X are denoted by x, and their corresponding outcomes are listed in Table 5.1. The random variable X can take on the following values:

- $x = 0$: No defective components (Outcome: NNN)
- $x = 1$: One defective component (Outcomes: NDN, NND, DNN)
- $x = 2$: Two defective components (Outcomes: NDD, DND, DDN)
- $x = 3$: Three defective components (Outcome: DDD).

Table 5.1: Possible Outcomes When Testing Three Electronic Components

Outcome	NNN	NDN	NND	DNN	NDD	DND	DDN	DDD
x	0	1	1	1	2	2	2	3

In this example, X is a discrete random variable because it can take on a countable number of distinct values. Each value of X corresponds to the number of defective components in the tested sample. There are two main types of random variables: discrete and continuous.

5.3 Discrete Random Variables

A **discrete random variable** can take on a countable number of possible values. Here are some examples of discrete random variables:

1. **Number of Heads in a Series of Coin Tosses**: When flipping a fair coin multiple times, the number of heads observed in the series is a discrete random variable. For example, if you flip a coin 10 times, the number of heads (0 to 10) is a discrete outcome.

2. **Number of Defective Items in a Batch**: In quality control, the number of defective items in a batch of products is a discrete random variable. For instance, if a factory produces 100 items in a day, the number of defective items could be any integer from 0 to 100.

3. **Number of Customers in a Queue**: The number of customers waiting in line at a service center or a bank is a discrete random variable. At any given time, this number could be 0, 1, 2, and so on.

4. **Roll of a Die**: When rolling a standard six-sided die, the outcome is a discrete random variable with possible values of 1, 2, 3, 4, 5, or 6.

5. **Number of Emails Received in a Day**: The number of emails a person receives in a day is a discrete random variable. It can take on any non-negative integer value (0, 1, 2, . . .).

6. **Number of Accidents at an Intersection**: The number of traffic accidents occurring at a particular intersection in a month is a discrete random variable. This count could be 0, 1, 2, and so on.

7. **Number of Children in a Family**: The number of children in a family is a discrete random variable, with possible values of 0, 1, 2, and so forth.

8. **Number of Sales Transactions in a Day**: The number of sales transactions processed by a retail store in a single day is a discrete random variable, representing the count of individual sales.

These examples illustrate various contexts in which discrete random variables are used to model and analyze real-world phenomena.

5.3.1 Probability Mass Function (pmf)

The probability distribution of a random variable describes how probabilities are distributed over the values of the random variable. For a discrete random variable, the probability distribution is described by the **probability mass function (pmf)**, which gives the probability that the random variable takes on a specific value.

> **Probability Mass Function (pmf):** The probability mass function (pmf) of a discrete random variable X is defined by a set of probabilities $p(x)$ assigned to each possible value x that the variable can take. These probabilities must satisfy the conditions $0 \leq p(x) \leq 1$ and $\sum_x p(x) = 1$. The probability that the random variable X takes the value x is denoted as $P(X = x) = p(x)$.

Example: Testing Electronic Components

Consider the example of **Testing Electronic Components** described in the previous section, where X is the random variable representing the **number of defective components** in the three tested electronic components. The probability mass function for X is shown below, and the graphical representation is presented in Figure 5.1.

Table 5.2: Probability mass function

Outcome	x	$P(X = x)$
{NNN}	0	$\frac{1}{8}$
{NDN, NND, DNN}	1	$\frac{3}{8}$
{NDD, DND, DDN}	2	$\frac{3}{8}$
{DDD}	3	$\frac{1}{8}$

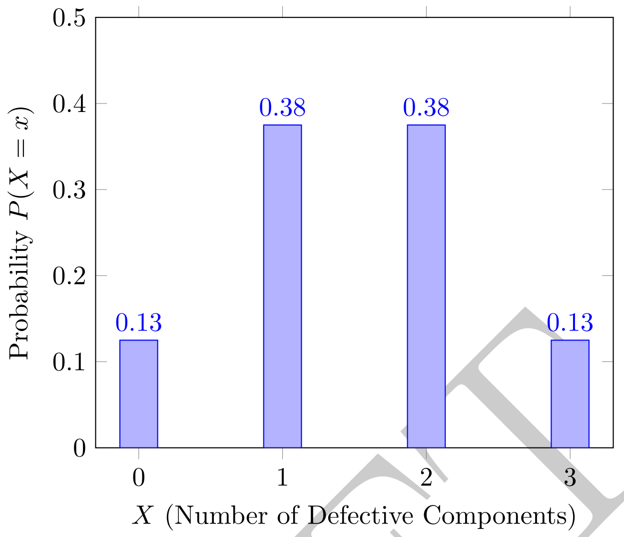


Figure 5.1: Probability Mass Function of X

5.3.2 Cumulative Distribution Function (cdf)

The **cumulative distribution function** (cdf) of a random variable X is a function that gives the probability that X will take a value less than or equal to x. For both discrete and continuous random variables,

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

The cdf is a non-decreasing function that ranges from 0 to 1.

For the discrete random variable X, the cumulative distribution function can then be calculated from the expression:

$$F(x) = \sum_{y:y\leq x} P(X = y).$$

Example: Testing Electronic Components

Consider the example of **Testing Electronic Components** described in the previous section, where X is the random variable representing the **number of defective components** in the three tested electronic components. The cumulative distribution function for X is shown below, and the graphical representation is presented in Figure 5.2.

Table 5.3: The cdf for the Number of Defective Components

x	$P(X = x)$	$F(x)$
0	$\frac{1}{8}$	0.125
1	$\frac{3}{8}$	0.5
2	$\frac{3}{8}$	0.875
3	$\frac{1}{8}$	1

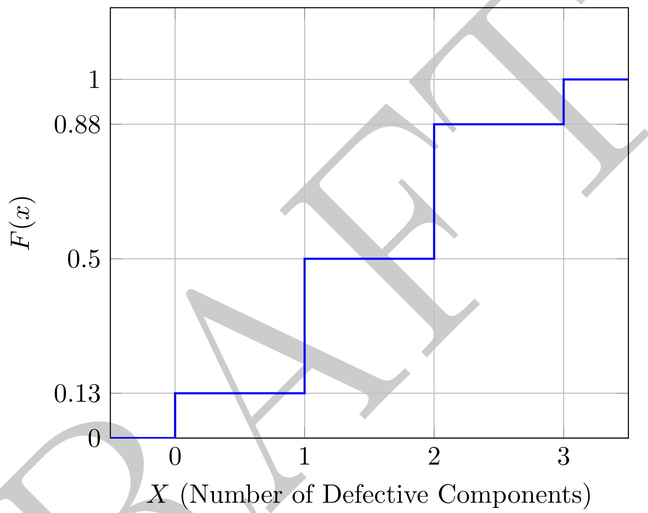


Figure 5.2: Cumulative Distribution Function of X

5.3.3 Properties of the Cumulative Distribution Function

The cdf of a random variable has several important properties:

1. **Non-decreasing**: The cdf $F(x)$ is a non-decreasing function. This means that if $x_1 \leq x_2$, then $F(x_1) \leq F(x_2)$. The probability that the random variable takes a value less than or equal to x does not decrease as x increases.

2. **Limits**:
   - $\lim_{x \to -\infty} F(x) = 0$; that is the minimum value of the cdf is 1.
   - $\lim_{x \to +\infty} F(x) = 1$; that is the maximum value of the cdf is 1.

3. **Right-Continuous**: The cdf $F(x)$ is right-continuous. This means that for any value x, the limit of $F(x)$ as t approaches x from the right ($t \to x^+$)

is equal to $F(x)$. Mathematically, this can be written as $\lim_{t\to x^+} F(t) = F(x)$.

4. **Range**: The cdf $F(x)$ takes values in the interval $[0, 1]$. For any real number x, $0 \leq F(x) \leq 1$. This reflects the fact that probabilities range from 0 to 1.

Problem 5.1. *An office has four copying machines, and the random variable X measures how many of them are in use at a particular moment in time. Suppose that $P(X = 0) = 0.08$, $P(X = 1) = 0.11$, $P(X = 2) = 0.27$, and $P(X = 3) = 0.33$.*

(a) What is $P(X = 4)$?

(b) Draw a line graph of the probability mass function.

(c) Construct and plot the cumulative distribution function.

Solution

(a) Since the sum of all probabilities must be 1, we have:

$$\begin{aligned} P(X = 4) &= 1 - (P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)) \\ &= 1 - (0.08 + 0.11 + 0.27 + 0.33) = 1 - 0.79 \\ &= 0.21 \end{aligned}$$

(b) The graphical presentation of the probability mass function is the following:

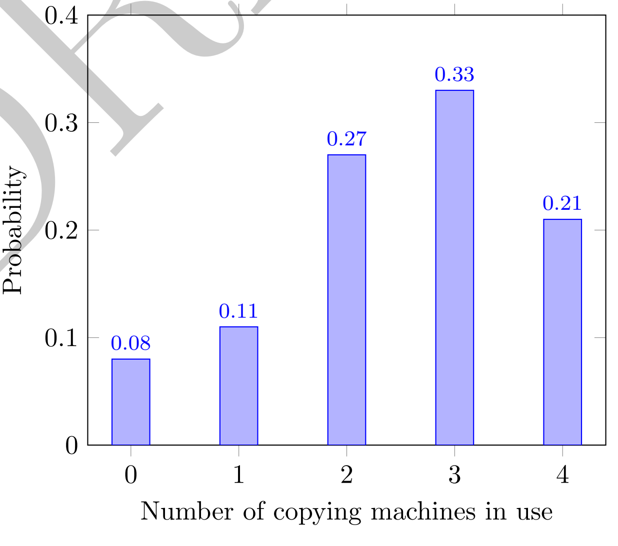


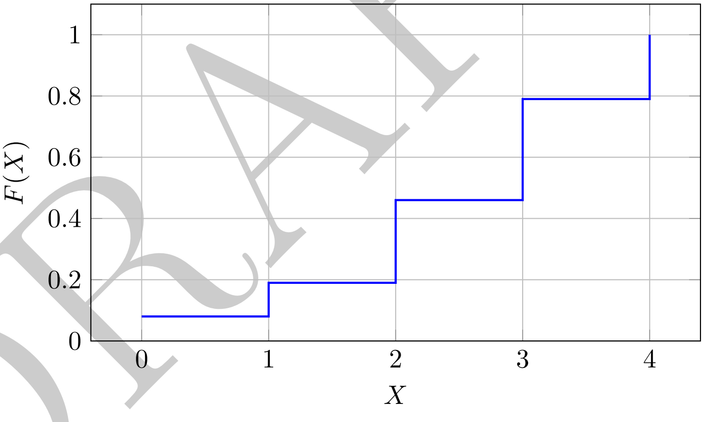
Figure 5.3: Probability Mass Function

(c) We konw, the cumulative distribution function $F(x)$ is defined as:

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

$$\begin{aligned}
F(0) &= P(X = 0) = 0.08 \\
F(1) &= P(X \leq 1) = P(X = 0) + P(X = 1) = 0.08 + 0.11 = 0.19 \\
F(2) &= P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) \\
&= 0.08 + 0.11 + 0.27 = 0.46 \\
F(3) &= P(X \leq 3) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) \\
&= 0.08 + 0.11 + 0.27 + 0.33 = 0.79 \\
F(4) &= P(X \leq 4) \\
&= P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) \\
&= 0.08 + 0.11 + 0.27 + 0.33 + 0.21 = 1.00
\end{aligned}$$

The graphical presentation of $F(x)$ is the following:



Problem 5.2. *Let the number of phone calls received by a switchboard during a 5-minute interval be a random variable X with probability function*

$$f(x) = \frac{e^{-2}2^x}{x!}, \quad \text{for } x = 0, 1, 2, \ldots$$

(a) Determine the probability that X equals 0, 1, 2, 3, 4, 5, and 6.

(b) Graph the probability mass function for these values of x.

(c) Determine the cumulative distribution function for these values of X.

Solution

(a) Probabilities

The probability function is given by

$$f(x) = \frac{e^{-2}2^x}{x!}$$

The probabilities for $X = 0, 1, 2, 3, 4, 5, 6$ are:

$$\begin{aligned}
P(X = 0) &= \frac{e^{-2}2^0}{0!} = e^{-2} \approx 0.1353 \\
P(X = 1) &= \frac{e^{-2}2^1}{1!} = 2e^{-2} \approx 0.2707 \\
P(X = 2) &= \frac{e^{-2}2^2}{2!} = 2^2e^{-2}/2 \approx 0.2707 \\
P(X = 3) &= \frac{e^{-2}2^3}{3!} = 2^3e^{-2}/6 \approx 0.1804 \\
P(X = 4) &= \frac{e^{-2}2^4}{4!} = 2^4e^{-2}/24 \approx 0.0902 \\
P(X = 5) &= \frac{e^{-2}2^5}{5!} = 2^5e^{-2}/120 \approx 0.0361 \\
P(X = 6) &= \frac{e^{-2}2^6}{6!} = 2^6e^{-2}/720 \approx 0.0120
\end{aligned}$$

(b) Graph of the Probability Mass Function

(c) Cumulative Distribution Function

The cumulative distribution function $F(x) = P(X \leq x)$ for $x = 0, 1, 2, 3, 4, 5, 6, \ldots$ is:

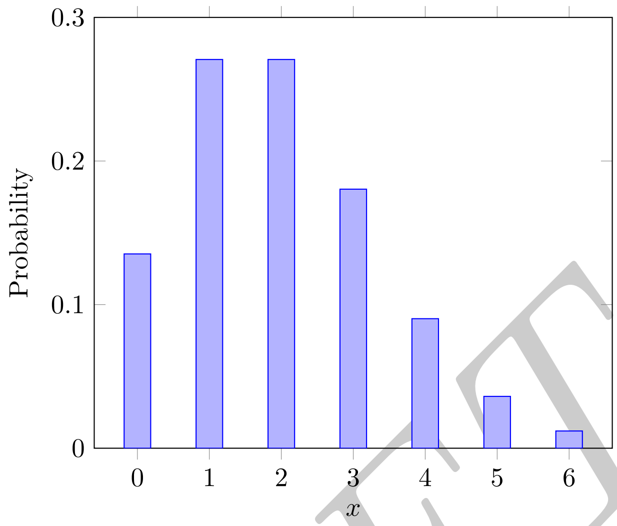


Figure 5.4: Probability Mass Function of X

$$
\begin{aligned}
F(0) &= P(X \leq 0) = P(X=0) = 0.1353\\
F(1) &= P(X \leq 1) = P(X=0) + P(X=1) = 0.1353 + 0.2707 = 0.4060\\
F(2) &= P(X \leq 2) = P(X=0) + P(X=1) + P(X=2)\\
&\quad = 0.4060 + 0.2707 = 0.6767\\
F(3) &= P(X \leq 3) = P(X=0) + P(X=1) + P(X=2) + P(X=3)\\
&\quad = 0.6767 + 0.1804 = 0.8571\\
F(4) &= P(X \leq 4) = P(X=0) + P(X=1) + P(X=2) + P(X=3) + P(X=4)\\
&\quad = 0.8571 + 0.0902 = 0.9473\\
F(5) &= P(X \leq 5) = P(X=0) + P(X=1) + P(X=2) + P(X=3)\\
&\quad + P(X=4) + P(X=5)\\
&= 0.9473 + 0.0361 = 0.9834\\
F(6) &= P(X \leq 6) = P(X=0) + P(X=1) + P(X=2) + P(X=3)\\
&\quad + P(X=4) + P(X=5) + P(X=6)\\
&= 0.9834 + 0.0120 = 0.9954
\end{aligned}
$$

Problem 5.3. *Let X be a discrete random variable with the following probability mass function :*

$$P(X = x) = \begin{cases} 0.2 & \textit{for } x = 1 \\ 0.5 & \textit{for } x = 2 \\ 0.3 & \textit{for } x = 3 \\ 0 & \textit{otherwise} \end{cases}$$

Compute $F(x)$.

Solution

The cumulative distribution function $F(x)$ is given by:

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

Let's compute $F(x)$ for different values of x:

- For $x < 1$:
$$F(x) = 0$$
- For $1 \leq x < 2$:
$$F(x) = P(X \leq 1) = P(X = 1) = 0.2$$
- For $2 \leq x < 3$:
$$F(x) = P(X \leq 2) = P(X = 1) + P(X = 2) = 0.2 + 0.5 = 0.7$$
- For $x \geq 3$:
$$F(x) = P(X \leq 3) = P(X = 1) + P(X = 2) + P(X = 3) = 0.2 + 0.5 + 0.3 = 1$$

So, the cdf $F(x)$ can be summarized as:

$$F(x) = \begin{cases} 0 & \text{for } x < 1 \\ 0.2 & \text{for } 1 \leq x < 2 \\ 0.7 & \text{for } 2 \leq x < 3 \\ 1 & \text{for } x \geq 3 \end{cases}$$

Problem 5.4. *Let X be a discrete random variable with the following probability mass function:*

$$P(X = x) = \begin{cases} 0.1 & \textit{for } x = 0 \\ 0.4 & \textit{for } x = 1 \\ 0.3 & \textit{for } x = 2 \\ 0.2 & \textit{for } x = 3 \\ 0 & \textit{otherwise} \end{cases}$$

Compute $F(x)$.

Solution

Let's compute $F(x)$ for different values of x:

- For $x < 0$:
$$F(x) = 0$$

- For $0 \leq x < 1$:
$$F(x) = P(X \leq 0) = P(X = 0) = 0.1$$

- For $1 \leq x < 2$:
$$F(x) = P(X \leq 1) = P(X = 0) + P(X = 1) = 0.1 + 0.4 = 0.5$$

- For $2 \leq x < 3$:
$$\begin{aligned} F(x) &= P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) \\ &= 0.1 + 0.4 + 0.3 = 0.8 \end{aligned}$$

- For $x \geq 3$:
$$\begin{aligned} F(x) &= P(X \leq 3) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) \\ &= 0.1 + 0.4 + 0.3 + 0.2 = 1 \end{aligned}$$

So, the cdf $F(x)$ is:

$$F(x) = \begin{cases} 0 & \text{for } x < 0 \\ 0.1 & \text{for } 0 \leq x < 1 \\ 0.5 & \text{for } 1 \leq x < 2 \\ 0.8 & \text{for } 2 \leq x < 3 \\ 1 & \text{for } x \geq 3. \end{cases}$$

5.3.4 Exercise (Solution in note 1)

1. An office has four copying machines, and the random variable X denotes how many of them are in use in a particular time. Suppose the probability mass function X is given below:

x	0	1	2	3	4
$\Pr(X = x)$	k	0.02	0.05	0.4	$(k + 0.3)$

(a) What is the value of k and draw the line graph of the probability mass function $\Pr(X = x)$.

(b) Find the Value of $\Pr(X \leq 2)$.

(c) Find the probability that at least two copying machines are in use.

(d) Find the cumulative Function $F(x)$ and draw the $F(x)$.

2. An office has five printers and the random variable Y measures how many of them are currently being used. Suppose that $P(Y = 0) = 0.05$, $P(Y = 1) = 0.10$, $P(Y = 2) = 0.20$, $P(Y = 3) = 0.30$, and $P(Y = 4) = 0.25$.

   (a) What is $P(Y = 5)$?
   (b) Draw a line graph of the probability mass function.
   (c) Construct and plot the cumulative distribution function.

3. A hospital has six emergency rooms and the random variable Z measures how many of them are occupied at a given time. Suppose that $P(Z = 0) = 0.04$, $P(Z = 1) = 0.10$, $P(Z = 2) = 0.20$, $P(Z = 3) = 0.25$, $P(Z = 4) = 0.20$, and $P(Z = 5) = 0.15$.

   (a) What is $P(Z = 6)$?
   (b) Draw a line graph of the probability mass function.
   (c) Construct and plot the cumulative distribution function.

4. A clinic has three doctors and the random variable W measures how many of them are available at a particular moment in time. Suppose that $P(W = 0) = 0.15$, $P(W = 1) = 0.20$, and $P(W = 2) = 0.30$.

   (a) What is $P(W = 3)$?
   (b) Draw a line graph of the probability mass function.
   (c) Construct and plot the cumulative distribution function.

5. A warehouse has seven forklifts and the random variable V measures how many of them are currently in operation. Suppose that $P(V = 0) = 0.02$, $P(V = 1) = 0.08$, $P(V = 2) = 0.18$, $P(V = 3) = 0.25$, $P(V = 4) = 0.20$, and $P(V = 5) = 0.15$.

   (a) What is $P(V = 6)$?
   (b) Draw a line graph of the probability mass function.
   (c) Construct and plot the cumulative distribution function.

6. A manufacturing plant has four assembly lines and the random variable U measures how many of them are operating at a given time. Suppose that $P(U = 0) = 0.10$, $P(U = 1) = 0.20$, and $P(U = 2) = 0.35$.

   (a) What is $P(U = 3)$?
   (b) Draw a line graph of the probability mass function.
   (c) Construct and plot the cumulative distribution function.

5.4 Continuous Random Variables

A **continuous random variable** can take on an uncountable number of possible values. Here are some examples of continuous random variables:

1. **Height of Individuals**: The height of a person is a continuous random variable because it can take any value within a given range. For example, the height could be 170.2 cm, 175.5 cm, etc.

2. **Time Taken to Complete a Task**: The time required to finish a task, such as running a marathon, is a continuous random variable. It can be measured in hours, minutes, seconds, and fractions of a second.

3. **Temperature**: The temperature at a specific location and time is a continuous random variable. It can take any value within the possible range of temperatures, such as 23.45°C, 37.8°C, etc.

4. **Weight of an Object**: The weight of an object is a continuous random variable. For example, a bag of flour might weigh 1.25 kg, 1.30 kg, etc.

5. **Amount of Rainfall**: The amount of rainfall in a day is a continuous random variable. It can be measured in millimeters or inches, and it can take any value within a range.

6. **Price of a Stock**: The price of a stock at any given moment is a continuous random variable. It can vary continuously and take on any value within the range of possible stock prices.

7. **Age of an Individual**: The age of a person can be considered a continuous random variable if measured precisely. For instance, someone could be 25.3 years old, 45.7 years old, etc.

8. **Voltage in an Electrical Circuit**: The voltage at a point in an electrical circuit is a continuous random variable. It can take any value within the possible voltage range.

These examples illustrate various contexts in which continuous random variables are used to model and analyze real-world phenomena.

5.4.1 Probability Density Function (pdf)

For a continuous random variable, the probability distribution is described by the **probability density function** (pdf), which specifies the probability density at each point in the random variable's range. For a continuous random variable X,

$$f(x)$$

is the pdf of X, where $f(x) \geq 0$ and the area under the pdf curve is equal to 1. Sometimes the density of X is denoted by $f_X(x)$ to indicate which random variable the function f corresponds to.

A probability density function $f(x)$ defines the probabilistic properties of a continuous random variable. It must satisfy $f(x) \geq 0$ and

$$\int_{\text{state space}} f(x)\,dx = 1.$$

The probability that the random variable lies between two values is obtained by integrating the probability density function between the two values.

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

It is useful to notice that the probability that a continuous random variable X takes any specific value a is always 0! Technically, this can be seen by noting that

$$P(X = a) = \int_a^a f(x)\,dx = 0.$$

5.4.2 Example: Metal Cylinder Production

Suppose that the diameter of a metal cylinder, denoted by X, has a probability density function

$$f(x) = \begin{cases} 1.5 - 6(x - 50.0)^2 & \text{for } 49.5 \leq x \leq 50.5 \\ 0 & \text{elsewhere.} \end{cases}$$

To determine if $f(x) = 1.5 - 6(x - 50.0)^2$ for $49.5 \leq x \leq 50.5$ and $f(x) = 0$ elsewhere is a valid probability pdf, we need to check two conditions:

1. **Non-negativity**: $f(x) \geq 0$ for all x.
2. **Normalization**: The total integral of $f(x)$ over all possible values must equal 1.

Non-negativity Check

We need to ensure that $f(x) \geq 0$ for $49.5 \leq x \leq 50.5$:

$$f(x) = 1.5 - 6(x - 50.0)^2$$

For $49.5 \leq x \leq 50.5$, let's calculate the minimum value of the quadratic function: $(x - 50.0)^2$ is minimized at $x = 50.0$, and $(x - 50.0)^2$ ranges from 0 to $(0.5)^2 = 0.25$.

$$f(x) = 1.5 - 6(x - 50.0)^2 \geq 1.5 - 6 \cdot 0.25 = 1.5 - 1.5 = 0.$$

Thus, $f(x) \geq 0$ for all x in the given interval.

Normalization Check

We need to integrate $f(x)$ over the interval $49.5 \leq x \leq 50.5$ and check if the integral equals 1:

$$
\begin{aligned}
\int_{49.5}^{50.5} f(x)\,dx &= \int_{49.5}^{50.5} \left(1.5 - 6(x-50.0)^2\right) dx \\
&= \int_{49.5}^{50.5} 1.5\,dx - \int_{49.5}^{50.5} 6(x-50.0)^2\,dx \\
&= 1.5 \times (50.5 - 49.5) - \int_{49.5}^{50.5} 6(x-50.0)^2\,dx \\
&= 1.5 - \int_{49.5}^{50.5} 6(x-50.0)^2\,dx. && (5.1) \\
& && (5.2)
\end{aligned}
$$

The integral

$$\int_{49.5}^{50.5} 6(x-50.0)^2\,dx$$

can be simplified by substitution. Let $u = x - 50.0$. Then $du = dx$, and the limits of integration change accordingly:

When $x = 49.5$, $u = -0.5$ and when $x = 50.5$, $u = 0.5$. So,

$$\int_{49.5}^{50.5} 6(x-50.0)^2\,dx = 6\int_{-0.5}^{0.5} u^2\,du$$

The integral of u^2 is:

$$
\begin{aligned}
\int_{-0.5}^{0.5} u^2\,du &= \left[\frac{u^3}{3}\right]_{-0.5}^{0.5} \\
&= \frac{(0.5)^3}{3} - \frac{(-0.5)^3}{3} = \frac{0.125}{3} - \left(-\frac{0.125}{3}\right) \\
&= \frac{0.125}{3} + \frac{0.125}{3} \\
&= \frac{1}{12}
\end{aligned}
$$

and

$$6\int_{-0.5}^{0.5} u^2\,du = 6 \times \frac{1}{12} = \frac{6}{12} = 0.5$$

Therefore, from the Equation (5.2), we have

$$\int_{49.5}^{50.5} f(x)\,dx = 1.5 - 0.5 = 1$$

Since both conditions are satisfied, $f(x)$ is indeed a valid probability density function.

The graphical presentation of the density $f(x)$ is presented in Figure 5.5.

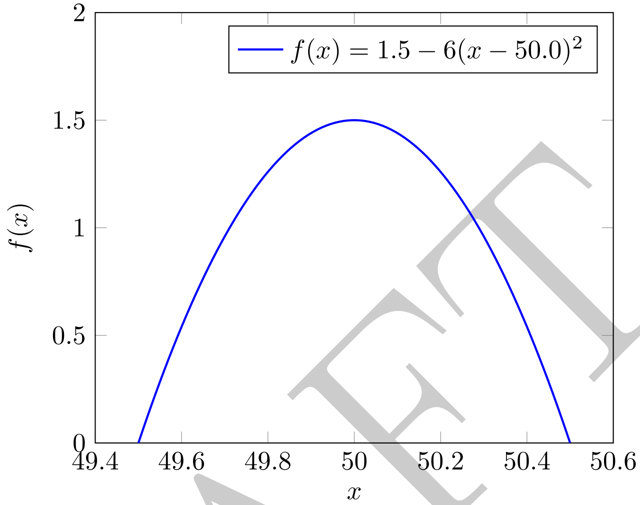


Figure 5.5: Density Plot of the pdf $f(x)$

For this example, we can find the probability that a metal cylinder has a diameter between 49.8 and 50.1 mm. This can be calculated to be

$$
\begin{aligned}
P(49.8 \leq X \leq 50.1) &= \int_{49.8}^{50.1} f(x)\,dx \\
&= \int_{49.8}^{50.1} \left(1.5 - 6(x - 50.0)^2\right)\,dx \\
&= \int_{49.8}^{50.1} 1.5\,dx - \int_{49.8}^{50.1} 6(x - 50.0)^2\,dx \\
&= 1.5\,[x]_{49.8}^{50.1} - 6\int_{-0.2}^{0.1} u^2\,du \quad [\text{let,} \quad u = x - 50.0] \\
&= 1.5(50.1 - 49.8) - 6\int_{-0.2}^{0.1} u^2\,du \\
&= 0.45 - \frac{(0.1)^3}{3} - \frac{(-0.2)^3}{3} \\
&= 0.45 - 0.018 = 0.432
\end{aligned}
$$

Thus, the probability that a metal cylinder has a diameter between 49.8 and 50.1 mm is 0.432 or 43.2%.

Height of a Plant Species

Suppose that the height of a particular plant species, denoted by H, has a probability density function

$$f(h) = \begin{cases} \frac{3}{40}(2 - 0.04(h-100)^2) & \text{for } 90 \le h \le 110 \\ 0 & \text{elsewhere} \end{cases}$$

where H is measured in centimeters.

5.4.3 Cumulative Distribution Function (cdf)

The **cumulative distribution function** (cdf) of a continuous random variable X is a function that gives the probability that X will take a value less than or equal to x. The cumulative distribution function can be calculated from the expression:

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(y)\,dy.$$

In practical applications, the lower integration limit of $-\infty$ can be replaced by the lower boundary of the state space, since the probability density function (pdf) is zero outside this region. The pdf can be obtained by differentiating the cumulative distribution function (cdf), which is given by:

$$f(x) = \frac{dF(x)}{dx}.$$

Cumulative Distribution Function: The cumulative distribution function $F_X(x)$ of a random variable X is defined as

$$F_X(x) = P(X \le x) = \int_{-\infty}^{x} f(y)\,dy.$$

For a continuous random variable X, the three properties mentioned in Section 5.3.3 are satisfied. In addition, the following property must hold: if the cdf $F(\cdot)$ is continuous at any $a \le x \le b$, then

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

5.4.4 Example: Metal Cylinder Production

Since $f(x) = 0$ outside the interval $[49.5, 50.5]$, we have three cases:

1. for $x < 49.5$, $F(x) = 0$.
2. for $49.5 \le x \le 50.5$:

$$F(x) = \int_{49.5}^{x} (1.5 - 6(t - 50.0)^2)\,dt$$

3. for $x > 50.5$, $F(x) = 1$.

For the second case, we integrate term-by-term:

$$\begin{aligned}
F(x) &= \int_{49.5}^{x} (1.5 - 6(t-50.0)^2)\,dt \\
&= \int_{49.5}^{x} 1.5\,dt - \int_{49.5}^{x} 6(t-50.0)^2\,dt \\
&= 1.5(x-49.5) - \int_{-0.5}^{x-50} u^2\,du \quad [\text{let},\quad u = x - 50.0] \\
&= 1.5(x-49.5) - 6\left[\frac{u^3}{3}\right]_{-0.5}^{x-50} \\
&= 1.5(x-49.5) - 2\left((x-50)^3 + 0.125\right) \\
&= 1.5(x-49.5) - 2(x-50)^3 - 0.25
\end{aligned}$$

Therefore, the cdf $F(x)$ is:

$$F(x) = \begin{cases} 0 & \text{for } x < 49.5 \\ 1.5(x-49.5) - 2(x-50)^3 - 0.25 & \text{for } 49.5 \le x \le 50.5 \\ 1 & \text{for } x > 50.5 \end{cases}$$

The graphical presentation of $F(x)$ is dipicted in Figure 5.6.

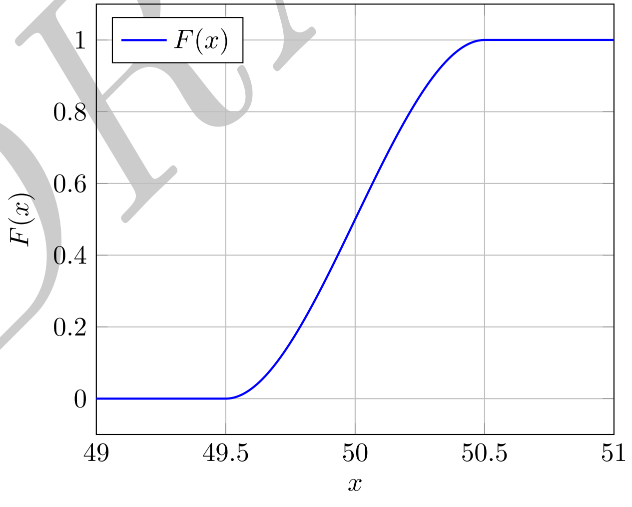


Figure 5.6: The cumulative distribution function of $f(x)$.

Problem 5.5. *Let X be a continuous random variable with the probability density function (pdf):*

$$f(x) = \begin{cases} 2x & \text{for } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Compute $F(x)$.

Solution

The cumulative distribution function $F(x)$ is given by:

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

Let's compute $F(x)$ for different values of x:

- for $x < 0$:
$$F(x) = 0$$
- for $0 \leq x \leq 1$:
$$F(x) = \int_0^x 2t\,dt = \left[t^2\right]_0^x = x^2$$
- for $x > 1$:
$$F(x) = \int_0^1 2t\,dt = \left[t^2\right]_0^1 = 1$$

So, the cdf $F(x)$ can be summarized as:

$$F(x) = \begin{cases} 0 & \text{for } x < 0 \\ x^2 & \text{for } 0 \leq x \leq 1 \\ 1 & \text{for } x > 1 \end{cases}$$

Problem 5.6. *Let X be a continuous random variable with the probability density function:*

$$f(x) = \begin{cases} 3x^2 & \text{for } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Compute $F(x)$.

Solution

Let's compute $F(x)$ for different values of x:

- for $x < 0$:
$$F(x) = 0$$

- for $0 \leq x \leq 1$:

$$F(x) = \int_0^x 3t^2\, dt = \left[t^3\right]_0^x = x^3$$

- for $x > 1$:

$$F(x) = \int_0^1 3t^2\, dt = \left[t^3\right]_0^1 = 1$$

So, the cdf $F(x)$ is:

$$F(x) = \begin{cases} 0 & \text{for } x < 0 \\ x^3 & \text{for } 0 \leq x \leq 1 \\ 1 & \text{for } x > 1 \end{cases}$$

Problem 5.7. *Let X be a continuous random variable with the pdf:*

$$f(x) = \begin{cases} \frac{1}{2} e^{-|x|} & \textit{for } -\infty < x < \infty \\ 0 & \textit{otherwise} \end{cases}$$

Compute $F(x)$.

Solution

Let's compute $F(x)$ for different values of x:

- for $x < -\infty$:

$$F(x) = 0$$

- for $-\infty < x < \infty$:

$$F(x) = \int_{-\infty}^x \frac{1}{2} e^{-|t|}\, dt$$

Since $e^{-|t|}$ can be split into two parts depending on the range of t:

$$\text{For } x < 0: \quad F(x) = \int_{-\infty}^x \frac{1}{2} e^t\, dt = \frac{1}{2}\left[e^t\right]_{-\infty}^x = \frac{1}{2}\left(e^x - 0\right) = \frac{1}{2} e^x$$

for $x \geq 0$:

$$\begin{aligned} F(x) &= \int_{-\infty}^0 \frac{1}{2} e^t\, dt + \int_0^x \frac{1}{2} e^{-t}\, dt \\ &= \frac{1}{2}\left[e^t\right]_{-\infty}^0 + \frac{1}{2}\left[-e^{-t}\right]_0^x = \frac{1}{2}(1 - 0) + \frac{1}{2}\left(1 - e^{-x}\right) \\ &= 1 - \frac{1}{2} e^{-x} \end{aligned}$$

So, the cdf $F(x)$ is:

$$F(x) = \begin{cases} \frac{1}{2} e^x & \text{for } x < 0 \\ 1 - \frac{1}{2} e^{-x} & \text{for } x \geq 0. \end{cases}$$

5.4.5 Exercises (Lecture note 1,2)

1. Consider a random variable measuring the following quantities. In each case, state with reasons whether you think it is more appropriate to define the random variable as discrete or continuous.

   (a) The number of books in a library
   (b) The duration of a phone call
   (c) The number of steps a person takes in a day
   (d) The amount of rainfall in a month
   (e) The number of languages a person speaks
   (f) The speed of a car on a highway

2. A random variable X takes values between 4 and 6 with a probability density function

$$f(x) = \frac{k}{x \ln(1.5)} \quad \text{for } 4 \le x \le 6.$$

   (a) What is the value of k?
   (b) Make a sketch of the probability density function.
   (c) Check that the total area under the probability density function is equal to 1.
   (d) What is $P(4.5 \le X \le 5.5)$?
   (e) Construct and sketch the cumulative distribution function.

3. A random variable Y takes values between 1 and 3 with a probability density function

$$g(y) = \frac{k}{(y+1)^2} \quad \text{for } 1 \le y \le 3.$$

   (a) Find the value of k and then make a sketch of the probability density function.
   (b) Check that the total area under the probability density function is equal to 1.
   (c) What is $P(1.5 \le Y \le 2.5)$?
   (d) Construct and sketch the cumulative distribution function.

4. A random variable Z takes values between 2 and 5 with a probability density function

$$h(z) = \frac{k}{(z+1)^3} \quad \text{for } 2 \le z \le 5.$$

(a) Find the value of k and then make a sketch of the probability density function.

(b) Check that the total area under the probability density function is equal to 1.

(c) What is $P(2.5 \leq Z \leq 4)$?

(d) Construct and sketch the cumulative distribution function.

5. A random variable X takes values between 0 and 4 with a cumulative distribution function

$$F(x) = \frac{x^2}{16} \quad \text{for } 0 \leq x \leq 4.$$

(a) Sketch the cumulative distribution function.

(b) What is $P(X \leq 2)$?

(c) What is $P(1 \leq X \leq 3)$?

(d) Construct and sketch the probability density function.

6. The resistance X of an electrical component has a probability density function

$$f(x) = Ax(130 - x^2) \quad \text{for resistance values in the range } 10 \leq x \leq 11.$$

(a) Calculate the value of the constant A.

(b) Calculate the cumulative distribution function.

(c) What is the probability that the electrical component has a resistance between 10.25 and 10.5?

5.5 The Expectation of a Random Variable

While the probability mass function or the probability density function provides complete information about the probabilistic properties of a random variable, it is often useful to use some summary measures of these properties. One of the most fundamental summary measures is the expectation or mean of a random variable, denoted by $E(X)$, which represents the "average" value of the random variable. Two random variables with the same expected value can be considered to have the same average value, even though their probability mass functions or probability density functions may differ significantly.

The **expected value** (or mean) of a random variable is a measure of its central tendency.

Expected Value of a Random Variable: For a discrete random variable X,

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

For a continuous random variable X,

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

5.5.1 Example: Testing Electronic Components

To find the expectation (expected value) $E[X]$ for the random variable X representing the number of defective components in the three tested electronic components, we use the definition of the expectation for a discrete random variable:

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

Given the probability mass function (pmf) for X in Table 5.2, the calculations for the $E[X]$ are shown in the following table.

x	$P(X = x)$	$x \cdot P(X = x)$
0	$\frac{1}{8}$	0
1	$\frac{3}{8}$	$1 \cdot \frac{3}{8} = \frac{3}{8}$
2	$\frac{3}{8}$	$2 \cdot \frac{3}{8} = \frac{6}{8}$
3	$\frac{1}{8}$	$3 \cdot \frac{1}{8} = \frac{3}{8}$
Total		$\frac{12}{8}$

Using the output in the above table

$$E[X] = \sum_x x \cdot P(X = x) = \frac{12}{8} = 1.5$$

Therefore, the expected number of defective components $E[X]$ is 1.5.

5.5.2 Example: Metal Cylinder Production

The probability density function of the diameter of a metal cylinder (X) is

$$f(x) = 1.5 - 6(x - 50.0)^2 \quad \text{for } 49.5 \leq x \leq 50.5$$

The expectation $E(X)$ is calculated as follows:

$$
\begin{aligned}
E(X) &= \int_{49.5}^{50.5} xf(x)\,dx = \int_{49.5}^{50.5} x\left(1.5 - 6(x-50.0)^2\right)\,dx \\
&= \int_{49.5}^{50.5} 1.5x\,dx - \int_{49.5}^{50.5} 6x(x-50)^2\,dx \\
&= 1.5\left[\frac{x^2}{2}\right]_{49.5}^{50.5} - \int_{49.5}^{50.5} 6x(x-50)^2\,dx \quad [\text{let},\quad u = x-50] \\
&= 1.5\left(\frac{50.5^2}{2} - \frac{49.5^2}{2}\right) - 6\int_{-0.5}^{0.5}(u^3 + 50u^2)\,du \\
&= 1.5\left(\frac{2550.25}{2} - \frac{2450.25}{2}\right) - 6\int_{-0.5}^{0.5} u^3\,du - 6\left(50\left[\frac{u^3}{3}\right]_{-0.5}^{0.5}\right) \\
&= 75 - 0 - 6\times 50\times\left(\frac{0.5^3}{3} + \frac{0.5^3}{3}\right) \\
&= 75 - 1.5\times 50 \\
&= 50
\end{aligned}
$$

Hence the expectation of the diameter of a metal cylinder is 50mm.

5.5.3 Exercises

1. Suppose the Laptop repair costs are \$50, \$200, and \$350 with respective probability values of 0.3, 0.2, and 0.5. What is the expected Laptop repair cost?

2. Suppose the daily sales of a small shop are \$100, \$150, and \$250 with respective probability values of 0.4, 0.3, and 0.3. What is the expected daily sales?

3. A game offers prizes of \$10, \$50, and \$100 with respective probability values of 0.6, 0.3, and 0.1. What is the expected prize amount?

4. Consider the waiting times (in minutes) at a bus stop: 5, 10, and 15 with respective probability values of 0.5, 0.3, and 0.2. What is the expected waiting time?

5. The lifetime (in years) of a certain type of light bulb is either 1, 3, or 5 with respective probability values of 0.2, 0.5, and 0.3. What is the expected lifetime of the light bulb?

6. The number of daily website visits for a company is either 200, 500, or 800 with respective probability values of 0.25, 0.5, and 0.25. What is the expected number of daily visits?

7. Let the temperature X in degrees Fahrenheit of a particular chemical reaction with density

$$f_X(x) = \frac{x - 190}{3600} \qquad 220 \le x \le 280.$$

Find the expectation of the temperature.

5.6 The Variance of a Random Variable

Another key summary measure of the distribution of a random variable is the variance, which quantifies the spread or variability in the values that the random variable can take. While the mean or expectation captures the central or average value of the random variable, the variance measures the dispersion or deviation of the random variable around its mean value. Specifically, the variance of a random variable is defined as

$$\text{Var}(X) = E((X - E(X))^2).$$

This means that the variance is the expected value of the squared deviations of the random variable values from the expected value $E(X)$. The variance is always positive, and larger values indicate a greater spread in the distribution of the random variable around the mean. An alternative and often simpler expression for calculating the variance is

$$\begin{aligned}
\text{Var}(X) &= E((X - E(X))^2) \\
&= E(X^2 - 2XE(X) + (E(X))^2) \\
&= E(X^2) - 2E(X)E(X) + (E(X))^2 \\
&= E(X^2) - (E(X))^2.
\end{aligned}$$

Variance: The variance of a random variable X is defined as

$$\text{Var}(X) = E((X - E(X))^2)$$

or equivalently

$$\text{Var}(X) = E(X^2) - (E(X))^2.$$

The variance is a positive measure that indicates the spread of the distribution of the random variable around its mean value. Larger values of the variance suggest that the distribution is more spread out.

It is typical to use the symbol μ to represent the mean or expectation of a random variable, and the symbol σ^2 to represent the variance. The standard deviation, denoted by σ, is the square root of the variance and is often used instead of the variance to describe the spread of the distribution.

Standard Deviation: The standard deviation of a random variable X is defined as the positive square root of the variance. The variance of a random variable is commonly denoted by σ^2, so σ represents the standard deviation.

The concept of variance can be illustrated graphically. Figure 5.7 shows two probability density functions with different mean values but identical variances. The variances are the same because the shape or spread of the density functions around their mean values is the same. In contrast, Figure 5.8 shows two probability density functions with the same mean values but different variances. The density function that is flatter and more spread out has the larger variance.

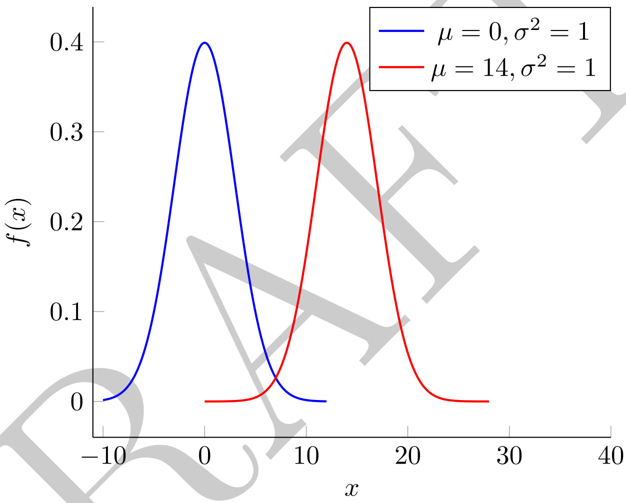


Figure 5.7: Two normal distributions with different means but identical variances.

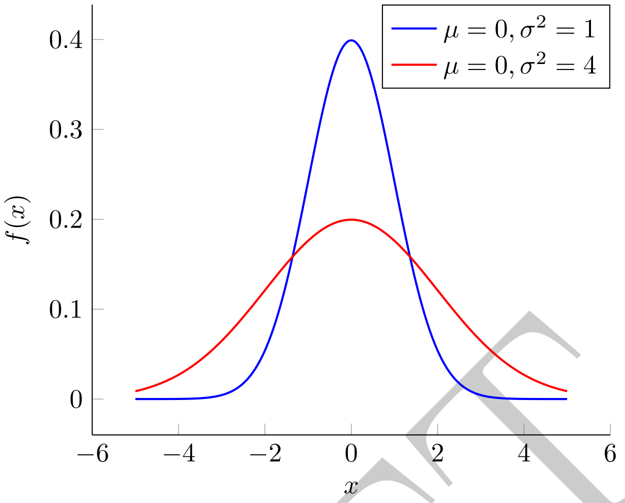


Figure 5.8: Two normal distributions with identical means but different variances.

It is important to note that the standard deviation has the same units as the random variable X, while the variance has units that are squared. For instance, if the random variable X is measured in seconds, then the standard deviation will also be in seconds, but the variance will be measured in seconds squared (seconds^2).

Example: Testing Electronic Components

We already found that $E[X] = 1.5$. To find the variance of the random variable X, we need to compute $E[X^2]$:

$$E[X^2] = \sum_x x^2 \cdot P(X = x)$$

Given the probability mass function (pmf) for X:

Table 5.4: Calculating $E[X]$ and $E[X^2]$

x	$P(X = x)$	$x \cdot P(X = x)$	$x^2 \cdot P(X = x)$
0	$\frac{1}{8}$	$0 \cdot \frac{1}{8} = 0$	$0^2 \cdot \frac{1}{8} = 0$
1	$\frac{3}{8}$	$1 \cdot \frac{3}{8} = \frac{3}{8}$	$1^2 \cdot \frac{3}{8} = \frac{3}{8}$
2	$\frac{3}{8}$	$2 \cdot \frac{3}{8} = \frac{6}{8}$	$2^2 \cdot \frac{3}{8} = \frac{12}{8}$
3	$\frac{1}{8}$	$3 \cdot \frac{1}{8} = \frac{3}{8}$	$3^2 \cdot \frac{1}{8} = \frac{9}{8}$
Total		$\frac{12}{8} = 1.5$	$\frac{24}{8} = 3$

Now, we can find the variance:

$$\mathrm{Var}(X) = E[X^2] - (E[X])^2$$

$$\mathrm{Var}(X) = 3 - (1.5)^2$$

$$\mathrm{Var}(X) = 3 - 2.25$$

$$\mathrm{Var}(X) = 0.75$$

Therefore, the variance of X is 0.75.

5.6.1 Example: Metal Cylinder Production

The probability density function of the diameter of a metal cylinder (X) is

$$f(x) = 1.5 - 6(x - 50.0)^2 \quad \text{for } 49.5 \le x \le 50.5$$

and the $E(X) = 50$. To find the variance $V(X)$, we need $E(X^2)$:

$$\begin{aligned}
E(X^2) &= \int_{49.5}^{50.5} x^2 f(x)\,dx = \int_{49.5}^{50.5} x^2 \left(1.5 - 6(x - 50.0)^2\right)\,dx \\
&= \int_{49.5}^{50.5} 1.5x^2\,dx - \int_{49.5}^{50.5} 6x^2(x-50)^2\,dx \\
&= 1.5\left[\frac{x^3}{3}\right]_{49.5}^{50.5} - 6\int_{-0.5}^{0.5} (u+50)^2 u^2\,du \quad [\text{let,} \quad u = x - 50] \\
&= 1.5\left(\frac{50.5^3}{3} - \frac{49.5^3}{3}\right) - 6\int_{-0.5}^{0.5} (u^4 + 100u^2 + 2500)\,du \\
&= 2500.05 \quad [\text{after simplification}]
\end{aligned}$$

Combining both parts:

$$E(X^2) = 3862.5 - 15049.9875 = 50$$

$$V(X) = E(X^2) - (E(X))^2 = 2500.05 - 2500 = 0.05$$

Thus, the variance $V(X) = 0.05$ and the standard deviation $sd(X) = \sqrt{0.05} = 0.2236$.

Problem 5.8. *Consider a random variable X representing the number of heads in three tosses of a fair coin. The possible values of X are 0, 1, 2, and 3. The pmf of X is given by:*

$$P(X = x) = \binom{3}{x}\left(\frac{1}{2}\right)^3$$

Find the expected value and variance of X.

5.6.2 Chebyshev's Inequality

Chebyshev's Inequality is a powerful tool for understanding the spread of data in scenarios where the distribution is unknown. It provides a way to make probabilistic statements about deviations from the mean, which is particularly useful in data science applications such as quality control and salary analysis.

Chebyshev's Inequality: Let X be a random variable with mean μ and variance σ^2. Chebyshev's Inequality states that for any $k > 0$,

$$P\left(|X-\mu| \geq k\sigma\right) \leq \frac{1}{k^2}$$

or equivalently,

$$P\left(\mu - k\sigma \leq X \leq \mu + k\sigma\right) \geq 1 - \frac{1}{k^2}.$$

This inequality indicates that the probability of a random variable deviating from its mean by more than k standard deviations is at most $\frac{1}{k^2}$.

5.6.3 Example: Blood Pressure Measurement

Consider a study on blood pressure measurements where the systolic blood pressure X of patients is known to have a mean of 120 mmHg and a standard deviation of 15 mmHg. We want to determine the range within which at least 90% of the measurements fall, according to Chebyshev's Inequality:

$$P\left(120 - k \times 15 \leq X \leq 120 + k \times 15\right) \geq 1 - \frac{1}{k^2}.$$

where k be the number of standard deviations from the mean. We want:

$$1 - \frac{1}{k^2} \geq 0.90$$

Solving for k:

$$\frac{1}{k^2} \leq 0.10$$

or,

$$k^2 \geq \frac{1}{0.10} = 10$$

$$\therefore k \geq \sqrt{10} \approx 3.16$$

So, at least 90% of systolic blood pressure measurements should fall within:

$$120 \pm 3.16 \times 15 = 120 \pm 47.4$$

In other words, the blood pressure measurements are expected to be within the range of:

5.6.4 Example: Employee Salaries

Consider a company where the average salary is \$60,000 with a standard deviation of \$5,000. If the company wants to guarantee that at least 80% of employees' salaries are within a certain range of the mean salary, we can use Chebyshev's Inequality to estimate this range.

$$P\left(60000 - k \times 5000 \leq X \leq 60000 + k \times 60000\right) \geq 1 - \frac{1}{k^2}.$$

To ensure at least 80% of salaries are within this range, we want:

$$1 - \frac{1}{k^2} \geq 0.80$$

$$\frac{1}{k^2} \leq 0.20 \implies k^2 \geq \frac{1}{0.20} = 5$$

$$k \geq \sqrt{5} \approx 2.24$$

Thus, at least 80% of salaries should fall within:

$$60,000 \pm 2.24 \times 5,000 \quad \text{or,} \quad 60,000 \pm 11,200$$

5.6.5 Quantiles of Random Variables

Quantiles are useful summary measures that provide insight into the spread or variability of a random variable's distribution. The p-th quantile of a random variable X, which has a cumulative distribution function $F(x)$, is the value x that satisfies

$$F(x) = p$$

meaning that there is a probability p that the random variable is less than the p-th quantile. The probability p is often expressed as a percentage, and the corresponding quantiles are known as percentiles. For instance, the 70th percentile is the value x for which $F(x) = 0.70$. It is important to note that the 50th percentile of a distribution is also known as the median.

Quantiles: The p-th quantile of a random variable X with a cumulative distribution function $F(x)$ is the value x such that

$$F(x) = p$$

This is also known as the $p \times 100$-th percentile of the random variable. The probability p signifies the chance that the random variable takes on a value less than the p-th quantile.

To understand the spread of a distribution, one can compute its quartiles. The upper quartile is the 75th percentile, and the lower quartile is the 25th percentile. Together with the median, these quartiles divide the range of the random variable into four equal parts, each with a probability of 0.25.

The interquartile range, which is the distance between the upper and lower quartiles as depicted in Figure 2.55, serves as an indicator of distribution spread similar to variance. A larger interquartile range suggests that the distribution of the random variable is more spread out.

Quartiles and Interquartile Range: The upper quartile of a distribution is the 75th percentile, and the lower quartile is the 25th percentile. The interquartile range, defined as the distance between these two quartiles, provides a measure of distribution spread analogous to variance.

5.6.6 Example: Metal Cylinder Production

The cumulative distribution function (cdf) for the diameters of the metal cylinders is given by

$$F(x) = 1.5x - 2(x - 50.0)^3 - 74.5$$

for $49.5 \leq x \leq 50.5$.

The upper quartile (Q_3) is found at the value of x where

$$F(x) = 0.75.$$

That is,

$$1.5x - 2(x - 50.0)^3 - 74.5 = 0.75$$

This equation can be solved numerically to find the precise value of x which corresponds to $Q_3 = 50.17$ mm.

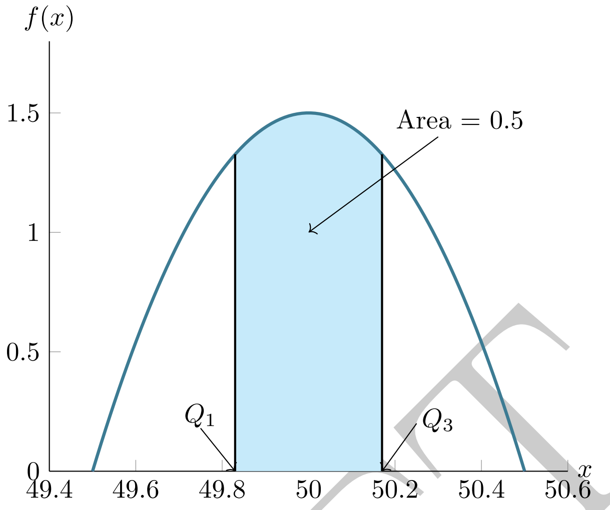


Figure 5.9: Interquartile range for metal cylinder diameters.

The lower quartile (Q_1) is the value where

$$F(x) = 0.25$$

resulting in $Q_1 = 49.83$ mm. Consequently, the interquartile range is calculated as

$$50.17 - 49.83 = 0.34 \text{ mm}$$

indicating that half of the cylinders will have diameters between 49.83 mm and 50.17 mm, as illustrated in Figure 5.9.

5.6.7 Exercises

1. Consider the Laptop repair costs discussed in Exercise 1 Calculate the variance and standard deviation of the number of copying machines in use at a particular moment.

2. A random variable X takes values between 4 and 6 with a probability density function

$$f(x) = \frac{1}{x \ln(1.5)} \quad \text{for } 4 \leq x \leq 6.$$

   (a) What is the variance of this random variable?
   (b) What is the standard deviation of this random variable?
   (c) Find the upper and lower quartiles of this random variable.

(d) What is the interquartile range?

3. A random variable X represents the time (in hours) until failure of a certain machine part, which is uniformly distributed between 100 and 200 hours.

$$f(x) = \frac{1}{100} \quad \text{for } 100 \leq x \leq 200.$$

(a) What is the variance of this random variable?

(b) What is the standard deviation of this random variable?

(c) Find the upper and lower quartiles of this random variable.

(d) What is the interquartile range?

4. Consider a random variable Y representing the strength of a material, which follows a normal distribution with mean 500 MPa and standard deviation 50 MPa.

(a) What is the probability that the strength is between 450 MPa and 550 MPa?

(b) What is the 95th percentile of the strength?

(c) Calculate the variance of the strength.

(d) What proportion of material samples have a strength greater than 600 MPa?

5. A random variable Z represents the systolic blood pressure (in mmHg) of a population, which is uniformly distributed between 90 and 140 mmHg.

$$f(z) = \frac{1}{50} \quad \text{for } 90 \leq z \leq 140.$$

(a) What is the variance of this random variable?

(b) What is the standard deviation of this random variable?

(c) Find the upper and lower quartiles of this random variable.

(d) What is the interquartile range?

6. A researcher is studying the cholesterol levels of a population of adults. The cholesterol levels are known to have a mean of 200 mg/dL and a standard deviation of 25 mg/dL.

(a) Using Chebyshev's Inequality, determine the minimum percentage of adults whose cholesterol levels are within 50 mg/dL of the mean.

(b) To guarantee that at least 85% of the population has cholesterol levels within a certain number of standard deviations from the mean, how many standard deviations from the mean are required?

(c) If a randomly selected adult has a cholesterol level of 250 mg/dL, what is the maximum probability that this level deviates from the mean by at least 50 mg/dL according to Chebyshev's Inequality?

7. The systolic blood pressure of a certain population is normally distributed with a mean of 120 mmHg and a standard deviation of 15 mmHg.

   (a) What is the probability that a randomly selected person has a systolic blood pressure less than 110 mmHg?
   (b) What is the probability that a randomly selected person has a systolic blood pressure between 110 mmHg and 130 mmHg?
   (c) Find the 95th percentile of the systolic blood pressure distribution.

5.7 Essential Generating Functions

In probability theory and statistics, generating functions are a powerful tool used to analyze and manipulate probability distributions. There are three main types of generating functions:

- Moment Generating Function (MGF): Used for both discrete and continuous random variables.
- Probability Generating Function (PGF): Used for discrete random variables.
- Characteristic Function (CF): It complements other generating functions, such as the MGF and PGF, and is closely related to the Fourier transform.

5.7.1 Moment Generating Function

The Moment Generating Function (MGF) is a powerful tool in probability theory and statistics used to summarize the properties of a probability distribution. It can help us find moments (for example, mean, variance), combine variables, and understand the distribution better. The Moment Generating Function of a random variable X is defined as:

$$M_X(t) = E[e^{tX}].$$

5.7.2 Key Properties of MGF

- **Higher Order Moments**:
  - The first moment (**mean**) is found by taking the first derivative of the MGF and evaluating it at $t = 0$:

$$\mu = \mathbb{E}[X] = M'_X(0)$$

  - The second moment is found by taking the second derivative of the MGF and evaluating it at $t = 0$:

    $$\mathbb{E}[X^2] = M_X''(0)$$

  - The nth moment is found by taking the nth derivative of the MGF and evaluating it at $t = 0$:

    $$\mathbb{E}[X^n] = \left.\frac{d^n M_X(t)}{dt^n}\right|_{t=0}$$

    Evaluate the Derivative at $t = 0$ After differentiating the MGF the required number of times, substitute $t = 0$ to obtain the moment.

  - The **variance** can be found by taking the second derivative of the MGF, evaluating at $t = 0$, and then using it with the mean:

    $$\mathrm{Var}(X) = M_X''(0) - (M_X'(0))^2$$

- **Combining Variables**:

  - If X and Y are independent random variables, the MGF of their sum $X + Y$ is the product of their individual MGFs:

    $$M_{X+Y}(t) = M_X(t) \cdot M_Y(t)$$

Problem 5.9 (**Discrete Random Variable**). *Consider a discrete random variable X that takes values 0 and 1 with the following probabilities:*

- $P(X = 0) = p$
- $P(X = 1) = 1 - p$

where $0 \leq p \leq 1$. Find the moment generating function of X, and hence find mean and variance.

Solution

The Moment Generating Function $M_X(t)$ is defined as:

$$M_X(t) = \mathbb{E}[e^{tX}]$$

To compute this, we use:

$$M_X(t) = e^{t\cdot 0} \cdot P(X = 0) + e^{t\cdot 1} \cdot P(X = 1)$$

Simplify it:

$$M_X(t) = e^0 \cdot p + e^t \cdot (1 - p)$$

$$M_X(t) = p + (1 - p)e^t$$

- **Mean (First Moment)**:
  Differentiate $M_X(t)$ once and evaluate at $t = 0$:

  $$M_X'(t) = \frac{d}{dt}[p + (1-p)e^t] = (1-p)e^t$$

  Evaluating at $t = 0$:

  $$M_X'(0) = (1-p)e^0 = 1-p$$

  Thus, the mean $\mathbb{E}[X] = 1 - p$.

- **Variance**:
  Compute $M_X''(t)$:

  $$M_X''(t) = \frac{d^2}{dt^2}[p + (1-p)e^t] = (1-p)e^t$$

  Evaluating at $t = 0$:

  $$M_X''(0) = (1-p)e^0 = 1-p$$

  Variance is given by:

  $$\mathrm{Var}(X) = M_X''(0) - (M_X'(0))^2$$

  $$\mathrm{Var}(X) = (1-p) - (1-p)^2$$

  $$\mathrm{Var}(X) = (1-p) - (1 - 2p + p^2)$$

  $$\mathrm{Var}(X) = p - p^2$$

  $$\mathrm{Var}(X) = p(1-p)$$

Problem 5.10 (Continuous Random Variable). *The pdf of X is given by:*

$$f_X(x) = \begin{cases} 1 & \text{if } 0 \le x \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

This is a density of the uniform distribution on $[0,1]$. Find the moment generating function of X, and hence find the mean and variance.

Solution

For this density the moment generating function is

$$M_X(t) = E\left[e^{tX}\right] = \int_0^1 e^{tx} \cdot 1\,dx = \left[\frac{e^{tx}}{t}\right]_0^1 = \frac{e^t - 1}{t}, \quad \text{for } t \neq 0.$$

For $t = 0$, $M_X(0) = 1$ (since $M_X(t)$ is always 1 at $t = 0$ for any distribution).

- The first derivative of $M_X(t)$ with respect to t is:

  $$M_X'(t) = \frac{d}{dt}\left(\frac{e^t - 1}{t}\right).$$

  Applying the quotient rule:

  $$M_X'(t) = \frac{t \cdot e^t - (e^t - 1)}{t^2} = \frac{te^t - e^t + 1}{t^2} = \frac{e^t(t-1) + 1}{t^2}.$$

  Evaluating at $t = 0$:

  $$M_X'(0) = \lim_{t \to 0} \frac{e^t(t-1) + 1}{t^2}.$$

  Using L'Hôpital's rule, we find:

  $$M_X'(0) = \frac{0 \cdot 1 + 1}{2 \cdot 0} = \frac{1}{2}.$$

  So, $\mathbb{E}[X] = \frac{1}{2}$.

- The second derivative of $M_X(t)$ with respect to t is:

  $$M_X''(t) = \frac{d}{dt}\left(\frac{e^t(t-1) + 1}{t^2}\right).$$

  After some calculation (similar to the first derivative), we find:

  $$M_X''(t) = \frac{e^t(t^2 - 2t + 2) - 2(t-1)e^t}{t^3}.$$

  Evaluating at $t = 0$:

  $$M_X''(0) = \lim_{t \to 0} \frac{e^t(t^2 - 2t + 2) - 2(t-1)e^t}{t^3} = \frac{1}{3}.$$

  So, $\mathbb{E}[X^2] = \frac{1}{3}$.

Thus, the first two moments for a uniform random variable X on $[0, 1]$ are:

$$\mathbb{E}[X] = \frac{1}{2} \text{ and } \mathbb{E}[X^2] = \frac{1}{3}.$$

5.7.3 Probability Generating Function (PGF)

A Probability Generating Function (PGF) is a tool used in probability theory and statistics to analyze discrete random variables. Specifically, it's used to encode the probability distribution of a random variable into a generating function, which can simplify the calculation of various properties of the distribution.

Let X be a discrete random variable with probability mass function $p_X(k) = P(X = k)$ for $k = 0, 1, 2, \ldots$. The PGF of X, denoted by $G_X(s)$, is defined as:

$$G_X(s) = E[s^X] = \sum_{k=0}^{\infty} p_X(k)s^k,$$

where s is a real or complex number for which the series converges.

5.7.4 Key Properties of PGF

Generating Probabilities

The probability mass function $p_X(k)$ can be obtained from the PGF using the k-th derivative:

$$p_X(k) = \frac{G_X^{(k)}(0)}{k!}.$$

Moments

The first derivative of the PGF is:

$$G'_X(s) = \frac{d}{ds}G_X(s) = E[Xs^{X-1}].$$

Evaluating at $s = 1$:

$$E[X] = G'_X(1).$$

The second derivative of the PGF is:

$$G''_X(s) = \frac{d^2}{ds^2}G_X(s) = \frac{d}{ds}\left(E[Xs^{X-1}]\right) = E[(X-1)Xs^{X-2}].$$

Evaluating at $s = 1$:

$$G''_X(1) = E[(X-1)X] = E[X^2 - X] = E[X^2] - E[X].$$

So,

$$E[X^2] = G''_X(1) + E[X] = G''_X(1) + G'_X(1).$$

This result combines the second derivative of the PGF with the first derivative, as shown in the derivation above. The n-th Derivative of the PGF is

$$G_X^{(n)}(s) = \frac{d^n}{ds^n}G_X(s).$$

Hence, the n-th moment of X can be derived from the PGF by differentiating:

$$E[X^n] = G_X^{(n)}(1).$$

Mean and Variance

The mean and variance of X can be calculated as follows:

$$E[X] = G'_X(1),$$

$$\text{Var}(X) = G''_X(1) + G'_X(1) - (G'_X(1))^2.$$

Theorem 5.1. *If X and Y are independent random variables, then the PGF of their sum is:*

$$G_{X+Y}(s) = G_X(s) \cdot G_Y(s).$$

Problem 5.11. *Consider a random variable X with parameter λ. The probability mass function of X is:*

$$p_X(k) = \frac{\lambda^k e^{-\lambda}}{k!} \quad k = 0, 2, 3, \ldots.$$

Find PGF, and hence mean and variance.

Solution

The PGF $G_X(s)$ is defined as:

$$G_X(s) = E[s^X] = \sum_{k=0}^{\infty} p_X(k) s^k.$$

Substituting the PMF $p_X(k)$:

$$G_X(s) = \sum_{k=0}^{\infty} \frac{\lambda^k e^{-\lambda}}{k!} s^k.$$

We can factor out $e^{-\lambda}$ as it is constant with respect to the sum:

$$G_X(s) = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda s)^k}{k!}.$$

Recognize that the sum is the Taylor series expansion of $e^{\lambda s}$:

$$\sum_{k=0}^{\infty} \frac{(\lambda s)^k}{k!} = e^{\lambda s}.$$

Therefore, the PGF is:

$$G_X(s) = e^{-\lambda} \cdot e^{\lambda s} = e^{\lambda(s-1)}.$$

Finding the Mean

The mean $\mathbb{E}[X]$ can be obtained by differentiating the PGF and evaluating at $s = 1$:

$$\mathbb{E}[X] = G'_X(1).$$

First, compute the derivative of $G_X(s)$:

$$G_X(s) = e^{\lambda(s-1)}.$$

$$G'_X(s) = \lambda e^{\lambda(s-1)}.$$

Evaluating at $s = 1$:

$$G'_X(1) = \lambda e^{\lambda(1-1)} = \lambda e^0 = \lambda.$$

Thus, the mean $\mathbb{E}[X]$ is λ.

Finding the Variance

The variance can be found using:

$$\mathrm{Var}(X) = G''_X(1) + G'_X(1) - (G'_X(1))^2.$$

Compute the second derivative of $G_X(s)$:

$$G''_X(s) = \lambda^2 e^{\lambda(s-1)}.$$

Evaluating at $s = 1$:

$$G''_X(1) = \lambda^2 e^{\lambda(1-1)} = \lambda^2 e^0 = \lambda^2.$$

Now, calculate the variance:

$$\mathrm{Var}(X) = G''_X(1) + G'_X(1) - (G'_X(1))^2.$$

Substitute the values:

$$\mathrm{Var}(X) = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

Thus, the variance $\mathrm{Var}(X)$ is λ.

Applications

PGFs are used in various fields including:

- **Queueing Theory**: To analyze the number of customers in a queue.
- **Reliability Engineering**: To model system lifetimes.
- **Genetics**: To study inheritance patterns.

The PGF is a compact and powerful tool for handling problems involving sums of random variables and their distributions.

5.7.5 Characteristic Function (CF)

The characteristic function (CF) of a random variable X is a fundamental tool in probability theory, and it is closely related to the moment generating function. The characteristic function provides an alternative way to describe the distribution of X, and it is particularly useful in the study of sums of independent random variables. It is important in data science, particularly in areas related to probability theory, statistical inference, and stochastic processes.

The characteristic function $\varphi_X(t)$ of a random variable X is defined as:

$$\varphi_X(t) = E\left[e^{itX}\right],$$

where i is the imaginary unit ($i^2 = -1$) and t is a real number.

5.7.6 Key Properties of Characteristic Functions

Let X be a random variable with characteristic function $\phi_X(t)$.

1. **Existence:** The characteristic function always exists and is well-defined for all real t.
2. **Normalization:**
$$\phi_X(0) = \mathbb{E}[e^{i\cdot 0\cdot X}] = \mathbb{E}[1] = 1.$$
3. **Uniqueness:** The characteristic function uniquely determines the distribution of X. If two random variables have the same characteristic function, they have the same distribution.
4. **Addition of Independent Random Variables:** If X and Y are independent, then:
$$\phi_{X+Y}(t) = \phi_X(t) \cdot \phi_Y(t).$$

5. **Moment Generating Function Relationship:** The moment generating function $M_X(t)$ is related to the characteristic function by:

$$M_X(t) = \mathbb{E}[e^{tX}] = \phi_X(-it).$$

6. **Inverse Relationship:** The distribution function $F_X(x)$ can be recovered from the characteristic function using the inverse Fourier transform:

$$F_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \phi_X(t) e^{-itx}\, dt.$$

7. **Derivatives and Moments:** The n-th moment of X, if it exists, is given by:

$$\mathbb{E}[X^n] = i^{-n} \left. \frac{d^n \phi_X(t)}{dt^n} \right|_{t=0}.$$

The Characteristic Function (CF) and the Moment Generating Function (MGF) are both tools used in probability theory to describe the distribution of random variables. While they share some similarities, they also have key differences. A detailed comparisons are presented in Table 5.5.

Table 5.5: Comparison between MGF and CF

Feature	**MGF** $M_X(t) = \mathbb{E}[e^{tX}]$	**CF** $\varphi_X(t) = \mathbb{E}[e^{itX}]$
Existence	Not guaranteed (finite only if $\mathbb{E}[e^{tX}]$ exists for all t)	Always exists (since $\lvert e^{itX} \rvert = 1$)
Range of t	Real t	Real t, but involves imaginary unit i
Moments	Differentiation gives moments directly	Differentiation gives moments with i adjustment
Uniqueness	Uniquely determines distribution if it exists	Uniquely determines distribution
Fourier Transform	No direct connection	Essentially the Fourier transform of the PDF
Application	Used for finding moments and cumulants, proving Central Limit Theorem	Used in distributional analysis, sums of random variables, proving Central Limit Theorem

Problem 5.12 (Discrete Random Variable). *Consider the Problem 5.9, the pmf of discrete random variable X is*

- $P(X = 0) = p$

- $P(X = 1) = 1 - p$

where $0 \leq p \leq 1$. *Find the characteristic function of* X, *and hence find mean and variance.*

Solution

1. Characteristic Function

The characteristic function $\varphi_X(t)$ of a discrete random variable X is defined as:

$$\varphi_X(t) = E[e^{itX}]$$

where E denotes the expectation and i is the imaginary unit.

For the given pmf:

$$\varphi_X(t) = E[e^{itX}] = \sum_x e^{itx} P(X = x)$$

Substituting the values for X:

$$\varphi_X(t) = e^{it\cdot 0} \cdot P(X = 0) + e^{it\cdot 1} \cdot P(X = 1)$$

$$\varphi_X(t) = e^0 \cdot p + e^{it} \cdot (1 - p)$$

$$\varphi_X(t) = p + (1 - p)e^{it}$$

2. Mean of X

The mean $E[X]$ can be derived from the characteristic function as follows:

$$E[X] = i\, \frac{d}{dt}\varphi_X(t)\bigg|_{t=0}$$

First, compute the derivative of $\varphi_X(t)$:

$$\frac{d}{dt}\varphi_X(t) = \frac{d}{dt}\left(p + (1 - p)e^{it}\right)$$

$$\frac{d}{dt}\varphi_X(t) = (1 - p) \cdot ie^{it}$$

Evaluate at $t = 0$:

$$\frac{d}{dt}\varphi_X(t)\bigg|_{t=0} = (1 - p) \cdot ie^{i\cdot 0} = (1 - p) \cdot i$$

$$E[X] = i \cdot (1 - p) \cdot i = 1 - p$$

3. Variance of X

To find the variance, we first need the second moment $E[X^2]$.

The second moment $E[X^2]$ can be derived from the characteristic function as follows:

$$E[X^2] = -\left.\frac{d^2}{dt^2}\varphi_X(t)\right|_{t=0}$$

Compute the second derivative of $\varphi_X(t)$:

$$\frac{d^2}{dt^2}\varphi_X(t) = \frac{d}{dt}\left((1-p)\cdot ie^{it}\right)$$

$$\frac{d^2}{dt^2}\varphi_X(t) = (1-p)\cdot i^2 e^{it} = -(1-p)e^{it}$$

Evaluate at $t = 0$:

$$\left.\frac{d^2}{dt^2}\varphi_X(t)\right|_{t=0} = -(1-p)e^{i\cdot 0} = -(1-p)$$

$$E[X^2] = -\left(-(1-p)\right) = 1-p$$

The variance $\mathrm{Var}(X)$ is given by:

$$\begin{aligned}
\mathrm{Var}(X) &= E[X^2] - (E[X])^2 \\
&= (1-p) - (1-p)^2 \\
&= (1-p) - (1-2p+p^2) \\
&= 1-p-1+2p-p^2 \\
&= p-p^2 \\
&= p(1-p)
\end{aligned}$$

Hence,

- Characteristic function: $\varphi_X(t) = p + (1-p)e^{it}$
- Mean: $E[X] = 1-p$
- Variance: $\mathrm{Var}(X) = p(1-p)$

Problem 5.13 (**Continuous Random Variable**). *Consider the Problem 5.10, the pdf of X is given by:*

$$f_X(x) = \begin{cases} 1 & \text{if } 0 \le x \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

Find the characteristic function of X, and hence find the mean and variance.

Solution

Let's find the characteristic function of a uniform random variable X on the interval $[0, 1]$.

The characteristic function is:

$$\varphi_X(t) = E\left[e^{itX}\right] = \int_{-\infty}^{\infty} e^{itx} f_X(x)\, dx,$$

where $f_X(x)$ is the pdf of X.

For the uniform distribution on $[0, 1]$, the pdf is:

$$f_X(x) = \begin{cases} 1 & \text{if } 0 \le x \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

Thus, the characteristic function becomes:

$$\varphi_X(t) = \int_0^1 e^{itx}\, dx.$$

The integral is straightforward to evaluate:

$$\varphi_X(t) = \int_0^1 e^{itx}\, dx = \left[\frac{e^{itx}}{it}\right]_0^1 = \frac{e^{it} - 1}{it}.$$

The characteristic function can be simplified as:

$$\varphi_X(t) = \frac{e^{it} - 1}{it}.$$

This can also be written as:

$$\varphi_X(t) = \frac{1}{t}\left(\frac{\sin(t)}{t} + i\frac{\cos(t) - 1}{t}\right), \quad \text{for } t \neq 0.$$

For $t = 0$, $\varphi_X(0) = 1$, which is consistent since the characteristic function always equals 1 at $t = 0$.

Mean

The mean $E[X]$ is:

$$E[X] = i\, \frac{d}{dt}\varphi_X(t)\bigg|_{t=0}$$

Compute the derivative of $\varphi_X(t)$:

$$\frac{d}{dt}\varphi_X(t) = \frac{d}{dt}\left(\frac{e^{it} - 1}{it}\right)$$

$$\frac{d}{dt}\varphi_X(t) = \frac{(it \cdot ie^{it}) - (e^{it} - 1)}{(it)^2}$$

$$\frac{d}{dt}\varphi_X(t) = \frac{-te^{it} + 1 - e^{it}}{-t^2}$$

$$\frac{d}{dt}\varphi_X(t) = \frac{1 - e^{it}(t+1)}{t^2}$$

Evaluate at $t = 0$:

$$\left.\frac{d}{dt}\varphi_X(t)\right|_{t=0} = \frac{1-1}{0^2} = 0$$

$$E[X] = 0$$

Variance

To find the variance, we first need the second moment $E[X^2]$:

$$E[X^2] = -\left.\frac{d^2}{dt^2}\varphi_X(t)\right|_{t=0}$$

Compute the second derivative:

$$\frac{d^2}{dt^2}\varphi_X(t) = \frac{d}{dt}\left(\frac{1 - e^{it}(t+1)}{t^2}\right)$$

$$\frac{d^2}{dt^2}\varphi_X(t) = \frac{t^2 \cdot (ie^{it}(t+1)) - (1 - e^{it}(t+1)) \cdot 2t}{t^4}$$

Evaluate at $t = 0$:

$$\left.\frac{d^2}{dt^2}\varphi_X(t)\right|_{t=0} = 2$$

$$E[X^2] = 2$$

The variance is:

$$\mathrm{Var}(X) = E[X^2] - (E[X])^2$$

$$\mathrm{Var}(X) = 2 - 0^2$$

$$\mathrm{Var}(X) = 2$$

Hence,

- Characteristic function: $\varphi_X(t) = \frac{e^{it}-1}{it}$
- Mean: $E[X] = 0$
- Variance: $\mathrm{Var}(X) = 2$

5.7.7 Exercises

1. Consider a discrete random variable Y with the following probability mass function (pmf):

$$P(Y=k)=\begin{cases}\frac{1}{2} & \text{for } k=1,\\ \frac{1}{2} & \text{for } k=2.\end{cases}$$

   (a) Find the moment generating function $M_Y(t)$ of Y.

   (b) Using the moment generating function, find the mean and variance of Y.

2. Let Z be a continuous random variable with the probability density function (pdf):

$$f_Z(z)=\begin{cases}\frac{2z}{\theta^2} & \text{if } 0\le z\le\theta,\\ 0 & \text{otherwise.}\end{cases}$$

   where $\theta>0$ is a parameter.

   (a) Find the moment generating function $M_Z(t)$ of Z.

   (b) Find the characteristic function $\varphi_Z(t)$ of Z.

   (c) Using the moment generating function, determine the mean and variance of Z.

3. Let X be a binomial random variable with parameters n and p, where n is the number of trials and p is the probability of success in each trial. The probability mass function of X is:

$$P(X=k)=\binom{n}{k}p^k(1-p)^{n-k},\quad k=0,1,\dots,n.$$

   (a) Find the moment generating function $M_X(t)$ of X.

   (b) Using the moment generating function, find the mean and variance of X.

4. Consider a discrete random variable X that takes non-negative integer values with the following probability mass function:

$$P(X=k)=\frac{3^k}{\sum_{i=0}^{\infty}3^i}\text{ for } k=0,1,2,\dots$$

   (a) Find the probability generating function (PGF) $G_X(s)$ of the random variable X.

   (b) Use the PGF to determine $\mathbb{E}[X]$ and $\mathrm{Var}(X)$.

   (c) Verify your results using the properties of the PGF.

5. Consider a continuous random variable W uniformly distributed over the interval $[0, a]$. The probability density function is:

$$f_W(w) = \begin{cases} \frac{1}{a} & \text{if } 0 \leq w \leq a, \\ 0 & \text{otherwise.} \end{cases}$$

   (a) Find the characteristic function $\varphi_W(t)$ of W.
   (b) Find the moment generating function $M_W(t)$ of W.
   (c) Use the moment generating function to find the mean and variance of W.

6. Let V be an exponential random variable with rate parameter λ. The probability density function is:

$$f_V(v) = \begin{cases} \lambda e^{-\lambda v} & \text{if } v \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

   (a) Find the moment generating function $M_V(t)$ of V.
   (b) Find the characteristic function $\varphi_V(t)$ of V.
   (c) Using the moment generating function, determine the mean and variance of V.

5.8 Jointly Distributed Random Variables

Jointly distributed random variables are crucial in data science as they allow for the modeling and analysis of relationships between multiple variables simultaneously. Understanding these relationships is essential for predicting outcomes, identifying correlations, and constructing probabilistic models that capture real-world complexities. Joint distributions enable informed decisions based on the combined behavior of multiple variables, which is vital for developing accurate and robust predictive models.

In probability theory, two or more random variables are jointly distributed if there is a joint probability distribution describing their behavior. For two random variables X and Y, the joint probability distribution provides the probability that X takes a specific value x and Y takes a specific value y simultaneously.

5.8.1 Joint Probability Mass Function (pmf)

For discrete random variables X and Y, the joint probability mass function $p_{X,Y}(x, y)$ is defined as:

$$p_{X,Y}(x, y) = P(X = x, Y = y).$$

The joint probability mass function must satisfy the condition:

$$\sum_{x} \sum_{y} p_{X,Y}(x, y) = 1.$$

The joint cumulative distribution function is defined as:

$$F(x, y) = P(X \leq x, Y \leq y).$$

For discrete random variables:

$$F(x, y) = \sum_{X \leq x} \sum_{Y \leq y} p_{X,Y}(x, y).$$

If $p_{ij} = P(X = i, Y = j)$, then

$$F(x, y) = P(X \leq x, Y \leq y) = \sum_{i=1}^{x} \sum_{j=1}^{y} p_{ij}$$

5.8.2 Example: Computer Maintenance

A company managing maintenance services for computer servers is interested in optimizing the scheduling of its technicians. Specifically, the company needs to understand how long a technician spends on-site, which primarily depends on the number of servers requiring maintenance.

Let the random variable X denote the maintenance time in hours at a location, taking values 1, 2, 3, and 4. Let the random variable Y represent the number of servers at the location, taking values 1, 2, and 3. These two random variables are considered jointly distributed.

The joint probability mass function p_{ij} for these variables is given in the table below:

		Number of Servers (Y)		
		1	2	3
Maintenance Time (X)	1	0.12 0.10	0.08	0.01
	2	0.08	0.15	0.01
	3	0.07	0.21	0.02
	4	0.05	0.13	0.07

For example, the table shows that there is a 0.10 probability that $X = 1$ and $Y = 1$, meaning a randomly selected location has one server that takes one hour to maintain. Similarly, the probability is 0.08 that a location with three servers requires four hours of maintenance. This is a valid probability mass function, as

$$\sum_x \sum_y p_{X,Y}(x,y) = \sum_i \sum_j p_{ij} = 0.10 + 0.07 + \cdots + 0.08 = 1.00$$

The joint cumulative distribution function is defined as:

$$F(x,y) = P(X \leq x, Y \leq y) = \sum_{i=1}^{x} \sum_{j=1}^{y} p_{ij}$$

For instance, the probability that a location has no more than two servers and that the maintenance time does not exceed two hours is:

$$F(2,2) = p_{11} + p_{12} + p_{21} + p_{22} = 0.10 + 0.07 + 0.12 + 0.03 = 0.32$$ wrong data

5.8.3 Joint Probability Density Function (pdf)

For continuous random variables X and Y, the joint probability density function $f_{X,Y}(x,y)$ is defined as:

$$f_{X,Y}(x,y) = \frac{\partial^2}{\partial x \partial y} P(X \leq x, Y \leq y).$$

The joint probability density function must satisfy the condition:

$$\iint_{\text{state space}} f(x,y)\, dx\, dy = 1.$$

The probability that $a \leq X \leq b$ and $c \leq Y \leq d$ is obtained from the joint probability density function as:

$$\int_{x=a}^{b} \int_{y=c}^{d} f(x,y)\, dy\, dx$$

For continuous random variables:

$$F(x,y) = \int_{w=-\infty}^{x} \int_{z=-\infty}^{y} f(w,z)\, dz\, dw$$

5.8.4 Example: Mineral Deposits

To evaluate the economic feasibility of mining in a specific region, a mining company collects ore samples from the site and measures their zinc and iron content. Let the random variable X represent the zinc content, ranging from 0.5 to 1.5, and the random variable Y represent the iron content, ranging from 20.0 to 35.0. Suppose the joint probability density function of X and Y is given by

$$f(x, y) = \frac{39}{400} - \frac{17(x-1)^2}{50} - \frac{(y-25)^2}{10,000}$$

for $0.5 \leq x \leq 1.5$ and $20.0 \leq y \leq 35.0$.

To verify the validity of this joint probability density function, we need to ensure that $f(x, y) \geq 0$ within the defined state space and that

$$\int_{0.5}^{1.5} \int_{20.0}^{35.0} f(x, y)\, dy\, dx = 1.$$

This joint probability density function provides comprehensive information about the joint probabilistic behavior of the random variables X and Y. For instance, the probability that a randomly selected ore sample has a zinc content between 0.8 and 1.0 and an iron content between 25 and 30 is given by

$$\int_{0.8}^{1.0} \int_{25.0}^{30.0} f(x, y)\, dy\, dx,$$

which evaluates to 0.092. Thus, only about 9% of the ore at this location has mineral levels within these specified ranges.

5.8.5 Marginal Distributions

The marginal distributions of X and Y can be obtained from the joint distribution.

For discrete variables:

$$p_X(x) = \sum_y p_{X,Y}(x, y)$$

$$p_Y(y) = \sum_x p_{X,Y}(x, y)$$

For continuous variables:

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

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

Using these marginal distributions, we can easily find the mean of X and Y.

5.8.6 Example: Computer Maintenance

To find the marginal distributions of X and Y for the example **Computer Maintenance**, as discussed in Section 5.8.2, we need to sum the probabilities across rows for X and across columns for Y.

		Number of Servers (Y)			
		1	2	3	$P_X(x)$
Maintenance Time (X)	1	0.12	0.08	0.01	0.21
	2	0.08	0.15	0.01	0.24
	3	0.07	0.21	0.02	0.30
	4	0.05	0.13	0.07	0.25
	$P_Y(y)$	0.32	0.57	0.11	1.00

Table 5.6: Joint probability mass function for server maintenance with marginal distributions.

Mean of X

$$\begin{aligned}\mu_X &= \sum_x x \cdot P_X(x)\\ &= 1 \cdot 0.21 + 2 \cdot 0.24 + 3 \cdot 0.30 + 4 \cdot 0.25\\ &= 2.59\end{aligned}$$

Expected Value of X^2

$$\begin{aligned}E(X^2) &= \sum_x x^2 \cdot P_X(x) = 1^2 \cdot 0.21 + 2^2 \cdot 0.24 + 3^2 \cdot 0.30 + 4^2 \cdot 0.25\\ &= 7.87\end{aligned}$$

Variance of X

$$\text{Var}(X) = E(X^2) - (\mu_X)^2 = 7.87 - (2.59)^2 = 1.1619$$

Standard Deviation of X

$$\sigma_X = \sqrt{\text{Var}(X)} = \sqrt{1.1619} \approx 1.0779$$

Mean of Y

$$\begin{aligned}\mu_Y &= \sum_y y \cdot P_Y(y) = 1 \cdot 0.32 + 2 \cdot 0.57 + 3 \cdot 0.11 \\ &= 1.79\end{aligned}$$

Expected Value of Y^2

$$E(Y^2) = \sum_y y^2 \cdot P_Y(y) = 1^2 \cdot 0.32 + 2^2 \cdot 0.57 + 3^2 \cdot 0.11 = 3.59$$

Variance of Y

$$\text{Var}(Y) = E(Y^2) - (\mu_Y)^2 = 3.59 - (1.79)^2 = 0.3859$$

Standard Deviation of Y

$$\sigma_Y = \sqrt{\text{Var}(Y)} = \sqrt{0.3859} \approx 0.6212$$

5.8.7 Example: Mineral Deposits

We consider the **Mineral Deposits**, as explain in 5.8.4. The marginal probability density function of X, representing the zinc content of the ore, is given by:

$$\begin{aligned}f_X(x) &= \int_{20.0}^{35.0} f(x,y)\,dy \\ &= \int_{20.0}^{35.0} \left(\frac{39}{400} - \frac{17(x-1)^2}{50} - \frac{(y-25)^2}{10,000}\right) dy \\ &= \left[\frac{39y}{400} - \frac{17y(x-1)^2}{50} - \frac{(y-25)^3}{30,000}\right]_{20.0}^{35.0} \\ &= \frac{57}{40} - \frac{51(x-1)^2}{10} \quad \text{for } 0.5 \le x \le 1.5.\end{aligned}$$

So, the expected zinc content $E(X)$ is:

$$\begin{aligned}E(X) &= \int_{0.5}^{1.5} x f_X(x)\,dx \\ &= \int_{0.5}^{1.5} x \left(\frac{57}{40} - \frac{51(x-1)^2}{10}\right) dx \\ &= \frac{57}{40}\int_{0.5}^{1.5} x\,dx - \frac{51}{10}\int_{0.5}^{1.5} x(x-1)^2\,dx \\ &= 1.\end{aligned}$$

Similarly, we can find $E(X^2)$ which is

$$E(X^2) = \int_{0.5}^{1.5} x^2 f_X(x)\,dx = 1.055$$

Therefore, the variance $V(X)$ is

$$V(X) = E(X^2) - (E(X))^2 = 1.055 - (1.00)^2 = 0.055$$

and the standard deviation is

$$\sigma_Y = \sqrt{\mathrm{Var}(Y)} = \sqrt{0.055} \approx 0.2345.$$

The probability that a sample of ore has a zinc content between 0.8 and 1.0 can be determined using the marginal probability density function. This probability is given by:

$$\begin{aligned}
P(0.8 \leq X \leq 1.0) &= \int_{0.8}^{1.0} f_X(x)\,dx \\
&= \int_{0.8}^{1.0} \left(\frac{57}{40} - \frac{51(x-1)^2}{10}\right) dx \\
&= \left[\frac{57x}{40} - \frac{17(x-1)^3}{10}\right]_{0.8}^{1.0} \\
&= [1.425] - [1.1536] \\
&= 0.2714
\end{aligned}$$

Therefore, approximately 27% of the ore has a zinc content within these limits.

The marginal probability density function of Y, the iron content of the ore, is given by:

$$\begin{aligned}
f_Y(y) &= \int_{0.5}^{1.5} f(x,y)\,dx \\
&= \int_{0.5}^{1.5} \left(\frac{39}{400} - \frac{17(x-1)^2}{50} - \frac{(y-25)^2}{10,000}\right) dx \\
&= \left[\frac{39x}{400} - \frac{17(x-1)^3}{150} - \frac{x(y-25)^2}{10,000}\right]_{0.5}^{1.5} \\
&= \left[\frac{39x}{400} - \frac{17(x-1)^3}{150} - \frac{x(y-25)^2}{10,000}\right]_{0.5}^{1.5} \\
&= \frac{83}{1200} - \frac{(y-25)^2}{10,000} \qquad \text{for } 20.0 \leq y \leq 35.0.
\end{aligned}$$

The expected iron content and the standard deviation of the iron content, which are $E(Y) = 27.36$ and $\sigma = 4.27$, respectively.

5.8.8 Conditional Distributions

Conditional distribution refers to the probability distribution of a random variable given the occurrence of another event or condition. It provides insights into how one variable behaves when another variable has a specific value or falls within a certain range. This concept is crucial in various fields such as economics, biology, and machine learning, where relationships between variables are studied under specific conditions or contexts.

Conditional Distributions: The conditional distribution of Y given $X = x$ for discrete variables:

$$p_{Y|X}(y|x) = \frac{p_{X,Y}(x,y)}{p_X(x)}$$

For continuous variables:

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$$

Using these conditional distributions, we can easily find the mean of X given Y and Y given X. Conditional distributions are often used to make predictions, assess risks, and uncover underlying patterns in data that may not be apparent from marginal distributions alone.

5.8.9 Example: Computer Maintenance

To find the conditional mean and variance of X given Y and Y given X, we use the definitions of conditional expectations and variances. Below, we derive these values based on the joint probability mass function provided in the Example 5.8.2.

Conditional Mean and Variance of X given Y

Table 5.7: The conditional distribution of X given $Y = 1$

x	1	2	3	4
$p_{X\|Y}(x\|1)$	0.375	0.250	0.219	0.156

.

For $Y = 1$:

$$E(X|Y = 1) = 1 \cdot 0.375 + 2 \cdot 0.250 + 3 \cdot 0.215 + 4 \cdot 0.156 = 2.1563$$

$$E(X^2|Y = 1) = 1^2 \cdot 0.375 + 2^2 \cdot 0.250 + 3^2 \cdot 0.215 + 4^2 \cdot 0.156 = 5.8438$$

$$\mathrm{Var}(X|Y=1) = 5.8438 - (2.1563)^2 \approx 1.1943$$

Hence, the standard deviation of X given $Y = 1$ is $\sqrt{1.1934} \approx 1.0924$.

Similarly, we can easily find the conditional distribution with its mean, variance, and standard deviation of X given $Y = 2$ and $Y = 3$. We can also find the conditional distribution with its mean, variance, and standard deviation of Y given different values of X.

5.8.10 Example: Mineral Deposits

Given a sample of ore with a zinc content of $X = 0.55$, what can be inferred about its iron content? The information regarding the iron content Y is encapsulated in the conditional probability density function, which is expressed as:

$$f_{Y|X=0.55}(y) = \frac{f(0.55, y)}{f_X(0.55)}$$

Here, the denominator represents the marginal distribution of the zinc content X evaluated at 0.55. Evaluating $f_X(0.55)$:

$$f_X(0.55) = \frac{57}{40} - \frac{51(0.55 - 1.00)^2}{10} = 0.39225$$

Thus, the conditional probability density function becomes:

$$f_{Y|X=0.55}(y) = \frac{f(0.55, y)}{0.39225} = \frac{\frac{39}{400} - \frac{17(0.55-1.00)^2}{50} - \frac{(y-25)^2}{10{,}000}}{0.39225}$$

Simplifying, we get:

$$f_{Y|X=0.55}(y) = 0.073 - \frac{(y - 25)^2}{3922.5}$$

for $20.0 \leq y \leq 35.0$. It can be easily find the conditional expectation of the iron content, which is calculated to be 27.14, and the conditional standard deviation, which is 4.14.

5.8.11 Independence and Covariance

Just as two events A and B are considered independent if they are "unrelated" to each other, two random variables X and Y are deemed independent if the value taken by one random variable is "unrelated" to the value taken by the other. Specifically, in the context of data science, random variables are independent if the distribution of one random variable does not depend on the value taken by the other random variable.

Independent Random Variables

- For discrete random variables, independence means that the joint probability mass function (pmf) can be expressed as the product of their individual pmf's:

$$p_{X,Y}(x, y) = p_X(x) \cdot p_Y(y).$$

- For continuous random variables, independence means that the joint probability density function (pdf) can be expressed as the product of their individual pdf's:

$$f_{X,Y}(x, y) = f_X(x) \cdot f_Y(y).$$

Example:

- Let X be the result of rolling a fair six-sided die, and Y be the result of flipping a fair coin, where X can take values 1 through 6, and Y can take values 0 (tails) and 1 (heads). The events are independent, so:

$$P(X = 3 \text{ and } Y = 1) = P(X = 3) \cdot P(Y = 1) = \frac{1}{6} \cdot \frac{1}{2} = \frac{1}{12}$$

Problem 5.14. *It is known that the ratio of gallium to arsenide does not affect the functioning of gallium-arsenide wafers, which are the main components of microchips. Let X denote the ratio of gallium to arsenide and Y denote the functional wafers retrieved during a 1-hour period. X and Y are independent random variables with the joint density function*

$$f(x, y) = \begin{cases} \frac{x(1+3y^2)}{4}, & 0 < x < 2,\ 0 < y < 1, \\ 0, & \textit{elsewhere.} \end{cases}$$

Show that X and Y are independent random variables.

Solution

To show that X and Y are independent random variables, we need to verify that the joint density function $f(x, y)$ can be factored into the product of the marginal density functions $f_X(x)$ and $f_Y(y)$. Specifically, X and Y are independent if and only if the joint density function $f(x, y)$ can be written as:

$$f(x, y) = f_X(x) \cdot f_Y(y).$$

Marginal density function $f_X(x)$:

To find $f_X(x)$, integrate the joint density function $f(x, y)$ over the possible values of y:

$$f_X(x) = \int_0^1 f(x, y)\, dy.$$

Given the joint density function $f(x, y) = \frac{x(1+3y^2)}{4}$ for $0 < x < 2$ and $0 < y < 1$, compute:

$$\begin{aligned}
f_X(x) &= \int_0^1 \frac{x(1+3y^2)}{4}\, dy \\
&= \frac{x}{4} \int_0^1 (1+3y^2)\, dy \\
&= \frac{x}{4} \left(\int_0^1 1\, dy + \int_0^1 3y^2\, dy \right) \\
&= \frac{x}{4}(1+1) \\
&= \frac{x}{2}.
\end{aligned}$$

So, the marginal density function for X is:

$$f_X(x) = \frac{x}{2}, \quad 0 < x < 2.$$

Marginal density function $f_Y(y)$:

To find $f_Y(y)$, integrate the joint density function $f(x, y)$ over the possible values of x:

$$\begin{aligned}
f_Y(y) &= \int_0^2 f(x, y)\, dx \\
&= \int_0^2 \frac{x(1+3y^2)}{4}\, dx \\
&= \frac{1+3y^2}{4} \int_0^2 x\, dx \\
&= \frac{1+3y^2}{4} \cdot 2 = \frac{1+3y^2}{2}.
\end{aligned}$$

Therefore, the marginal density function for Y is:

$$f_Y(y) = \frac{1+3y^2}{2}, \quad 0 < y < 1.$$

Verify independence

Check if $f(x, y)$ can be written as $f_X(x) \cdot f_Y(y)$:

$$f_X(x) \cdot f_Y(y) = \left(\frac{x}{2}\right) \cdot \left(\frac{1+3y^2}{2}\right) = \frac{x(1+3y^2)}{4}.$$

This matches the given joint density function $f(x, y)$.

Since $f(x, y) = f_X(x) \cdot f_Y(y)$, the random variables X and Y are independent.

Thus, we have shown that X and Y are independent random variables.

5.8.12 Example: Computer Maintenance

Check Independence

To check the independent, we consider the example **Computer Maintenance**, as discussed in Section 5.8.2.

For $X = 1$ and $Y = 1$:

$$P(X = 1, Y = 1) = 0.10 \quad \text{and} \quad P(X = 1) \cdot P(Y = 1) = 0.21 \cdot 0.41 = 0.0861$$

So, they are not independent, We can check others.

To find the covariance of X and Y for the example **Computer Maintenance**, as discussed in Section 5.8.2, we need to find $E(XY)$.

5.8.13 Covariance and Correlation

Covariance

Covariance is essential for understanding and quantifying relationships between variables, which is a cornerstone of many data science techniques and analyses. It measures the joint variability of two random variables.

Covariance: For two random variables X and Y, the covariance is defined as:

$$\text{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])]$$

or equivalently

$$\text{Cov}(X, Y) = E[XY] - E(X)E(Y).$$

If X and Y are independent, then $\text{Cov}(X, Y) = 0$. However, a covariance of zero does not necessarily imply independence.

Correlation

Correlation is a normalized form of covariance that measures the strength and direction of the linear relationship between two random variables. This normalization makes correlation a more interpretable metric, useful for understanding the strength and direction of the relationship between variables. It's widely used in statistical analysis, machine learning, and data visualization to reveal and quantify relationships that might otherwise be obscured by differences in scale or units.

Correlation: For two random variables X and Y, the correlation is defined as:

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

where σ_X and σ_Y are the standard deviations of X and Y, respectively.

The correlation coefficient $\rho_{X,Y}$ ranges from -1 to $+1$. A value of $+1$ implies a perfect positive linear relationship, -1 implies a perfect negative linear relationship, and 0 implies no linear relationship.

5.8.14 Example: Computer Maintenance

Covariance

The covariance of X and Y is defined as

$$\text{Cov}(X, Y) = E[XY] - E(X)E(Y).$$

We already have $E(X) = \mu_X = 2.5$ and $E(Y) = \mu_Y = 3.05$. Now we need to compute $E(XY)$. Therefore, we need to sum the products of x, y, and their corresponding joint probabilities:

$$E(XY) = \sum_x \sum_y x \cdot y \cdot P(X = x, Y = y)$$

Calculating each term:

$$\begin{aligned} E(XY) &= 1 \cdot 1 \cdot 0.12 + 1 \cdot 2 \cdot 0.08 + 1 \cdot 3 \cdot 0.01 \\ &\quad + 2 \cdot 1 \cdot 0.08 + 2 \cdot 2 \cdot 0.15 + 2 \cdot 3 \cdot 0.01 \\ &\quad + 3 \cdot 1 \cdot 0.07 + 3 \cdot 2 \cdot 0.21 + 3 \cdot 3 \cdot 0.02 \\ &\quad + 4 \cdot 1 \cdot 0.05 + 4 \cdot 2 \cdot 0.13 + 4 \cdot 3 \cdot 0.07 \\ &= 4.86 \end{aligned}$$

Therefore, the expected value $E(XY)$ is 4.86.

$$\text{Cov}(X, Y) = E[XY] - E(X)E(Y) = 4.86 - (2.59 \times 1.79) = 0.2239.$$

The positive value of 0.2239 indicates that there is an positive relationship between X and Y. As X increases, Y tends to increases, and vice versa. The magnitude of the covariance gives an idea of the strength of the relationship. However, because covariance is not standardized, it is difficult to assess the strength of the relationship without additional context such as the variances of X and Y.

Correlation

To get a more standardized measure of the relationship between X and Y, we can compute the correlation coefficient:

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

Given the standard deviations $\sigma_X = 1.0779$ and $\sigma_Y = 06212$, we can find the correlation coefficient $\rho_{X,Y}$ using the covariance $\text{Cov}(X,Y) = 0.2239$.

Substitute the given values:

$$\rho_{X,Y} = \frac{0.2239}{1.0779 \times 0.6212} \approx 0.3344$$

The correlation of 0.3344 suggests that there is a tendency for more servers to need maintenance as the maintenance time increases.

3 method to identify independence, in note

5.8.15 Example: Continuous Random Variables

Consider two continuous random variables X and Y with the following joint probability density function:

$$f_{X,Y}(x,y) = \begin{cases} 4xy & \text{if } 0 \leq x \leq 1 \text{ and } 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

We can compute the marginal pdf's by integrating the joint pdf over the range of the other variable.

For the marginal pdf of X:

$$f_X(x) = \int_0^1 4xy\, dy = 2x \quad \text{for } 0 \leq x \leq 1.$$

For the marginal pdf of Y:

$$f_Y(y) = \int_0^1 4xy\, dx = 2y \quad \text{for } 0 \leq y \leq 1.$$

The fact that $f_{X,Y}(x,y) = f_X(x) f_Y(y)$ confirms that the random variables X and Y are independent.

To find the covariance and correlation between the random variables X and Y, we use the given information. Given the joint probability density function $f(x,y)$ and the marginal distributions $f_X(x)$ and $f_Y(y)$, we can calculate these quantities.

Covariance

Given the independence of X and Y (as indicated by $f(x, y) = f_X(x)f_Y(y)$), we have:

$$E[XY] = E[X]E[Y].$$

Then the covariance between X and Y is

$$\mathrm{Cov}(X, Y) = E[XY] - E[X]E[Y] = E[X]E[Y] - E[X]E[Y] = 0.$$

Correlation

Since $\mathrm{Cov}(X, Y) = 0$ (due to independence),

$$\rho_{XY} = \frac{0}{\sigma_X \sigma_Y} = 0.$$

Thus, the correlation coefficient is also zero.

5.8.16 Linear Functions of a Random Variable

We will now explore some properties that will simplify calculating the means and variances of random variables discussed in later chapters. These properties allow us to express expectations using other parameters that are either known or easily computed. The results presented are applicable to both discrete and continuous random variables, though proofs are provided only for the continuous case. We start with a theorem and two corollaries that should be intuitively understandable to the reader.

Theorem 5.2. *If a and b are constants, then*

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

and the variance is

$$\mathit{Var}(aX + b) = a^2\,\mathit{Var}(X).$$

Proof. By the definition of expected value,

$$E(aX + b) = \int_{-\infty}^{\infty} (ax + b) f(x)\, dx.$$

This can be rewritten as

$$E(aX + b) = a\int_{-\infty}^{\infty} x f(x)\, dx + b\int_{-\infty}^{\infty} f(x)\, dx.$$

The first integral on the right is $E(X)$ and the second integral equals 1. Therefore, we have

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

Compute Var$(aX+b)$**:**

$$\begin{aligned}
\text{Var}(aX+b) &= E[(aX+b-E(aX+b))^2] \\
&= E[(aX+b-(aE(X)+b))^2] \\
&= E[(aX+b-aE(X)-b)^2] \\
&= E[(aX-aE(X))^2] \\
&= E[a^2(X-E(X))^2] \\
&= a^2E[(X-E(X))^2] \\
& a^2\text{Var}(X).
\end{aligned}$$

Thus, we have shown that:

$$\text{Var}(aX+b) = a^2\text{Var}(X)$$

□

Problem 5.15. *Applying Theorem 5.2 to the continuous random variable*

$$Y = 1.1X + 0.3,$$

rework Example 5.5.2.

For Example 5.5.2 and 5.6.1, it is obtained $E(X) = 50$ and $V(X) = 0.05$. We may use Theorem 5.2 to write

$$\begin{aligned}
E[Y] &= 1.1E[X] + 0.3 \\
&= 1.1 \times 50 - 4.5 \\
&= 50.5
\end{aligned}$$

and

$$\text{Var}(Y) = (1.1)^2\text{Var}(X) = (1.1)^2 \times 0.05 = 0.0605$$

Problem 5.16. *Suppose that a temperature has a mean of* $110°F$ *and a standard deviation of* $2.2°F$. *The conversion formula from Fahrenheit to Centigrade is given by:*

$$F = \frac{9C}{5} + 32$$

where F *is the temperature in Fahrenheit and* C *is the temperature in Centigrade. What are the mean and the standard deviation in degrees Centigrade?*

Solution

To find the mean temperature in Centigrade, we use:

$$C_{\text{mean}} = \frac{5}{9}(F_{\text{mean}} - 32)$$

Substitute $F_{\text{mean}} = 110$:

$$C_{\text{mean}} = \frac{5}{9}(110 - 32) = \frac{5}{9} \times 78 = 43.\overline{3}^{\circ}\text{C}$$

To find the standard deviation in Centigrade, we use:

$$\sigma_C = \frac{5}{9}\sigma_F$$

Substitute $\sigma_F = 2.2$:

$$\sigma_C = \frac{5}{9} \times 2.2 = \frac{11}{9} \approx 1.22^{\circ}\text{C}$$

Thus, the mean temperature is approximately $43.\overline{3}^{\circ}$C and the standard deviation is approximately 1.22°C.

Theorem 5.3. *The expected value of the sum or difference of two or more functions of a random variable X is the sum or difference of the expected values of the functions. That is,*

$$E[g(X) \pm h(X)] = E[g(X)] \pm E[h(X)].$$

Problem 5.17. *Let X be a random variable with probability distribution as follows:*

x	0	1	2	3
$f(x)$	$\frac{1}{3}$	$\frac{1}{2}$	0	$\frac{1}{6}$

Find the expected value of $Y = (X-1)^2$ and variance of X.

Solution

Applying Theorem 5.3 to the function $Y = (X-1)^2$, we can write

$$E[(X-1)^2] = E(X^2 - 2X + 1) = E(X^2) - 2E(X) + E(1).$$

From Theorem 5.2, $E(1) = 1$, and by direct computation,

$$E(X) = (0)\left(\frac{1}{3}\right) + (1)\left(\frac{1}{2}\right) + (2)(0) + (3)\left(\frac{1}{6}\right) = 1,$$

and

$$E(X^2) = (0)\left(\frac{1}{3}\right) + (1)\left(\frac{1}{2}\right) + (4)(0) + (9)\left(\frac{1}{6}\right) = 2.$$

Hence,

$$E[(X-1)^2] = 2 - (2)(1) + 1 = 1.$$

Now, to calculate the variance of Y, we need $E[Y^2]$. Since $Y = (X-1)^2$,

$$Y^2 = (X-1)^4.$$

We need to compute $E[(X-1)^4]$:

$$\begin{aligned}
E[(X-1)^4] &= \sum_x (x-1)^4 f(x) \\
&= 0-1)^4 \cdot \frac{1}{3} + (1-1)^4 \cdot \frac{1}{2} + (2-1)^4 \cdot 0 + (3-1)^4 \cdot \frac{1}{6} \\
&= \frac{1}{3} + 0 + 0 + \frac{8}{3} \\
&= 3.
\end{aligned}$$

Therefore,

$$\mathrm{Var}(Y) = E[Y^2] - (E[Y])^2 = 3 - 1^2 = 2.$$

Problem 5.18. *The weekly demand for a particular drink, measured in thousands of liters, at a chain of convenience stores is a continuous random variable* $g(X) = X^2 + X - 2$, *where* X *has the following density function:*

$$f(x) = \begin{cases} 2(x-1), & 1 < x < 2, \\ 0, & \textit{elsewhere.} \end{cases}$$

Solution

To find the expected value of the weekly demand for the drink, we use Theorem 5.2:

$$E(X^2 + X - 2) = E(X^2) + E(X) - E(2).$$

From Theorem 5.2, $E(2) = 2$. By direct integration, we find:

$$E(X) = \int_1^2 2x(x-1)\,dx = \frac{5}{3},$$

and

$$E(X^2) = \int_1^2 2x^2(x-1)\,dx = \frac{17}{6}.$$

Thus,

$$E(X^2 + X - 2) = \frac{17}{6} + \frac{5}{3} - 2 = \frac{5}{2}.$$

Therefore, the average weekly demand for the drink at this chain of convenience stores is 2500 liters.

Example: Test Score Standardization

Suppose that the raw scores X from a particular testing procedure are distributed between -5 and 20 with an expected value of 10 and a variance of 7. In order to standardize the scores so that they lie between 0 and 100, the linear transformation

$$Y = 4X + 20$$

is applied to the scores. This means, for example, that a raw score of $x = 12$ corresponds to a standardized score of $y = (4 \times 12) + 20 = 68$.

The expected value of the standardized scores is then known to be

$$E(Y) = 4E(X) + 20 = (4 \times 10) + 20 = 60$$

with a variance of

$$\text{Var}(Y) = 4^2\text{Var}(X) = 4^2 \times 7 = 112$$

The standard deviation of the standardized scores is $\sigma_Y = \sqrt{112} = 10.58$, which is $4 \times \sigma_X = 4 \times \sqrt{7}$.

5.8.17 Linear Combinations of Random Variables

When dealing with two random variables, X_1 and X_2, it is often beneficial to analyze the random variable formed by their sum. A general principle states that:

$$E(X_1 + X_2) = E(X_1) + E(X_2)$$

This means the expected value of the sum of two random variables is equal to the sum of their individual expected values.

In addition:

$$\text{Var}(X_1 + X_2) = \text{Var}(X_1) + \text{Var}(X_2) + 2\text{Cov}(X_1, X_2)$$

Note that if the two random variables are **independent**, their covariance is zero, simplifying the variance of their sum to the sum of their variances:

$$\text{Var}(X_1 + X_2) = \text{Var}(X_1) + \text{Var}(X_2)$$

Thus, the variance of the sum of two independent random variables is equal to the sum of their individual variances.

These results are straightforward, but it's crucial to remember that while the expected value of the sum of two random variables always equals the sum of their expected values, the variance of the sum only equals the sum of their variances if the random variables are independent.

Sums of Random Variables: If X_1 and X_2 are two random variables, then:

$$E(X_1 + X_2) = E(X_1) + E(X_2)$$

and

$$\text{Var}(X_1 + X_2) = \text{Var}(X_1) + \text{Var}(X_2) + 2\text{Cov}(X_1, X_2)$$

If X_1 and X_2 are independent random variables such that $\text{Cov}(X_1, X_2) = 0$, then:

$$\text{Var}(X_1 + X_2) = \text{Var}(X_1) + \text{Var}(X_2)$$

Now, consider a sequence of random variables $X_1, \ldots, X_n$ along with constants $a_1, \ldots, a_n$ and b. Define a new random variable Y as the linear combination:

$$Y = a_1 X_1 + \cdots + a_n X_n + b$$

Linear combinations of random variables are important in various contexts, and deriving general results for them is useful. The expectation of the linear combination is:

$$E(Y) = a_1 E(X_1) + \cdots + a_n E(X_n) + b$$

which is simply the linear combination of the expectations of the random variables X_i. Additionally, if the random variables $X_1, \ldots, X_n$ are independent, then:

$$\text{Var}(Y) = a_1^2 \text{Var}(X_1) + \cdots + a_n^2 \text{Var}(X_n)$$

Note that the constant b does not affect the variance of Y, and the coefficients a_i are squared in this expression.

Theorem 5.4. *If $X_1, \ldots, X_n$ is a sequence of random variables and $a_1, \ldots, a_n$ and b are constants, then*

$$E(a_1 X_1 + \cdots + a_n X_n + b) = a_1 E(X_1) + \cdots + a_n E(X_n) + b.$$

If, in addition, the random variables are independent, then

$$Var(a_1 X_1 + \cdots + a_n X_n + b) = a_1^2\, Var(X_1) + \cdots + a_n^2\, Var(X_n).$$

Problem 5.19. *Suppose that $X_1, \ldots, X_n$ is a sequence of independent random variables each with an expectation μ and a variance σ^2. Consider the sample mean $\bar{X}$ defined as:*

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

Find the mean and variance of sample mean $\bar{X}$.

Solution

Mean of the Sample Mean: Using the linearity of expectation:

$$E(\bar{X}) = E\left(\frac{1}{n} \sum_{i=1}^{n} X_i\right) = \frac{1}{n} \sum_{i=1}^{n} E(X_i) = \frac{1}{n} \sum_{i=1}^{n} \mu = \frac{n\mu}{n} = \mu$$

Variance of the Sample Mean: Since the X_i are independent and each has a variance σ^2:

$$\begin{aligned}
\mathrm{Var}(\bar{X}) &= \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \\
&= \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Var}(X_i) \\
&= \frac{1}{n^2}\sum_{i=1}^{n} \sigma^2 \\
&= \frac{n\sigma^2}{n^2} \\
&= \frac{\sigma^2}{n}.
\end{aligned}$$

Therefore, the mean and variance of the sample mean $\bar{X}$ are:

$$E(\bar{X}) = \mu$$

$$\mathrm{Var}(\bar{X}) = \frac{\sigma^2}{n}$$

Example: Test Score Standardization

Suppose that in a particular exam process, candidates are required to complete two tests. Let X_1 represent the score on test 1 and X_2 represent the score on test 2. Assume that the scores for test 1 range from 0 to 30, with an expected value of 18 and a variance of 24, while the scores for test 2 range from -10 to 50, with an expected value of 30 and a variance of 60.

The examination board aims to standardize each test score to a range of 0 to 100, and then calculate a final score out of 100, weighted 2/3 from test 1 and 1/3 from test 2. The standardized scores for the two tests are:

$$Y_1 = \frac{10}{3}X_1 \quad \text{and} \quad Y_2 = \frac{5}{3}X_2 + \frac{50}{3}$$

Thus, the final score is:

$$Z = \frac{2}{3}Y_1 + \frac{1}{3}Y_2 = \frac{20}{9}X_1 + \frac{5}{9}X_2 + \frac{50}{9}$$

For instance, a candidate scoring $x_1 = 11$ on test 1 and $x_2 = 2$ on test 2 would have a final score of:

$$z = \left(\frac{20}{9} \times 11\right) + \left(\frac{5}{9} \times 2\right) + \frac{50}{9} = 31.11$$

The expected value of the final score is:

$$E(Z) = \frac{20}{9}E(X_1) + \frac{5}{9}E(X_2) + \frac{50}{9} = \left(\frac{20}{9} \times 18\right) + \left(\frac{5}{9} \times 30\right) + \frac{50}{9} = 62.22$$

To determine the variance of the final score, it is crucial to consider the independence of the two test scores. The scores are independent only if the tests measure unrelated attributes. For example, if test 1 evaluates probability skills and test 2 assesses statistics skills, the scores might not be independent, and calculating the variance of the final score would require the covariance between the two scores.

However, if test 1 evaluates probability skills and test 2 measures athletic abilities, it is reasonable to assume the scores are independent. In this case, the variance of the final score is:

$$\begin{aligned}\mathrm{Var}(Z) = \mathrm{Var}\left(\frac{20}{9}X_1 + \frac{5}{9}X_2 + \frac{50}{9}\right) &= \left(\frac{20}{9}\right)^2 \mathrm{Var}(X_1) + \left(\frac{5}{9}\right)^2 \mathrm{Var}(X_2)\\ &= \left(\frac{20}{9}\right)^2 \times 24 + \left(\frac{5}{9}\right)^2 \times 60 = 137.04\end{aligned}$$

Thus, the standard deviation of the final score is:

$$\sigma_Z = \sqrt{137.04} = 11.71$$

5.9 Exercises

1. Suppose that the random variables X, Y, and Z are independent with $E(X) = 3$, $\mathrm{Var}(X) = 4$, $E(Y) = -4$, $\mathrm{Var}(Y) = 2$, $E(Z) = 7$, and $\mathrm{Var}(Z) = 7$. Calculate the expectation and variance of the following random variables.

   (a) $3X + 7$
   (b) $5X - 9$
   (c) $2X + 6Y$
   (d) $4X - 3Y$
   (e) $5X - 9Z + 8$
   (f) $-3Y - Z - 5$
   (g) $X + 2Y + 3Z$
   (h) $6X + 2Y - Z + 16$

2. Suppose that items from a manufacturing process are subject to three separate evaluations, and that the results of the first evaluation X_1 have a mean value of 59 with a standard deviation of 10, the results of the second evaluation X_2 have a mean value of 67 with a standard deviation

of 13, and the results of the third evaluation X_3 have a mean value of 72 with a standard deviation of 4. In addition, suppose that the results of the three evaluations can be taken to be independent of each other.

(a) If a final evaluation score is obtained as the average of the three evaluations $X = \frac{X_1+X_2+X_3}{3}$, what are the mean and the standard deviation of the final evaluation score?

(b) If a final evaluation score is obtained as the weighted average of the three evaluations $X = 0.4X_1 + 0.4X_2 + 0.2X_3$, what are the mean and the standard deviation of the final evaluation score?

3. A machine part is assembled by fastening two components of type A and one component of type B end to end. Suppose that the lengths of components of type A have an expectation of 37.0 mm and a standard deviation of 0.7 mm, whereas the lengths of components of type B have an expectation of 24.0 mm and a standard deviation of 0.3 mm. What are the expectation and variance of the length of the machine part?

4. A product is assembled by linking four components of type C and one component of type D sequentially. The lengths of components of type C have an average of 50.0 mm and a standard deviation of 0.8 mm, while the lengths of components of type D have an average of 20.0 mm and a standard deviation of 0.4 mm. Determine the average and variance of the length of the product.

5. A system is constructed by connecting five components of type G and two components of type H end to end. Assume that the lengths of components of type G have an expected value of 40.0 mm and a standard deviation of 1.5 mm, and the lengths of components of type H have an expected value of 22.0 mm and a standard deviation of 0.7 mm. What are the expectation and variance of the total length of the system?

6. A person's cholesterol level C can be measured by three different tests. Test-α returns a value X_α with a mean C and a standard deviation of 1.2, test-β returns a value X_β with a mean C and a standard deviation of 2.4, and test-γ returns a value X_γ with a mean C and a standard deviation of 3.1. Suppose that the three test results are independent. If a doctor decides to use the weighted average $0.5X_\alpha + 0.3X_\beta + 0.2X_\gamma$, what is the standard deviation of the cholesterol level obtained by the doctor?

7. Suppose that the impurity levels of water samples taken from a particular source are independent with a mean value of 3.87 and a standard deviation of 0.18.

(a) What are the mean and the standard deviation of the sum of the impurity levels from two water samples?

(b) What are the mean and the standard deviation of the sum of the impurity levels from three water samples?

(c) What are the mean and the standard deviation of the average of the impurity levels from four water samples?

(d) If the impurity levels of two water samples are averaged, and the result is subtracted from the impurity level of a third sample, what are the mean and the standard deviation of the resulting value?

5.10 Python Functions for Statistical Distributions

In the analysis of statistical distributions, Python provides a variety of functions to work with different types of distributions. These functions can be used to perform tasks such as generating random variates, computing probability mass functions (pmf), cumulative density functions (cdf), and more. The following Table 5.8 summarizes some of the key functions available for discrete and continuous distributions in Python.

These functions are typically part of the `'scipy.stats'` module, which includes a wide range of probability distributions and statistical functions. The table below lists each function along with a brief explanation of its purpose:

Python function	**Function explanation**
rvs(p, loc=0, size=1)	Random variates.
pmf(x, p, loc=0)	Probability mass function.
logpmf(x, p, loc=0)	Log of the probability mass function.
cdf(x, p, loc=0)	Cumulative density function.
logcdf(x, p, loc=0)	Log of the cumulative density function.
sf(x, p, loc=0)	Survival function (1-cdf — sometimes more accurate).
logsf(x, p, loc=0)	Log of the survival function.
ppf(q, p, loc=0)	Percent point function (inverse of cdf — percentiles).
isf(q, p, loc=0)	Inverse survival function (inverse of sf).
stats(p, loc=0, moments='mv')	Mean ('m'), variance ('v'), skew ('s'), and/or kurtosis ('k').
entropy(p, loc=0)	(Differential) entropy of the RV.
expect(func, p, loc=0, lb=None, ub=None, conditional=False)	Expected value of a function (of one argument) with respect to the distribution.
median(p, loc=0)	Median of the distribution.
mean(p, loc=0)	Mean of the distribution.
var(p, loc=0)	Variance of the distribution.
std(p, loc=0)	Standard deviation of the distribution.
interval(alpha, p, loc=0)	Endpoints of the range that contains alpha percent of the distribution.

Table 5.8: Summary of Python functions for statistical distributions.

5.11 Concluding Remarks

In this chapter, we have established a comprehensive foundation for understanding random variables and their fundamental properties. We began by defining random variables and distinguishing between discrete and continuous types, examining their respective probability functions and cumulative distribution functions. We then delved into the crucial concepts of expectation and variance, illustrating their applications through various examples. The discussion on

Chebyshev's Inequality highlighted its utility in providing probabilistic bounds without assuming a specific distribution. Additionally, we explored jointly distributed random variables, emphasizing the importance of understanding independence, covariance, and correlation. With these essential concepts and tools in place, we are now well-equipped to explore specific discrete probability distributions in the following chapter, where we will extend our knowledge to model and analyze discrete data more effectively.

5.12 Chapter Exercises

1. A study records the number of adverse reactions to a new drug among a group of patients. The probability distribution of the number of adverse reactions is given by:

$$P(X = x) = \begin{cases} 0.5 & \text{if } x = 0 \\ 0.3 & \text{if } x = 1 \\ 0.2 & \text{if } x = 2 \end{cases}$$

   (a) Find the expected number of adverse reactions.

   (b) Calculate the variance and standard deviation of the number of adverse reactions.

2. Let the lifetime T in hours of a certain type of electronic device have the probability density function

$$f_T(t) = \begin{cases} \frac{1}{100}e^{-\frac{t}{100}} & \text{for } t \geq 0 \\ 0 & \text{elsewhere} \end{cases}$$

   Find the expectation and variance of the lifetime.

3. Let the height H in centimeters of a particular species of plant have the probability density function

$$f_H(h) = \begin{cases} \frac{3}{64}(h-120)^2 & \text{for } 120 \leq h \leq 124 \\ 0 & \text{elsewhere} \end{cases}$$

   Calculate the cumulative distribution function $F_H(h)$ and find the probability that a plant's height is between 121 and 123 cm.

   (a) Write down the probability mass function $P(Y = y)$.

   (b) Calculate the expectation and variance of Y.

   (c) Find the probability that there are exactly 2 defective items in a batch.

4. Let the length L in meters of a certain type of fish have the probability density function

$$f_L(l) = \begin{cases} 0.2l & \text{for } 0 \le l \le 2 \\ 0 & \text{elsewhere} \end{cases}$$

Find the expectation, variance, and cumulative distribution function $F_L(l)$ of the length.

5. The random variable X measures the concentration of ethanol in a chemical solution, and the random variable Y measures the acidity of the solution. They have a joint probability density function

$$f(x, y) = A(20 - x - 2y)$$

for $0 \le x \le 5$ and $0 \le y \le 5$ and $f(x, y) = 0$ elsewhere.

(a) What is the value of A? What is $P(1 \le X \le 2, 2 \le Y \le 3)$?

(b) Construct the marginal probability density function $f_X(x)$.

(c) What are the expectation and the variance of the ethanol concentration?

Chapter 6

Some Discrete Probability Distributions

6.1 Introduction

In the field of data science, understanding discrete probability distributions is crucial for analyzing and modeling data that can be categorized into distinct outcomes. These distributions help data scientists interpret and predict the likelihood of various events based on historical data, which can be essential for making informed decisions and developing predictive models.

This chapter focuses on three fundamental discrete probability distributions: the Bernoulli distribution, the Binomial distribution, and the Poisson distribution. Each of these distributions plays a vital role in data science applications, ranging from binary classification problems to event counting and rate modeling.

Throughout this chapter, we will delve into each distribution's mathematical properties, including expected value, variance, moment generating function, and characteristic function. We will also present practical examples and exercises to illustrate how these distributions can be applied to real-world data science problems.

6.2 Bernoulli Distribution

Consider a simple experiment where we flip a fair coin. The outcome of this experiment can be either "Heads" or "Tails." We can assign a value of 1 to "Heads" and 0 to "Tails." This experiment is an example of a Bernoulli trial, which is a random experiment with exactly two possible outcomes. It is named after **Jacob Bernoulli**, a Swiss mathematician.

Suppose we are interested in modeling the probability of getting "Heads" in a single coin flip. If the coin is fair, the probability of getting "Heads" (success) is $p = 0.5$ and the probability of getting "Tails" (failure) is $1-p = 0.5$. However, in general, the probability of success in a Bernoulli trial can be any value p such that $0 \leq p \leq 1$.

Definition: A random variable X is said to have a Bernoulli distribution with parameter p if it takes the value 1 with probability p and the value 0 with probability $1-p$. The probability mass function (pmf) of X is given by:

$$P(X = x) = \begin{cases} p & \text{if } x = 1 \\ 1-p & \text{if } x = 0 \end{cases}$$

or more compactly, if $X \sim \text{Bernoulli}(0.1)$ then the pmf is

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

6.2.1 Expected Value (Mean)

The mean or expected value $E(X)$ of a Bernoulli distributed random variable X can be calculated as follows:

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

For a Bernoulli random variable X:

$$E(X) = 1 \cdot P(X = 1) + 0 \cdot P(X = 0)$$

Since $P(X = 1) = p$ and $P(X = 0) = 1 - p$, we have:

$$E(X) = 1 \cdot p + 0 \cdot (1-p) = p$$

So, the mean of a Bernoulli distribution is:

$$E(X) = p$$

6.2.2 Variance

The variance $\text{Var}(X)$ of a Bernoulli distributed random variable X is defined as:

$$\text{Var}(X) = E\left[(X - E(X))^2\right]$$

First, we calculate $E(X^2)$:

$$E(X^2) = \sum_x x^2 \cdot P(X = x)$$

For a Bernoulli random variable X:

$$E(X^2) = 1^2 \cdot P(X = 1) + 0^2 \cdot P(X = 0) = p$$

Now, using the formula for variance:

$$\text{Var}(X) = E(X^2) - [E(X)]^2$$

Substitute the values we have calculated:

$$\text{Var}(X) = p - p^2 = p(1 - p)$$

So, the variance of a Bernoulli distribution is:

$$\text{Var}(X) = p(1 - p)$$

Properties

- **Mean:** The expected value (mean) of a Bernoulli random variable X is given by:
$$E(X) = p$$
- **Variance:** The variance of a Bernoulli random variable X is given by:
$$\text{Var}(X) = p(1 - p)$$
- **Standard Deviation:** The standard deviation of a Bernoulli random variable X is:
$$\sigma = \sqrt{p(1 - p)}$$

Problem 6.1. *A factory produces light bulbs, and each bulb has a 95% chance of passing the quality control test. Define a random variable X such that $X = 1$ if a light bulb passes the quality control test (success) and $X = 0$ if it fails (failure).*

(a). *What is the probability that a randomly selected light bulb passes the quality control test?*

(b). *What is the expected value (mean) of X?*

(c). *What is the variance of X?*

Solution
Let's define the random variable X as follows:

$$X = \begin{cases} 1 & \text{with probability } p = 0.95 \\ 0 & \text{with probability } 1 - p = 0.05 \end{cases}$$

(a). Probability of Passing the Quality Control Test

The probability that a randomly selected light bulb passes the quality control test is given by $P(X = 1)$.

$$P(X = 1) = p = 0.95$$

So, the probability that a light bulb passes the quality control test is 0.95, or 95%.

(b). Expected Value (Mean) of X

The expected value $E(X)$ of a Bernoulli distributed random variable X is given by:

$$E(X) = p$$

Substituting the value of p:

$$E(X) = 0.95$$

So, the expected value of X is 0.95.

(c). Variance of X

The variance $\mathrm{Var}(X)$ of a Bernoulli distributed random variable X is given by:

$$\mathrm{Var}(X) = p(1 - p)$$

Substituting the value of p:

$$\mathrm{Var}(X) = 0.95 \times (1 - 0.95) = 0.95 \times 0.05 = 0.0475$$

So, the variance of X is 0.0475.

Problem 6.2. *A new vaccine is being tested for its effectiveness. In clinical trials, it was found that the vaccine successfully immunizes 90% of the participants. Define a random variable X such that $X = 1$ if a participant is successfully immunized (success) and $X = 0$ if not (failure).*

(i). What is the probability that a randomly selected participant is successfully immunized?

(ii). What is the expected value (mean) of X?

(iii). What is the variance of X?

(iv). In a group of 10 participants, what is the expected number of participants that will be successfully immunized?

Solution

Let's define the random variable X as follows:

$$X = \begin{cases} 1 & \text{with probability } p = 0.90 \\ 0 & \text{with probability } 1 - p = 0.10 \end{cases}$$

(i) Probability of Successful Immunization

The probability that a randomly selected participant is successfully immunized is given by $P(X = 1)$.

$$P(X = 1) = p = 0.90$$

So, the probability that a participant is successfully immunized is 0.90, or 90%.

(ii) Expected Value (Mean) of X

The expected value $E(X)$ of a Bernoulli distributed random variable X is given by:

$$E(X) = p = 0.90$$

So, the expected value of X is 0.90.

(iii) Variance of X

The variance $\text{Var}(X)$ of a Bernoulli distributed random variable X is given by:

$$\text{Var}(X) = p(1 - p) = 0.90 \times (1 - 0.90) = 0.90 \times 0.10 = 0.09$$

So, the variance of X is 0.09.

(iv) Expected Number of Successful Immunizations in a Group of 10 Participants

Let Y be the total number of participants successfully immunized in a group of 10. Y follows a Binomial distribution with parameters $n = 10$ and $p = 0.90$. The expected value $E(Y)$ of a Binomial random variable is given by:

$$E(Y) = n \cdot p$$

Substituting the values of n and p:

$$E(Y) = 10 \times 0.90 = 9$$

So, the expected number of participants successfully immunized in a group of 10 is 9.

6.2.3 Moment Generating Function (MGF)

The moment generating function (MGF) $M_X(t)$ of a random variable X is defined as:

$$M_X(t) = E(e^{tX})$$

For a Bernoulli distributed random variable X with probability p of success (i.e., $X = 1$ with probability p and $X = 0$ with probability $1 - p$):

$$\begin{aligned} M_X(t) &= E(e^{tX}) = e^{t \cdot 0} \cdot P(X = 0) + e^{t \cdot 1} \cdot P(X = 1) \\ &= e^0 \cdot (1 - p) + e^t \cdot p \\ &= 1 \cdot (1 - p) + e^t \cdot p \\ &= (1 - p) + pe^t \end{aligned}$$

So, the moment generating function of a Bernoulli distributed random variable X is:

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

6.2.4 Characteristic Function

The characteristic function $\varphi_X(t)$ of a Bernoulli distributed random variable X is defined as:

$$\begin{aligned} \varphi_X(t) &= E(e^{itX}) = e^{it \cdot 0} \cdot P(X = 0) + e^{it \cdot 1} \cdot P(X = 1) \\ &= e^{i \cdot 0} \cdot (1 - p) + e^{it} \cdot p \\ &= 1 \cdot (1 - p) + e^{it} \cdot p \\ &= (1 - p) + pe^{it}. \end{aligned}$$

6.2.5 Probability Generating Function

For a Bernoulli random variable X with parameter p, the probability generating function (PGF) is given by:

$$G_X(s) = E[s^X]$$

Since X can take values 0 and 1, we have:

$$\begin{aligned} G_X(s) = E[s^X] &= \sum_x P(X = x) \cdot s^x \\ &= (1 - p) \cdot s^0 + p \cdot s^1 \\ &= (1 - p) + p \cdot s \end{aligned}$$

Therefore, the PGF of a Bernoulli random variable X with parameter p is:

$$G_X(s) = 1 - p + p \cdot s$$

6.2.6 Example

Let's consider a biased coin where the probability of getting "Heads" is $p = 0.7$. The random variable X representing the outcome of a single coin flip follows a Bernoulli distribution with parameter $p = 0.7$.

The pmf of X is:

$$P(X = x) = \begin{cases} 0.7 & \text{if } x = 1 \\ 0.3 & \text{if } x = 0 \end{cases}$$

The mean and variance of X are:

$$E(X) = 0.7$$

$$\text{Var}(X) = 0.7 \times (1 - 0.7) = 0.21$$

Thus, we can model and analyze the outcomes of a single coin flip using the Bernoulli distribution.

Problem 6.3. *A factory produces light bulbs, and each light bulb is tested for quality. The probability that a light bulb is defective is $p = 0.1$. Let X be a random variable that represents whether a randomly selected light bulb is defective (1 if defective, 0 if not defective).*

1. *What is the probability that a randomly selected light bulb is defective?*
2. *What is the probability that a randomly selected light bulb is not defective?*
3. *Compute the expected value and variance of X.*

Solution

Let X follow a Bernoulli distribution with parameter $p = 0.1$, i.e., $X \sim \text{Bernoulli}(0.1)$.

1. The probability that a randomly selected light bulb is defective is given by $P(X = 1)$:
$$P(X = 1) = p = 0.1$$

2. The probability that a randomly selected light bulb is not defective is given by $P(X = 0)$:
$$P(X = 0) = 1 - p = 1 - 0.1 = 0.9$$

3. For a Bernoulli random variable X with parameter p:
   - The expected value $E[X]$ is:
   $$E[X] = p = 0.1$$
   - The variance $\text{Var}(X)$ is:
   $$\text{Var}(X) = p(1 - p) = 0.1 \cdot (1 - 0.1) = 0.1 \cdot 0.9 = 0.09$$

□

6.2.7 Applications

The Bernoulli distribution is used to model binary outcomes in various scenarios, such as:

- **Quality Control in Manufacturing**

  In manufacturing, the Bernoulli distribution is used to model the probability of a defect in a production process. For example, a factory producing electronic components might use the Bernoulli distribution to determine the likelihood that a randomly selected component is defective.

- **Clinical Trials in Medicine**

  The Bernoulli distribution can model the outcome of a clinical trial for a new drug, where $X = 1$ represents a successful treatment (e.g., patient recovery) and $X = 0$ represents an unsuccessful treatment. This helps in estimating the effectiveness of the drug.

- **A/B Testing in Marketing**

  In digital marketing, A/B testing is used to compare two versions of a webpage or advertisement. The Bernoulli distribution models the probability of a user clicking on an ad or making a purchase, where $X = 1$ indicates a click or purchase and $X = 0$ indicates no click or purchase.

- **Sports Performance Analysis**

  The Bernoulli distribution can be applied to model the probability of a successful outcome in sports, such as a basketball player making a free throw or a soccer player scoring a penalty kick. Here, $X = 1$ represents a successful attempt, and $X = 0$ represents a failure.

- **Insurance Risk Assessment**

  In insurance, the Bernoulli distribution is used to model the occurrence of certain events, such as accidents or claims. For instance, $X = 1$ could represent a policyholder filing a claim within a year, and $X = 0$ could represent no claim filed.

- **Genetics**

  The Bernoulli distribution is used in genetics to model the inheritance of a particular gene. For example, $X = 1$ might represent the presence of a specific gene in an offspring, and $X = 0$ represents its absence, assuming a certain probability of inheritance.

6.2.8 Python Code for Bernoulli Distribution

In Python, we can calculate various characteristics of the Bernoulli distribution using the ‘`scipy.stats`‘ module. Below, we detail how to compute the Probability Mass Function, Cumulative Distribution Function, Mean (Expected Value), Variance, and Probability Generating Function, etc.

Python Code

Here's how we can compute these characteristics using Python:

```python
import numpy as np
from scipy.stats import bernoulli

# Define the parameter p
p = 0.25

# Bernoulli distribution
dist = bernoulli(p)

# 1. Probability Mass Function (PMF)
x_values = [0, 1]  # Possible values for a Bernoulli random variable
pmf_values = dist.pmf(x_values)
print("PMF values for x = 0 and x = 1:", pmf_values)

# 2. Cumulative Distribution Function (CDF)
cdf_values = dist.cdf(x_values)
print("CDF values for x = 0 and x = 1:", cdf_values)
```

```

# 3. Mean (Expected Value)
mean = dist.mean()
print("Mean (Expected Value):", mean)

# 4. Variance
variance = dist.var()
print("Variance:", variance)

# 5. Probability Generating Function (PGF)
def pgf(t, p):
  return (1 - p) + p * t

# PGF values at t = 0 and t = 1
pgf_values = [pgf(t, p) for t in [0, 1]]
print("PGF values for t = 0 and t = 1:", pgf_values)
```

Explanations

- **Probability Mass Function (PMF):**
  - The PMF gives the probability of each outcome (0 or 1). Use `dist.pmf(x_values)` to compute these probabilities.
- **Cumulative Distribution Function (CDF):**
  - The CDF gives the cumulative probability up to each outcome. Use `dist.cdf(x_values)` to compute these values.
- **Mean (Expected Value):**
  - The mean is simply the parameter p of the Bernoulli distribution. Use `dist.mean()` to get this value.
- **Variance:**
  - The variance of a Bernoulli distribution is $p \cdot (1 - p)$. Use `dist.var()` to compute this.
- **Probability Generating Function (PGF):**
  - The PGF is calculated using the formula $G_X(t) = 1 - p + p \cdot t$. Define a function `pgf(t, p)` to compute PGF values for specific t values (e.g., 0 and 1).

6.2.9 Exercises

1. A medical test for a certain disease has a 98% chance of correctly identifying a diseased person (true positive) and a 2% chance of incorrectly

identifying a healthy person as diseased (false positive). Define a random variable X such that $X = 1$ if the test result is positive (either true positive or false positive) and $X = 0$ if the test result is negative.

(a) What is the probability that a randomly selected test result is positive?

(b) What is the expected value (mean) of X?

(c) What is the variance of X?

(d) In a group of 50 people who took the test, what is the expected number of positive test results?

A genetic trait is passed on to the next generation with a probability of 25%. Define a random variable X such that $X = 1$ if the trait is passed on (success) and $X = 0$ if it is not (failure).

2. (a) What is the probability that a randomly selected offspring inherits the trait?

   (b) What is the expected value (mean) of X?

   (c) What is the variance of X?

   (d) In a group of 40 offspring, what is the expected number of offspring that will inherit the trait?

3. In a clinical trial, a new drug is found to be effective in 85% of the patients. Define a random variable X such that $X = 1$ if a patient responds positively to the drug (success) and $X = 0$ if not (failure).

   (a) What is the probability that a randomly selected patient responds positively to the drug?

   (b) What is the expected value (mean) of X?

   (c) What is the variance of X?

   (d) In a sample of 20 patients, what is the expected number of patients who will respond positively to the drug?

4. A diagnostic test has a 92% chance of correctly detecting a condition when it is present (true positive rate) and an 8% chance of detecting the condition when it is not present (false positive rate). Define a random variable X such that $X = 1$ if the test result is positive (either true positive or false positive) and $X = 0$ if the test result is negative.

   (a) What is the probability that a randomly selected test result is positive?

   (b) What is the expected value (mean) of X?

   (c) What is the variance of X?

   (d) In a group of 100 people who took the test, what is the expected number of positive test results?

5. In a study of a certain disease, it is found that 70% of the subjects have a particular gene variant that increases susceptibility to the disease. Define a random variable X such that $X = 1$ if a subject has the gene variant (success) and $X = 0$ if not (failure).

   (a) What is the probability that a randomly selected subject has the gene variant?

   (b) What is the expected value (mean) of X?

   (c) What is the variance of X?

   (d) In a sample of 30 subjects, what is the expected number of subjects who have the gene variant?

6.3 Binomial Distribution

Consider the simple experiment of flipping a fair coin, where the outcome can be either "Heads" (success) or "Tails" (failure). We assign a value of 1 to "Heads" and 0 to "Tails," making this a Bernoulli trial. Suppose we flip the coin n times. We are interested in modeling the probability of obtaining exactly k "Heads" (successes) in n flips.

This problem is an example of a Binomial experiment, which consists of n independent Bernoulli trials, each with the same probability of success p. To calculate the probability of getting exactly k successes in n trials, consider a specific sequence of trials where exactly k trials are successes and $n - k$ trials are failures. The probability of such a specific sequence is given by:

$$p^k \times (1-p)^{n-k}$$

where p is the probability of success and $1 - p$ is the probability of failure.

We need to account for all possible sequences of n trials that result in exactly k successes. The number of ways to choose k positions for successes out of n positions is given by the Binomial coefficient:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

where $\binom{n}{k}$ represents the number of combinations of n items taken k at a time.

The total probability of having exactly k successes in n trials is the product of the probability of any specific sequence and the number of such sequences. Thus, the probability mass function of the Binomial distribution is:

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

where X is the random variable representing the number of successes.

The Binomial distribution is a discrete probability distribution. This distribution has the following conditions:

1. **Fixed Number of Trials:** The experiment is repeated a fixed number of times, denoted as n.
2. **Two Possible Outcomes:** Each trial results in one of two outcomes, often referred to as "success" and "failure."
3. **Constant Probability:** The probability of success in each trial is constant and denoted by p. Consequently, the probability of failure is $1 - p$.
4. **Independence:** The outcome of one trial is independent of the outcomes of other trials.

If these conditions are met, the number of successes in n trials follows a Binomial distribution with parameters n and p.

Definition: A random variable X that represents the number of successes in n independent Bernoulli trials, each with a probability of success p, is said to follow a Binomial distribution if its probability mass function (pmf) is given by:

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{for} \quad k = 0, 1, 2, \ldots, n,$$

where $\binom{n}{k}$ is the Binomial coefficient. We write this as:

$$X \sim \mathrm{B}(n, p).$$

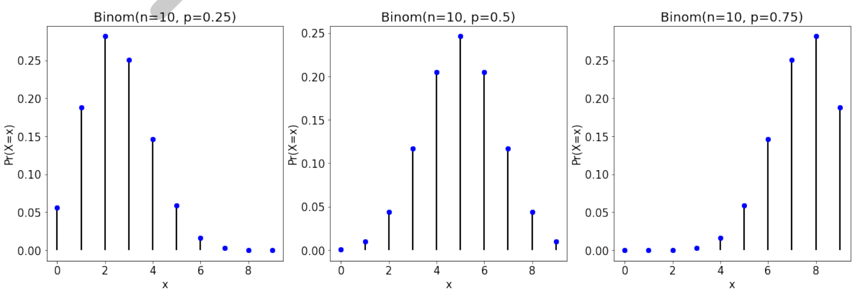


Figure 6.1: Graphical Presentation of pmf of Binomial Distribution.

The graphical presentation of the Binomial distribution for $n = 10$ and $p = 0.25$, $p = 0.5$, and $p = 0.75$ is depicted in Figure 6.1. The Python code used to generate these figures is provided below.

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

# Parameters
n = 10
p = 0.5

# Values
k = np.arange(0, n+1)
pmf = binom.pmf(k, n, p)

# Plot
plt.bar(k, pmf, color='blue', edgecolor='black')
plt.xlabel('Number of Successes')
plt.ylabel('Probability')
plt.title('Binomial PMF (n=10, p=0.5)')
plt.xticks(k)
plt.grid(True)
plt.show()
```

Symmetric Binomial Distributions A $\mathrm{B}(n, 0.5)$ distribution is a symmetric probability distribution for any value of the parameter n. The distribution is symmetric about the expected value $n/2$.

6.3.1 Expected Value

The expected value $E[X]$ is:

$$
\begin{aligned}
E[X] &= \sum_{k=0}^{n} k \cdot P(X = k) \\
&= \sum_{k=0}^{n} k \cdot \binom{n}{k} p^k (1-p)^{n-k}.
\end{aligned}
$$

By recognizing that X can be written as the sum of n independent Bernoulli trials X_i:

$$X = X_1 + X_2 + \cdots + X_n$$

Each $X_i \sim \text{Bernoulli}(p)$ with:

$$E[X_i] = p$$

Using the linearity of expectation:

$$E[X] = E[X_1 + X_2 + \cdots + X_n] = E[X_1] + E[X_2] + \cdots + E[X_n] = n \cdot p$$

The expected value $E[X]$ is:

$$E[X] = n \cdot p$$

6.3.2 Variance and Standard Deviation

Since $X_1, X_2, \ldots, X_n$ are independent random variables, we can use Theorem 5.4. Thus, the variance $\mathrm{Var}(X)$ can be computed as follows:

$$\begin{aligned}
\sigma^2 = \mathrm{Var}(X) &= \mathrm{Var}(X_1 + X_2 + \cdots + X_n) \\
&= \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + \cdots + \mathrm{Var}(X_n) \\
&= p(1-p) + p(1-p) + \cdots + p(1-p) \\
&= np(1-p).
\end{aligned}$$

The standard deviation σ is:

$$\sigma = \sqrt{\mathrm{Var}(X)} = \sqrt{np(1-p)}$$

Properties

- **Mean:** The expected value (mean) of a Binomial random variable X is given by:
$$E(X) = np$$
- **Variance:** The variance of a Binomial random variable X is given by:
$$\mathrm{Var}(X) = np(1-p)$$
- **Standard Deviation:** The standard deviation of a Binomial random variable X is:
$$\sigma = \sqrt{np(1-p)}$$

6.3.3 Example

Let's consider a biased coin where the probability of getting "Heads" is $p = 0.7$. Suppose we flip this coin $n = 10$ times. The random variable X representing the number of "Heads" in 10 flips follows a Binomial distribution with parameters $n = 10$ and $p = 0.7$.

The pmf of X is:

$$P(X = k) = \binom{10}{k} (0.7)^k (0.3)^{10-k}$$

for $k = 0, 1, 2, \ldots, 10$.

The mean and variance of X are:

$$E(X) = 10 \times 0.7 = 7$$

$$\mathrm{Var}(X) = 10 \times 0.7 \times (1 - 0.7) = 2.1$$

Thus, we can model and analyze the number of "Heads" in 10 coin flips using the Binomial distribution.

Problem 6.4. *A factory produces light bulbs, and 5% of them are defective. Suppose a quality control inspector randomly selects 8 bulbs for testing.*

1. *What is the probability that exactly 2 of the 8 bulbs are defective?*
2. *What is the probability that at most 2 bulbs are defective?*

Solution

Let X be the number of defective bulbs in the sample. Here, X follows a binomial distribution with parameters $n = 8$ and $p = 0.05$:

$$X \sim \mathrm{B}(8, 0.05)$$

1. Probability of Exactly 2 Defective Bulbs

$$P(X = 2) = \binom{8}{2}(0.05)^2(1 - 0.05)^{8-2} = 28 \times (0.05)^2 \times (0.95)^6 \approx 0.042$$

Thus, the probability that exactly 2 bulbs are defective is approximately 0.042.

2. Probability of At Most 2 Defective Bulbs

$$P(X \leq 2) = \sum_{k=0}^{2} P(X = k)$$

Compute each term:

$$P(X = 0) = \binom{8}{0}(0.05)^0(0.95)^8 \approx 0.663$$
$$P(X = 1) = \binom{8}{1}(0.05)^1(0.95)^7 \approx 0.232$$
$$P(X = 2) = \binom{8}{2}(0.05)^2(0.95)^6 \approx 0.042$$

Sum these probabilities:

$$P(X \leq 2) \approx 0.663 + 0.232 + 0.042 = 0.937$$

Thus, the probability that at most 2 bulbs are defective is approximately 0.937.

Problem 6.5. *A school administers a math test to 15 students, and the probability that a student passes the test is 0.7.*

1. *What is the probability that exactly 10 students pass the test?*
2. *What is the probability that at least 10 students pass the test?*

Solution

Let Y be the number of students who pass the test. Here, Y follows a binomial distribution with parameters $n = 15$ and $p = 0.7$:

$$Y \sim \mathrm{B}(15, 0.7)$$

1. Probability of Exactly 10 Students Passing

$$P(Y = 10) = \binom{15}{10}(0.7)^{10}(1 - 0.7)^{15-10} = 3003 \times (0.7)^{10} \times (0.3)^5 \approx 0.205$$

Thus, the probability that exactly 10 students pass the test is approximately 0.205.

2. Probability of At Least 10 Students Passing

$$P(Y \geq 10) = \sum_{k=10}^{15} P(Y = k)$$

This involves calculating and summing the probabilities for

$$k = 10, 11, 12, 13, 14, 15.$$

Using statistical software or tables, you get:

$$P(Y \geq 10) \approx 0.542$$

Thus, the probability that at least 10 students pass the test is approximately 0.542.

Problem 6.6. *A pharmaceutical company is testing a new drug to see if it improves recovery rates. The probability that a patient responds positively to the drug is $p = 0.25$. Suppose the company tests the drug on 10 patients. Let X be the number of patients who respond positively to the drug.*

1. *What is the probability that exactly 3 patients respond positively?*
2. *What is the probability that at most 3 patients respond positively?*
3. *Calculate the expected number of patients who respond positively and the variance of X.*
4. *If Y represents the number of patients who do not respond positively, find the probability distribution of Y and its mean and variance.*

Solution

Let X be the number of patients who respond positively out of 10 patients, with each patient responding positively with probability $p = 0.25$. Then X follows a binomial distribution:

$$X \sim \mathrm{B}(n = 10, p = 0.25)$$

1. To find the probability that exactly 3 patients respond positively:

$$P(X = 3) = \binom{10}{3} p^3 (1-p)^{10-3}$$

   Substituting $p = 0.25$:

$$P(X = 3) = \binom{10}{3} (0.25)^3 (0.75)^7 = 120 \times (0.25)^3 \times (0.75)^7 \approx 0.2503$$

2. To find the probability that at most 3 patients respond positively:

$$\begin{aligned} P(X \leq 3) &= \sum_{k=0}^{3} \binom{10}{k} p^k (1-p)^{10-k} \\ &= 0.0563 + 0.1877 + 0.2816 + 0.2503 \\ &= 0.7759 \end{aligned}$$

   Thus, the probability that at least 10 students pass the test is approximately 0.542.

3. The expected number of patients who respond positively and the variance of X are:

$$E[X] = n \cdot p = 10 \cdot 0.25 = 2.5$$

$$\mathrm{Var}(X) = n \cdot p \cdot (1-p) = 10 \cdot 0.25 \cdot 0.75 = 1.875$$

4. Let Y represent the number of patients who do not respond positively. Then Y follows a binomial distribution with parameters $n = 10$ and $q = 1 - p = 0.75$:

$$Y \sim \mathrm{B}(n = 10, q = 0.75)$$

   The mean and variance of Y are:

$$E[Y] = n \cdot q = 10 \cdot 0.75 = 7.5$$

$$\mathrm{Var}(Y) = n \cdot q \cdot (1-q) = 10 \cdot 0.75 \cdot 0.25 = 1.875$$

6.3.4 Python Code for Binomial Distribution

In Python, you can compute various characteristics of the Binomial distribution using the ‘scipy.stats‘ module. Below is a demonstration of how to compute these characteristics.

Python Code

Here’s how you can calculate various characteristics of a Binomial distribution:

```python
import numpy as np
from scipy.stats import binom

# Define parameters
n = 10  # number of trials
p = 0.25  # probability of success

# Binomial distribution
dist = binom(n, p)

# 1. Probability Mass Function (PMF)
x_values = np.arange(0, n + 1)  # possible values for the
  random variable
pmf_values = dist.pmf(x_values)
print("PMF values for x = 0 to 10:", pmf_values)

# 2. Cumulative Distribution Function (CDF)
cdf_values = dist.cdf(x_values)
print("CDF values for x = 0 to 10:", cdf_values)

# 3. Mean (Expected Value)
mean = dist.mean()
print("Mean (Expected Value):", mean)

# 4. Variance
variance = dist.var()
print("Variance:", variance)

# 5. Probability Generating Function (PGF)
def pgf(t, n, p):
  return (1 - p + p * t) ** n

# PGF values at t = 0 and t = 1
pgf_values = [pgf(t, n, p) for t in [0, 1]]
print("PGF values for t = 0 and t = 1:", pgf_values)
```

Explanations

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

  - The PMF provides the probability of each number of successes. Compute these probabilities using `dist.pmf(x_values)`.

- **Cumulative Distribution Function (CDF):**
  - The CDF provides the cumulative probability up to a certain number of successes. Compute these values using `dist.cdf(x_values)`.

- **Mean (Expected Value):**
  - The mean is given by $n \cdot p$. Compute this using `dist.mean()`.

- **Variance:**
  - The variance is given by $n \cdot p \cdot (1 - p)$. Compute this using `dist.var()`.

- **Probability Generating Function (PGF):**
  - The PGF is given by $G_X(t) = (1 - p + p \cdot t)^n$. Define a function `pgf(t, n, p)` to compute PGF values for specific t values.

6.3.5 Exercises

1. Define the Binomial distribution. Include the conditions that must be met for a random variable to follow a Binomial distribution.

2. Prove that the expected value (mean) of a Binomially distributed random variable X with parameters n and p is $E(X) = np$. Also, prove that the variance of X is given by $\mathrm{Var}(X) = np(1 - p)$.

3. Prove that the sum of the probabilities of all possible outcomes of a Binomial random variable X with parameters n and p is equal to 1. That is, show that
$$\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = 1.$$

4. A fair coin is flipped 5 times. What is the probability of getting exactly 3 heads?

5. In a class, the probability that a student passes an exam is 0.8 If 15 students are randomly selected, calculate the expected number of students who pass the exam.

6. Show that
$$\binom{n}{k} = \binom{n}{n-k}$$
and use this property to demonstrate that the sum of the Binomial coefficients for a given n is symmetric around $k = \frac{n}{2}$.

7. Find the moment generating function (MGF) of the Binomial distribution and use it to compute the first two moments (mean and variance).

8. A factory produces widgets, with a 1% defect rate. If a sample of 50 widgets is tested, what is the probability that at least 3 widgets are defective?

9. A medical test has a 95% sensitivity and a 90% specificity. If 10 patients are tested who all have the disease, calculate the probability that exactly 9 patients test positive.

10. A soccer player has a 60% chance of scoring a goal in a penalty kick. If the player takes 12 penalty kicks, find the probability that the player scores more than 8 goals.

11. In a marketing campaign, the probability of converting a lead into a customer is 10%. If 25 leads are contacted, find the probability of converting at most 5 leads.

12. In a clinical trial, the number of successes (patients who show improvement) follows a Binomial distribution with $n = 10$ and $p = 0.7$.

    (a) What is the probability of exactly 7 successes?

    (b) What is the probability of at least 8 successes?

    (c) Find the expected number of successes.

13. Prove the following Binomial coefficient identity:

$$\binom{n}{k} + \binom{n}{k-1} = \binom{n+1}{k}.$$

    Use a combinatorial argument or algebraic manipulation to justify this identity.

6.4 Poisson Distribution

Imagine you are managing a small coffee shop. You notice that, on average, 4 customers come into the shop every hour. You are interested in understanding how likely it is to see a specific number of customers in the shop during a given hour. For example, what is the probability of having exactly 2 customers or exactly 6 customers in an hour? In this scenario, the number of customers X arriving at the coffee shop in an hour follows a **Poisson distribution** with parameter $\lambda = 4$, where λ represents the average rate of customers per hour.

The Poisson distribution is a discrete probability distribution that models the number of events occurring within a **fixed interval of time** or **space**, given the following conditions:

(i). Each event happens independently of the others. (For example, if we're modeling the number of customers arriving at a coffee shop, the number of customers arriving in one hour does not affect the number of customers arriving in the next hour.)

(ii). The average rate (mean number of events) λ is constant over time or space. This means that the expected number of events in any interval of the same length is the same.

(iii). The events are relatively rare in the given interval. More specifically, the probability of more than one event occurring in a very short interval is negligible.

(iv). The number of events can only be whole numbers (0, 1, 2, ...). You can't have fractional events.

(v). The number of events is counted over a fixed interval of time or space. The intervals are non-overlapping, meaning events in different intervals do not influence each other.

When these conditions are met, the number of events occurring in a fixed interval follows a Poisson distribution. Formally, a random variable X follows a Poisson distribution with parameter λ, denoted $X \sim \text{Poisson}(\lambda)$, if its probability mass function is given by:

$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!} \qquad x = 0, 1, 2, \ldots$$

where:

- λ is the average rate or mean number of events in the interval.
- e is the base of the natural logarithm, approximately equal to 2.71828.

It is observe that the series expansion of e^{λ} guarantees that the total probability sums to 1, as shown below:

$$\begin{aligned}\sum_{x=0}^{\infty} P(X = x) &= \sum_{x=0}^{\infty} e^{-\lambda} \frac{\lambda^x}{x!} = e^{-\lambda} \left(\frac{1}{0!} + \frac{\lambda}{1!} + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \cdots \right) \\ &= e^{-\lambda} \cdot e^{\lambda} = 1\end{aligned}$$

Additionally, for a random variable X that follows a Poisson distribution with parameter λ (denoted $X \sim \text{Poisson}(\lambda)$), it holds that:

$$E(X) = \text{Var}(X) = \lambda$$

Definition: A random variable X is said to follow a Poisson distribution with parameter λ, written as $X \sim \text{Poisson}(\lambda)$, if its probability mass function is given by:

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!} \quad \text{for } x = 0, 1, 2, 3, \ldots.$$

The Poisson distribution is particularly useful for modeling the number of occurrences of a certain event within a specified unit of time, distance, or volume, with both its mean and variance equal to λ.

The probability mass function (pmf) of a Poisson random variable with parameter $\lambda = 2$ is given by:

$$P(X = x) = \frac{e^{-2} \cdot 2^x}{x!}$$

and we plot this pmf for integer values of x from 0 to 10 in Figure 6.2.

Figures 6.2 and 6.3 compare the probability mass functions and cumulative distribution functions of Poisson distributions with parameters $\lambda = 2$ and $\lambda = 5$. These figures demonstrate that, given that the mean and variance of a Poisson distribution both equal the parameter value, a distribution with a higher parameter value will have a greater expected value and exhibit a wider spread.

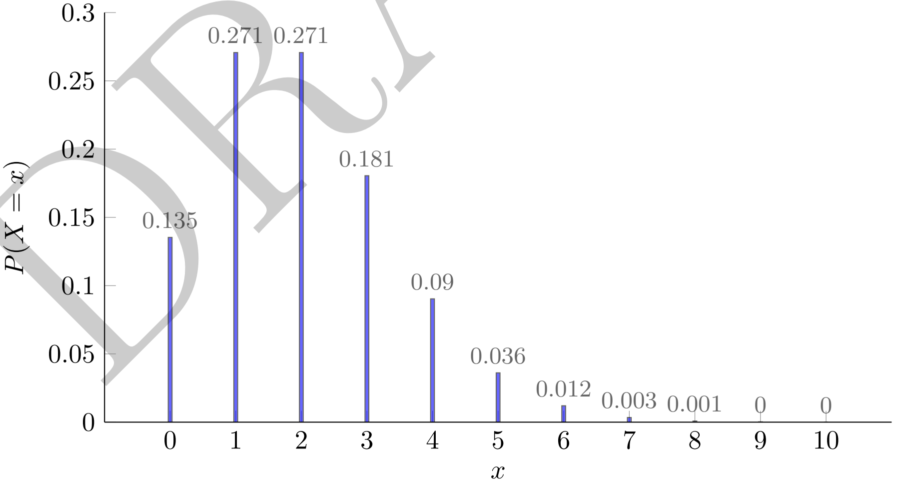


Figure 6.2: Probability Mass Function of a Poisson Distribution with $\lambda = 2$.

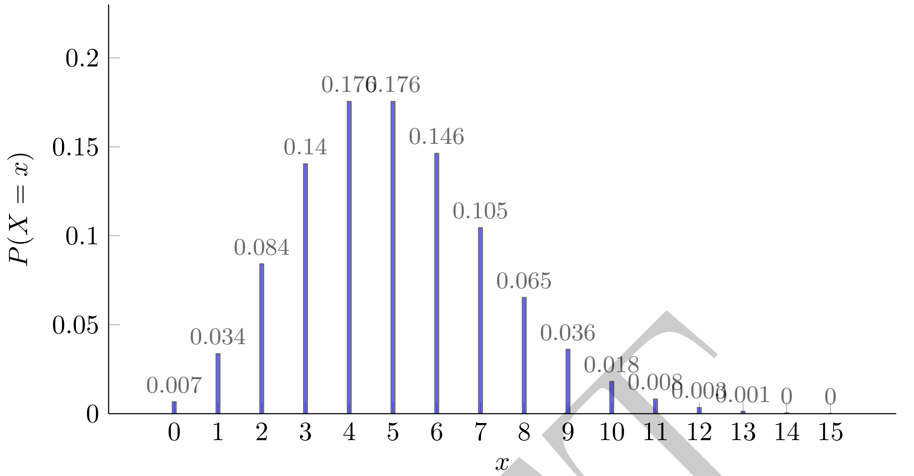


Figure 6.3: Probability Mass Function of a Poisson Distribution with $\lambda = 5$.

It is worth noting that the Poisson distribution can serve as an approximation for the Binomial distribution $B(n, p)$ under certain conditions. Specifically, this approximation is useful when n is quite large (e.g., greater than 150) and p is very small (e.g., less than 0.01). To achieve this approximation, the parameter $\lambda = np$ should be employed in the Poisson distribution to match the expected value of the Binomial distribution.

The Poisson distribution is used to model the **number of rare events** occurring within a fixed interval of time or space, under the assumption that these events occur independently and at a constant average rate. This distribution helps answer questions about the likelihood of observing a specific number of events given an average rate, such as predicting the number of genetic mutations in bacterial cultures or call arrivals in a call center.

Problem 6.7. *A quality inspector at a glass manufacturing company checks each glass sheet for imperfections. Suppose the number of flaws in each sheet follows a Poisson distribution with a parameter $\lambda = 0.5$, which indicates that the expected number of flaws per sheet is 0.5.*

(a). Determine the probability that a glass sheet has no flaws.

(b). Sheets with two or more flaws are scrapped by the company Estimate the percentage of glass sheets that need to be scrapped and recycled.

Solution

(a). Probability of No Flaws

The probability that a glass sheet has no flaws $(X = 0)$ is given by:

$$P(X = 0) = \frac{e^{-0.5} \cdot 0.5^0}{0!} = e^{-0.5} \approx 0.607$$

Thus, approximately 61% of the glass sheets are in "perfect" condition.

(b). Probability of Two or More Flaws

The probability of having two or more flaws ($X \geq 2$) can be computed as:

$$P(X \geq 2) = 1 - P(X = 0) - P(X = 1)$$

Where

$$P(X = 1) = \frac{e^{-0.5} \cdot 0.5^1}{1!} = e^{-0.5} \cdot 0.5 \approx 0.305$$

Therefore:

$$P(X \geq 2) = 1 - e^{-0.5} - (e^{-0.5} \cdot 0.5) \approx 1 - 0.607 - 0.305 = 0.090$$

Hence, about 9% of the glass sheets have two or more flaws and need to be scrapped.

6.4.1 Expected Value

The expected value $E(X)$ of a discrete random variable X is defined as:

$$\begin{aligned} E(X) &= \sum_{x=0}^{\infty} x \cdot P(X = x) \\ &= \sum_{x=0}^{\infty} x \cdot \frac{\lambda^x e^{-\lambda}}{x!} \\ &= \sum_{x=0}^{\infty} \frac{\lambda^x e^{-\lambda}}{(x-1)!}. \end{aligned}$$

To adjust the index, let $x' = x - 1$. Then, when x starts from 1, x' starts from 0.

Rewriting the sum in terms of x':

$$\begin{aligned} E(X) &= e^{-\lambda} \sum_{x'=0}^{\infty} \frac{\lambda^{x'+1}}{x'!} \\ &= \lambda e^{-\lambda} \sum_{x'=0}^{\infty} \frac{\lambda^{x'}}{x'!}. \end{aligned}$$

Recognize that the sum is the Taylor series expansion of e^{λ}:

$$\sum_{x'=0}^{\infty} \frac{\lambda^{x'}}{x'!} = e^{\lambda}$$

Thus:

$$E(X) = \lambda e^{-\lambda} \cdot e^{\lambda} = \lambda$$

Problem 6.8. *A researcher is studying the occurrence of a rare genetic mutation in a population of individuals. The number of individuals with the mutation in a randomly selected sample of 100 individuals follows a Poisson distribution with a parameter* $\lambda = 2$.

(a). Calculate the probability that exactly one individual in the sample has the mutation.

(b). Determine the probability that at least three individuals in the sample have the mutation.

(c). Estimate the percentage of samples in which two or more individuals are expected to have the mutation.

Solution (a). Probability of Exactly One Individual with the Mutation

The probability that exactly one individual has the mutation is given by:

$$P(X = 1) = \frac{e^{-2} \cdot 2^1}{1!} = 2e^{-2} \approx 0.2707$$

(b). Probability of At Least Three Individuals with the Mutation

The probability of at least three individuals having the mutation is:

$$P(X \geq 3) = 1 - P(X < 3) = 1 - [P(X = 0) + P(X = 1) + P(X = 2)]$$

Where:

$$P(X = 0) = e^{-2} \approx 0.1353$$
$$P(X = 1) = 2e^{-2} \approx 0.2707$$
$$P(X = 2) = \frac{2^2 \cdot e^{-2}}{2!} \approx 0.2707$$

Therefore:

$$P(X \geq 3) = 1 - (0.1353 + 0.2707 + 0.2707) = 1 - 0.6767 = 0.3233$$

(c). Percentage of Samples with Two or More Individuals Having the Mutation

To estimate the percentage of samples where two or more individuals have the mutation:

$$P(X \geq 2) = 1 - P(X < 2) = 1 - [P(X = 0) + P(X = 1)]$$

Thus:

$$P(X \geq 2) = 1 - (0.1353 + 0.2707) = 1 - 0.4060 = 0.5940$$

The percentage is:

$$0.5940 \times 100\% = 59.40\%$$

6.4.2 Variance

To find the variance, we first need to calculate $E(X^2)$. We use:

$$
\begin{aligned}
E(X^2) &= \sum_{x=0}^{\infty} x^2 \cdot P(X = x) \\
&= \sum_{x=0}^{\infty} x^2 \cdot \frac{\lambda^x e^{-\lambda}}{x!} \\
&= e^{-\lambda} \sum_{x=0}^{\infty} x \cdot \frac{\lambda^x}{(x-1)!} \\
&= \sum_{x=0}^{\infty} \frac{x\lambda^x e^{-\lambda}}{(x-1)!}
\end{aligned}
$$

Change the index of summation. Let $x' = x - 1$. Therefore, $x = x' + 1$, and when x starts from 1, x' starts from 0.

Rewriting the sum:

$$
\begin{aligned}
\underline{E(X^2)} &= e^{-\lambda} \sum_{x'=0}^{\infty} \frac{(x'+1)\lambda^{x'+1}}{x'!} \\
&= e^{-\lambda} \left(\sum_{x'=0}^{\infty} \frac{x'\lambda^{x'+1}}{x'!} + \sum_{x'=0}^{\infty} \frac{\lambda^{x'+1}}{x'!} \right) \\
&= e^{-\lambda} \left(\lambda \sum_{x'=0}^{\infty} \frac{x'\lambda^{x'}}{x'!} + \lambda \sum_{x'=0}^{\infty} \frac{\lambda^{x'}}{x'!} \right) \\
&= e^{-\lambda} \left(\lambda \cdot \lambda e^{\lambda} + \lambda e^{\lambda} \right) \\
&= \underline{\lambda^2 + \lambda.}
\end{aligned}
$$

Now, using the definition of variance:

$$
\begin{aligned}
\underline{V(X)} &= E(X^2) - (E(X))^2 \\
&= (\lambda^2 + \lambda) - \lambda^2 \\
&= \underline{\lambda.}
\end{aligned}
$$

6.4.3 Moment Generating Function

The moment generating function (MGF) $M_X(t)$ of X is defined as:

$$\begin{aligned}
M_X(t) = E[e^{tX}] &= \sum_{x=0}^{\infty} e^{tx} \cdot P(X = x) \\
&= e^{-\lambda} \sum_{x=0}^{\infty} \frac{(e^t\lambda)^x}{x!} \\
&= e^{-\lambda} \cdot e^{e^t\lambda} \\
&= e^{\lambda(e^t-1)}.
\end{aligned}$$

Thus, the moment generating function $M_X(t)$ of a Poisson random variable X with parameter λ is:

$$M_X(t) = e^{\lambda(e^t-1)}.$$

6.4.4 Characteristic Function

The characteristic function $\phi_X(t)$ of X is defined as:

$$\begin{aligned}
\phi_X(t) = E[e^{tX}] &= \sum_{x=0}^{\infty} e^{itx} \cdot P(X = x) \\
&= e^{-\lambda} \sum_{x=0}^{\infty} \frac{(e^{it}\lambda)^x}{x!} \\
&= e^{-\lambda} \cdot e^{e^{it}\lambda} \\
&= e^{\lambda(e^{it}-1)}.
\end{aligned}$$

Thus, the characteristic function $\phi_X(t)$ of a Poisson random variable X with parameter λ is:

$$\phi_X(t) = e^{\lambda(e^{it}-1)}.$$

6.4.5 Approximation of Binomial Distribution Using Poisson Distribution

In statistical theory, the Poisson distribution can be used as an approximation to the Binomial distribution under certain conditions. Specifically, when dealing with scenarios where the number of trials n is very large and the probability of success p is very small, while the product np (which represents the mean of the Binomial distribution) remains constant, the Binomial distribution can be approximated by the Poisson distribution.

Theorem 6.1. ***Approximation of Binomial Distribution by Poisson Distribution:*** *Let X be a Binomial random variable with probability distribution $b(x; n, p)$. When $n \to \infty$, $p \to 0$, and $np \to \lambda$ remains constant, the*

Binomial distribution $b(x; n, p)$ converges to the Poisson distribution $P(x; \lambda)$ as $n \to \infty$. That is,

$$B(x; n, p) \xrightarrow{n \to \infty} P(x; \lambda)$$

where:

$$\lambda = np$$

Problem 6.9. *A box contains 500 electrical switches, each one of which has a probability of 0.005 of being defective. Use the Poisson distribution to make an approximate calculation of the probability that the box contains no more than 3 defective switches.*

Solution:

- Calculate the parameter λ for the Poisson distribution:

  Given:

  $$n = 500 \quad \text{and} \quad p = 0.005$$

  The parameter λ for the Poisson distribution is:

  $$\lambda = np = 500 \times 0.005 = 2.5$$

- Determine the probability of having no more than 3 defective switches using the Poisson distribution with $\lambda = 2.5$:

  The probability that the number of defective switches X is at most 3 is:

  $$P(X \le 3) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)$$

  where X follows a Poisson distribution with parameter $\lambda = 2.5$. The probability mass function of the Poisson distribution is:

  $$P(X = k) = \frac{e^{-\lambda} \lambda^k}{k!}$$

  Therefore:

  $$P(X = 0) = \frac{e^{-2.5} \cdot 2.5^0}{0!} = e^{-2.5}$$
  $$P(X = 1) = \frac{e^{-2.5} \cdot 2.5^1}{1!} = 2.5e^{-2.5}$$
  $$P(X = 2) = \frac{e^{-2.5} \cdot 2.5^2}{2!} = \frac{2.5^2 e^{-2.5}}{2}$$
  $$P(X = 3) = \frac{e^{-2.5} \cdot 2.5^3}{3!} = \frac{2.5^3 e^{-2.5}}{6}$$

  Summing these probabilities gives:

  $$P(X \le 3) = e^{-2.5} + 2.5e^{-2.5} + \frac{2.5^2 e^{-2.5}}{2} + \frac{2.5^3 e^{-2.5}}{6}$$

6.4.6 Python Code for Poisson Distribution

The Poisson distribution models the number of occurrences of an event in a fixed interval of time or space, given a constant mean rate of occurrence. In Python, you can calculate various characteristics of the Poisson distribution using the 'scipy.stats' module. Below is a demonstration of how to compute these characteristics.

Python Code

Here's how you can calculate various characteristics of the Poisson distribution:

```python
import numpy as np
from scipy.stats import poisson

# Define the parameter lambda (mean rate of occurrence)
lambda_ = 4

# Poisson distribution
dist = poisson(mu=lambda_)

# 1. Probability Mass Function (PMF)
x_values = np.arange(0, 11)  # Possible values for the
  random variable
pmf_values = dist.pmf(x_values)
print("PMF values for x = 0 to 10:", pmf_values)

# 2. Cumulative Distribution Function (CDF)
cdf_values = dist.cdf(x_values)
print("CDF values for x = 0 to 10:", cdf_values)

# 3. Mean (Expected Value)
mean = dist.mean()
print("Mean (Expected Value):", mean)

# 4. Variance
variance = dist.var()
print("Variance:", variance)

# 5. Probability Generating Function (PGF)
def pgf(t, lambda_):
  return np.exp(lambda_ * (np.exp(t) - 1))

# PGF values at t = 0 and t = 1
pgf_values = [pgf(t, lambda_) for t in [0, 1]]
print("PGF values for t = 0 and t = 1:", pgf_values)
```

Explanations

- **Probability Mass Function (PMF):**
  - The PMF provides the probability of observing a certain number of events. Compute these probabilities using `dist.pmf(x_values)`.
- **Cumulative Distribution Function (CDF):**
  - The CDF provides the cumulative probability up to each number of events. Compute these cumulative probabilities using `dist.cdf(x_values)`.
- **Mean (Expected Value):**
  - The mean of a Poisson distribution is λ. Compute this using `dist.mean()`.
- **Variance:**
  - The variance of a Poisson distribution is also λ. Compute this using `dist.var()`.
- **Probability Generating Function (PGF):**
  - The PGF for a Poisson distribution is given by $G_X(t) = \exp(\lambda \cdot (\exp(t) - 1))$. Define a function `pgf(t, lambda_)` to compute PGF values for specific t values (e.g., 0 and 1).

6.4.7 Exercises

1. The number of patients arriving at a clinic follows a Poisson distribution with a mean of 3 patients per hour.
   (a) What is the probability that exactly 5 patients will arrive in an hour?
   (b) What is the probability that at most 2 patients will arrive in an hour?
   (c) What is the expected number of patients arriving in 3 hours?
2. A call center receives an average of 10 calls per hour. Assume the number of calls follows a Poisson distribution.
   (a) What is the probability that the call center receives exactly 12 calls in an hour?
   (b) Calculate the probability of receiving fewer than 5 calls in an hour.
   (c) Determine the probability of receiving more than 15 calls in an hour.
3. On a particular stretch of highway, the number of traffic accidents follows a Poisson distribution with an average rate of 3 accidents per month.

   (a) What is the probability of exactly 2 accidents occurring in a month?
   (b) Find the probability that there will be no accidents in a given month.
   (c) Calculate the probability of having 4 or more accidents in a month.

4. A rare disease affects an average of 0.2 patients per 1000 individuals in a population. Assume the number of affected individuals follows a Poisson distribution.
   (a) What is the probability of finding exactly 1 patient with the disease in a sample of 1000 individuals?
   (b) Determine the probability of finding no patients with the disease in a sample of 1000 individuals.
   (c) Find the probability of discovering 2 or more patients with the disease in a sample of 1000 individuals.
5. An employee receives an average of 8 emails per day. Assume the number of emails follows a Poisson distribution.
   (a) What is the probability of receiving exactly 10 emails in a day?
   (b) Calculate the probability of receiving fewer than 6 emails in a day.
   (c) Determine the probability of receiving more than 12 emails in a day.
6. A retail store has an average of 20 customers arriving per hour. The number of customer arrivals follows a Poisson distribution.
   (a) What is the probability that exactly 25 customers will arrive in an hour?
   (b) Find the probability that fewer than 15 customers arrive in an hour.
   (c) Calculate the probability of having 30 or more customers in an hour.
7. A factory produces an average of 2 defective items per 1000 items produced. Assume the number of defective items follows a Poisson distribution.
   (a) What is the probability of finding exactly 3 defective items in a batch of 1000 items?
   (b) Determine the probability of finding no defective items in a batch of 1000 items.
   (c) Calculate the probability of finding 1 or more defective items in a batch of 1000 items.
8. In a laboratory experiment, rare events occur at an average rate of 0.4 events per hour. Assume these events follow a Poisson distribution.
   (a) What is the probability of observing exactly 2 events in an hour?

(b) Find the probability of observing no events in an hour.

(c) Determine the probability of observing more than 1 event in an hour.

9. A software program encounters an average of 5 errors per week. Assume the number of errors follows a Poisson distribution.

   (a) What is the probability of encountering exactly 7 errors in a week?

   (b) Calculate the probability of encountering fewer than 4 errors in a week.

   (c) Find the probability of encountering 8 or more errors in a week.

10. A library has an average of 12 book checkouts per day. The number of checkouts follows a Poisson distribution.

   (a) What is the probability of exactly 10 book checkouts in a day?

   (b) Determine the probability of having 15 or more book checkouts in a day.

   (c) Calculate the probability of having fewer than 8 book checkouts in a day.

11. In a small town, a rare disease has an average occurrence rate of 1 case per month. Assume the number of cases follows a Poisson distribution.

   (a) What is the probability of having exactly 2 cases of the disease in a month?

   (b) Find the probability of having no cases of the disease in a month.

   (c) Calculate the probability of having at least 1 case of the disease in a month.

12. In a certain industrial facility, accidents occur infrequently. It is known that the probability of an accident on any given day is 0.005 and accidents are independent of each other.

   (a) What is the probability that in any given period of 400 days there will be an accident on one day?

   (b) What is the probability that there are at most three days with an accident?

13. In a manufacturing process where glass products are made, defects or bubbles occur, occasionally rendering the piece undesirable for marketing. It is known that, on average, 1 in every 1000 of these items produced has one or more bubbles. What is the probability that a random sample of 8000 will yield fewer than 7 items possessing bubbles?

6.5 Concluding Remarks

In this chapter, we have examined key discrete probability distributions that are integral to data science: the Bernoulli, Binomial, and Poisson distributions. Understanding these distributions equips data scientists with powerful tools for analyzing categorical data and modeling discrete events.

Mastering these discrete distributions enhances our ability to model and interpret data effectively, leading to more accurate predictions and insights. As we transition to continuous probability distributions in the next chapter, we will expand our analytical toolkit to handle a broader range of data types and modeling scenarios.

Chapter 7

Some Continuous Probability Distributions

7.1 Introduction

In probability theory and statistics, continuous probability distributions play a fundamental role in modeling and analyzing real-world phenomena. Unlike discrete distributions, which are defined for countable outcomes, continuous distributions are used to describe outcomes that can take on any value within a given range. This chapter delves into some of the most widely used continuous probability distributions, including the Uniform, Exponential, and Normal distributions.

Continuous probability distributions are integral to various fields, such as engineering, economics, and the natural sciences, due to their ability to represent diverse processes and events accurately. Understanding these distributions enables us to calculate probabilities and make inferences about populations based on sample data.

We begin with the Uniform distribution, which serves as a simple model for random variables that have equally likely outcomes over a specific interval. Following this, we explore the Exponential distribution, commonly used to model the time between events in a Poisson process. We then delve into the Normal distribution, arguably the most important distribution in statistics, due to the Central Limit Theorem's implication that it approximates many natural phenomena.

Each section will provide a detailed definition of the distribution, its properties, and practical examples to illustrate its application. Additionally, exercises are included to reinforce the concepts and allow for hands-on practice in calculating probabilities and understanding the distribution's behavior.

7.2 Uniform Distribution

The **uniform distribution** is a type of probability distribution in which all outcomes are equally likely. There are two main types of uniform distributions: (i) Discrete Uniform Distribution and (ii) Continuous Uniform Distribution.

In the discrete uniform distribution, the number of outcomes is finite, and each outcome is equally likely. The probability mass function of the discrete uniform distribution is

$$P(X = x) = \frac{1}{n} \quad ; \quad x = 0, 1, 2, \ldots, n$$

where n is the number of possible outcomes.

In the continuous uniform distribution, the number of outcomes is infinite within a certain interval, where every outcome within that interval is equally likely.

A random variable X is said to follow a continuous uniform distribution on the interval $[a, b]$, denoted $X \sim U(a, b)$, if its probability density function (pdf) is:

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{for } a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$

The plot of $X \sim U(a, b)$ where X is uniformly distributed between a and b are presented in Figure 7.1.

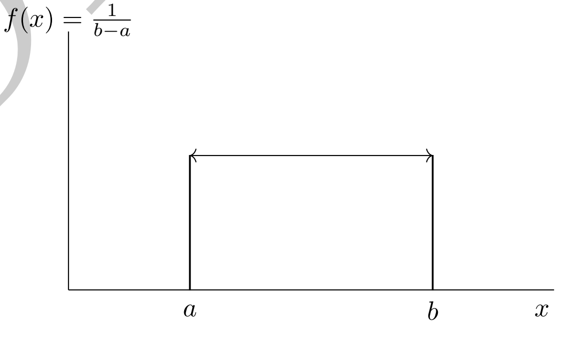


Figure 7.1: The plot of $X \sim U(a, b)$.

Mean

Expected Value Formula: The expected value $E(X)$ is given by:

$$
\begin{aligned}
E(X) &= \int_{-\infty}^{\infty} x f_X(x)\,dx = \int_a^b x \cdot \frac{1}{b-a}\,dx. \\
&= \frac{1}{b-a}\int_a^b x\,dx = \frac{b^2}{2} - \frac{a^2}{2} \\
&= \frac{(b-a)(b+a)}{2(b-a)} = \frac{b+a}{2}.
\end{aligned}
$$

Hence,

$$
\text{Mean} = E(X) = \frac{a+b}{2}
$$

Variance and Standard Deviation

Expected Value: We already have the expected value $E(X)$ for a uniform random variable X over $[a,b]$:

$$
E(X) = \frac{a+b}{2}.
$$

Expected Value of X^2: To find the variance, we first need $E(X^2)$. This is given by:

$$
\begin{aligned}
E(X^2) &= \int_a^b x^2 f_X(x)\,dx = \frac{1}{b-a}\int_a^b x^2\,dx \\
&= \frac{1}{b-a}\left[\frac{x^3}{3}\right]_a^b = \frac{b^3}{3} - \frac{a^3}{3} \\
&= \frac{1}{b-a}\left(\frac{b^3}{3} - \frac{a^3}{3}\right) = \frac{b^3-a^3}{3(b-a)} \\
&= \frac{b^2+ab+a^2}{3}.
\end{aligned}
$$

Hence, the variance is

$$
\begin{aligned}
\text{Var}(X) &= \frac{b^2+ab+a^2}{3} - \left(\frac{a+b}{2}\right)^2 \\
&= \frac{4(b^2+ab+a^2) - 3(a^2+2ab+b^2)}{12} \\
&= \frac{(b-a)^2}{12}
\end{aligned}
$$

and the standard deviation is

$$\text{Standard Deviation} = \frac{b-a}{\sqrt{12}}.$$

Problem 7.1. *Suppose the time (in minutes) it takes for a customer to be served at a coffee shop follows a continuous uniform distribution between 5 and 15 minutes.*

(a). What is the probability that a customer is served within 10 minutes?

(b). What is the expected time for a customer to be served?

(c). What is the variance of the time for a customer to be served?

Solution (a). **Probability that a customer is served within 10 minutes:**

Let X be the time taken to be served, and $X \sim \text{Uniform}(5, 15)$.

To find $P(X \leq 10)$:

The pdf of X is given by:

$$f(x) = \frac{1}{15-5} = \frac{1}{10} \text{ for } 5 \leq x \leq 15$$

The probability is given by the integral of the pdf from 5 to 10:

$$P(X \leq 10) = \int_5^{10} \frac{1}{10}\,dx = \frac{10-5}{10} = 0.5$$

(b). **Expected time:**

The expected value for a continuous uniform distribution $X \sim \text{Uniform}(a, b)$ is:

$$E(X) = \frac{a+b}{2}$$

Here, $a = 5$ and $b = 15$:

$$E(X) = \frac{5+15}{2} = 10 \text{ minutes}$$

(c). **Variance of the time:**

The variance for a continuous uniform distribution $X \sim \text{Uniform}(a, b)$ is:

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

Here, $a = 5$ and $b = 15$:

$$\text{Var}(X) = \frac{(15-5)^2}{12} = \frac{100}{12} \approx 8.33$$

Problem 7.2. *Suppose you are designing a random number generator that outputs a number between 1 and 100. Each number in this range is equally likely to be selected.*

(a). What is the probability that the generator outputs a number between 20 and 50?

(b). Determine the mean and variance of the numbers generated by this random number generator.

(c). If you generate 10,000 numbers, what is the expected number of times a number between 20 and 50 is generated?

Solution

(a). **Probability Calculation**

Since the numbers generated are uniformly distributed between 1 and 100, we can model this using a continuous uniform distribution $U(a, b)$ with $a = 1$ and $b = 100$.

The probability density function (pdf) is:

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{for } a \leq x \leq b, \\ 0 & \text{otherwise.} \end{cases}$$

Here, $f(x) = \frac{1}{100-1} = \frac{1}{99}$ for $1 \leq x \leq 100$.

To find the probability that the generator outputs a number between 20 and 50, we calculate:

$$\begin{aligned} P(20 \leq X \leq 50) &= \int_{20}^{50} f(x)\, dx = \int_{20}^{50} \frac{1}{99}\, dx \\ &= \frac{1}{99} \int_{20}^{50} dx \\ &= \frac{50-20}{99} \\ &\approx 0.303. \end{aligned}$$

(b). **Mean and Variance**

For a uniform distribution $U(a, b)$:

$$\text{Mean} = \mu = \frac{a+b}{2} = \frac{1+100}{2} = 50.5,$$

$$\text{Variance} = \sigma^2 = \frac{(b-a)^2}{12} = \frac{(100-1)^2}{12} = \frac{99^2}{12} = \frac{9801}{12} \approx 816.75.$$

(c). **Expected Number of Times a Number Between 20 and 50 is Generated**

The expected number of times a number between 20 and 50 is generated out of 10,000 numbers can be found by multiplying the probability by the total number of trials:

$$E(\text{number of times}) = P(20 \leq X \leq 50) \times 10000 = 0.303 \times 10000 = 3030.$$

Thus, we expect the number generator to output a number between 20 and 50 approximately 3030 times out of 10,000 trials.

7.2.1 Python Code for Uniform Distribution Characteristics

The Uniform distribution models scenarios where all outcomes are equally likely within a given interval. Below is Python code demonstrating how to compute various characteristics for both Continuous and Discrete Uniform distributions.

Continuous Uniform Distribution

```
import numpy as np
from scipy.stats import uniform

# Define the parameters
a = 0  # Lower bound
b = 10  # Upper bound

# Define the Continuous Uniform distribution
dist_continuous = uniform(loc=a, scale=b-a)

# 1. Probability Density Function (PDF)
x_values = np.linspace(a, b, 100)
pdf_values = dist_continuous.pdf(x_values)
print("PDF values for x from {} to {}:".format(a, b),
  pdf_values)

# 2. Cumulative Distribution Function (CDF)
cdf_values = dist_continuous.cdf(x_values)
print("CDF values for x from {} to {}:".format(a, b),
  cdf_values)

# 3. Mean (Expected Value)
mean_continuous = dist_continuous.mean()
print("Mean (Expected Value):", mean_continuous)

```

```python
# 4. Variance
variance_continuous = dist_continuous.var()
print("Variance:", variance_continuous)

# 5. Standard Deviation
std_dev_continuous = dist_continuous.std()
print("Standard Deviation:", std_dev_continuous)

# 6. Quantiles
quantiles_continuous = dist_continuous.ppf([0.25, 0.5, 0.75])  # 25th, 50th (median), and 75th percentiles
print("Quantiles at 0.25, 0.5, and 0.75:", quantiles_continuous)

# 7. Percentiles
percentiles_continuous = dist_continuous.ppf([0.1, 0.9])  # 10th and 90th percentiles
print("Percentiles at 0.1 and 0.9:", percentiles_continuous)
```

Discrete Uniform Distribution

```python
import numpy as np

# Define the Discrete Uniform parameters
a_discrete = 1
b_discrete = 10

def pmf_discrete(x, a, b):
  if a <= x <= b:
    return 1 / (b - a + 1)
  else:
    return 0

def cdf_discrete(x, a, b):
  if x < a:
    return 0
  elif a <= x <= b:
    return (x - a + 1) / (b - a + 1)
  else:
    return 1

# PMF values
x_discrete_values = np.arange(a_discrete, b_discrete + 1)
pmf_values_discrete = [pmf_discrete(x, a_discrete, b_discrete) for x in x_discrete_values]
print("PMF values for x from {} to {}:".format(a_discrete, b_discrete), pmf_values_discrete)
```

```python

# CDF values
cdf_values_discrete = [cdf_discrete(x, a_discrete, b_discrete) for x in x_discrete_values]
print("CDF values for x from {} to {}:".format(a_discrete, b_discrete), cdf_values_discrete)

# Mean (Expected Value)
mean_discrete = (a_discrete + b_discrete) / 2
print("Mean (Expected Value):", mean_discrete)

# Variance
variance_discrete = ((b_discrete - a_discrete + 1) ** 2 - 1) / 12
print("Variance:", variance_discrete)

# Standard Deviation
std_dev_discrete = np.sqrt(variance_discrete)
print("Standard Deviation:", std_dev_discrete)
```

7.2.2 Exercises

1. Consider a discrete uniform distribution where X takes values $\{1, 2, 3, \ldots, n\}$. Show that the expected value $E(X)$ is $\frac{n+1}{2}$.

2. Find the cumulative distribution function (cdf) of a continuous uniform random variable $X \sim U(a, b)$.

3. The cholesterol level of adults in a certain region follows a uniform distribution between 150 mg/dL and 250 mg/dL.

4. If $X \sim U(0, 1)$, find the distribution of $Y = a + (b - a)X$.

5. Suppose the waiting time for a bus is uniformly distributed between 0 and 30 minutes. What is the probability that a person will wait more than 20 minutes?

6. A factory produces items with weights that are uniformly distributed between 50 grams and 150 grams. What is the probability that a randomly chosen item weighs between 80 grams and 120 grams?

7. The cholesterol level of adults in a certain region follows a uniform distribution between 150 mg/dL and 250 mg/dL.

   (a) Write the probability density function $f(x)$ of the cholesterol level.

   (b) What is the probability that a randomly selected adult has a cholesterol level between 180 mg/dL and 220 mg/dL?

   (c) Find the mean and variance of the cholesterol level.

8. A machine produces metal rods that are uniformly distributed in length between 98 cm and 102 cm. What is the probability that a randomly selected rod is between 99 cm and 101 cm in length?

9. In a quality control process, the time to inspect a product is uniformly distributed between 1 and 5 minutes. Find the probability that the inspection time for a randomly chosen product is more than 4 minutes.

10. The download speed of a certain internet connection is uniformly distributed between 10 Mbps and 100 Mbps. What is the probability that the download speed at any given time is less than 50 Mbps?

11. The delivery time for a package from a warehouse to a customer is uniformly distributed between 2 and 7 days. What is the probability that a package will be delivered in less than 4 days?

12. The time a wildlife photographer waits to see a particular bird is uniformly distributed between 30 minutes and 3 hours. Find the probability that the wait time is more than 2 hours.

13. The fuel efficiency of a car is uniformly distributed between 15 and 25 miles per gallon. What is the probability that the car's fuel efficiency is between 18 and 22 miles per gallon?

7.3 Exponential Distribution

The **Exponential Distribution** is a continuous probability distribution often used to model the time between events in a Poisson process, where events occur continuously and independently at a constant average rate. It is widely applied in various fields, including data science, to analyze the duration or time between events.

Definition: The probability density function (pdf) of an exponential distribution is given by:

$$f(x;\lambda) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0, \\ 0 & x < 0, \end{cases} \tag{7.1}$$

where $\lambda > 0$ is the rate parameter.

The pdf of the exponential random variable is given in Equation (7.1) and the density plot is present in Figure 7.2 for value of $\lambda = 0.04$.

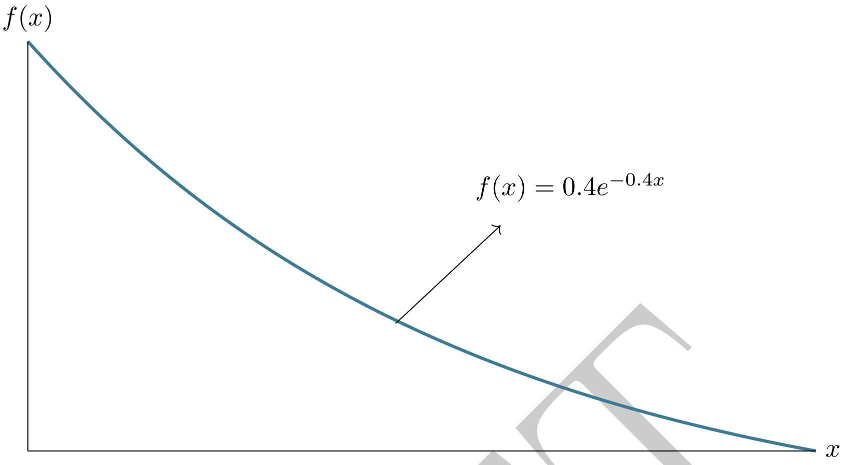


Figure 7.2: The probability density function $f(x) = 0.4e^{0.4x}$.

This probability is the area under the probability density function between the points $a = 1$ and $b = 2$ as illustrated in Figure 7.3.

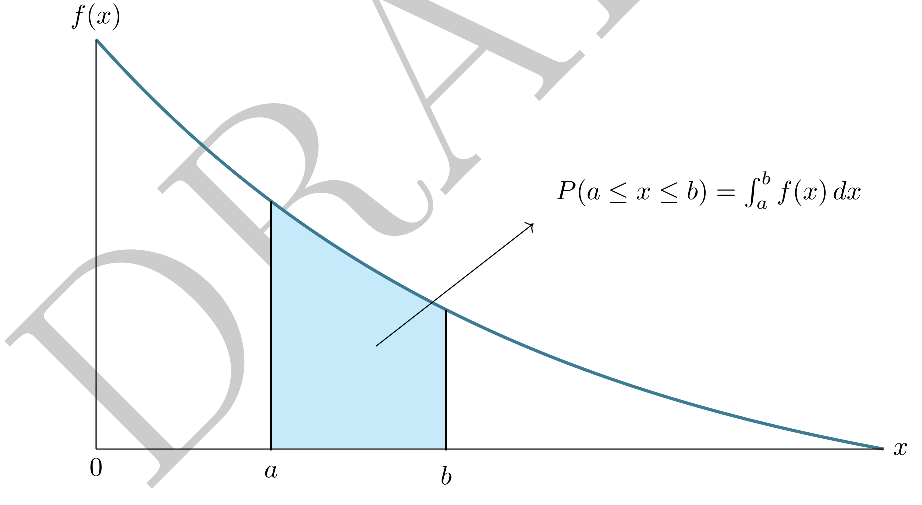


Figure 7.3: The area under the probability density function $f(x)$ between a and b.

The cdf is obtained by integrating the pdf:

$$\begin{aligned} F(x) &= \int_0^x \lambda e^{-\lambda t}\, dt \\ &= -e^{-\lambda t}\Big|_0^x \\ &= -e^{-\lambda x} + e^0 \\ &= 1 - e^{-\lambda x}. \end{aligned}$$

Thus, the CDF $F(x)$ for the exponential distribution is:

$$F(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-\lambda x} & x \geq 0 \end{cases}$$

The CDF of the exponential distribution for $\lambda = 0.4$ is $F(x) = 1 - e^{-0.4x}$ and the graphical presentation is presented in Figure 7.4.

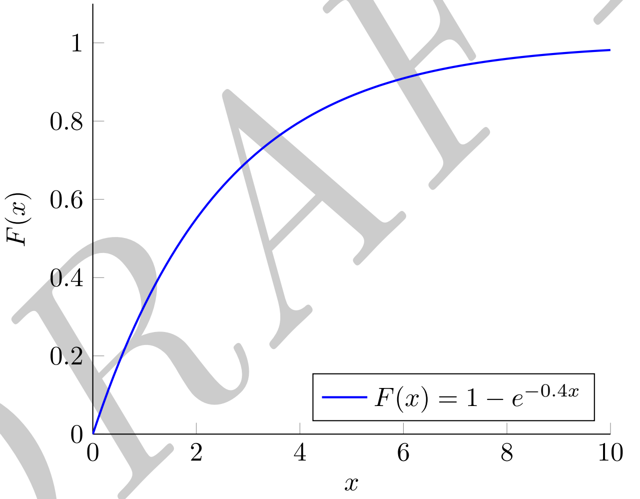


Figure 7.4: CDF of the Exponential Distribution with $\lambda = 0.4$

Example

Consider a scenario where a data center experiences server failures, and the time between failures follows an exponential distribution with a mean of 10 days. The rate parameter λ can be calculated as:

$$\lambda = \frac{1}{\text{Mean}} = \frac{1}{10} = 0.1.$$

The probability that a server will fail within the next 5 days is given by the CDF:

$$P(X \leq 5) = 1 - e^{-0.1 \times 5} = 1 - e^{-0.5} \approx 0.3935.$$

Problem 7.3. *In a hospital, the time between arrivals of patients in the emergency room follows an exponential distribution with an average time of 15 minutes.*

(a) What is the probability that the time between two successive arrivals is more than 20 minutes?

(b) What is the probability that the time between two successive arrivals is less than 10 minutes?

(c) Calculate the expected time between two successive arrivals and its standard deviation.

Solution 1. Let X be the time between arrivals, which follows an exponential distribution with parameter λ. The rate parameter λ is the reciprocal of the mean, so $\lambda = \frac{1}{15}$ per minute. The probability that the time between two successive arrivals is more than 20 minutes is calculated as follows:

$$P(X > 20) = 1 - P(X \leq 20) = 1 - F_X(20) = 1 - \left(1 - e^{-\lambda \cdot 20}\right) = e^{-\lambda \cdot 20}$$

Substituting $\lambda = \frac{1}{15}$:

$$P(X > 20) = e^{-\frac{20}{15}} = e^{-\frac{4}{3}} \approx 0.2636$$

Thus, the probability that the time between two successive arrivals is more than 20 minutes is approximately 0.2636.

2. The probability that the time between two successive arrivals is less than 10 minutes is calculated as follows:

$$P(X < 10) = F_X(10) = 1 - e^{-\lambda \cdot 10}$$

Substituting $\lambda = \frac{1}{15}$:

$$P(X < 10) = 1 - e^{-\frac{10}{15}} = 1 - e^{-\frac{2}{3}} \approx 0.4866$$

Thus, the probability that the time between two successive arrivals is less than 10 minutes is approximately 0.4866.

3. The expected time between two successive arrivals (the mean of the exponential distribution) is given by:

$$E(X) = \frac{1}{\lambda} = 15 \text{ minutes}$$

The standard deviation of the time between two successive arrivals is the same as the mean for an exponential distribution, so:

$$\text{Standard deviation} = \frac{1}{\lambda} = 15 \text{ minutes}$$

Therefore, the expected time between two successive arrivals is 15 minutes, and the standard deviation is also 15 minutes.

7.3.1 Properties of the Exponential Distribution

Let X be an exponential random variable with rate parameter λ. The following are the key properties of the exponential distribution:

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

$$F_X(x) = \begin{cases} 1 - e^{-\lambda x} & x \geq 0, \\ 0 & x < 0. \end{cases}$$

- **Mean (Expected Value):**

$$E[X] = \int_0^\infty x \cdot \lambda e^{-\lambda x}\, dx = \frac{1}{\lambda}.$$

- **Variance:**

$$\mathrm{Var}(X) = E(X^2) - [E(X)]^2 = \frac{1}{\lambda^2}.$$

- **Standard Deviation:**

$$\sigma_X = \frac{1}{\lambda}.$$

- **Memoryless Property:** The exponential distribution has the memoryless property, which states that:

$$P(X > s + t \mid X > s) = P(X > t) \quad \text{for all } s, t \geq 0.$$

- **Moment Generating Function (MGF):**

$$M_X(t) = E[e^{tX}] = \frac{\lambda}{\lambda - t}, \quad \text{for } t < \lambda.$$

- **Characteristic Function:**

$$\varphi_X(t) = E[e^{itX}] = \frac{\lambda}{\lambda - it}, \quad \text{for } t \in \mathbb{R}.$$

- **Quantile Function:** The quantile function (inverse of the CDF) for $0 < p < 1$ is given by:

$$Q(p) = F_X^{-1}(p) = -\frac{1}{\lambda} \ln(1-p).$$

- **Relationship with the Poisson Process:** If $X \sim \text{Exponential}(\lambda)$, it can be interpreted as the waiting time between events in a Poisson process with rate λ.

- **Sum of Independent Exponential Variables:** The sum of n independent exponential random variables with the same rate parameter λ follows a Gamma distribution:

$$Y = \sum_{i=1}^{n} X_i \sim \text{Gamma}(n, \lambda).$$

  For $n = 1$, the Gamma distribution reduces to the exponential distribution.

- **Relationship with Other Distributions:** - If $X \sim \text{Exponential}(\lambda)$, then $X \sim \text{Gamma}(1, \lambda)$. - If $X \sim \text{Exponential}(\lambda)$, then $Y = \frac{X}{\beta} \sim \text{Exponential}\left(\frac{\lambda}{\beta}\right)$ for $\beta > 0$.

Theorem 7.1. *The characteristic function of an exponential random variable X with rate parameter λ is given by:*

$$\varphi_X(t) = \frac{\lambda}{\lambda - it}, \quad \textit{for } t \in \mathbb{R}.$$

Solution
To find the characteristic function $\varphi_X(t)$, we need to calculate the expected value $E[e^{itX}]$:

$$\varphi_X(t) = E[e^{itX}] = \int_0^\infty e^{itx} \lambda e^{-\lambda x}\, dx = \lambda \int_0^\infty e^{-(\lambda - it)x}\, dx.$$

The integral is of the form $\int_0^\infty e^{-ax}\, dx = \frac{1}{a}$ for $\Re(a) > 0$:

$$\varphi_X(t) = \lambda \left[\frac{1}{\lambda - it}\right] = \frac{\lambda}{\lambda - it}.$$

Thus, the characteristic function of X is:

$$\varphi_X(t) = \frac{\lambda}{\lambda - it}.$$

Theorem 7.2. *The moment generating function (MGF) of an exponential random variable X with rate parameter λ is given by:*

$$M_X(t) = \frac{\lambda}{\lambda - t}, \quad \text{for } t < \lambda.$$

Solution
To find the moment generating function $M_X(t)$, we need to calculate the expected value $E[e^{tX}]$:

$$M_X(t) = E[e^{tX}] = \int_0^\infty e^{tx} \lambda e^{-\lambda x}\, dx = \lambda \int_0^\infty e^{-(\lambda - t)x}\, dx = \lambda \left[\frac{1}{\lambda - t} \right].$$

Thus, the moment generating function of X is:

$$M_X(t) = \frac{\lambda}{\lambda - t}, \quad \text{for } t < \lambda.$$

7.3.2 Memoryless Property

In probability and statistics, the concept of memorylessness refers to a characteristic of specific probability distributions. It indicates scenarios where the duration already waited for an event does not influence the remaining waiting time. To accurately represent memoryless situations, the past state of the system must be ignored – the probabilities remain independent of the process's history.

Only two types of distributions exhibit the memoryless property: geometric and exponential probability distributions. The exponential distribution has the memoryless property, which states that the probability of an event occurring in the next t units of time is independent of how much time has already elapsed. Mathematically, this can be expressed in the following theorem.

Theorem 7.3 (Memoryless Property)**.** *Let X be an exponentially distributed random variable with rate parameter $\lambda > 0$. Then, for all $s, t \geq 0$,*

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

Proof. The proof of the memoryless property relies on the definition of conditional probability and the exponential distribution's probability density function.

By definition of conditional probability, we have:

$$P(X > s + t \mid X > s) = \frac{P(X > s + t \text{ and } X > s)}{P(X > s)} = \frac{P(X > s + t)}{P(X > s)}.$$

Since X is exponentially distributed with rate parameter λ, the cumulative distribution function (CDF) is:

$$F_X(x) = 1 - e^{-\lambda x} \quad \text{for } x \geq 0,$$

and the survival function (which gives the probability that X is greater than a certain value) is:

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

Using the survival function, we can rewrite the conditional probability:

$$P(X > s+t \mid X > s) = \frac{P(X > s+t)}{P(X > s)} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} = P(X > t).$$

Thus, the memoryless property is proved. □

Problem 7.4. *Suppose that the waiting time for a bus at a certain bus stop is exponentially distributed with a mean waiting time of 10 minutes. Let X denote the waiting time.*

Given that a person has already waited for 5 minutes, the probability that they will have to wait at least an additional 10 minutes is:

$$P(X > 15 \mid X > 5) = P(X > 10) = e^{-10/10} = e^{-1} \approx 0.3679.$$

Problem 7.5. *Assume that the lifetime of a light bulb follows an exponential distribution with a mean lifetime of 1000 hours. Let X be the lifetime of the light bulb.*

If a light bulb has already been used for 800 hours, the probability that it will last at least an additional 500 hours is:

$$P(X > 1300 \mid X > 800) = P(X > 500) = e^{-500/1000} = e^{-0.5} \approx 0.6065.$$

Problem 7.6. *Consider a system with a component whose time to failure is exponentially distributed with a rate parameter $\lambda = 0.01$ failures per hour. Let X represent the time to failure of the component.*

If the component has been operational for 100 hours without failure, the probability that it will operate for at least another 50 hours is:

$$P(X > 150 \mid X > 100) = P(X > 50) = e^{-0.01\times 50} = e^{-0.5} \approx 0.6065.$$

Applications in Data Science

The important applications of exponential distribution are mentioned in the following:

1. **Survival Analysis**: Exponential distributions are used to model the time until an event of interest (e.g., failure of a machine, death of a patient). The constant hazard rate assumption simplifies the analysis and is often used in survival analysis.

2. **Queuing Theory**: In scenarios like customer service or network traffic, the exponential distribution models the time between arrivals of customers or data packets.

3. **Reliability Engineering**: It helps in estimating the lifespan of products and systems, providing insights into maintenance schedules and warranty analysis.

4. **Markov Processes**: As part of continuous-time Markov chains, exponential distributions model the time spent in each state before transitioning to another state.

7.3.3 Python Code for Exponential Distribution Characteristics

The Exponential distribution models the time between events in a Poisson process. Below is Python code demonstrating how to compute various characteristics of the Exponential distribution.

Python Code

```python
import numpy as np
from scipy.stats import expon

# Define the parameter lambda (rate of occurrence)
lambda_ = 1  # Rate parameter
scale = 1 / lambda_  # Scale parameter for scipy's expon

# Exponential distribution
dist = expon(scale=scale)

# 1. Probability Density Function (PDF)
x_values = np.linspace(0, 10, 100)  # Values for the random variable
pdf_values = dist.pdf(x_values)
print("PDF values for x from 0 to 10:", pdf_values)

# 2. Cumulative Distribution Function (CDF)
cdf_values = dist.cdf(x_values)
print("CDF values for x from 0 to 10:", cdf_values)

# 3. Mean (Expected Value)
mean = dist.mean()
print("Mean (Expected Value):", mean)

# 4. Variance
variance = dist.var()
print("Variance:", variance)

# 5. Standard Deviation
std_dev = dist.std()
print("Standard Deviation:", std_dev)
```

```

# 6. Quantiles
quantiles = dist.ppf([0.25, 0.5, 0.75])  # 25th, 50th (
  median), and 75th percentiles
print("Quantiles at 0.25, 0.5, and 0.75:", quantiles)

# 7. Percentiles
percentiles = dist.ppf([0.1, 0.9])  # 10th and 90th
  percentiles
print("Percentiles at 0.1 and 0.9:", percentiles)

# 8. Moment Generating Function (MGF)
def mgf(t, lambda_):
    return 1 / (1 - t / lambda_)  # MGF of Exponential
  distribution for t < lambda_

# MGF values at t = 0.1 and t = 0.5
mgf_values = [mgf(t, lambda_) for t in [0.1, 0.5]]
print("MGF values for t = 0.1 and t = 0.5:", mgf_values)

# 9. Probability Generating Function (PGF)
# Note: The PGF is not typically used for continuous
  distributions like the Exponential distribution.
# Here, we will use the MGF as a substitute.
```

Explanations

- **Probability Density Function (PDF):**
  - The PDF provides the likelihood of each value. Compute these values using `dist.pdf(x_values)`.
- **Cumulative Distribution Function (CDF):**
  - The CDF gives the probability that the random variable is less than or equal to a given value. Compute these values using `dist.cdf(x_values)`.
- **Mean (Expected Value):**
  - The mean (expected value) is $\frac{1}{\lambda}$. Compute this using `dist.mean()`.
- **Variance:**
  - The variance is $\frac{1}{\lambda^2}$. Compute this using `dist.var()`.
- **Standard Deviation:**
  - The standard deviation is $\frac{1}{\lambda}$. Compute this using `dist.std()`.

- **Quantiles:**
  - Quantiles are values below which a given proportion of data falls. Compute these using `dist.ppf()` for the desired quantile probabilities.
- **Percentiles:**
  - Percentiles are specific quantiles, such as the 10th and 90th percentiles. Compute these using `dist.ppf()` for the desired percentiles.
- **Moment Generating Function (MGF):**
  - The MGF for the Exponential distribution is $M_X(t) = \frac{1}{1-t/\lambda}$ for $t < \lambda$. Define a function `mgf(t, lambda_)` to compute MGF values.
- **Probability Generating Function (PGF):**
  - The PGF is generally used for discrete distributions. For continuous distributions like the Exponential, the MGF serves a similar purpose.

7.3.4 Exercises

1. The lifetime of a particular brand of lightbulb is exponentially distributed with a mean of 1000 hours.
   (a) What is the probability that a lightbulb lasts more than 1200 hours?
   (b) What is the probability that a lightbulb lasts between 800 and 1200 hours?
2. A machine in a factory breaks down on average once every 500 hours. The time between breakdowns is exponentially distributed.
   (a) What is the probability that the machine will operate for at least 1000 hours without a breakdown?
   (b) Determine the probability that the machine will break down within the next 200 hours.
3. The waiting time for a specific genetic test result from a laboratory follows an exponential distribution with a mean of 2 days.
   (a) Find the probability that the test result will be available in less than 1 day.
   (b) Find the probability that the test result will take more than 3 days.

4. The time until failure of a critical component in a medical device follows an exponential distribution with a mean of 5 years.
   (a) Calculate the probability that the component will fail within the first 3 years.
   (b) Determine the probability that the component will last more than 7 years.
5. In a call center, the time between consecutive calls follows an exponential distribution with a mean time of 4 minutes.
   (a) What is the probability that the next call comes within 2 minutes?
   (b) What is the probability that the next call will not come for at least 6 minutes?
6. The time between arrivals of buses at a particular bus stop follows an exponential distribution with an average time of 20 minutes.
   (a) What is the probability that a bus will arrive in the next 5 minutes?
   (b) Calculate the probability that you will have to wait more than 25 minutes for the next bus.
7. The lifespan of a certain species of bacteria follows an exponential distribution with a mean of 24 hours.
   (a) Find the probability that a bacterium lives longer than 30 hours.
   (b) Determine the probability that a bacterium lives between 10 and 20 hours.
8. The response time of a server to a network request is exponentially distributed with an average time of 0.5 seconds.
   (a) What is the probability that the server responds in less than 0.3 seconds?
   (b) Calculate the probability that the server takes more than 1 second to respond.
9. The time it takes for a chemical reaction to complete in a lab experiment follows an exponential distribution with a mean of 45 minutes.
   (a) What is the probability that the reaction will complete in less than 30 minutes?
   (b) What is the probability that the reaction will take more than 60 minutes to complete?
10. The lifespan of a certain species of laboratory mice follows an exponential distribution with a mean lifespan of 2.5 years.

   (a) Determine the probability density function (pdf) of the lifespan.
   (b) What is the median lifespan of the mice?
   (c) Calculate the variance and standard deviation of the lifespan.
   (d) What is the probability that a randomly selected mouse lives more than 3 years?

11. The time (in hours) until recovery from a certain disease follows an exponential distribution with a mean of 2 hours.
   (a) What is the probability that a patient will recover in less than 1 hour?
   (b) What is the probability that a patient will recover between 1 and 3 hours?
   (c) Find the median recovery time.

7.4 Normal Distribution

In the field of data science, the **normal distribution** plays a pivotal role due to its ubiquitous nature and the mathematical properties that simplify analysis. Often referred to as the **Gaussian distribution**, the normal distribution is characterized by its **symmetric**, **bell-shaped curve**. It is instrumental in various statistical methods, including hypothesis testing, regression analysis, and many machine learning algorithms.

Consider a scenario where we need to understand the distribution of heights in a population. Heights of individuals in a large population tend to cluster around a central value, with fewer individuals having heights significantly shorter or taller than the average. To analyze this, we need a mathematical model that can accurately describe this distribution of heights. Let's assume we have measured the heights of a large number of individuals in a population and want to find a suitable model to describe this data. We observe that most individuals have heights close to the average, with fewer individuals having extremely short or tall heights. This pattern suggests that the heights may follow a normal distribution.

The normal distribution is particularly valuable because of the **Central Limit Theorem**, which states that the sum of a large number of independent, identically distributed variables tends toward a normal distribution, regardless of the original distribution of the variables. This makes the normal distribution a powerful tool for modeling real-world phenomena and for making inferences about populations based on sample data.

The normal distribution is characterized by its bell-shaped curve, which is symmetric about its mean. The parameters of the normal distribution are the mean (μ) and the standard deviation (σ). The mean represents the central value of the distribution, while the standard deviation measures the spread of the distribution.

7.4.1 Definition of the Normal Distribution

Normal Distribution: A random variable X is said to follow a normal distribution if its probability density function (pdf) is given by:

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

where μ is the mean and σ is the standard deviation of the distribution. The graphical presentation of a normal distribution presented in Figure 7.5. The probability density function is a bell-shaped curve that is symmetric about μ. The notation

$$X \sim N(\mu, \sigma^2)$$

denotes that the random variable X has a normal distribution with mean μ and variance σ^2.

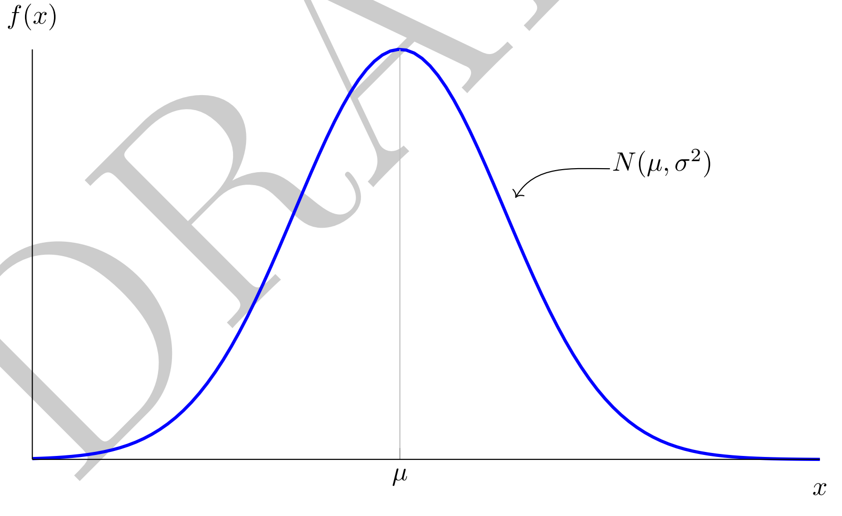


Figure 7.5: The graphical presentation of a normal distribution.

The mean and the variance

$$E(X) = \mu \quad \text{and} \quad \text{Var}(X) = \sigma^2$$

of the distribution.

The probability density function of a normal random variable is symmetric around the mean value μ and exhibits a "bell-shaped" curve. Figure 7.6 displays the probability density functions of normal distributions with $\mu = 5, \sigma = 2$ and $\mu = 10, \sigma = 2$. It illustrates that while altering the mean value μ shifts the location of the density function, it does not affect its shape. In contrast, Figure 7.7 presents the probability density functions of normal distributions with $\mu = 5, \sigma = 2$ and $\mu = 5, \sigma = 0.5$. Here, the central position of the density function remains the same, but its shape changes. Larger values of the variance σ^2 lead to wider, flatter bell-shaped curves, whereas smaller values of the variance σ^2 produce narrower, sharper bell-shaped curves.

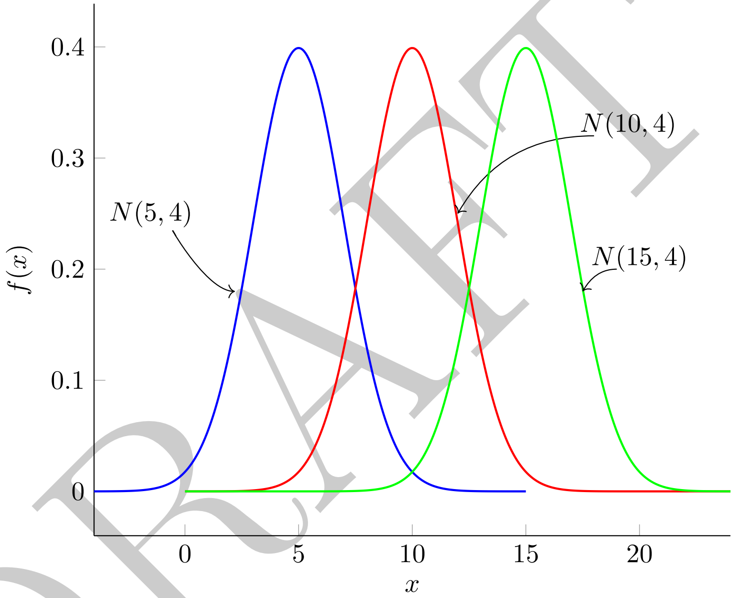


Figure 7.6: The effect of changing the mean of a normal distribution.

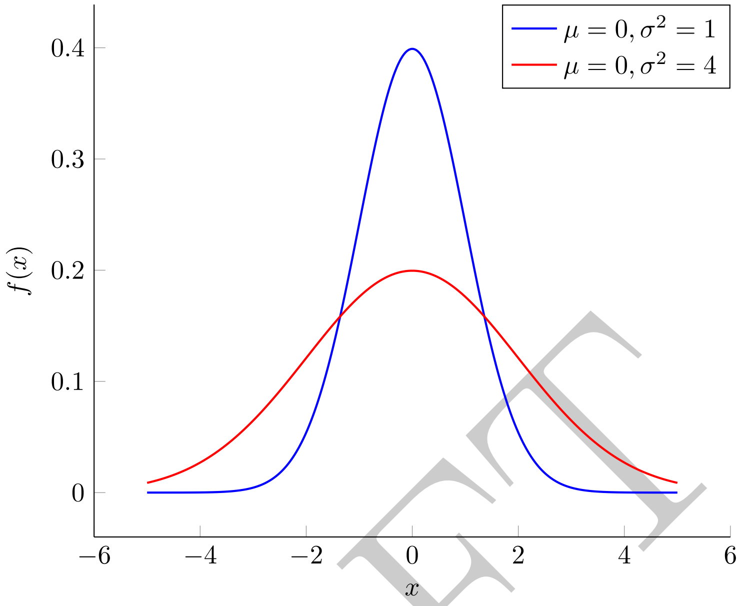


Figure 7.7: The effect of changing the variance of a normal distribution.

7.4.2 Properties of the Normal Distribution

The normal distribution, also known as the Gaussian distribution, is one of the most important probability distributions in statistics. The normal distribution has several important properties:

1. **Symmetry:** The normal distribution is symmetric about its mean μ. That is $f(\mu - x) = f(\mu + x)$. This means that the left half of the distribution is a mirror image of the right half.

2. **Bell-Shaped Curve:** The pdf of the normal distribution forms a bell-shaped curve that is unimodal, meaning it has a single peak at the mean μ.

3. **Mean, Median, and Mode:** For a normal distribution, the mean, median, and mode are all equal and located at μ.

4. **Total Area:** The total area under the curve and above the horizontal axis is equal to 1.

5. **Inflection Points:** The points at which the curve changes concavity are located at $\mu - \sigma$ and $\mu + \sigma$.

6. **Empirical Rule (68-95-99.7 Rule):** Approximately 68% of the data lies within one standard deviation of the mean ($\mu \pm \sigma$), about 95% within

two standard deviations ($\mu \pm 2\sigma$), and about 99.7% within three standard deviations ($\mu \pm 3\sigma$). See the Figure 7.8.

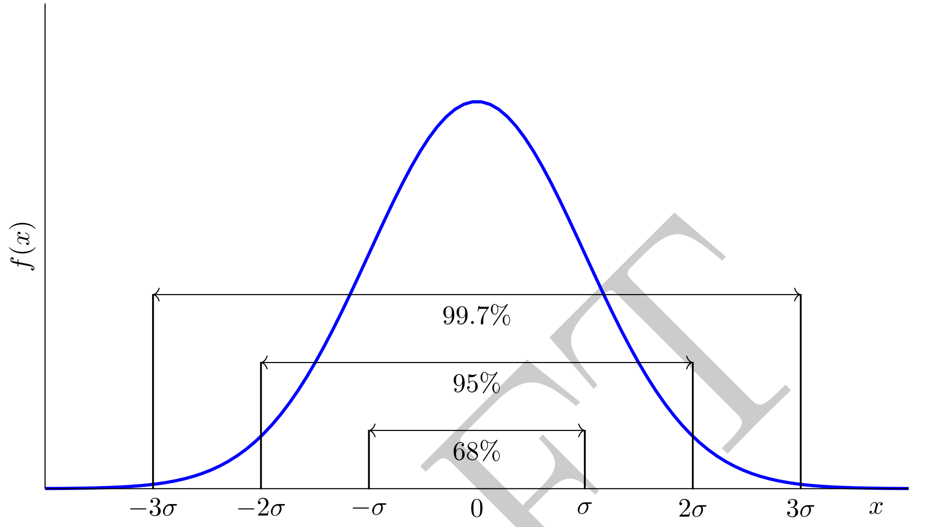


Figure 7.8: The Empirical Rule for Normal Distribution

7. **Asymptotic Behavior:** The tails of the normal distribution approach the horizontal axis but never touch it. This implies that every value, no matter how extreme, has a non-zero probability of occurring.

8. **Linear Transformations:** If X is normally distributed with mean μ and standard deviation σ, then a linear transformation of X given by $Y = aX + b$ is also normally distributed with mean $a\mu + b$ and standard deviation $|a|\sigma$.

9. **Additivity:** The sum of two independent normal random variables is also normally distributed. Specifically, if $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$, then $X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

10. **Moment Generating Function:** The moment generating function $M_X(t)$ of a normal random variable $X \sim N(\mu, \sigma^2)$ is given by:

$$M_X(t) = \exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right).$$

11. **Characteristic Function:** The characteristic function $\varphi_X(t)$ of a normal random variable $X \sim N(\mu, \sigma^2)$ is given by:

$$\varphi_X(t) = \exp\left(i\mu t - \frac{1}{2}\sigma^2 t^2\right).$$

12. **Cumulative Distribution Function (cdf):** The cumulative distribution function (CDF) of the normal distribution with mean μ and standard deviation σ is given by:

$$\Phi(x) = \Pr(X \le x) = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right]$$

where $\operatorname{erf}(z)$ is the error function, defined as:

$$\operatorname{erf}(z) = \frac{2}{\sqrt{\pi}}\int_0^z e^{-t^2}\, dt.$$

Theorem 7.4. *If $X \sim N(\mu, \sigma^2)$, then the mean and standard deviation are μ and σ^2, respectively.*

Proof. To evaluate the mean, we first calculate

$$E(X-\mu) = \int_{-\infty}^{\infty} (x-\mu)\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx.$$

Setting $z = \frac{x-\mu}{\sigma}$ and $dx = \sigma dz$, we obtain

$$E(X-\mu) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} z e^{-\frac{1}{2}z^2} dz = 0,$$

since the integrand above is an odd function of z. Using Theorem 4.5 on page 128, we conclude that

$$E(X) = \mu.$$

The variance of the normal distribution is given by

$$E[(X-\mu)^2] = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} (x-\mu)^2 e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx.$$

Again setting $z = \frac{x-\mu}{\sigma}$ and $dx = \sigma dz$, we obtain

$$E[(X-\mu)^2] = \sigma^2 \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} z^2 e^{-\frac{1}{2}z^2} dz.$$

Integrating by parts with $u = z$ and $dv = ze^{-z^2/2}dz$ so that $du = dz$ and $v = -e^{-z^2/2}$, we find that

$$E[(X-\mu)^2] = \sigma^2\left[\frac{-ze^{-z^2/2}}{\sqrt{2\pi}}\Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-z^2/2}dz\right] = \sigma^2(0+1) = \sigma^2.$$

□

Theorem 7.5. *For a normal distribution with mean μ and standard deviation σ, the inflection points occur at $x = \mu - \sigma$ and $x = \mu + \sigma$.*

Proof. To find the inflection points, we need to determine where the second derivative of the density function changes sign. **First Derivative**
The first derivative of $f(x)$ with respect to x is:

$$f'(x) = \frac{d}{dx}\left(\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}\right)$$

Using the chain rule:

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

Simplifying:

$$f'(x) = -\frac{(x-\mu)}{\sigma^3\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Second Derivative
The second derivative of $f(x)$ with respect to x is:

$$f''(x) = \frac{d}{dx}\left(-\frac{(x-\mu)}{\sigma^3\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}\right)$$

Again, using the product rule and chain rule:

$$f''(x) = -\frac{1}{\sigma^3\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}} \cdot \left(1 - \frac{(x-\mu)^2}{\sigma^2}\right)$$

Simplifying:

$$f''(x) = \frac{1}{\sigma^5\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}\left((x-\mu)^2 - \sigma^2\right)$$

Setting the Second Derivative to Zero
To find the inflection points, we set $f''(x) = 0$:

$$(x-\mu)^2 - \sigma^2 = 0$$

Solving for x:

$$(x-\mu)^2 = \sigma^2$$

$$x - \mu = \pm\sigma$$

$$x = \mu \pm \sigma$$

Thus, the inflection points of the normal distribution are at $x = \mu - \sigma$ and $x = \mu + \sigma$. □

7.4.3 Standard Normal Distribution

The standard normal distribution is a special case of the normal distribution with mean $\mu = 0$ and standard deviation $\sigma = 1$. It is denoted by Z and its properties are as follows:

- **Probability Density Function (pdf):** The pdf of the standard normal distribution is given by:

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

  as illustrated in Figure 7.9.

- **Cumulative Distribution Function (cdf):** The cdf of the standard normal distribution is denoted by $\Phi(z)$ and is defined as:

$$\Phi(z) = \Pr(Z \leq z) = \frac{1}{2}\left[1 + \text{erf}\left(\frac{z}{\sqrt{2}}\right)\right]$$

  where $\text{erf}(z)$ is the error function:

$$\text{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2}\, dt$$

Below is a plot of the standard normal distribution's pdf.

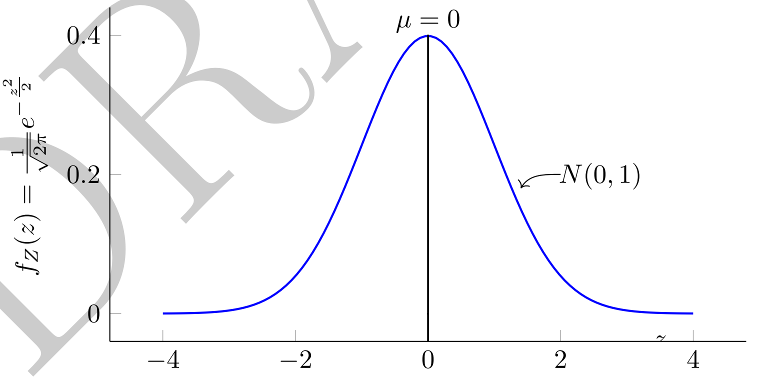


Figure 7.9: The standard normal distribution with mean $\mu = 0$ and standard deviation $\sigma = 1$.

The symmetry of the standard normal distribution about 0 implies that if the random variable Z has a standard normal distribution, then

$$1 - \Phi(z) = \Pr(Z \geq z) = \Pr(Z \leq -z) = \Phi(-z),$$

as illustrated in Figure **??**. This equation can be rearranged to provide the easily remembered relationship

$$\Phi(z) + \Phi(-z) = 1.$$

The plot presented in Figure 7.10, illustrates the cumulative distribution functions $\Phi(z)$ and $\Phi(-z)$ of the standard normal distribution. The symmetry of the standard normal distribution is evident from these plots.

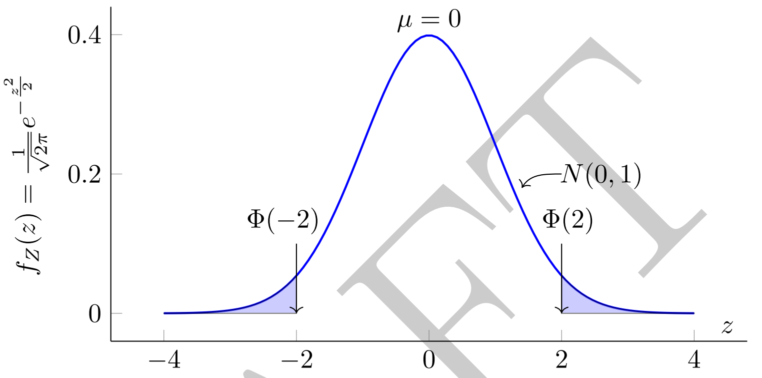


Figure 7.10: Standard normal distribution with shaded areas for $\Phi(-2)$ and $\Phi(2)$.

7.4.4 Finding Areas under the Normal Curve

It is a well-known property that the total area under the curve and above the horizontal axis of a probability density function is equal to 1. However, in this section, we aim to compute the probability that $a \leq X \leq b$.

Finding the Probability $P(a \leq X \leq b)$

Given a normally distributed random variable $X \sim N(\mu, \sigma^2)$, we want to find the probability $P(a \leq X \leq b)$ as illustrated in Figure 7.11.

Steps to Find the Probability

1. **Standardize the variable**: Convert the normal variable X to the standard normal variable Z, where $Z \sim N(0, 1)$.

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

2. **Standardize the limits**: Convert the limits a and b to their corresponding z-values.

Table 7.1: A.2: Cumulative Distribution Function of the Standard Normal Distribution

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
−3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
−3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
−3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
−3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
−2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
−2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
−2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
−2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
−2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
−2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
−2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
−2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
−2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
−2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
−1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
−1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
−1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
−1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
−1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
−1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
−1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
−1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
−1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
−1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
−0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
−0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
−0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
−0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
−0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
−0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
−0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
−0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
−0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641

Table 7.2: A.3: Cumulative Distribution Function of the Standard Normal Distribution

z	**.00**	**.01**	**.02**	**.03**	**.04**	**.05**	**.06**	**.07**	**.08**	**.09**
0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5200	0.5240	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5754
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7258	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7518	0.7549
0.7	0.7580	0.7612	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7996	0.8023	0.8051	0.8079	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8600	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9430	0.9441
1.6	0.9452	0.9463	0.9474	0.9485	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9600	0.9616	0.9625	0.9633	0.9641
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9700	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9762	0.9767
2.0	0.9773	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9924	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9942	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9958	0.9959	0.9960	0.9961	0.9962	0.9963
2.7	0.9964	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973
2.8	0.9974	0.9975	0.9976	0.9977	0.9978	0.9979	0.9980	0.9981	0.9982	0.9983
2.9	0.9984	0.9985	0.9986	0.9987	0.9988	0.9989	0.9990	0.9991	0.9992	0.9993
3.0	0.9994	0.9995	0.9996	0.9997	0.9998	0.9999	1.0000	1.0001	1.0002	1.0003

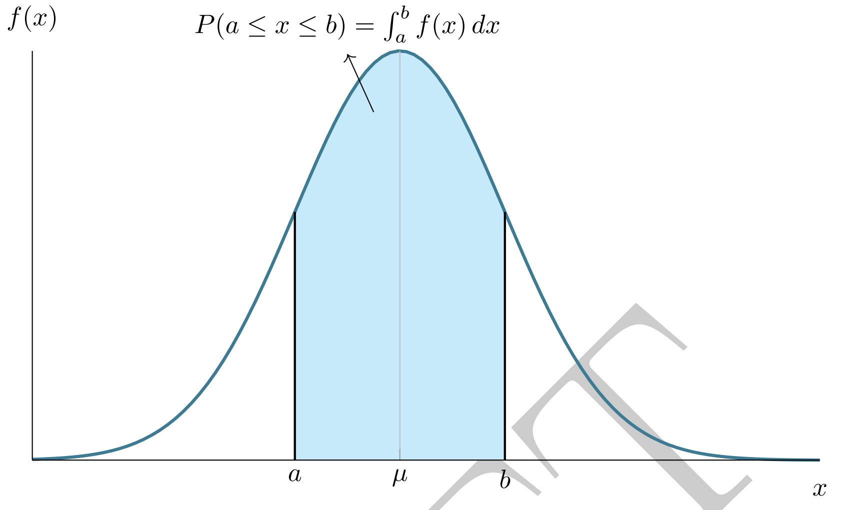


Figure 7.11: The area under the probability density function $f(x)$ between a and b.

$$z_a = \frac{a-\mu}{\sigma}$$
$$z_b = \frac{b-\mu}{\sigma}$$

3. **Find the cumulative probabilities**: Use the cumulative distribution function (CDF) of the standard normal distribution, denoted as $\Phi(z)$, to find the cumulative probabilities at z_a and z_b.

$$\Phi(z_a) = \text{CDF of the standard normal distribution at } z_a$$
$$\Phi(z_b) = \text{CDF of the standard normal distribution at } z_b$$

4. **Calculate the probability**: The probability $P(a \leq X \leq b)$ is the difference between the cumulative probabilities at z_b and z_a. That is,

$$\begin{aligned} P(a \leq X \leq b) = \Phi(z_b) - \Phi(z_a) &= P\left(\frac{a-\mu}{\sigma} \leq \frac{X-\mu}{\sigma} \leq \frac{b-\mu}{\sigma}\right) \\ &= P(z_a \leq Z \leq z_b) \\ &= P(Z \leq z_b) - P(Z \leq z_a) \\ &= \Phi(z_b) - \Phi(z_a). \end{aligned}$$

Finding the Probability $P(a \leq X \leq b)$: If $X \sim N(\mu, \sigma^2)$, then the probability $P(a \leq X \leq b)$ can be computed as

$$P(a \leq X \leq b) = P\left(\frac{a-\mu}{\sigma} \leq \frac{X-\mu}{\sigma} \leq \frac{b-\mu}{\sigma}\right)$$
$$= \Phi(z_b) - \Phi(z_a).$$

Problem 7.7. *Suppose $X \sim N(10, 4)$. Find $P(8 \leq X \leq 12)$.*

Solution 1. Standardize the limits:

$$z_8 = \frac{8-10}{2} = -1$$
$$z_{12} = \frac{12-10}{2} = 1$$

2. Find the cumulative probabilities:

$$\Phi(-1) \approx 0.1587$$
$$\Phi(1) \approx 0.8413$$

3. Calculate the probability:

$$P(8 \leq X \leq 12) = \Phi(1) - \Phi(-1) \approx 0.8413 - 0.1587 = 0.6826$$

Thus, the probability that X lies between 8 and 12 is approximately 0.6826.

Problem 7.8. *Assume the heights of adult males in a certain population are normally distributed with a mean of 70 inches and a standard deviation of 3 inches. To find the probability that a randomly selected adult male has a height between 67 and 73 inches, we standardize the values and use the standard normal distribution.*

Standardizing:

$$z_1 = \frac{67-70}{3} = -1, \quad z_2 = \frac{73-70}{3} = 1$$

Using the standard normal distribution table, we find:

$$\Phi(1) - \Phi(-1) \approx 0.8413 - 0.1587 = 0.6826$$

Thus, approximately 68.26% of adult males have heights between 67 and 73 inches.

Problem 7.9. *Suppose the heights of a large population of adult males in a certain country follow a normal distribution with a mean height of 175 centimeters and a standard deviation of 6 centimeters. Calculate the height below which the shortest 10% of adult males in this population fall.*

Solution
Let X be the height of an adult male, which follows a normal distribution:

$$X \sim N(175, 6^2)$$

We need to find the height x such that the cumulative probability up to x is 0.10. In other words, we want to find x for which:

$$P(X \leq x) = 0.10$$

To solve this, we first convert the problem to the standard normal distribution Z. The standard normal variable Z is related to X by:

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

where $\mu = 175$ cm and $\sigma = 6$ cm. We need to find the corresponding z-score for $P(Z \leq z) = 0.10$.

From standard normal distribution tables or using a statistical software, we find:

$$z_{0.10} \approx -1.28$$

Next, we use this z-score to find the corresponding height x:

$$z = \frac{x - 175}{6}$$

Solving for x:

$$-1.28 = \frac{x - 175}{6}$$

or,

$$x \approx 167.32$$

Thus, the height below which the shortest 10% of adult males in this population fall is approximately $\boxed{167.32}$ centimeters.

Problem 7.10. *The Wall Street Journal Interactive Edition spend average of 27 hours per week using the computer at work. Assume the normal distribution applies and that the standard deviation is 8 hours.*

(a). What is the probability a randomly selected subscriber spends less than 10 hours using the computer at work?

(b). What percentage of the subscribers spends more than 35 hours per week using the computer at work?

(c). A person is classified as a heavy user if he or she is in the upper 20% in terms of hours of usage. How many hours must a subscriber use the computer in order to be classified as a heavy user?

Solution
Let X be the number of hours a randomly selected subscriber spends using the computer at work per week. Assume X follows a normal distribution:

$$X \sim \mathcal{N}(27, 8^2)$$

where $\mu = 27$ hours and $\sigma = 8$ hours.

(a). To find the probability that a subscriber spends less than 10 hours on the computer, we need to calculate $P(X < 10)$. First, convert this to the standard normal variable Z:

$$Z = \frac{X - \mu}{\sigma} = \frac{10 - 27}{8} = \frac{-17}{8} = -2.125$$

Using standard normal distribution tables or software, find:

$$P(Z < -2.125) \approx 0.0169$$

Thus, the probability that a randomly selected subscriber spends less than 10 hours using the computer is approximately $\boxed{0.0169}$ or 1.69%.

(b). To find the percentage of subscribers who spend more than 35 hours per week, we need to calculate $P(X > 35)$. Convert this to the standard normal variable Z:

$$Z = \frac{X - \mu}{\sigma} = \frac{35 - 27}{8} = \frac{8}{8} = 1$$

Using standard normal distribution tables or software, find:

$$P(Z > 1) = 1 - P(Z \leq 1) \approx 1 - 0.8413 = 0.1587$$

Thus, the percentage of subscribers who spend more than 35 hours per week is approximately $\boxed{15.87\%}$.

(c). To classify as a heavy user, a subscriber must be in the upper 20% of usage. This corresponds to the 80th percentile of the normal distribution. Find the z-score for the 80th percentile:

$$z_{0.80} \approx 0.84$$

Convert this z-score to the corresponding number of hours x:

$$x = \mu + z\sigma = 27 + 0.84 \times 8$$

$$x = 27 + 6.72 = 33.72$$

Therefore, a subscriber must use the computer for at least $\boxed{33.72}$ hours per week to be classified as a heavy user.

7.4.5 Central Limit Theorem

The Central Limit Theorem (CLT) is a fundamental theorem in probability theory and statistics. It states that the distribution of the sum (or average) of a large number of independent, identically distributed (i.i.d.) random variables approaches a normal distribution, regardless of the original distribution of the variables. This theorem underpins many statistical methods and justifies the use of the normal distribution in inferential statistics.

Theorem 7.6 (Central Limit Theorem)**.** *Let $X_1, X_2, \ldots, X_n$ be a sequence of i.i.d. random variables with mean μ and variance σ^2. Let $\bar{X}_n$ denote the sample mean:*

$$\bar{X}_n = \frac{1}{n} \sum_{i=1} n X_i$$

Then, as n approaches infinity, the distribution of the standardized sample mean approaches the standard normal distribution:

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} N(0, 1)$$

or equivalently,

$$\bar{X}_n \xrightarrow{d} N(\mu, \sigma^2/n)$$

where $\xrightarrow{d}$ denotes convergence in distribution.

Example: Central Limit Theorem

Consider a fair die. The random variable X representing the outcome of a single roll has a mean $\mu = 3.5$ and variance $\sigma^2 = \frac{35}{12}$. Suppose we roll the die 30 times and compute the sample mean $\bar{X}_{30}$. According to the CLT, the distribution of $\bar{X}_{30}$ can be approximated by a normal distribution with mean 3.5 and standard deviation $\frac{\sigma}{\sqrt{30}}$.

$$\bar{X}_{30} \xrightarrow{d} N\left(3.5, \frac{35}{12 \times 30}\right).$$

Problem 7.11. *A researcher is studying the systolic blood pressure levels in a population of adults. It is known that the systolic blood pressure levels are normally distributed with a mean (μ) of 120 mmHg and a standard deviation (σ) of 15 mmHg.*

(a). What proportion of the population has systolic blood pressure levels between 110 mmHg and 130 mmHg?

(b). What is the probability that a randomly selected individual from this population has a systolic blood pressure level above 140 mmHg?

(c). If the researcher takes a random sample of 25 adults, what is the probability that the sample mean systolic blood pressure is less than 115 mmHg?

Solution (a). Proportion of the Population between 110 mmHg and 130 mmHg
We need to find $P(110 \leq X \leq 130)$ where X is the systolic blood pressure level.

$$\begin{aligned}
P(110 \leq X \leq 130) &= P\left(\frac{110-120}{15} \leq \frac{X-\mu}{\sigma} \leq \frac{130-120}{15}\right) \\
&= P(-0.67 \leq Z \leq 0.67) \\
&= P(Z \leq 0.67) - P(Z \leq -0.67) \\
&= \Phi(0.67) - \Phi(-0.67) \\
&= 0.7486 - 0.2514 \\
&= 0.4972
\end{aligned}$$

So, approximately 49.72% of the population has systolic blood pressure levels between 110 mmHg and 130 mmHg.

(b). Probability of Blood Pressure Above 140 mmHg
We need to find $P(X > 140)$.

$$\begin{aligned}
P(X > 140) &= P\left(\frac{X-\mu}{\sigma} > \frac{140-120}{15}\right) \\
&= P(Z \leq 1.33) \\
&\approx 0.9082
\end{aligned}$$

So, the probability that a randomly selected individual has a systolic blood pressure level above 140 mmHg is approximately 0.0918 or 9.18%.

(c). Probability of Sample Mean Less Than 115 mmHg
For a sample of $n = 25$, the distribution of the sample mean $\bar{X}$ is normally distributed with mean $\mu_{\bar{X}} = \mu$ and standard deviation $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$.

Given:

$$\mu_{\bar{X}} = 120, \quad \sigma_{\bar{X}} = \frac{15}{\sqrt{25}} = 3$$

We need to find $P(\bar{X} < 115)$.

Standardize the value to the standard normal distribution Z:

$$Z = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} = \frac{115 - 120}{3} = \frac{-5}{3} \approx -1.67$$

Using the standard normal distribution table or a calculator, we find:

$$P(Z \leq -1.67) \approx 0.0475$$

So, the probability that the sample mean systolic blood pressure of 25 adults is less than 115 mmHg is approximately 0.0475 or 4.75%.

Problem 7.12. *A study is conducted to measure the cholesterol levels in a population of adults. It is known that the cholesterol levels are normally distributed with a mean (μ) of 200 mg/dL and a standard deviation (σ) of 25 mg/dL.*

(a). What percentage of the population has cholesterol levels between 175 mg/dL and 225 mg/dL?

(b). What is the probability that a randomly selected individual has a cholesterol level below 180 mg/dL?

(c). If a sample of 36 adults is taken, what is the probability that the sample mean cholesterol level is greater than 210 mg/dL?

Solution

(a). Percentage of the Population between 175 mg/dL and 225 mg/dL

We need to find $P(175 \leq X \leq 225)$ where X is the cholesterol level.

First, we standardize the values to the standard normal distribution Z:

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

For $X = 175$:

$$Z = \frac{175 - 200}{25} = \frac{-25}{25} = -1$$

For $X = 225$:

$$Z = \frac{225 - 200}{25} = \frac{25}{25} = 1$$

Using the standard normal distribution table or a calculator, we find:

$$P(Z \leq 1) \approx 0.8413$$

$$P(Z \leq -1) \approx 0.1587$$

Therefore,

$$P(175 \leq X \leq 225) = P(Z \leq 1) - P(Z \leq -1) = 0.8413 - 0.1587 = 0.6826$$

So, approximately 68.26% of the population has cholesterol levels between 175 mg/dL and 225 mg/dL.

(b). Probability of Cholesterol Level Below 180 mg/dL

We need to find $P(X < 180)$.

Standardize the value to the standard normal distribution Z:

$$Z = \frac{180 - 200}{25} = \frac{-20}{25} = -0.8$$

Using the standard normal distribution table or a calculator, we find:

$$P(Z \leq -0.8) \approx 0.2119$$

So, the probability that a randomly selected individual has a cholesterol level below 180 mg/dL is approximately 0.2119 or 21.19%.

(c). Probability of Sample Mean Greater Than 210 mg/dL

For a sample of $n = 36$, the distribution of the sample mean $\bar{X}$ is normally distributed with mean $\mu_{\bar{X}} = \mu$ and standard deviation $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$.

Given:

$$\mu_{\bar{X}} = 200, \quad \sigma_{\bar{X}} = \frac{25}{\sqrt{36}} = \frac{25}{6} \approx 4.17$$

We need to find $P(\bar{X} > 210)$.

Standardize the value to the standard normal distribution Z:

$$Z = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} = \frac{210 - 200}{4.17} \approx \frac{10}{4.17} \approx 2.40$$

Using the standard normal distribution table or a calculator, we find:

$$P(Z \leq 2.40) \approx 0.9918$$

Therefore,

$$P(\bar{X} > 210) = 1 - P(Z \leq 2.40) = 1 - 0.9918 = 0.0082$$

So, the probability that the sample mean cholesterol level of 36 adults is greater than 210 mg/dL is approximately 0.0082 or 0.82%.

Problem 7.13. *Given a standard normal distribution, find the value of k such that* *(a).* $P(Z > k) = 0.3015$

(b). $P(k < Z < -0.18) = 0.4197$.

Solution (a). **Finding k for $P(Z > k) = 0.3015$**

The area to the right of k is 0.3015. Therefore, the area to the left of k is:

$$1 - 0.3015 = 0.6985$$

Using the standard normal distribution table (Table 7.3), we look up the value that corresponds to an area of 0.6985 to the left. This value is:

$$k = 0.52$$

(b). **Finding k for $P(k < Z < -0.18) = 0.4197$**

The total area to the left of -0.18 is:

$$P(Z < -0.18) = 0.4286$$

The area between k and -0.18 is 0.4197. Therefore, the area to the left of k is:

$$0.4286 - 0.4197 = 0.0089$$

Using the standard normal distribution table (Table 7.3), we look up the value that corresponds to an area of 0.0089 to the left. This value is:

$$k = -2.37$$

7.4.6 Python Code for Normal Distribution Characteristics

The Normal (Gaussian) distribution is fundamental in statistics and is used to model continuous random variables. Below is Python code that demonstrates how to compute various characteristics of the Normal distribution.

Python Code

```python
import numpy as np
from scipy.stats import norm

# Define the parameters
mu = 0          # Mean
sigma = 1       # Standard deviation

# Normal distribution
```

```python
dist = norm(loc=mu, scale=sigma)

# 1. Probability Density Function (PDF)
x_values = np.linspace(-5, 5, 100)  # Values for the random variable
pdf_values = dist.pdf(x_values)
print("PDF values for x from -5 to 5:", pdf_values)

# 2. Cumulative Distribution Function (CDF)
cdf_values = dist.cdf(x_values)
print("CDF values for x from -5 to 5:", cdf_values)

# 3. Mean (Expected Value)
mean = dist.mean()
print("Mean (Expected Value):", mean)

# 4. Variance
variance = dist.var()
print("Variance:", variance)

# 5. Standard Deviation
std_dev = dist.std()
print("Standard Deviation:", std_dev)

# 6. Quantiles
quantiles = dist.ppf([0.25, 0.5, 0.75])  # 25th, 50th (median), and 75th percentiles
print("Quantiles at 0.25, 0.5, and 0.75:", quantiles)

# 7. Percentiles
percentiles = dist.ppf([0.1, 0.9])  # 10th and 90th percentiles
print("Percentiles at 0.1 and 0.9:", percentiles)

# 8. Moment Generating Function (MGF)
def mgf(t, mu, sigma):
  return np.exp(mu * t + 0.5 * (sigma**2) * t**2)

# MGF values at t = 0 and t = 1
mgf_values = [mgf(t, mu, sigma) for t in [0, 1]]
print("MGF values for t = 0 and t = 1:", mgf_values)

```

Explanations

- **Probability Density Function (PDF):**
  - The PDF provides the likelihood of each value. Compute these values using `dist.pdf(x_values)`.

- **Cumulative Distribution Function (CDF):**
  - The CDF gives the probability that the variable takes on a value less than or equal to a given value. Compute this using `dist.cdf(x_values)`.
- **Mean (Expected Value):**
  - The mean (expected value) is μ. Compute this using `dist.mean()`.
- **Variance:**
  - The variance is σ^2. Compute this using `dist.var()`.
- **Standard Deviation:**
  - The standard deviation is σ. Compute this using `dist.std()`.
- **Quantiles:**
  - Quantiles are values at specific probabilities. Compute these using `dist.ppf()` for the desired quantile probabilities.
- **Percentiles:**
  - Percentiles are specific types of quantiles. Compute these using `dist.ppf()` for the desired percentile probabilities.
- **Moment Generating Function (MGF):**
  - The MGF for a Normal distribution is given by $M_X(t) = \exp(\mu \cdot t + 0.5 \cdot \sigma^2 \cdot t^2)$. Define a function `mgf(t, mu, sigma)` to compute MGF values.

7.5 Concluding Remarks

In this chapter, we have examined several essential continuous probability distributions, highlighting their definitions, properties, and applications. The Uniform distribution provided a foundation for understanding equal likelihood over a range of values, while the Exponential distribution illustrated the modeling of time intervals between events in a Poisson process.

The Normal distribution, with its profound significance in statistical theory and practice, was explored in detail. We discussed its properties, the concept of the Standard Normal distribution, methods for finding areas under the normal curve, and the Central Limit Theorem, which underscores the Normal distribution's ubiquitous presence in statistical analysis.

Understanding these continuous distributions equips us with powerful tools for modeling and analyzing data in numerous disciplines. By mastering these concepts, we can better interpret real-world phenomena, make informed decisions, and contribute to advancements in various fields.

As we continue our journey through probability and statistics, the knowledge gained from studying continuous distributions will serve as a crucial foundation for more complex analyses and applications.

7.6 Exercises

1. Given a normal distribution $X \sim N(\mu, \sigma^2)$, answer the following questions:
   (a) If $\mu = 10$ and $\sigma = 2$, what is the probability that X is less than 8?
   (b) Find the probability that X lies between 8 and 12.
   (c) Determine the value a such that $P(X \leq a) = 0.975$.
2. The standard normal distribution is a special case of the normal distribution where $\mu = 0$ and $\sigma = 1$. Use the standard normal distribution table (z-table) to answer the following:
   (a) Find $P(Z \leq 1.645)$ for $Z \sim N(0, 1)$.
   (b) Determine $P(-1.96 \leq Z \leq 1.96)$.
   (c) Calculate the value z such that $P(Z \geq z) = 0.05$.
3. Given $X \sim N(20, 25)$, transform X to the standard normal distribution Z and solve the following:
   (a) Find the probability that X is less than 15.
   (b) Determine the probability that X lies between 18 and 22.
   (c) Find the value x such that $P(X \leq x) = 0.90$.
4. Answer the following application-based questions:
   (a) The heights of adult men in a certain population are normally distributed with a mean of 175 cm and a standard deviation of 10 cm. What percentage of men are taller than 190 cm?
   (b) A factory produces light bulbs with lifetimes that are normally distributed with a mean of 1200 hours and a standard deviation of 100 hours. What is the probability that a randomly selected light bulb lasts between 1100 and 1300 hours?
   (c) The scores on a standardized test are normally distributed with a mean of 500 and a standard deviation of 100. What is the minimum score needed to be in the top 5

5. Consider a sample mean $\bar{X}$ from a normal distribution with population mean μ and variance σ^2. Assume the sample size is $n = 36$.
   (a) If $\mu = 50$ and $\sigma = 12$, find the probability that the sample mean is greater than 52.
   (b) Calculate the probability that the sample mean lies between 48 and 51.
   (c) Determine the value of $\bar{x}$ such that $P(\bar{X} \leq \bar{x}) = 0.95$.

6. Given a population with mean $\mu = 60$ and standard deviation $\sigma = 15$:
   (a) If a sample of size 50 is taken, what is the expected value and standard deviation of the sample mean?
   (b) Using the Central Limit Theorem, find the probability that the sample mean is less than 58.
   (c) Calculate the probability that the sample mean lies between 59 and 62.

7. A study measures the cholesterol levels (in mg/dL) of a group of patients, which are found to follow a normal distribution with a mean of 200 mg/dL and a standard deviation of 20 mg/dL.
   (a) What is the probability that a randomly selected patient has a cholesterol level between 180 mg/dL and 220 mg/dL?
   (b) What is the 95th percentile of the cholesterol levels?
   (c) Calculate the variance and standard deviation of the cholesterol levels.
   (d) If a cholesterol level above 240 mg/dL is considered high, what proportion of the patients have high cholesterol levels?

8. Consider a population where the heights of adult women are approximately normally distributed with a mean of 65 inches and a standard deviation of 4 inches.
   (a) Using Chebyshev's Inequality, find the minimum percentage of women whose heights are within 10 inches of the mean height.
   (b) Suppose you want to ensure that at least 95% of the women fall within a certain number of standard deviations from the mean. Using Chebyshev's Inequality, determine the minimum number of standard deviations required for this guarantee.
   (c) In the same population, if the height of a randomly selected woman is 70 inches, what is the probability that her height deviates from the mean by at least 5 inches, according to Chebyshev's Inequality?

9. Suppose X and Y are independent random variables representing the weights (in kg) of two different species of fish in a lake. X follows a normal distribution with mean 2 kg and variance 0.25 kg^2, and Y follows a normal distribution with mean 3 kg and variance 0.36 kg^2.

   (a) What is the probability that a fish of the first species weighs more than 2.5 kg?

   (b) What is the probability that a fish of the second species weighs between 2.5 kg and 3.5 kg?

   (c) What is the expected value of the difference in weights $D = X - Y$?

   (d) What is the variance of the difference in weights D?

Appendix

Table 7.3: A.2: Cumulative Distribution Function of the Standard Normal Distribution

z	**.00**	**.01**	**.02**	**.03**	**.04**	**.05**	**.06**	**.07**	**.08**	**.09**
−3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
−3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
−3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
−3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
−3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
−2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
−2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
−2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
−2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
−2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
−2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
−2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
−2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
−2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
−2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
−1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
−1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
−1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
−1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
−1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
−1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
−1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
−1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
−1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
−1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
−0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
−0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
−0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
−0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
−0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
−0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
−0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
−0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
−0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641

Table 7.4: A.3: Cumulative Distribution Function of the Standard Normal Distribution

z	**.00**	**.01**	**.02**	**.03**	**.04**	**.05**	**.06**	**.07**	**.08**	**.09**
0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5200	0.5240	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5754
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7258	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7518	0.7549
0.7	0.7580	0.7612	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7996	0.8023	0.8051	0.8079	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8600	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9430	0.9441
1.6	0.9452	0.9463	0.9474	0.9485	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9600	0.9616	0.9625	0.9633	0.9641
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9700	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9762	0.9767
2.0	0.9773	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9924	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9942	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9958	0.9959	0.9960	0.9961	0.9962	0.9963
2.7	0.9964	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973
2.8	0.9974	0.9975	0.9976	0.9977	0.9978	0.9979	0.9980	0.9981	0.9982	0.9983
2.9	0.9984	0.9985	0.9986	0.9987	0.9988	0.9989	0.9990	0.9991	0.9992	0.9993
3.0	0.9994	0.9995	0.9996	0.9997	0.9998	0.9999	1.0000	1.0001	1.0002	1.0003

Table 7.5: A.4: Critical points of the t-distribution with its degrees of freedom (ν)

ν	α						
	0.10	**0.05**	**0.025**	**0.01**	**0.005**	**0.001**	**0.0005**
1	3.078	6.314	12.706	31.821	63.657	318.31	636.62
2	1.886	2.920	4.303	6.965	9.925	22.326	31.598
3	1.638	2.353	3.182	4.541	5.841	10.213	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	1.319	1.714	2.069	2.500	2.807	3.485	3.767
24	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	1.296	1.671	2.000	2.390	2.660	3.232	3.460
120	1.289	1.658	1.980	2.358	2.617	3.160	3.373
∞	1.282	1.645	1.960	2.326	2.576	3.090	3.291

Bibliography

[1] David R Anderson, Thomas A Williams, and James J Cochran. *Statistics for business & economics.* Cengage Learning, 2020.

[2] Sheldon Ross. Probability and statistics for engineers and scientists. *Elsevier, New Delhi*, 16:32–33, 2009.

[3] Ronald E Walpole, Raymond H Myers, Sharon L Myers, and Keying Ye. *Probability and statistics for engineers and scientists*, volume 5. Macmillan New York, 1993.