International Baccalaureate (IB) Diploma Programme Mathematics Analysis and Approaches Higher Level

Probability & Statistics

The IB 7-Scorer's Ultimate Guide

Crafted Exclusively for High-Achieving IB Mathematics Students: April 2025 Edition

Mathematics Elevate Academy

Excellence in Advanced Mathematics Education

Rishabh Kumar

Founder, Mathematics Elevate Academy

Elite Mentor for IB Mathematics

Alumnus of Indian Institute of Technology Guwahati & Indian Statistical Institute

Visit me, Access Free Resources & Apply For Exclusive Personalized Mentorship

www.mathematicselevateacademy.com

www.linkedin.com/in/rishabh-kumar-iitg-isi/

Disclaimer

Introduction

The IB 7-Scorer's Ultimate Guide — April 2025 Edition is thoughtfully designed for IB DP Mathematics students striving to excel in IB DP Mathematics AA SL/HL, with a special emphasis on Probability & Statistics conceptual mastery. This guide offers a wealth of expertly crafted high-level Probability & Statistics problems, conceptual challenges, and much more.

Explore examiner-style solutions, detailed marking scheme breakdowns, and insightful commentary on common errors to refine your problem-solving skills. Each problem is designed to test your grasp of Probability & Statistics concepts.

This guide goes beyond the IB syllabus, offering enrichment problems that challenge your mathematical thinking and prepare you for Olympiads and university-level mathematics. The solutions are presented with step-by-step clarity, expert insights, and advanced techniques, ensuring a comprehensive and engaging learning experience.

For answers or detailed solutions, keep following me — they will be available soon! For personalized learning, book a one-on-one mentorship session with me to receive customized guidance on mastering IB DP Mathematics AA/AI SL/HL Probability & Statistics, or even Olympiad-level problems. Together, we will build the confidence and skills you need to excel.

Check Your Understanding!

1 Probability and Statistics: Sampling

Problem 1.1: Identifying Continuous and Discrete Data

Problem Statement

1. Identify whether the following data are continuous or discrete:

1. The number of students in a classroom.
2. The height of a person.
3. The number of cars in a parking lot.
4. The time taken to complete a race.
5. The weight of a bag of rice.

2. Explain the difference between continuous and discrete data, providing examples for each.

—

Problem 1.2: Identifying Population, Sample, and Randomness

Problem Statement

1. In the following contexts, identify the population, the sample, and whether the sample is random:

1. A survey of 100 students in a school to determine their favorite subject.
2. A study of 50 randomly selected households in a city to measure electricity usage.
3. A poll of 200 voters in a district to predict the outcome of an election.
4. A quality check of 10 items from a batch of 500 products.

2. Explain the importance of randomness in sampling and how it affects the reliability of results.

—

Problem 1.3: Identifying Bias and Reliability in Sampling

Problem Statement

1. Identify potential sources of bias in the following sampling methods:

1. Surveying only people in a shopping mall to determine the average income of a city.
2. Asking only students in a math class about their favorite subject.
3. Selecting only morning commuters to study public transportation usage.

2. Evaluate the reliability of the following data sources:

1. A government census.
2. An online poll on a social media platform.
3. A scientific study published in a peer-reviewed journal.

3. Explain how missing data or errors in recording data can affect the reliability of results and suggest ways to handle such issues.

—

Problem 1.4: Interpretation of Outliers

Problem Statement

1. For the following data sets, determine if there are any outliers using the rule that an outlier is more than $1.5 \times$ IQR from the nearest quartile:

1. $[5, 7, 8, 10, 12, 15, 18, 20, 25]$
2. $[2, 3, 3, 4, 5, 6, 7, 8, 50]$

2. Suggest how to determine whether an outlier should be removed from a sample and explain the potential impact of removing or retaining outliers on the results.

—

Problem 1.5: Sampling Techniques and Their Effectiveness

Problem Statement

1. Identify the sampling technique used in the following scenarios:

1. Selecting every 10th person from a list of names.
2. Asking for volunteers to participate in a survey.
3. Dividing a population into age groups and randomly selecting individuals from each group.
4. Interviewing people who are easily accessible, such as those in a shopping mall.
5. Selecting a fixed number of individuals from each category, such as gender or income level.

2. Evaluate the effectiveness of the following sampling techniques:

- Simple random sampling.
- Convenience sampling.
- Systematic sampling.
- Quota sampling.
- Stratified sampling.

3. Calculate the number of data items in each category of a stratified sample:

- A population of 1,000 people is divided into three groups: 40% in Group A, 30% in Group B, and 30% in Group C. A stratified sample of 200 people is taken. How many people should be selected from each group?

Key Concepts and Definitions

Key Concepts and Definitions

1. **Population and Sample**:
   - The population is the entire group being studied.
   - A sample is a subset of the population used to make inferences about the population.
2. **Continuous vs. Discrete Data**:
   - Continuous data can take any value within a range (e.g., height, weight).
   - Discrete data can only take specific values (e.g., number of students, number of cars).
3. **Outliers**: An outlier is a data point that is more than $1.5 \times$ IQR from the nearest quartile.
4. **Sampling Techniques**:
   - **Simple Random Sampling**: Every individual has an equal chance of being selected.
   - **Convenience Sampling**: Individuals are selected based on ease of access.
   - **Systematic Sampling**: Every nth individual is selected.
   - **Quota Sampling**: A fixed number of individuals is selected from each category.
   - **Stratified Sampling**: The population is divided into groups, and a random sample is taken from each group.

—

Marking Guidelines

Marking Scheme

Problem 1.1: Identifying Continuous and Discrete Data

- Correct identification of data type [2 marks per part]
- Valid explanation of the difference between continuous and discrete data [2 marks]

Problem 1.2: Identifying Population, Sample, and Randomness

- Correct identification of population, sample, and randomness [2 marks per part]
- Valid explanation of the importance of randomness [2 marks]

Problem 1.3: Identifying Bias and Reliability

- Correct identification of bias [2 marks per part]
- Valid evaluation of data reliability [2 marks per part]
- Explanation of handling missing data or errors [2 marks]

Problem 1.4: Interpretation of Outliers

- Correct calculation of IQR and identification of outliers [2 marks per part]
- Valid explanation of the impact of outliers [2 marks]

Problem 1.5: Sampling Techniques and Effectiveness

- Correct identification of sampling techniques [2 marks per part]
- Valid evaluation of sampling techniques [2 marks per part]
- Accurate calculation of stratified sample sizes [2 marks]

Additional Points

- Clear presentation of solutions [1 mark]
- Logical reasoning in calculations [1 mark]

2 Probability and Statistics: Statistical Diagrams

Problem 2.1: Frequency Distributions

Problem Statement

1. Interpret the following frequency distribution table:

Class Interval	Frequency
$0 - 10$	5
$10 - 20$	8
$20 - 30$	12
$30 - 40$	10
$40 - 50$	5

1. What is the total number of data points?
2. What is the modal class?
3. Calculate the midpoint of each class interval.

2. Explain how to construct a frequency distribution table from raw data.

—

Problem 2.2: Histograms

Problem Statement

1. Interpret the following histogram:

- Identify the class interval with the highest frequency.
- Estimate the total number of data points.
- Explain how the height of each bar relates to the frequency.

2. Construct a histogram for the following data:

Class Interval	Frequency
$0 - 5$	4
$5 - 10$	6
$10 - 15$	10
$15 - 20$	8
$20 - 25$	2

3. Explain the difference between a histogram and a bar chart.

—

Problem 2.3: Cumulative Frequency Graphs

Problem Statement

1. Interpret the following cumulative frequency graph:

- Find the median.
- Find the lower quartile, upper quartile, and interquartile range.
- Estimate the 90th percentile.

2. Construct a cumulative frequency graph for the following data:

Class Interval	Cumulative Frequency
$0-10$	5
$10-20$	13
$20-30$	25
$30-40$	35
$40-50$	40

3. Explain how to use a cumulative frequency graph to find the range, interquartile range, and percentiles.

—

Problem 2.4: Box and Whisker Plots

Problem Statement

1. Produce a box and whisker plot for the following data:

Data: $5, 7, 8, 10, 12, 15, 18, 20, 25$

- Find the minimum, lower quartile, median, upper quartile, and maximum.
- Draw the box and whisker plot.

2. Interpret the following box and whisker plot:

- Identify the range and interquartile range.
- Determine if the data is symmetric or skewed.
- Suggest whether the data could follow a normal distribution.

3. Explain how box and whisker plots can be used to compare distributions.

—

Key Concepts and Definitions

—

Marking Guidelines

Marking Scheme

Problem 2.1: Frequency Distributions

- Correct interpretation of the frequency table [2 marks per part]
- Valid explanation of how to construct a frequency table [2 marks]

Problem 2.2: Histograms

- Correct interpretation of the histogram [2 marks per part]
- Accurate construction of the histogram [2 marks per part]
- Valid explanation of the difference between histograms and bar charts [2 marks]

Problem 2.3: Cumulative Frequency Graphs

- Correct interpretation of the graph [2 marks per part]
- Accurate construction of the cumulative frequency graph [2 marks per part]
- Valid explanation of how to use the graph to find key statistics [2 marks]

Problem 2.4: Box and Whisker Plots

- Correct calculation of the five-number summary [2 marks per part]
- Accurate construction of the box and whisker plot [2 marks per part]
- Valid interpretation of the plot [2 marks per part]

Additional Points

- Clear presentation of solutions [1 mark]
- Logical reasoning in calculations [1 mark]

3 Probability and Statistics: Summary Statistics

Problem 3.1: Measures of Central Tendency

Problem Statement

1. Calculate the mean, median, and mode for the following data set:

$$\text{Data: } 5, 7, 8, 10, 12, 15, 18, 20, 25$$

2. For the following frequency distribution table, calculate the mean:

Value (x)	Frequency (f)
1	4
2	6
3	8
4	5
5	2

Use the formula:

$$\bar{x} = \frac{\sum_{i=1}^{k} f_i x_i}{\sum_{i=1}^{k} f_i}$$

3. Explain the difference between the mean, median, and mode, and provide examples of when each measure is most appropriate.

—

Problem 3.2: Estimation of Mean from Grouped Data

Problem Statement

1. Estimate the mean for the following grouped data using mid-interval values:

Class Interval	Frequency (f)
$0 - 10$	5
$10 - 20$	8
$20 - 30$	12
$30 - 40$	10
$40 - 50$	5

2. Explain why mid-interval values are used to estimate the mean for grouped data and discuss the limitations of this method.

—

Problem 3.3: Modal Class for Grouped Data

Problem Statement

1. Identify the modal class for the following grouped data:

Class Interval	Frequency (f)
$0 - 10$	5
$10 - 20$	8
$20 - 30$	12
$30 - 40$	10
$40 - 50$	5

2. Explain how the modal class is determined from a frequency table or histogram and discuss its significance in summarizing data.

—

Problem 3.4: Measures of Dispersion

Problem Statement

1. Use technology to calculate the interquartile range (IQR), standard deviation, and variance for the following data set:

$$\text{Data: } 5, 7, 8, 10, 12, 15, 18, 20, 25$$

2. Explain the significance of the IQR, standard deviation, and variance in describing the spread of data.
3. Discuss how outliers can affect the standard deviation and variance.

—

Problem 3.5: Effect of Constant Changes on Data

Problem Statement

1. A data set has a mean of 10 and a standard deviation of 2. Calculate the new mean and standard deviation if:

1. Each data point is increased by 5.
2. Each data point is multiplied by 3.

2. Explain how adding or multiplying a constant affects the mean and standard deviation of a data set.

—

Problem 3.6: Quartiles of Discrete Data

Problem Statement

1. Use technology to find the quartiles for the following data set:

Data: $5, 7, 8, 10, 12, 15, 18, 20, 25$

2. Explain the significance of quartiles in summarizing data and how they are used to calculate the interquartile range (IQR).

—

Key Concepts and Definitions

Key Concepts and Definitions

1. **Measures of Central Tendency**:
   - **Mean**: The average of the data, calculated as:
   $$\bar{x} = \frac{\sum_{i=1}^{k} f_i x_i}{\sum_{i=1}^{k} f_i}$$
   - **Median**: The middle value when the data is ordered.
   - **Mode**: The most frequently occurring value in the data.
2. **Grouped Data**:
   - Use mid-interval values to estimate the mean.
   - The modal class is the class interval with the highest frequency.
3. **Measures of Dispersion**:
   - **Interquartile Range (IQR)**: The difference between the upper quartile (Q3) and lower quartile (Q1).
   - **Standard Deviation**: A measure of the spread of data around the mean.
   - **Variance**: The square of the standard deviation.
4. **Effect of Constant Changes**:
   - Adding a constant to all data points increases the mean by the same constant but does not affect the standard deviation.
   - Multiplying all data points by a constant multiplies the mean and standard deviation by the same constant.

—

Marking Guidelines

Marking Scheme

Problem 3.1: Measures of Central Tendency

- Correct calculation of mean, median, and mode [2 marks per part]
- Valid explanation of the differences between the measures [2 marks]

Problem 3.2: Estimation of Mean from Grouped Data

- Correct use of mid-interval values [2 marks per part]
- Valid explanation of the method and its limitations [2 marks]

Problem 3.3: Modal Class for Grouped Data

- Correct identification of the modal class [2 marks per part]
- Valid explanation of its significance [2 marks]

Problem 3.4: Measures of Dispersion

- Correct calculation of IQR, standard deviation, and variance [2 marks per part]
- Valid explanation of their significance [2 marks]

Problem 3.5: Effect of Constant Changes on Data

- Correct calculation of new mean and standard deviation [2 marks per part]
- Valid explanation of the effects of constant changes [2 marks]

Problem 3.6: Quartiles of Discrete Data

- Correct calculation of quartiles [2 marks per part]
- Valid explanation of their significance [2 marks]

Additional Points

- Clear presentation of solutions [1 mark]
- Logical reasoning in calculations [1 mark]

4 Probability and Statistics: Correlation and Regression

Problem 4.1: Pearson's Product Moment Correlation Coefficient

Problem Statement

1. Use technology to calculate the Pearson's product moment correlation coefficient, r, for the following bivariate data:

x	y
1	2
2	4
3	6
4	8
5	10

2. Interpret the value of r in the context of the data. Discuss whether the correlation implies causation.
3. Explain the range of r and what different values of r (e.g., $r = 1$, $r = 0$, $r = -1$) indicate about the strength and direction of the linear relationship.

—

Problem 4.2: Scatter Diagrams and Line of Best Fit

Problem Statement

1. Plot a scatter diagram for the following data:

x	y
1	3
2	5
3	7
4	9
5	11

2. Estimate the line of best fit by eye, ensuring that it passes through the mean point $(\bar{x}, \bar{y})$.
3. Explain how a scatter diagram can be used to visually assess the strength and direction of a linear relationship.

—

Problem 4.3: Equation of the Regression Line of y on x

Problem Statement

1. Use technology to calculate the equation of the regression line of y on x for the following data:

x	y
1	2
2	4
3	6
4	8
5	10

2. Write the equation in the form $y = a + bx$, where a is the intercept and b is the slope.
3. Explain the meaning of the parameters a and b in the context of the data.

—

Problem 4.4: Using the Regression Line for Prediction

Problem Statement

1. For the regression line $y = 2 + 3x$, predict the value of y when:

1. $x = 4$
2. $x = 10$

2. Discuss the dangers of extrapolation when using the regression line for prediction.
3. Explain when a y-on-x regression line is appropriate and why it may not be suitable for predicting x from y.

—

Problem 4.5: Piecewise Linear Models

Problem Statement

1. Create a piecewise linear model for the following data:

x	y
1	2
2	4
3	6
4	8
5	10
6	15
7	20
8	25

Divide the data into two sections: $1 \leq x \leq 5$ and $6 \leq x \leq 8$. Find the equation of the regression line for each section.

2. Explain how piecewise linear models can be used to model non-linear relationships.

3. Discuss the limitations of piecewise linear models.

—

Key Concepts and Definitions

Key Concepts and Definitions

1. **Pearson's Product Moment Correlation Coefficient (r)**:
   - Measures the strength and direction of the linear relationship between two variables.
   - r ranges from -1 to 1:
     - $r = 1$: Perfect positive linear correlation.
     - $r = -1$: Perfect negative linear correlation.
     - $r = 0$: No linear correlation.
2. **Scatter Diagram**: A graphical representation of bivariate data, where each point represents a pair of values (x, y). It is used to visually assess the relationship between the variables.
3. **Regression Line of y on x**: The line of best fit that minimizes the sum of squared vertical distances between the data points and the line. The equation is:

   $$y = a + bx$$

   where:
   - a is the intercept (value of y when $x = 0$).
   - b is the slope (rate of change of y with respect to x).
4. **Extrapolation**: Using the regression line to predict values outside the range of the data. This can be unreliable as the relationship may not hold outside the observed range.
5. **Piecewise Linear Models**: A model that uses different linear equations for different sections of the data. It is useful for modeling non-linear relationships in a piecewise manner.

Marking Guidelines

Marking Scheme

Problem 4.1: Pearson's Product Moment Correlation Coefficient

- Correct calculation of r using technology [2 marks per part]
- Valid interpretation of r in context [2 marks per part]
- Explanation of the range and meaning of r [2 marks]

Problem 4.2: Scatter Diagrams and Line of Best Fit

- Accurate plotting of the scatter diagram [2 marks per part]
- Correct estimation of the line of best fit [2 marks per part]
- Valid explanation of the visual assessment of correlation [2 marks]

Problem 4.3: Equation of the Regression Line of y on x

- Correct calculation of the regression line equation [2 marks per part]
- Valid interpretation of the parameters a and b [2 marks per part]

Problem 4.4: Using the Regression Line for Prediction

- Correct prediction using the regression line [2 marks per part]
- Valid discussion of the dangers of extrapolation [2 marks]
- Explanation of when a y-on-x regression line is appropriate [2 marks]

Problem 4.5: Piecewise Linear Models

- Correct creation of piecewise linear models [2 marks per part]
- Valid explanation of their use and limitations [2 marks per part]

Additional Points

- Clear presentation of solutions [1 mark]
- Logical reasoning in calculations [1 mark]

5 Probability: Definitions and Basic Concepts

Problem 5.1: Concepts of Trial, Outcome, and Sample Space

Problem Statement

1. Define the following terms and provide an example for each:

1. Trial
2. Outcome
3. Sample space
4. Event
5. Equally likely outcomes
6. Relative frequency

2. A die is rolled. List the sample space and identify the following events:

1. Event A: Rolling an even number.
2. Event B: Rolling a number greater than 4.
3. Event C: Rolling a 3.

3. Explain how relative frequency can be used to estimate the probability of an event.

—

Problem 5.2: Theoretical Probability

Problem Statement

1. A bag contains 3 red balls, 2 blue balls, and 5 green balls. Find the probability of:

1. Drawing a red ball.
2. Drawing a blue ball.
3. Drawing a ball that is not green.

2. A coin is flipped twice. List all possible outcomes and find the probability of:

1. Getting exactly one head.
2. Getting at least one tail.
3. Getting no heads.

3. Explain the formula for theoretical probability:

$$P(A) = \frac{n(A)}{n(U)}$$

where $n(A)$ is the number of favorable outcomes and $n(U)$ is the total number of outcomes.

—

Problem 5.3: Complementary Events

Problem Statement

1. A card is drawn from a standard deck of 52 cards. Find the probability of:

1. Drawing a heart.
2. Not drawing a heart.

2. A die is rolled. Find the probability of:

1. Rolling a number less than 4.
2. Rolling a number that is not less than 4.

3. Explain the relationship between complementary events A and A' and the formula:

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

—

Problem 5.4: Expected Number of Occurrences

Problem Statement

1. A coin is flipped 100 times. Calculate the expected number of:

1. Heads.
2. Tails.

2. A die is rolled 60 times. Calculate the expected number of times:

1. A 6 is rolled.
2. An even number is rolled.

3. Explain the formula for the expected number of occurrences:

$$\text{Expected number} = \text{Number of trials} \times \text{Probability of the event.}$$

—

Key Concepts and Definitions

Key Concepts and Definitions

1. **Trial**: A single performance of an experiment (e.g., rolling a die).
2. **Outcome**: A possible result of a trial (e.g., rolling a 4).
3. **Sample Space**: The set of all possible outcomes of a trial (e.g., $\{1, 2, 3, 4, 5, 6\}$ for a die roll).
4. **Event**: A subset of the sample space (e.g., rolling an even number).
5. **Equally Likely Outcomes**: Outcomes that have the same probability of occurring (e.g., flipping a fair coin).
6. **Relative Frequency**: The ratio of the number of times an event occurs to the total number of trials.
7. **Theoretical Probability**:

$$P(A) = \frac{n(A)}{n(U)}$$

   where $n(A)$ is the number of favorable outcomes and $n(U)$ is the total number of outcomes.
8. **Complementary Events**: Events A and A' are complementary if A' represents all outcomes not in A. The relationship is:

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

9. **Expected Number of Occurrences**: The expected number of times an event occurs is given by:

$$\text{Expected number} = \text{Number of trials} \times \text{Probability of the event.}$$

—

Marking Guidelines

Marking Scheme

Problem 5.1: Concepts of Trial, Outcome, and Sample Space

- Correct definitions and examples [2 marks per part]
- Accurate listing of sample space and identification of events [2 marks per part]
- Valid explanation of relative frequency [2 marks]

Problem 5.2: Theoretical Probability

- Correct calculation of probabilities [2 marks per part]
- Accurate listing of all possibilities [2 marks per part]
- Valid explanation of the formula for theoretical probability [2 marks]

Problem 5.3: Complementary Events

- Correct calculation of probabilities for complementary events [2 marks per part]
- Valid explanation of the relationship between $P(A)$ and $P(A')$ [2 marks]

Problem 5.4: Expected Number of Occurrences

- Correct calculation of expected numbers [2 marks per part]
- Valid explanation of the formula for expected occurrences [2 marks]

Additional Points

- Clear presentation of solutions [1 mark]
- Logical reasoning in calculations [1 mark]

6 Probability: Techniques and Concepts

Problem 6.1: Venn Diagrams

Problem Statement

1. Use the following Venn diagram to calculate probabilities:

Region	Description	Number of Elements
$A \cap B$	$AandB$	5
$A \setminus B$	$Aonly$	8
$B \setminus A$	$Bonly$	7
$A^c \cap B^c$	$NeitherAnorB$	10

1. $P(A)$
2. $P(B)$
3. $P(A \cup B)$
4. $P(A \cap B)$

2. Explain how Venn diagrams can be used to organize information and calculate probabilities.

—

Problem 6.2: Tree Diagrams

Problem Statement

1. A bag contains 3 red balls and 2 blue balls. A ball is drawn, its color is noted, and it is replaced. Draw a tree diagram to represent the situation and calculate the probability of:

1. Drawing two red balls.
2. Drawing one red ball and one blue ball (in any order).
3. Drawing at least one blue ball.

2. Explain the rules for using tree diagrams:

- Multiply along the branches.
- Add between the branches.

—

Problem 6.3: Sample Space Diagrams

Problem Statement

1. A die is rolled, and a coin is flipped. Use a sample space diagram to list all possible outcomes and calculate the probability of:

1. Rolling a 4 and flipping heads.
2. Rolling an even number.
3. Flipping tails.

2. Explain how sample space diagrams can be used to organize information and calculate probabilities.

—

Problem 6.4: Tables of Outcomes

Problem Statement

1. Two dice are rolled. Use a table of outcomes to calculate the probability of:

1. Rolling a sum of 7.
2. Rolling doubles (e.g., 1 and 1, 2 and 2, etc.).
3. Rolling a sum greater than 9.

2. Explain how tables of outcomes can be used to organize information and calculate probabilities.

—

Problem 6.5: Combined Events

Problem Statement

1. Given $P(A) = 0.4$, $P(B) = 0.5$, and $P(A \cap B) = 0.2$, calculate:

1. $P(A \cup B)$
2. $P(A^c)$
3. $P(B^c)$

2. Explain the formula for combined events:

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

and its significance in probability calculations.

—

Problem 6.6: Mutually Exclusive Events

Problem Statement

1. Two events, A and B, are mutually exclusive. If $P(A) = 0.3$ and $P(B) = 0.4$, calculate:

1. $P(A \cap B)$
2. $P(A \cup B)$

2. Explain the concept of mutually exclusive events and the formula:

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

when $P(A \cap B) = 0$.

—

Problem 6.7: Conditional Probability

Problem Statement

1. A card is drawn from a standard deck of 52 cards. Find the probability that the card is a heart given that it is red.
2. A bag contains 4 red balls and 6 blue balls. Two balls are drawn without replacement. Find the probability that:

1. The second ball is red given that the first ball is red.
2. The second ball is blue given that the first ball is red.

3. Explain the formula for conditional probability:

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

and how it can be applied using Venn diagrams, tree diagrams, or tables of outcomes.

—

Problem 6.8: Independent Events

Problem Statement

1. Two events, A and B, are independent. If $P(A) = 0.5$ and $P(B) = 0.6$, calculate:

1. $P(A \cap B)$
2. $P(A \cup B)$

2. Explain the concept of independent events and the formula:

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

and how it differs from mutually exclusive events.

—

Key Concepts and Definitions

Key Concepts and Definitions

1. **Venn Diagrams**: A visual representation of events and their relationships, used to calculate probabilities.
2. **Tree Diagrams**: A branching diagram that represents all possible outcomes of a sequence of events. Multiply along branches and add between branches.
3. **Sample Space Diagrams**: A grid or list showing all possible outcomes of an experiment.
4. **Tables of Outcomes**: A tabular representation of all possible outcomes, often used for two-stage experiments.
5. **Combined Events**:
$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$
6. **Mutually Exclusive Events**: Events that cannot occur simultaneously, so $P(A \cap B) = 0$.
7. **Conditional Probability**: The probability of A given B:
$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$
8. **Independent Events**: Events that do not affect each other, so:
$$P(A \cap B) = P(A)P(B)$$

—

Marking Guidelines

Marking Scheme

Problem 6.1: Venn Diagrams

- Correct calculation of probabilities [2 marks per part]
- Valid explanation of how Venn diagrams are used [2 marks]

Problem 6.2: Tree Diagrams

- Accurate construction of the tree diagram [2 marks]
- Correct calculation of probabilities [2 marks per part]
- Valid explanation of the rules for tree diagrams [2 marks]

Problem 6.3: Sample Space Diagrams

- Correct listing of all outcomes [2 marks]
- Accurate calculation of probabilities [2 marks per part]
- Valid explanation of how sample space diagrams are used [2 marks]

Problem 6.4: Tables of Outcomes

- Correct construction of the table [2 marks]
- Accurate calculation of probabilities [2 marks per part]
- Valid explanation of how tables of outcomes are used [2 marks]

Problem 6.5: Combined Events

- Correct calculation of probabilities using the formula [2 marks per part]
- Valid explanation of the formula for combined events [2 marks]

Problem 6.6: Mutually Exclusive Events

- Correct calculation of probabilities [2 marks per part]
- Valid explanation of mutually exclusive events [2 marks]

Problem 6.7: Conditional Probability

- Correct calculation of conditional probabilities [2 marks per part]
- Valid explanation of the formula for conditional probability [2 marks]

Problem 6.8: Independent Events

- Correct calculation of probabilities [2 marks per part]
- Valid explanation of independent events [2 marks]

Additional Points

- Clear presentation of solutions [1 mark]
- Logical reasoning in calculations [1 mark]

7 Probability: Discrete Random Variables

Problem 7.1: Probability Distributions

Problem Statement

1. A fair six-sided die is rolled. Create the probability distribution for the random variable X, where X represents the number rolled.
2. A bag contains 3 red balls, 2 blue balls, and 1 green ball. A ball is drawn at random, and the random variable X represents the number of red balls drawn. Create the probability distribution for X.
3. Explain the concept of a discrete random variable and how its probability distribution is created from context.

—

Problem 7.2: Total Probability in a Distribution

Problem Statement

1. Verify that the following is a valid probability distribution:

X	$P(X)$
1	0.2
2	0.3
3	0.4
4	0.1

2. A random variable X has the following probabilities:

$$P(X = 1) = 0.25, \quad P(X = 2) = 0.35, \quad P(X = 3) = 0.15.$$

Find $P(X = 4)$ if the total probability must equal 1.
3. Explain why the total probability in a probability distribution must equal 1.

—

Problem 7.3: Expected Value (Mean) of Discrete Data

Problem Statement

1. Calculate the expected value $E(X)$ for the following probability distribution:

X	$P(X)$
1	0.2
2	0.3
3	0.4
4	0.1

2. A game involves rolling a fair six-sided die. The random variable X represents the winnings, where:

$$X = \begin{cases} 5 & \text{if a 6 is rolled,} \\ 0 & \text{otherwise.} \end{cases}$$

Find the expected value $E(X)$.

3. Explain the formula for the expected value:

$$E(X) = \sum xP(X = x)$$

and its significance in probability.

—

Problem 7.4: Applications of Probability Distributions

Problem Statement

1. A factory produces light bulbs, and the probability distribution for the number of defective bulbs in a batch of 5 is given by:

X	$P(X)$
0	0.5
1	0.3
2	0.15
3	0.05

1. Find the expected number of defective bulbs in a batch.
2. What is the probability of having at least 2 defective bulbs in a batch?

2. A game involves flipping a fair coin three times. The random variable X represents the number of heads obtained. Create the probability distribution for X and calculate $E(X)$.
3. Explain how probability distributions can be used to answer questions in context.

—

Problem 7.5: Fair Games and Expected Value

Problem Statement

1. A game involves rolling a fair six-sided die. The player wins $10ifa6isrolledandloses2$ otherwise. Let X represent the gain of the player. Find $E(X)$ and determine if the game is fair.
2. A spinner has 4 equal sections labeled 1, 2, 3, and 4. A player wins $5ifthespinnerlandson4andloses1$ otherwise. Let X represent the gain of the player. Find $E(X)$ and determine if the game is fair.
3. Explain the concept of a fair game and how $E(X) = 0$ indicates fairness.

—

Key Concepts and Definitions

Key Concepts and Definitions

1. **Discrete Random Variable**: A variable that takes on a finite or countable number of values, each with an associated probability.
2. **Probability Distribution**: A table or function that assigns probabilities to each possible value of a discrete random variable. The total probability must equal 1.
3. **Expected Value (Mean)**:

$$E(X) = \sum xP(X = x)$$

The expected value represents the long-term average value of the random variable.

4. **Fair Game**: A game is fair if the expected value of the player's gain is 0, i.e., $E(X) = 0$.

—

Marking Guidelines

Marking Scheme

Problem 7.1: Probability Distributions

- Correct creation of probability distributions [2 marks per part]
- Valid explanation of discrete random variables [2 marks]

Problem 7.2: Total Probability in a Distribution

- Correct verification of total probability [2 marks per part]
- Accurate calculation of missing probabilities [2 marks per part]
- Valid explanation of why total probability equals 1 [2 marks]

Problem 7.3: Expected Value (Mean) of Discrete Data

- Correct calculation of $E(X)$ [2 marks per part]
- Valid explanation of the formula for expected value [2 marks]

Problem 7.4: Applications of Probability Distributions

- Correct calculation of probabilities and expected values [2 marks per part]
- Valid explanation of how distributions are used in context [2 marks]

Problem 7.5: Fair Games and Expected Value

- Correct calculation of $E(X)$ [2 marks per part]
- Valid determination of whether the game is fair [2 marks per part]
- Explanation of the concept of a fair game [2 marks]

Additional Points

- Clear presentation of solutions [1 mark]
- Logical reasoning in calculations [1 mark]

8 Probability: Binomial Distribution

Problem 8.1: Recognizing a Binomial Distribution

Problem Statement

1. For each of the following scenarios, determine whether the situation can be modeled using a binomial distribution. Justify your answer.

1. A coin is flipped 10 times, and the number of heads is recorded.
2. A die is rolled 20 times, and the number of times a 6 is rolled is recorded.
3. A survey is conducted to determine whether people prefer tea or coffee, and 50 people are surveyed.
4. A student takes a multiple-choice test with 10 questions, each with 4 options, and guesses on each question.
5. A factory produces 100 light bulbs, and the number of defective bulbs is recorded.

2. Explain the conditions for a binomial distribution:

- Fixed number of trials.
- Two possible outcomes (success and failure).
- Fixed probability of success.
- Independent trials.

—

Problem 8.2: Calculating Binomial Probabilities

Problem Statement

1. Use technology to calculate the following binomial probabilities:

1. $P(X = 3)$ where $X \sim \text{Bin}(n = 5, p = 0.6)$.
2. $P(X \leq 2)$ where $X \sim \text{Bin}(n = 8, p = 0.4)$.
3. $P(X \geq 4)$ where $X \sim \text{Bin}(n = 10, p = 0.7)$.

2. A basketball player has a 75% chance of making a free throw. If the player takes 6 free throws, calculate the probability of:

1. Making exactly 4 free throws.
2. Making at least 5 free throws.
3. Making no more than 2 free throws.

3. Explain how technology (e.g., a calculator or software) can be used to calculate binomial probabilities efficiently.

—

Problem 8.3: Mean and Variance of the Binomial Distribution

Problem Statement

1. For each of the following binomial distributions, calculate the mean $E(X)$ and variance $\text{Var}(X)$:

1. $X \sim \text{Bin}(n = 10, p = 0.5)$
2. $X \sim \text{Bin}(n = 20, p = 0.3)$
3. $X \sim \text{Bin}(n = 15, p = 0.8)$

2. A factory produces light bulbs, and 5% of the bulbs are defective. If a random sample of 50 bulbs is taken, calculate:

1. The expected number of defective bulbs.
2. The variance in the number of defective bulbs.

3. Explain the formulas for the mean and variance of a binomial distribution:

$$E(X) = np, \quad \text{Var}(X) = np(1 - p)$$

and their significance in understanding the distribution.

—

Key Concepts and Definitions

Key Concepts and Definitions

1. **Binomial Distribution**: A discrete probability distribution that models the number of successes in n independent trials, each with a probability p of success. The conditions for a binomial distribution are:
   - Fixed number of trials (n).
   - Two possible outcomes (success and failure).
   - Fixed probability of success (p).
   - Independent trials.
2. **Binomial Probability Formula**: The probability of exactly k successes in n trials is given by:
   $$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$
   where $\binom{n}{k} = \frac{n!}{k!(n-k)!}$.
3. **Mean and Variance**: For a binomial random variable $X \sim \text{Bin}(n, p)$:
   $$E(X) = np, \quad \text{Var}(X) = np(1-p)$$
   The mean represents the expected number of successes, and the variance measures the spread of the distribution.

—

Marking Guidelines

Marking Scheme

Problem 8.1: Recognizing a Binomial Distribution

- Correct identification of whether the situation is binomial [2 marks per part]
- Valid explanation of the conditions for a binomial distribution [2 marks]

Problem 8.2: Calculating Binomial Probabilities

- Correct calculation of binomial probabilities using technology [2 marks per part]
- Valid explanation of how technology is used [2 marks]

Problem 8.3: Mean and Variance of the Binomial Distribution

- Correct calculation of mean and variance [2 marks per part]
- Valid explanation of the formulas for mean and variance [2 marks]

Additional Points

- Clear presentation of solutions [1 mark]
- Logical reasoning in calculations [1 mark]

9 Probability: Normal Distribution

Problem 9.1: Properties of the Normal Distribution

Problem Statement

1. Explain the key properties of the normal distribution:

- Symmetry about the mean.
- The total area under the curve equals 1.
- Approximately 68% of the data lies within one standard deviation of the mean, 95% within two standard deviations, and 99.7% within three standard deviations.

2. A dataset is approximately normally distributed with a mean of 50 and a standard deviation of 5. Use the 68-95-99.7 rule to estimate the percentage of data that lies:

1. Between 45 and 55.
2. Between 40 and 60.
3. Outside the range 35 to 65.

3. Discuss why many natural phenomena (e.g., heights, test scores) are well modeled by a normal distribution.

—

Problem 9.2: Diagrammatic Representation of the Normal Distribution

Problem Statement

1. Sketch a normal distribution curve for a dataset with a mean of 100 and a standard deviation of 15. Label the mean and the points at one, two, and three standard deviations from the mean.
2. Explain how the area under the curve represents probability and why the curve is symmetric about the mean.
3. For a normal distribution with mean $\mu = 70$ and standard deviation $\sigma = 10$, shade the region representing the probability of the random variable being between 60 and 80.

—

Problem 9.3: Normal Probability Calculations

Problem Statement

1. A random variable X follows a normal distribution with mean $\mu = 50$ and standard deviation $\sigma = 10$. Use technology to calculate the probability that:

1. $X < 40$
2. $X > 60$
3. $40 < X < 60$

2. The heights of a group of students are normally distributed with a mean of 170 cm and a standard deviation of 8 cm. Find the probability that a randomly selected student has a height:

1. Less than 160 cm.
2. Greater than 180 cm.
3. Between 165 cm and 175 cm.

3. Explain how to use the standard normal distribution (Z-scores) to calculate probabilities for any normal distribution.

—

Problem 9.4: Inverse Normal Calculations

Problem Statement

1. A random variable X follows a normal distribution with mean $\mu = 100$ and standard deviation $\sigma = 15$. Use technology to find:

1. The value of X such that $P(X < x) = 0.25$.
2. The value of X such that $P(X > x) = 0.10$.
3. The values of X that enclose the middle 90% of the distribution.

2. The weights of apples in a farm are normally distributed with a mean of 150 g and a standard deviation of 20 g. Find:

1. The weight below which 5% of the apples fall.
2. The weight above which 10% of the apples fall.
3. The range of weights that contains the middle 80% of the apples.

3. Explain the concept of inverse normal calculations and how they are used to find boundaries for given probabilities.

—

Key Concepts and Definitions

Key Concepts and Definitions

1. **Normal Distribution**: A continuous probability distribution that is symmetric about the mean and follows a bell-shaped curve. It is defined by two parameters:
   - Mean (μ): The center of the distribution.
   - Standard deviation (σ): The spread of the distribution.
2. **68-95-99.7 Rule**: For a normal distribution:
   - 68% of the data lies within one standard deviation of the mean.
   - 95% of the data lies within two standard deviations of the mean.
   - 99.7% of the data lies within three standard deviations of the mean.
3. **Standard Normal Distribution**: A normal distribution with mean $\mu = 0$ and standard deviation $\sigma = 1$. Z-scores are used to standardize any normal distribution:

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

4. **Inverse Normal Calculations**: The process of finding the value of X for a given probability. This is the reverse of normal probability calculations.

—

Marking Guidelines

Marking Scheme

Problem 9.1: Properties of the Normal Distribution

- Correct explanation of the properties of the normal distribution [2 marks]
- Accurate use of the 68-95-99.7 rule [2 marks per part]
- Valid discussion of why natural phenomena follow a normal distribution [2 marks]

Problem 9.2: Diagrammatic Representation of the Normal Distribution

- Accurate sketch of the normal distribution curve [2 marks]
- Correct labeling of mean and standard deviations [2 marks]
- Valid explanation of the area under the curve representing probability [2 marks]

Problem 9.3: Normal Probability Calculations

- Correct calculation of probabilities using technology [2 marks per part]
- Valid explanation of how Z-scores are used [2 marks]

Problem 9.4: Inverse Normal Calculations

- Correct calculation of boundaries for given probabilities [2 marks per part]
- Valid explanation of inverse normal calculations [2 marks]

Additional Points

- Clear presentation of solutions [1 mark]
- Logical reasoning in calculations [1 mark]

10 Statistics: X-on-Y Regression

Problem 10.1: Equation of the Regression Line of x on y

Problem Statement

1. Use your GDC (Graphical Display Calculator) to find the regression line of x on y for the following data set:

y	x
1	2
2	4
3	6
4	8
5	10

Write the equation of the regression line in the form:

$$x = a + by$$

2. For the following data set, calculate the regression line of x on y using your GDC:

y	x
10	15
20	25
30	35
40	45
50	55

3. Explain the difference between the regression line of x on y and the regression line of y on x.

—

Problem 10.2: Using the Regression Line for Prediction

Problem Statement

1. For the regression line $x = 2 + 3y$, predict the value of x when:

1. $y = 5$
2. $y = 10$

2. For the regression line $x = -1 + 0.5y$, predict the value of x when:

1. $y = 8$
2. $y = 20$

3. Discuss the dangers of extrapolation when using the regression line of x on y for prediction purposes.

4. Explain when it is appropriate to use the regression line of x on y instead of the regression line of y on x.

—

Key Concepts and Definitions

Key Concepts and Definitions

1. **Regression Line of x on y**: The regression line of x on y is used to predict the value of x for a given value of y. It minimizes the sum of squared horizontal deviations between the data points and the line.

2. **Equation of the Regression Line**: The regression line of x on y is written in the form:
$$x = a + by$$
where:
   - a is the intercept (the value of x when $y = 0$).
   - b is the slope (the rate of change of x with respect to y).

3. **Prediction Using the Regression Line**: The regression line can be used to predict the value of x for a given value of y. However, predictions should be made cautiously, especially when extrapolating beyond the range of the data.

4. **Extrapolation**: Using the regression line to predict values outside the range of the observed data. This can lead to unreliable predictions as the relationship may not hold outside the observed range.

—

Marking Guidelines

Marking Scheme

Problem 10.1: Equation of the Regression Line of x on y

- Correct use of the GDC to calculate the regression line [2 marks per part]
- Accurate equation of the regression line in the form $x = a + by$ [2 marks per part]
- Valid explanation of the difference between x-on-y and y-on-x regression lines [2 marks]

Problem 10.2: Using the Regression Line for Prediction

- Correct prediction of x for given values of y [2 marks per part]
- Valid discussion of the dangers of extrapolation [2 marks]
- Explanation of when to use the regression line of x on y [2 marks]

Additional Points

- Clear presentation of solutions [1 mark]
- Logical reasoning in calculations [1 mark]

11 Probability: Formal Conditional Probability

Problem 11.1: Using the Conditional Probability Formula

Problem Statement

1. Given $P(A) = 0.6$, $P(B) = 0.4$, and $P(A \cap B) = 0.2$, calculate:

1. $P(A|B)$
2. $P(B|A)$

Show all steps using the formula:

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

2. In a group of students:

- 60% study mathematics
- 40% study physics
- 25% study both mathematics and physics

Calculate:

1. The probability that a student studies mathematics given that they study physics.
2. The probability that a student studies physics given that they study mathematics.

3. Explain the meaning of conditional probability and how it differs from joint probability.

—

Problem 11.2: Testing for Independence

Problem Statement

1. Two events A and B are independent if $P(A|B) = P(A)$. Use this to test whether the following events are independent:

1. Given $P(A) = 0.3$, $P(B) = 0.4$, and $P(A \cap B) = 0.12$
2. Given $P(A) = 0.5$, $P(B) = 0.6$, and $P(A \cap B) = 0.3$

2. A card is drawn from a standard deck of 52 cards. Let:

- A be the event of drawing a heart
- B be the event of drawing a red card

Determine whether events A and B are independent.
3. Explain the three equivalent conditions for independence:

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

—

Problem 11.3: Applications of Conditional Probability

Problem Statement

1. A medical test for a disease has the following probabilities:

- $P(D)$ = probability of having the disease = 0.01
- $P(+|D)$ = probability of testing positive given disease = 0.95
- $P(+|\neg D)$ = probability of testing positive given no disease = 0.02

Calculate:

1. $P(D \cap +)$
2. $P(+)$
3. $P(D|+)$ (the probability of having the disease given a positive test)

2. Explain how conditional probability is used in real-world applications such as medical testing, weather forecasting, or quality control.

—

Key Concepts and Definitions

Key Concepts and Definitions

1. **Conditional Probability**: The probability of event A occurring given that event B has occurred:

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

2. **Independent Events**: Events A and B are independent if any of the following equivalent conditions hold:

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

3. **Joint Probability**: The probability of both events occurring:

$$P(A \cap B)$$

4. **Multiplication Rule**: For any two events:

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

—

Marking Guidelines

Marking Scheme

Problem 11.1: Using the Conditional Probability Formula

- Correct calculation of conditional probabilities [2 marks per part]
- Clear use of the conditional probability formula [2 marks per part]
- Valid explanation of conditional probability [2 marks]

Problem 11.2: Testing for Independence

- Correct testing for independence [2 marks per part]
- Valid explanation of the conditions for independence [2 marks]
- Clear reasoning in determining independence [2 marks]

Problem 11.3: Applications of Conditional Probability

- Correct calculation of probabilities [2 marks per part]
- Valid explanation of real-world applications [2 marks]
- Clear understanding of the context [2 marks]

Additional Points

- Clear presentation of solutions [1 mark]
- Logical reasoning in calculations [1 mark]

12 Probability: Standardizing Normal Variables

Problem 12.1: Finding Z-Values

Problem Statement

1. For a normal distribution with mean $\mu = 50$ and standard deviation $\sigma = 10$, calculate the z-value for:

1. $x = 60$
2. $x = 40$
3. $x = 50$

2. Explain the meaning of a z-value and how it represents the number of standard deviations a value is from the mean.
3. A student scores 85 on a test where the mean score is 70 and the standard deviation is 10. Find the z-value for the student's score and interpret its meaning.
4. Derive the formula for standardizing a normal variable:

$$z = \frac{x - \mu}{\sigma}$$

and explain each term in the formula.

—

Problem 12.2: Inverse Normal Calculations with Unknown Mean and Standard Deviation

Problem Statement

1. A random variable X is normally distributed. It is known that:

- $P(X < 70) = 0.25$
- $P(X > 90) = 0.10$

Use the inverse normal function on your GDC to find the mean μ and standard deviation σ of X.

2. The weights of apples in a farm are normally distributed. It is known that:

- 5% of the apples weigh less than 150 g.
- 10% of the apples weigh more than 200 g.

Find the mean μ and standard deviation σ of the weights of the apples.

3. Explain how the inverse normal function and z-values are used to find unknown parameters of a normal distribution.

—

Key Concepts and Definitions

Key Concepts and Definitions

1. **Z-Value (Standard Score)**: A z-value represents the number of standard deviations a value x is from the mean μ. It is calculated using the formula:

$$z = \frac{x - \mu}{\sigma}$$

where:

- x is the value of the random variable.
- μ is the mean of the distribution.
- σ is the standard deviation of the distribution.

2. **Standard Normal Distribution**: A normal distribution with mean $\mu = 0$ and standard deviation $\sigma = 1$. Z-scores are used to standardize any normal distribution to the standard normal distribution.

3. **Inverse Normal Calculations**: The process of finding the value of x or the parameters μ and σ for a given probability. This is done using the inverse normal function on a GDC and the relationship between z-values and probabilities.

4. **Key Properties**:

- $z > 0$: The value x is above the mean.
- $z < 0$: The value x is below the mean.
- $z = 0$: The value x is equal to the mean.

—

Marking Guidelines

Marking Scheme

Problem 12.1: Finding Z-Values

- Correct calculation of z-values using the formula [2 marks per part]
- Valid explanation of the meaning of z-values [2 marks]
- Clear derivation of the standardization formula [2 marks]

Problem 12.2: Inverse Normal Calculations with Unknown Mean and Standard Deviation

- Correct use of the inverse normal function on the GDC [2 marks per part]
- Accurate calculation of the mean μ and standard deviation σ [2 marks per part]
- Valid explanation of the process and its applications [2 marks]

Additional Points

- Clear presentation of solutions [1 mark]
- Logical reasoning in calculations [1 mark]

13 Probability: Bayes' Theorem

Problem 13.1: Two-Event Bayes' Theorem

Problem Statement

1. Given:

- $P(B) = 0.4$
- $P(A|B) = 0.3$
- $P(A|\neg B) = 0.2$

Calculate $P(B|A)$ using:

$$P(B|A) = \frac{P(B)P(A|B)}{P(B)P(A|B) + P(\neg B)P(A|\neg B)}$$

2. A medical test for a disease has the following probabilities:

- Probability of having the disease: $P(D) = 0.01$
- Probability of a positive test given disease: $P(+|D) = 0.95$
- Probability of a positive test given no disease: $P(+|\neg D) = 0.02$

Calculate the probability of having the disease given a positive test result using:

1. The formula method
2. A tree diagram method

3. Explain the relationship between Bayes' theorem and conditional probability.

—

Problem 13.2: Three-Event Bayes' Theorem

Problem Statement

1. A factory has three machines (A, B, and C) that produce identical items. The probabilities are:

- Machine A produces 30% of items: $P(A) = 0.3$
- Machine B produces 45% of items: $P(B) = 0.45$
- Machine C produces 25% of items: $P(C) = 0.25$

The probability of a defective item from each machine is:

- Machine A: $P(D|A) = 0.02$
- Machine B: $P(D|B) = 0.03$
- Machine C: $P(D|C) = 0.04$

Calculate:

1. The probability that a randomly selected item is defective.
2. The probability that a defective item was produced by Machine B.

2. Use the formula:

$$P(B_i|A) = \frac{P(B_i)P(A|B_i)}{P(B_1)P(A|B_1) + P(B_2)P(A|B_2) + P(B_3)P(A|B_3)}$$

to solve a three-event problem and compare the result with the tree diagram method.

3. Explain the advantages and disadvantages of using:

1. The formula method
2. The tree diagram method

—

Key Concepts and Definitions

Key Concepts and Definitions

1. **Bayes' Theorem (Two Events)**:

$$P(B|A) = \frac{P(B)P(A|B)}{P(B)P(A|B) + P(\neg B)P(A|\neg B)}$$

or using the conditional probability formula:

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

2. **Bayes' Theorem (Three Events)**:

$$P(B_i|A) = \frac{P(B_i)P(A|B_i)}{P(B_1)P(A|B_1) + P(B_2)P(A|B_2) + P(B_3)P(A|B_3)}$$

3. **Tree Diagram Method**:
   - Draw branches for each event and their complements.
   - Write probabilities along branches.
   - Multiply along branches for joint probabilities.
   - Add across branches for total probabilities.

4. **Applications**: Bayes' theorem is used in:
   - Medical diagnosis
   - Quality control
   - Machine learning
   - Decision making under uncertainty

—

Marking Guidelines

Marking Scheme

Problem 13.1: Two-Event Bayes' Theorem

- Correct calculation using the formula method [2 marks per part]
- Accurate construction and use of tree diagram [2 marks per part]
- Valid explanation of the relationship with conditional probability [2 marks]

Problem 13.2: Three-Event Bayes' Theorem

- Correct calculation of probabilities [2 marks per part]
- Accurate use of the three-event formula [2 marks]
- Valid comparison of methods [2 marks]
- Clear explanation of advantages and disadvantages [2 marks]

Additional Points

- Clear presentation of solutions [1 mark]
- Logical reasoning in calculations [1 mark]

14 Probability: Random Variables

Problem 14.1: Variance of a Discrete Random Variable

Problem Statement

1. For the following probability distribution, calculate the variance Var(X):

X	$P(X)$
1	0.2
2	0.3
3	0.4
4	0.1

Use the formula:

$$\text{Var}(X) = \sum x^2 P(X = x) - \mu^2, \quad \text{where } \mu = E(X).$$

2. A game involves rolling a fair six-sided die. The random variable X represents the winnings, where:

$$X = \begin{cases} 5 & \text{if a 6 is rolled,} \\ 0 & \text{otherwise.} \end{cases}$$

Find the variance Var(X).

—

Problem 14.2: Continuous Random Variables and Probability Density Functions (PDFs)

Problem Statement

1. A continuous random variable X has the probability density function:

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

1. Find the value of k such that $\int_{-\infty}^{\infty} f(x)dx = 1$.
2. Calculate $P(0.5 \leq X \leq 1.5)$.

2. Explain the properties of a probability density function (PDF):

- $f(x) \geq 0$ for all x.
- $\int_{-\infty}^{\infty} f(x)dx = 1$.

—

Problem 14.3: Piecewise Defined PDFs

Problem Statement

1. A continuous random variable X has the following piecewise PDF:

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

1. Verify that $f(x)$ is a valid PDF.
2. Find $P(0.5 \leq X \leq 1.5)$.
3. Identify the interval in which the median lies and calculate the median m such that:
$$\int_{-\infty}^{m} f(x)dx = \frac{1}{2}.$$

—

Problem 14.4: Mean, Variance, and Standard Deviation of Continuous Random Variables

Problem Statement

1. A continuous random variable X has the PDF:

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

1. Find the mean $E(X)$ using:

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

2. Find $E(X^2)$ using:

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

3. Calculate the variance $\text{Var}(X)$ using:

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

2. Explain how to calculate the mean and variance for a piecewise PDF by splitting the integrals into separate parts.

—

Problem 14.5: Linear Transformations of Random Variables

Problem Statement

1. A random variable X has $E(X) = 5$ and $\text{Var}(X) = 4$. Find $E(2X + 3)$ and $\text{Var}(2X + 3)$ using:

$$E(aX + b) = aE(X) + b, \quad \text{Var}(aX + b) = a^2\text{Var}(X).$$

2. A game involves rolling a fair six-sided die. The random variable X represents the winnings, where:

$$X = \begin{cases} 10 & \text{if a 6 is rolled,} \\ -2 & \text{otherwise.} \end{cases}$$

Find the value of b such that the game is fair, i.e., $E(X + b) = 0$.

—

Key Concepts and Definitions

Key Concepts and Definitions

1. **Variance of a Discrete Random Variable**:

$$\text{Var}(X) = \sum x^2 P(X = x) - \mu^2, \quad \text{where } \mu = E(X).$$

2. **Continuous Random Variables**: A continuous random variable X is represented by a probability density function (PDF) $f(x)$, which satisfies:

   - $f(x) \geq 0$ for all x.
   - $\int_{-\infty}^{\infty} f(x)dx = 1$.

3. **Mean and Variance of a Continuous Random Variable**:

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx, \quad E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)dx, \quad \text{Var}(X) = E(X^2) - [E(X)]^2.$$

4. **Linear Transformations**: For a random variable X:

$$E(aX + b) = aE(X) + b, \quad \text{Var}(aX + b) = a^2\text{Var}(X).$$

5. **Median and Mode of a Continuous Random Variable**:

   - The mode corresponds to the maximum value of $f(x)$.
   - The median m satisfies:

$$\int_{-\infty}^{m} f(x)dx = \frac{1}{2}.$$

—

Marking Guidelines

Marking Scheme

Problem 14.1: Variance of a Discrete Random Variable

- Correct calculation of $E(X)$ and Var(X) [2 marks per part]
- Valid explanation of the formula for variance [2 marks]

Problem 14.2: Continuous Random Variables and PDFs

- Correct verification of PDF properties [2 marks]
- Accurate calculation of probabilities [2 marks per part]

Problem 14.3: Piecewise Defined PDFs

- Correct verification of PDF validity [2 marks]
- Accurate calculation of probabilities and median [2 marks per part]

Problem 14.4: Mean, Variance, and Standard Deviation of Continuous Random Variables

- Correct calculation of $E(X)$, $E(X^2)$, and Var(X) [2 marks per part]
- Valid explanation of splitting integrals for piecewise PDFs [2 marks]

Problem 14.5: Linear Transformations of Random Variables

- Correct calculation of $E(aX + b)$ and Var$(aX + b)$ [2 marks per part]
- Accurate determination of b for a fair game [2 marks]

Additional Points

- Clear presentation of solutions [1 mark]
- Logical reasoning in calculations [1 mark]

Conclusion

Mathematics is not just about understanding theory; it is about applying concepts to solve problems effectively. This guide has provided you with a collection of expertly crafted practice problems focused on Probability, designed to challenge your understanding and enhance your problem-solving skills.
For detailed solutions and answers, keep following me — they will be available soon! If you're looking for personalized guidance, book a one-on-one mentorship session with me to deepen your understanding of IB Mathematics AA/AI HL, Probability & Statistics, or even Olympiad-level problems. Together, we can build the confidence and skills you need to excel in mathematics.

As you prepare for your exams, remember:

- **Practice is the key to success**: The more problems you solve, the more confident and efficient you become. Focus on understanding the logic behind each solution rather than memorizing formulas.
- **Learn from mistakes**: Every mistake is an opportunity to grow. Analyze where you went wrong and refine your approach.
- **Time management is crucial**: Simulate exam conditions to improve your speed and accuracy under pressure.

If you're aiming for a guaranteed improvement and want to elevate your performance to the next level, consider applying for my **exclusive personalized mentorship program**. As an alumnus of IIT Guwahati and ISI, with over 5 years of teaching experience from the school level to university students, now mentoring high-achieving IB students, I specialize in:

- **Tailored guidance**: Customized study plans and strategies based on your strengths and weaknesses.
- **Exam-focused preparation**: Insights into examiner expectations and tips to maximize your score.
- **Beyond IB HL Problem-Solving**: My mentorship is not limited to IB HL Mathematics. I will enrich your mathematical thinking to push you toward **Olympiad-level problem-solving** and help you excel in **quantitative aptitude**, preparing you for competitive exams and real-world challenges.
- **One-on-one mentorship**: Direct support to clarify doubts, build confidence, and achieve your goals.

Join the ranks of students who have transformed their performance and achieved top scores with my mentorship. Visit www.mathematicselevateacademy.com to access free resources, book a session, or apply for the program. Let's work together to make your IB Mathematics journey a success!

"Success in mathematics comes not from the number of problems you've solved, but from the confidence you've gained in solving them."

- Rishabh Kumar
Founder, Mathematics Elevate Academy
Elite Mentor for IB Mathematics
Alumnus of IIT Guwahati & Indian Statistical Institute

Thank You!

Rishabh Kumar
Mathematics Elevate Academy
www.mathematicselevateacademy.com