Probability and Statistics

Lecture Notes

HAMZA BERAHMA

Student

December 26, 2025

For

my idea of beautiful

Contents

Chapter 1

Foundations of Probability

1.1 Probability and Intuition

Sometimes our intuition misleads us. To be more accurate, we need to recognize that intuition can be wrong for mainly two reasons:

1. Misunderstanding the terminology of the problem.
2. Biases from past experiences.

Intuition. Intuition about probability often fails because our minds rely on past experience or oversimplified reasoning. Being aware of this can help us approach problems more carefully.

Problem. Consider the following statements. Decide whether they are true or false, and explain why.

Example (Fair Coin Toss). **Statement:** When we toss a fair coin, the chance of heads is $\frac{1}{2}$ and the chance of tails is $\frac{1}{2}$.

Answer: True. This is the definition of a fair coin.

Example (Possible Outcomes). **Statement:** When we toss a coin, there are 2 possible outcomes: H or T.

Answer: Mostly true, but it depends on the experiment's definition. What if the coin lands on its edge? We typically simplify the model to include only Heads and Tails.

Example (Mean vs. Median). **Statement:** We have a set of 1000 numbers. Their average is A. Half the numbers are $\leq A$, and half are $> A$.

Answer: False. This describes the **median**, not the **average** (mean). Consider the set $\{1, 2, 97\}$. The average is $A = 33.3$, but two-thirds of the numbers are less than A.

Example (Random Number Probability). **Statement:** I will randomly pick a number from $\{1, 2, 3, 4\}$. What is $P(1)$?

Answer: It is $\frac{1}{4}$, but only if "randomly" implies that each outcome is equally likely (a uniform distribution). The term "random" alone is not sufficient.

Example (Average Number of Legs). **Statement:** What are the chances that the next person who walks into this room has more than the average number of legs?

Answer: Extremely high (close to 100%). The vast majority of people have two legs. A very small number of people have one or zero legs. The average number of legs for a human population will be slightly less than 2 (e.g., 1.999...). Therefore, almost everyone has more than the average.

Remark. Always clarify definitions and assumptions in probability problems. Words like "random" or "average" can be misleading if interpreted naively.

1.2 Random Experiments and Sample Spaces

To formalize probability, we begin with a few core definitions.

Definition 1 (Random Experiment). A well-defined procedure that results in an uncertain outcome.

Definition 2 (Outcome). A single result of a random experiment.

Definition 3 (Sample Space). The set of all possible outcomes of a random experiment, denoted S.

Definition 4 (Trial). Each attempt or performance of a random experiment is called a trial.

Definition 5 (Event). Any subset of the sample space ($A \subseteq S$). An event is a set of outcomes.

Remark. The most critical step in solving a probability problem is often defining the sample space correctly. A poorly defined sample space can lead to incorrect probability calculations.

Example (Examples of Sample Spaces). Here are several random experiments and their corresponding sample spaces:

E_1: Roll a fair die and record the top face.
$S_1 = \{1, 2, 3, 4, 5, 6\}$

E_2: Toss a coin twice and record the sequence of outcomes.
$S_2 = \{HH, HT, TH, TT\}$

E_3: Toss a die twice and record the sum of the faces.
$S_3 = \{2, 3, \ldots, 12\}$

E_4: Toss a 3-sided die twice and record the ordered pair of results.
$S_4 = \{(i, j) : i, j \in \{1, 2, 3\}\} = \{(1, 1), (1, 2), (1, 3), (2, 1), \ldots, (3, 3)\}$

E_5: Toss a 3-sided die twice and record the set of results (order ignored).
$S_5 = \{\{i, j\} : i, j \in \{1, 2, 3\}\} = \{\{1, 1\}, \{1, 2\}, \{1, 3\}, \{2, 2\}, \{2, 3\}, \{3, 3\}\}$

E_6: Toss a coin until the first head appears, recording the number of tosses.
$S_6 = \{1, 2, 3, \ldots\} = \mathbb{N}^+$

E_7: Pick a random English word and record its number of letters.
$S_7 = \{1, 2, 3, \ldots\}$

1.3 Probability Measures

A probability measure (or probability distribution) is a function that assigns a likelihood to each event.

Definition 6 (Probability Measure). A probability measure is a function

$$P : \mathscr{P}(S) \to [0, 1]$$

that maps events in the sample space S to a real number between 0 and 1. It must satisfy the following three axioms:

1. **Non-negativity:** For any event $A \subseteq S$, $P(A) \geq 0$.
2. **Normalization:** The probability of the entire sample space is 1, i.e., $P(S) = 1$.
3. **Additivity:** For any finite collection of pairwise disjoint events.

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

(meaning $A_i \cap A_j = \emptyset$ for $i \neq j$),
the probability of their union is the sum of their individual probabilities:

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

Remark. Intuitively, a probability measure quantifies how likely an event is to occur. The axioms ensure probabilities are consistent: they cannot be negative, the total probability is 1, and disjoint events' probabilities add up.

1.4 Computing Probabilities

Let's apply the concepts of probability measures to compute probabilities in various scenarios.

Example (Fair 12-sided Die). Throw a fair 12-sided die. The sample space is $S = \{1, 2, \dots, 12\}$. Each outcome has probability $\frac{1}{12}$. Let E be the event of rolling an even number and F be the event of rolling a number ≤ 4:

$$E = \{2, 4, 6, 8, 10, 12\}, \quad F = \{1, 2, 3, 4\}.$$

The probabilities are:

- $P(E) = \frac{|E|}{|S|} = \frac{6}{12} = \frac{1}{2}$
- $P(F) = \frac{|F|}{|S|} = \frac{4}{12} = \frac{1}{3}$
- $P(E \cap F) = P(\{2, 4\}) = \frac{2}{12} = \frac{1}{6}$
- $P(F^c) = 1 - P(F) = 1 - \frac{1}{3} = \frac{2}{3}$

Example (Sum of Two Dice). Throw a fair 6-sided die twice and record the sum. The sample space of ordered pairs is $S' = \{(i, j) : i, j \in \{1, \dots, 6\}\}$, so $|S'| = 36$. The sample space for the sum is $S = \{2, 3, \dots, 12\}$. Outcomes in S are not equally likely. Let E be the event that the sum is even.

$$P(\text{Sum} = 2) = \frac{1}{36}, \quad P(\text{Sum} = 3) = \frac{2}{36}, \quad P(\text{Sum} = 4) = \frac{3}{36}.$$

Sum	2	4	6	8	10	12
# Ways	1	3	5	5	3	1

So,

$$P(E) = \frac{1 + 3 + 5 + 5 + 3 + 1}{36} = \frac{18}{36} = \frac{1}{2}.$$

Example (Weighted Die). Consider a 5-sided die where $P(k) = c \cdot k$. Using normalization $\sum_{k=1}^{5} P(k) = 1$:

$$c(1+2+3+4+5) = 15c = 1 \Rightarrow c = \frac{1}{15}.$$

Then $P(k) = \frac{k}{15}$. The probability of rolling an even face is

$$P(\text{even}) = P(\{2,4\}) = \frac{2}{15} + \frac{4}{15} = \frac{6}{15} = \frac{2}{5}.$$

Example (Permutations). Consider all permutations of $\{1,2,3\}$ with equal probability. Let E be "starts with 1" and F be "2 comes before 3":

$$E = \{123, 132\} \Rightarrow P(E) = \frac{2}{6} = \frac{1}{3}, \quad F = \{123, 213, 231\} \Rightarrow P(F) = \frac{3}{6} = \frac{1}{2}.$$

Example (Weighted Permutations). Consider permutations of $\{1,2,3,4\}$. Let $P(abcd) \propto a$, the first element. Sum of weights:
$6(1) + 6(2) + 6(3) + 6(4) = 60 \Rightarrow P(x) = \frac{a}{60}$.

Let $E = \{abcd \in S_4 \mid a < b \text{ and } c < d\} = \{1234, 1324, 1423, 2314, 2413, 3412\}$. Then

$$P(E) = \frac{1+1+1+2+2+3}{60} = \frac{10}{60} = \frac{1}{6}.$$

Remark. Weighted probabilities allow modeling biased or non-uniform scenarios. Always normalize probabilities to ensure the total sum equals 1.

1.5 Conditional Probability

Conditional probability measures the likelihood of an event occurring given that another event has already occurred. It is a fundamental concept that allows us to update our beliefs in light of new evidence.

Definition 7 (Conditional Probability). Given two events A and B from a sample space S, with $P(B) > 0$, the **conditional probability of A given B** is defined as:

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

This formula re-scales the probability of the intersection of A and B by the probability of the given event, B. In essence, we are treating B as our new, smaller sample space.

Example (Medical Testing). Consider a medical test for a disease. Let S be the event that a person has the disease, and $\bar{S}$ the event they do not. Let `pos` and `neg` denote positive and negative test results, respectively.

We are given:

- The probability of having the disease (prevalence): $P(S) = 0.20$ ($P(\bar{S}) = 0.80$).
- Probability of a positive test if the patient has the disease (sensitivity): $P(\text{pos}|S) = 0.97$.
- Probability of a negative test if the patient does not have the disease (specificity): $P(\text{neg}|\bar{S}) = 0.90$.

From these, we deduce:

- False Negative Rate: $P(\text{neg}|S) = 1 - 0.97 = 0.03$.
- False Positive Rate: $P(\text{pos}|\bar{S}) = 1 - 0.90 = 0.10$.

Step 1: Compute Joint Probabilities Using the multiplication rule $P(A \cap B) = P(B)P(A|B)$:

$$
\begin{aligned}
P(S \cap \text{pos}) &= 0.20 \times 0.97 = 0.194, \\
P(S \cap \text{neg}) &= 0.20 \times 0.03 = 0.006, \\
P(\bar{S} \cap \text{pos}) &= 0.80 \times 0.10 = 0.080, \\
P(\bar{S} \cap \text{neg}) &= 0.80 \times 0.90 = 0.720.
\end{aligned}
$$

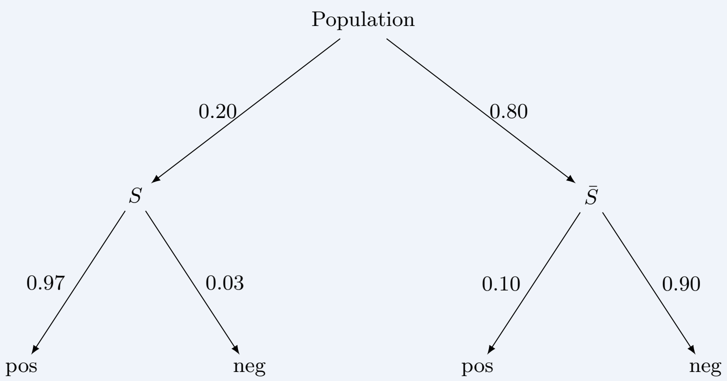


Figure 1.1: Tree diagram illustrating the relationships between disease status and test results.

Step 2: Compute the Marginal Probability of a Positive Test

$$
P(\text{pos}) = P(S \cap \text{pos}) + P(\bar{S} \cap \text{pos}) = 0.194 + 0.080 = 0.274.
$$

Step 3: Compute the Posterior Probability

$$P(S|\text{pos}) = \frac{P(S \cap \text{pos})}{P(\text{pos})} = \frac{0.194}{0.274} \approx 0.708.$$

Intuition (Understanding Bayes' Theorem). Even with a highly accurate test, a positive result does not guarantee the presence of the disease. This is because the overall prevalence of the disease in the population affects the probability. Bayes' theorem allows us to "update" our belief by combining the test's accuracy with prior information (prevalence).

Example (Blood Type Distribution). The joint probability distribution $P(\text{Type} \cap \text{Category})$ for individuals classified by blood type and category (P, Q, R, S, T) is given in Table 1.1.

Blood Type	P	Q	R	S	T
A	0.30	0.40	0.20	0.45	0.20
B	0.30	0.30	0.10	0.45	0.20
AB	0.20	0.10	0.40	0.05	0.30
O	0.20	0.20	0.30	0.05	0.30

Table 1.1: Rescaled joint distribution where each category column sums to 1.

(a) Probability of blood type A given category T

$$P(A \mid T) = \frac{P(A \cap T)}{P(T)} = \frac{0.20}{1.00} = 0.20.$$

(b) Probability of blood type B given category Q

$$P(B \mid Q) = \frac{P(B \cap Q)}{P(Q)} = \frac{0.30}{1.00} = 0.30.$$

(c) Probability of blood type A or B given category Q

$$P((A \cup B) \mid Q) = \frac{P(A \cap Q) + P(B \cap Q)}{P(Q)} = \frac{0.40 + 0.30}{1.00} = 0.70.$$

Remark. Conditional probability allows reasoning about events under additional information. Tree diagrams and tables are effective tools for visualizing and computing these probabilities.

1.6 Independence of Events

1.6.1 Joint Probability

For any two events A and B, the general rule is:

$$P(A \cap B) \neq P(A) \cdot P(B)$$

However, there is an important exception: if events A and B are independent, then

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

Proof. This multiplication rule applies only when the events do not influence each other.

1.6.2 Definition of Independence

Definition 8 (Independence). Events A and B are said to be **independent** if and only if:

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

Proof. Knowing the outcome of event A provides no information about the likelihood of event B, and vice versa.

1.6.3 Example 1: Fair Die Tosses

Example. Consider tossing a fair 6-sided die twice. We want to find the probability that:

the first face shows a "2" and the second face is odd.

Since the two tosses are independent, we can apply the **multiplication rule**:

$$\begin{aligned} P(\text{first face} = 2 \text{ and second face odd}) &= P(\text{first face} = 2) \cdot P(\text{second face odd}) \\ &= \frac{1}{6} \cdot \frac{3}{6} \\ &= \frac{1}{12}. \end{aligned}$$

Therefore, the probability is:

$$\frac{1}{12} \approx 0.0833 \text{ or } 8.33\%.$$

1.6.4 Example 2: Testing Independence Numerically

Example (Population Data). Consider selecting a Moroccan person at random:

- A: person is female
- B: person is a nurse
- C: person is a school teacher

Given Data:

$$P(A) = 0.50, \quad P(B) = 0.05, \quad P(C) = 0.01$$

80% of nurses are women, 50% of school teachers are women.

Step 1: Check A and B

$$P(A \cap B) = 0.80 \times 0.05 = 0.04, \quad P(A) \cdot P(B) = 0.50 \times 0.05 = 0.025$$

Observe. Since $0.04 \neq 0.025$, A and B are **not independent**.

Step 2: Check A and C

$$P(A \cap C) = 0.50 \times 0.01 = 0.005, \quad P(A) \cdot P(C) = 0.50 \times 0.01 = 0.005$$

Observe. Since $0.005 = 0.005$, A and C **are independent**.

Intuition (Independence Intuition)**.** Two events are independent if knowing one occurs does not change the probability of the other. Numerical checks reveal hidden correlations or biases.

1.6.5 Union of Two Events

Definition 9 (Union of Events). For events A and B, the union $A \cup B$ represents the event that at least one of A or B occurs:

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

Proof. This is the principle of inclusion–exclusion: subtract the intersection so that outcomes are not double-counted.

1.6.6 Example: Students at AUI

Example (Sports Participation). A university has 3500 students. The distribution of sports is:

- 800 students play football (F)

- 300 students swim (S)
- 120 students do both

Step 0: Visual Representation

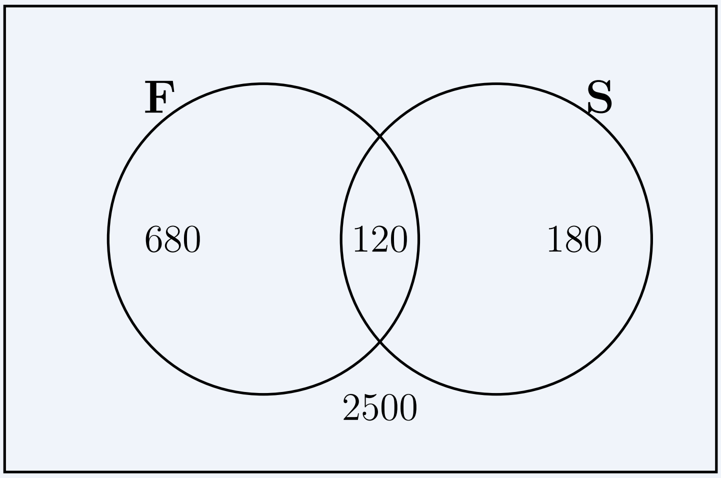


Figure 1.2: Venn diagram showing football (F) and swimming (S) students at AUI.

Step a) Probability student plays football

$$P(F) = \frac{800}{3500} = \frac{8}{35}$$

Step b) Probability student swims

$$P(S) = \frac{300}{3500} = \frac{3}{35}$$

Step c) Probability student plays both football and swimming

$$P(F \cap S) = \frac{120}{3500} = \frac{6}{175}$$

Step d) Probability student plays football or swims

$$P(F \cup S) = P(F) + P(S) - P(F \cap S) = \frac{7}{25}$$

Step e) Probability student plays football given that they swim

$$P(F \mid S) = \frac{P(F \cap S)}{P(S)} = \frac{2}{5}$$

Step f) Probability student plays football given that they do not swim

$$P(F \cap S^c) = P(F) - P(F \cap S) = \frac{34}{175}, \quad P(S^c) = 1 - P(S) = \frac{32}{35}$$

$$P(F \mid S^c) = \frac{P(F \cap S^c)}{P(S^c)} = \frac{17}{80}$$

Step g) Probability student plays neither football nor swimming

$$P(F^c \cap S^c) = 1 - P(F \cup S) = \frac{18}{25}$$

Step h) Probability student plays both football and swimming given they play at least one

$$P(F \cap S \mid F \cup S) = \frac{6}{49}$$

Remark. Conditional and joint probabilities, combined with independence checks, allow us to solve real-life problems such as sports participation by properly counting overlapping events.

1.7 Venn Diagram Probability Problems

1.7.1 Three-Set Venn Diagram

Definition 10 (Venn Diagram for Three Sets). A Venn diagram represents three sets A, B, and C and their relationships, including all possible intersections and complements. It is a visual tool to solve probability problems involving unions, intersections, and conditional probabilities.

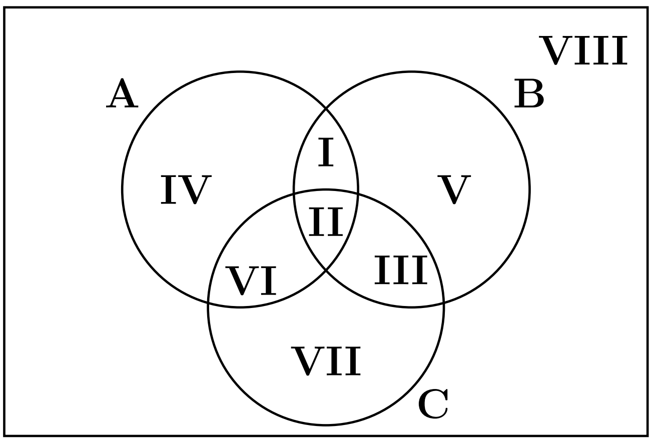


Figure 1.3: Three-set Venn diagram with labeled regions I–VIII.

1.7.2 Region Labeling

Notation. We assign Roman numerals I–VIII to each region in the Venn diagram as follows:

- I: $A \cap B \cap C^c$ (A and B, not C)
- II: $A \cap B \cap C$ (all three sets)
- III: $A^c \cap B \cap C$ (B and C, not A)
- IV: $A \cap B^c \cap C^c$ (only A)
- V: $A^c \cap B \cap C^c$ (only B)
- VI: $A \cap B^c \cap C$ (A and C, not B)
- VII: $A^c \cap B^c \cap C$ (only C)
- VIII: $A^c \cap B^c \cap C^c$ (outside all sets)

1.7.3 Probability Solutions

Example (Venn Diagram Probabilities). Let us solve the following probability problems using the labeled regions.

(a) Probability of A

$$P(A) = \mathrm{I} + \mathrm{II} + \mathrm{IV} + \mathrm{VI}$$

Proof. We sum all regions that include A.

(b) Probability of $A \cap C$

$$P(A \cap C) = \mathrm{II} + \mathrm{VI}$$

Proof. Only the regions that belong to both A and C are counted.

(c) Probability of $A \cup C$

$$P(A \cup C) = \mathrm{I} + \mathrm{II} + \mathrm{III} + \mathrm{IV} + \mathrm{VI} + \mathrm{VII}$$

(d) Conditional Probability $P(A \mid B)$

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{\mathrm{I} + \mathrm{II}}{\mathrm{I} + \mathrm{II} + \mathrm{III} + \mathrm{V}}$$

(e) Conditional Probability $P(A \mid \overline{B \cap C})$

$$P(A \mid \overline{B \cap C}) = \frac{P(A \cap \overline{B \cap C})}{P(\overline{B \cap C})} = \frac{\text{I} + \text{IV} + \text{VI}}{\text{I} + \text{IV} + \text{V} + \text{VI} + \text{VII} + \text{VIII}}$$

(f) Probability exactly one of A, B, or C occurs

$$P(\text{exactly 1 of A, B, C}) = \text{IV} + \text{V} + \text{VII}$$

(g) Conditional Probability $P(A \mid \overline{B} \cap C)$

$$P(A \mid \overline{B} \cap C) = \frac{P(A \cap \overline{B} \cap C)}{P(\overline{B} \cap C)} = \frac{\text{VI}}{\text{VI} + \text{VII}}$$

(h) Conditional Probability of only A given $A \cup B$

$$P(\text{only A} \mid A \cup B) = \frac{P(\text{only A})}{P(A \cup B)} = \frac{\text{IV}}{\text{I} + \text{II} + \text{III} + \text{IV} + \text{V} + \text{VI}}$$

Intuition (Venn Diagram Approach). Labeling regions helps systematically account for overlapping areas. By referring to regions with Roman numerals, we avoid double-counting and make conditional probabilities clear.

1.8 Important Probability Theorems

1.8.1 Total Probability Theorem

Theorem 1 (Total Probability Theorem). Let $\{B_1, B_2, \ldots, B_n\}$ form a partition of the sample space S (i.e., mutually exclusive and exhaustive). Then for any event A:

$$P(A) = \sum_{i=1}^{n} P(A \mid B_i) \cdot P(B_i)$$

Proof. Since $\{B_1, B_2, \ldots, B_n\}$ is a partition of S, we have

$$A = A \cap S = A \cap \bigcup_{i=1}^{n} B_i = \bigcup_{i=1}^{n} (A \cap B_i)$$

where the sets $A \cap B_i$ are mutually exclusive. By the **additivity of probability**:

$$P(A) = \sum_{i=1}^{n} P(A \cap B_i)$$

Using the definition of conditional probability, $P(A \cap B_i) = P(A \mid B_i) \cdot P(B_i)$.

Substituting gives:

$$P(A) = \sum_{i=1}^{n} P(A \mid B_i) \cdot P(B_i)$$

□

Proof. The theorem allows us to compute the probability of A by considering all "paths" B_i through which A can occur, then summing their contributions.

Example (Special Case: Two Events)**.** For B and its complement B^c:

$$P(A) = P(A \mid B) \cdot P(B) + P(A \mid B^c) \cdot P(B^c)$$

Example (Venn Diagram Example)**.** Using labeled Venn diagram regions:

$$P(A) = \frac{\mathrm{I} + \mathrm{II}}{\mathrm{I} + \mathrm{II} + \mathrm{III} + \mathrm{V}} \cdot P(B) + \frac{\mathrm{IV} + \mathrm{VI}}{\mathrm{IV} + \mathrm{VI} + \mathrm{VII} + \mathrm{VIII}} \cdot P(B^c)$$

Intuition (Understanding Total Probability)**.** Think of the partition as splitting the sample space into mutually exclusive "paths" through which event A can occur. Summing the contributions gives the total probability of A.

1.8.2 Bayes' Theorem

Theorem 2 (Bayes' Theorem)**.** For any events A and B with $P(A) > 0$:

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

Proof. By the definition of conditional probability:

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

Using the definition of conditional probability again, $P(A \cap B) = P(A \mid B) \cdot P(B)$. Substituting gives:

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

□

Proof. Bayes' Theorem updates our belief about B after observing A. It "flips" conditional probabilities, weighting by the prior probability of B.

Example (Using Total Probability in Bayes). If $\{B, B^c\}$ is a partition:

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

Example (General Partition). For a partition $\{B_1, \ldots, B_n\}$:

$$P(B_i \mid A) = \frac{P(A \mid B_i) \cdot P(B_i)}{\sum_{j=1}^{n} P(A \mid B_j) \cdot P(B_j)}$$

Example (Venn Diagram Interpretation). Using Venn diagram regions:

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} = \frac{\mathrm{I} + \mathrm{II}}{\mathrm{I} + \mathrm{II} + \mathrm{IV} + \mathrm{VI}}$$

Intuition (Bayes Intuition)**.** Bayes' Theorem allows us to answer: "Given we observed A, what is the probability that B occurred?" It converts prior knowledge and likelihoods into updated beliefs.

Chapter 2

Counting

2.1 Combinatorics

Probability problems often reduce to counting problems. By systematically enumerating outcomes, we can compute probabilities rigorously.

2.1.1 Counting Principles

Multiplication Principle

Proof. If a process consists of successive tasks, each of which can be performed in multiple ways, the total number of possible outcomes is the product of the number of ways to perform each task. This is sometimes called the *Fundamental Counting Principle*.
If task 1 can be done in m_1 ways, task 2 in m_2 ways, and task 3 in m_3 ways, then the total number of ways is $m_1 \cdot m_2 \cdot m_3$.

Example (Dice Example)**.** Throw a four-sided die, then a six-sided die. Total outcomes:

$$4 \times 6 = 24$$

Proof. We first choose the outcome of the first die (4 possibilities), then the second die (6 possibilities). Each choice of the first die can be paired with any of the second die outcomes, giving $4 \cdot 6 = 24$ total possibilities.

Addition Principle

Proof. If a task can be completed by choosing one option from disjoint sets, the total number of ways is the sum of the number of ways in each set. This is essentially a "choose one among mutually exclusive options" principle.

$$\text{If sets have } m_1, m_2, m_3 \text{ choices: } m_1 + m_2 + m_3$$

Example (Course Selection). A student can choose a language course from either set A (3 courses) or set B (2 courses). Total choices:

$$3 + 2 = 5$$

Generalized Addition Principle (Inclusion–Exclusion)

Proof. When sets are not disjoint, the *Inclusion–Exclusion Principle* ensures we do not double-count overlapping elements.

$$|A \cup B| = |A| + |B| - |A \cap B|$$

$$|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|$$

Proof. Each element in $A \cap B$ is counted twice when summing $|A| + |B|$. Subtracting $|A \cap B|$ corrects the overcount. This generalizes to three or more sets with alternating inclusion and exclusion. □

Example (License Plates). License plates can have format `nnllll` or `nlllll`, where n is a digit (0–9) and l is a lowercase letter (a–z).

$$\texttt{nnllll} : 10^2 \cdot 26^4, \quad \texttt{nlllll} : 10 \cdot 26^5$$

Total possibilities (disjoint formats):

$$10^2 \cdot 26^4 + 10 \cdot 26^5$$

Subtraction Principle (Complement Principle)

Proof. Sometimes it's easier to count what does *not* happen and subtract from the total. This is often called the *complement principle*.

$$|S| = |U| - |\bar{S}|$$

Example (Strings Example). Random string of length 6 from $\{a, b, c, d, e, f, g\}$. Probability it contains at least one of $\{a, b, c\}$?

$$|U| = 7^6$$

$$|\bar{S}| = 4^6 \quad \text{(strings with no } a, b, c\text{)}$$

$$|S| = |U| - |\bar{S}| = 7^6 - 4^6$$

$$P(S) = \frac{|S|}{|U|} = \frac{7^6 - 4^6}{7^6}$$

Proof. Instead of counting all strings that contain a, b, or c (complicated), we count strings without a, b, c (simpler) and subtract from total.

Division Principle

Proof. Used when objects are grouped into identical classes, or arrangements are indistinguishable in some way. If each object in A corresponds to k objects in B:

$$|A| = \frac{|B|}{k}$$

Example (Seating Around a Table). 5 people around a circular table. Linear arrangements: $5! = 120$, but rotations are identical. Divide by 5:

$$\frac{5!}{5} = 24$$

2.1.2 Common Discrete Structures

1. Strings Sequence of symbols. Number of strings of length l from an alphabet of size m:

$$m^l$$

2. Permutations

- **All n objects:**

$$n! = n \cdot (n-1) \cdot \dots \cdot 1$$

- **k objects out of n:**

$$P(n, k) = \frac{n!}{(n-k)!} = n \cdot (n-1) \cdot \dots \cdot (n-k+1)$$

- **Example:** Arrange 3 objects from $\{a, b, c, d, e\}$:

$$5 \cdot 4 \cdot 3 = 60 \text{ sequences}$$

Proof. First pick the first object (5 options), then second (4 options), then third (3 options). Multiplication principle applies.

3. Combinations / Selections

Proof. Number of ways to select k elements from n distinct objects where order does not matter:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Example (Selecting Students). Choose 3 from $\{a, b, c, d, e\}$:

$$\binom{5}{3} = 10$$

Proof. Order does not matter; $\{a, b, c\} = \{c, b, a\}$, so divide permutations 3! to remove overcounting.

4. Permutations with Repetition

Proof. When objects repeat, divide by factorials of repeated counts to avoid counting indistinguishable arrangements multiple times.

Example (Name `Hamza`). Letters: H,A,M,Z,A (two A's):

$$\frac{5!}{2!} = 60$$

Proof. n objects with $n_1, n_2, \ldots, n_k$ repetitions:

$$\frac{n!}{n_1! \cdot n_2! \cdot \cdots \cdot n_k!}$$

5. Combinations with Repetition

Proof. Selecting k elements from n types, allowing repetition:

$$\binom{n+k-1}{k} = \frac{(n+k-1)!}{k!(n-1)!}$$

Example (Fruits Example). Buy 4 fruits from 3 types (apples, oranges, bananas):

$$\binom{3+4-1}{4} = \binom{6}{4} = 15$$

Some possibilities: {4 apples, 0 oranges, 0 bananas}, {3 apples, 1 orange, 0 bananas}, {2 apples, 2 oranges, 0 bananas}, etc.

Proof. Let $C(n, k)$ denote the number of ways to select k elements from n types with repetition.

Recursive idea: Consider the first type of element (say apples). We can choose $0, 1, 2, \ldots, k$ of them.

- If we choose i of the first type, we have $k - i$ elements left to choose from the remaining $n - 1$ types. - By recursion, the number of ways for this case is $C(n - 1, k - i)$.

Hence, the recursion is:

$$C(n, k) = \sum_{i=0}^{k} C(n - 1, k - i)$$

Base cases:

$$C(1, k) = 1 \quad \text{(all elements must be the only type)}, \qquad C(n, 0) = 1 \quad \text{(choose nothing)}$$

Closed form via induction: Assume the formula holds for $n-1$ types. Then:

$$C(n, k) = \sum_{i=0}^{k} \binom{(n-1) + (k-i) - 1}{k - i} = \sum_{i=0}^{k} \binom{n + k - i - 3}{k - i}$$

Using the standard identity

$$\sum_{i=0}^{k} \binom{m + i}{i} = \binom{m + k + 1}{k},$$

we get

$$C(n, k) = \binom{n + k - 1}{k}.$$

□

Remark. Intuitively, each selection of k items from n types can be represented as k stars and $n - 1$ bars separating the types. Counting sequences of k stars and $n - 1$ bars gives exactly $\binom{n+k-1}{k}$.

Remark (Counting Matrix for Sampling). Here is a compact overview of the four classic counting cases in combinatorics:

	Without Repetition	**With Repetition**
Ordered	$P(n, r) = \dfrac{n!}{(n-r)!}$	n^r
Unordered	$\binom{n}{r} = \dfrac{n!}{r!(n-r)!}$	$\binom{n+r-1}{r}$

Exercise 1 (Permutation Probability). For a random permutation of $\{1, 2, 3, 4, 5\}$, what is the probability that the first two digits sum to 5?

The starting pairs summing to 5 are $(1, 4), (4, 1), (2, 3)$, and $(3, 2)$. There are 4 such pairs. The remaining 3 digits can be arranged in 3! ways.

$$P(\text{sum is } 5) = \frac{4 \times 3!}{5!} = \frac{24}{120} = \frac{1}{5}$$

Exercise 2 (Committee Selection). A committee of 6 is chosen from 20 Moroccan and 10 foreign professors. Find the probability of selecting: a) exactly 2 foreign professors, and b) at least 2 foreign professors.

a) For exactly 2 foreign (and 4 Moroccan) professors:

$$P(\text{exactly 2 foreign}) = \frac{\binom{10}{2}\binom{20}{4}}{\binom{30}{6}} \approx 0.3672$$

b) For at least 2 foreign professors, we use the complement rule:

$$P(\geq 2) = 1 - P(0 \text{ or } 1 \text{ foreign}) = 1 - \frac{\binom{10}{0}\binom{20}{6} + \binom{10}{1}\binom{20}{5}}{\binom{30}{6}} \approx 0.6736$$

Exercise 3 (Weighted String Probability). A 4-digit string is made from $\{1, 2, 3\}$. The probability of a string x is proportional to its first digit x_1. Find the probability that the string is: a) all the same digit, and b) a palindrome.

The probability of any string x is $P(x) = x_1/162$.

a) All digits are the same ('1111', '2222', '3333'):

$$P(\text{all same}) = \frac{1}{162} + \frac{2}{162} + \frac{3}{162} = \frac{6}{162} = \frac{1}{27}$$

b) The string is a palindrome (form $x_1x_2x_2x_1$):

$$P(\text{palindrome}) = \frac{3 \times 1 + 3 \times 2 + 3 \times 3}{162} = \frac{3(1 + 2 + 3)}{162} = \frac{18}{162} = \frac{1}{9}$$

Exercise 4. Consider picking a permutation of 12233 with equal likelihood

$$P(\text{perm has 2 symbols adding up to 4})$$

Exercise 5. Consider all solutions to

$$x_1 + x_2 + x_3 + x_4 = 10$$

Pick any one of them with equal likelihood

1. P(exactly 3 of the x_i'have the same value) $=$

2. P(all 4 of them have the same value) $=$

3. $P(x_1 = 3) =$

Chapter 3

Random Variables

3.1 Basic Concepts

Definition 11 (Random Variable). A *random variable* is a function

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

from the sample space $\mathcal{S}$ of an experiment to the set of real numbers $\mathbb{R}$.

Example. Suppose we choose 3 elements from the set $\{1, 2, \dots, 6\}$ uniformly at random. Then the sample space is

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

and

$$|\mathcal{S}| = \binom{6}{3} = 20.$$

(a) Define a random variable

$$X : \mathcal{S} \to \mathbb{R}, \quad X(\{a, b, c\}) = \max(a, b, c).$$

Then

$$\mathrm{Range}(X) = \{3, 4, 5, 6\}.$$

(b) Define another random variable

$$Y : \mathcal{S} \to \mathbb{R}, \quad Y(\{a, b, c\}) = \frac{a + b + c}{3}.$$

Then

$$\mathrm{Range}(Y) = \{2, 2.5, 3, 3.5, 4, 4.5, 5\}.$$

Example. Consider all permutations of the numbers $1, 2, 3, 4, 5$. Define the random variable

$$Z(\pi) = \#\{i \in \{1, \dots, 5\} : \pi(i) \neq i\},$$

the number of elements that are *not* in their original position (i.e., the number of displaced elements).

For example:

$$Z(12345) = 0, \qquad Z(12453) = 2.$$

Definition 12 (Probability Mass Function)**.** Let X be a discrete random variable with range

$$R_X = \{x_1, x_2, x_3, \dots\},$$

where R_X is finite or countably infinite. The *probability mass function (PMF)* of X is the function

$$p_X(x) = \Pr(X = x), \quad x \in \mathbb{R}.$$

The PMF satisfies:

$$p_X(x) \geq 0 \quad \forall x, \qquad \sum_{x \in R_X} p_X(x) = 1.$$

Example. Continuing from Example 1, let $X = \max(a, b, c)$. To find its PMF, note that:

- For $X = 3$: the only possible subset is $\{1, 2, 3\}$, so $\Pr(X = 3) = \frac{1}{20}$.
- For $X = 4$: we must have 4 included and the other two from $\{1, 2, 3\}$, giving $\binom{3}{2} = 3$ subsets, so $\Pr(X = 4) = \frac{3}{20}$.
- For $X = 5$: we must have 5 included and the other two from $\{1, 2, 3, 4\}$, giving $\binom{4}{2} = 6$ subsets, so $\Pr(X = 5) = \frac{6}{20}$.
- For $X = 6$: we must have 6 included and the other two from $\{1, 2, 3, 4, 5\}$, giving $\binom{5}{2} = 10$ subsets, so $\Pr(X = 6) = \frac{10}{20} = \frac{1}{2}$.

Hence,

$$p_X(x) = \begin{cases} \dfrac{1}{20}, & x = 3, \\ \dfrac{3}{20}, & x = 4, \\ \dfrac{6}{20}, & x = 5, \\ \dfrac{10}{20}, & x = 6, \\ 0, & \text{otherwise.} \end{cases}$$

We can verify normalization:

$$\frac{1+3+6+10}{20} = \frac{20}{20} = 1.$$

Example. You have 5 cameras, of which 2 work and 3 do not. You test them one by one until you find a working one. Let W be the number of cameras tested.

The range is $\text{Range}(W) = \{1, 2, 3\}$ (at most 3 tests needed). The PMF is:

$$p_W(w) = \begin{cases} \dfrac{2}{5}, & w = 1, \\ \dfrac{3}{5} \cdot \dfrac{2}{4} = \dfrac{3}{10}, & w = 2, \\ \dfrac{3}{5} \cdot \dfrac{2}{4} \cdot \dfrac{1}{3} = \dfrac{1}{10}, & w = 3, \\ 0, & \text{otherwise.} \end{cases}$$

Verification: $\frac{2}{5} + \frac{3}{10} + \frac{1}{10} = 1$.

Example. Throw a 4-sided die until you get a "3". Let the random variable X be the number of throws required to get the first "3".

$$\text{Range}(X) = \mathbb{N} = \{1, 2, 3, \dots\}.$$

The probability of success (getting a "3") on any single throw is $p = \dfrac{1}{4}$, and the probability of failure is $q = 1 - p = \dfrac{3}{4}$.

(a) For $X = 8$:

$$p_X(8) = P(X = 8) = q^{8-1}p = \left(\frac{3}{4}\right)^7 \left(\frac{1}{4}\right).$$

(b) In general, for the geometric distribution:

$$p_X(k) = P(X = k) = q^{k-1}p = \left(\frac{3}{4}\right)^{k-1} \left(\frac{1}{4}\right), \quad k = 1, 2, 3, \dots$$

(c) To verify that p_X is a valid PMF:

$$\sum_{k=1}^{\infty} p_X(k) = \sum_{k=1}^{\infty} \left(\frac{3}{4}\right)^{k-1} \left(\frac{1}{4}\right) = \frac{1}{4} \sum_{k=0}^{\infty} \left(\frac{3}{4}\right)^k = \frac{1}{4} \cdot \frac{1}{1 - \frac{3}{4}} = 1.$$

3.2 Cumulative Distribution Function

Definition 13 (Cumulative Distribution Function). The *cumulative distribution function (CDF)* of a random variable X is defined as

$$F_X(x) = P(X \leq x), \quad \forall x \in \mathbb{R}.$$

3.2.1 Discrete Random Variables

Example. Let Y be a discrete random variable with probability mass function $p_Y(y)$ given by

$$p_Y(y) = \begin{cases} \frac{1}{15}, & y = -1, \\ \frac{3}{15}, & y = 0, \\ \frac{4}{15}, & y = 2, \\ \frac{2}{15}, & y = 6, \\ \frac{5}{15}, & y = 10, \\ 0, & \text{otherwise.} \end{cases}$$

(a) Verify that this is a valid PMF:

$$\frac{1}{15} + \frac{3}{15} + \frac{4}{15} + \frac{2}{15} + \frac{5}{15} = \frac{15}{15} = 1.$$

(b) Compute the CDF $F_Y(y) = P(Y \leq y)$:

$$F_Y(y) = \begin{cases} 0, & y < -1, \\ \frac{1}{15}, & -1 \leq y < 0, \\ \frac{4}{15}, & 0 \leq y < 2, \\ \frac{8}{15}, & 2 \leq y < 6, \\ \frac{10}{15}, & 6 \leq y < 10, \\ 1, & y \geq 10. \end{cases}$$

Remark. The probability mass function (PMF) and the cumulative distribution function (CDF) describe a random variable in related but distinct ways.

Aspect	PMF	CDF
Symbol	$p_X(x)$	$F_X(x)$
Definition	$p_X(x) = P(X = x)$	$F_X(x) = P(X \leq x)$
Domain	Discrete x	Real x
Range	$[0, 1]$	$[0, 1]$
Relation	$F_X(x) = \sum_{t \leq x} p_X(t)$	$p_X(x) = F_X(x) - F_X(x^-)$
Nature	Stepwise (non-continuous)	Non-decreasing, right-continuous

3.2.2 Continuous Random Variables

Definition 14 (Probability Density Function). For a continuous random variable X, the *probability density function (PDF)* $f_X(x)$ satisfies:

$$P(a \leq X \leq b) = \int_a^b f_X(x)\,dx, \quad \text{and} \quad \int_{-\infty}^{\infty} f_X(x)\,dx = 1.$$

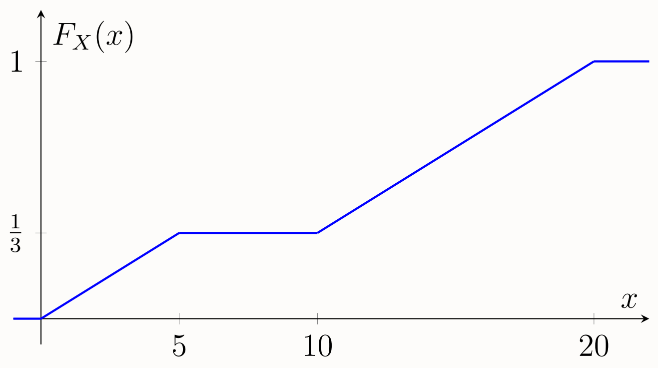
The CDF and PDF are related by:

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt, \quad \text{and} \quad f_X(x) = \frac{d}{dx} F_X(x) \text{ (where differentiable).}$$

Exercise 6. X is uniform over $[0, 5] \cup [10, 20]$. Find $F_X(x)$ and draw it.

Proof. Total length of support is 15. The CDF is:

$$F_X(x) = \begin{cases} 0 & x < 0 \\ \frac{x}{15} & 0 \leq x < 5 \\ \frac{1}{3} & 5 \leq x < 10 \\ \frac{x-5}{15} & 10 \leq x < 20 \\ 1 & x \geq 20 \end{cases}$$

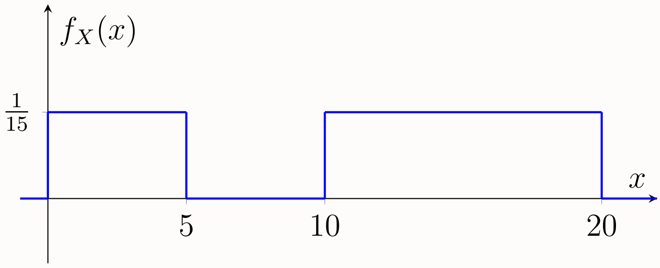


□

Exercise 7. X is uniform over $[0,5] \cup [10,20]$. Find $f_X(x)$ and draw it.

Proof. Total length is 15. For uniform distribution, $f_X(x) = \frac{1}{15}$ on the support:

$$f_X(x) = \begin{cases} \frac{1}{15} & 0 \le x \le 5 \text{ or } 10 \le x \le 20 \\ 0 & \text{otherwise} \end{cases}$$



□

Exercise 8. Y has range $[0,10] \cup [20,25]$. It is uniform over $[0,10]$ and uniform over $[20,25]$ with $P(20 \le Y \le 25) = 3 \cdot P(0 \le Y \le 10)$. Find $F_Y(y)$.

Proof. Let $p_1 = P(0 \le Y \le 10)$ and $p_2 = P(20 \le Y \le 25)$. Given $p_1 + p_2 = 1$ and $p_2 = 3p_1$, we get $p_1 = \frac{1}{4}$ and $p_2 = \frac{3}{4}$.

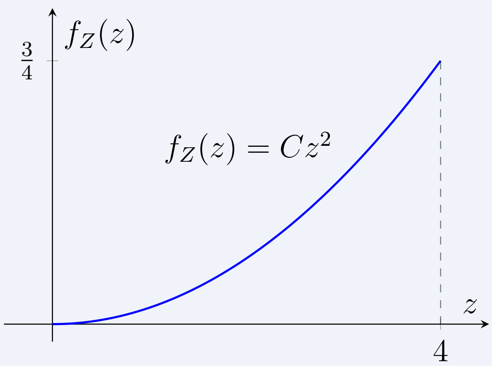
For uniform densities: $c_1 = \frac{p_1}{10} = \frac{1}{40}$ on $[0,10]$, and $c_2 = \frac{p_2}{5} = \frac{3}{20}$ on $[20,25]$.

The CDF is:

$$F_Y(y) = \begin{cases} 0 & y < 0 \\ \frac{y}{40} & 0 \le y < 10 \\ \frac{1}{4} & 10 \le y < 20 \\ \frac{1}{4} + \frac{3(y-20)}{20} & 20 \le y < 25 \\ 1 & y \ge 25 \end{cases}$$

□

Exercise 9. Random variable Z has range $[0,4]$ with $f_Z(z) = Cz^2$. Find C.

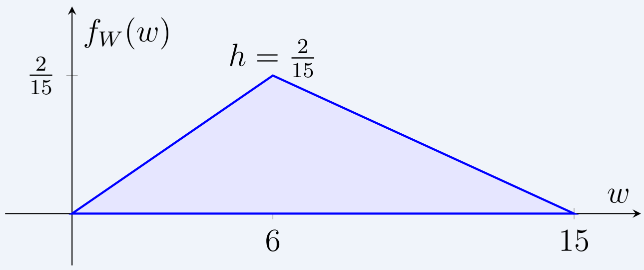


Proof. For $f_Z(z)$ to be a valid pdf, $\int_0^4 Cz^2\,dz = 1$. We have:

$$C\left[\frac{z^3}{3}\right]_0^4 = C \cdot \frac{64}{3} = 1 \Rightarrow C = \frac{3}{64}.$$

□

Exercise 10. Random variable W has the following pdf. Find $f_W(w)$.

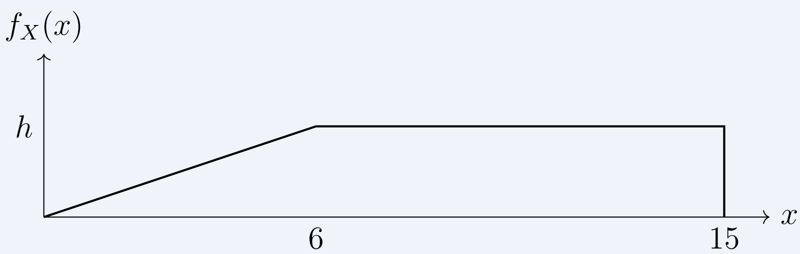


Proof. The area of the triangle is $\frac{1}{2} \cdot 15 \cdot h = 1$, so $h = \frac{2}{15}$.
For $0 \le w < 6$: line through $(0,0)$ and $(6, 2/15)$ gives $f_W(w) = \frac{w}{45}$.
For $6 \le w \le 15$: line through $(6, 2/15)$ and $(15,0)$ gives $f_W(w) = \frac{2}{9} - \frac{2w}{135}$.
Therefore:

$$f_W(w) = \begin{cases} \frac{w}{45} & 0 \le w < 6 \\ \frac{2}{9} - \frac{2w}{135} & 6 \le w \le 15 \\ 0 & \text{otherwise} \end{cases}$$

□

Exercise 11. Consider the density sketched below: it is linear from $x = 0$ to $x = 6$ reaching height h at $x = 6$, constant equal to h on $[6, 15]$, and zero elsewhere.



(a) Write down $f_X(x)$.

(b) Calculate the following probabilities:

(i) $P(X \le 6)$

(ii) $P(X \ge 6)$

(iii) $P(4 \le X \le 12)$

(iv) $P(|X-6| \leq 2)$

Solution. **(a)** The total area must equal 1. The area is the triangle on $[0,6]$ plus the rectangle on $[6,15]$:

$$\frac{1}{2} \cdot 6 \cdot h + 9 \cdot h = 12h = 1 \quad \Longrightarrow \quad h = \frac{1}{12}.$$

On $[0,6]$, the density is linear: $f_X(x) = \frac{h}{6}x = \frac{x}{72}$. On $[6,15]$, $f_X(x) = h = \frac{1}{12}$. Thus:

$$f_X(x) = \begin{cases} \frac{x}{72}, & 0 \leq x \leq 6, \\ \frac{1}{12}, & 6 < x \leq 15, \\ 0, & \text{otherwise.} \end{cases}$$

(b) (i) $P(X \leq 6) = \frac{1}{2} \cdot 6 \cdot h = \frac{1}{4}$.
(ii) $P(X \geq 6) = 9h = \frac{3}{4}$.
(iii) $P(4 \leq X \leq 12) = \int_4^6 \frac{x}{72}\,dx + \int_6^{12} \frac{1}{12}\,dx = \frac{5}{36} + \frac{1}{2} = \frac{23}{36}$.
(iv) $P(|X-6| \leq 2) = P(4 \leq X \leq 8) = \frac{5}{36} + \frac{1}{6} = \frac{11}{36}$. □

3.3 Expectation of Random Variables

Definition 15 (Expected Value - Discrete Case)**.** For a discrete random variable X with range $R_X = \{x_1, x_2, x_3, \ldots\}$ and probability mass function $p_X(x)$, the *expected value* (or *expectation* or *mean*) of X is defined as

$$E(X) = \sum_{x \in R_X} x \cdot p_X(x),$$

provided this sum converges absolutely.

Example. Consider a fair 6-sided die. Let X be the outcome of a single roll.
The probability mass function is

$$p_X(x) = \frac{1}{6}, \quad x \in \{1,2,3,4,5,6\}.$$

The expected value is

$$E(X) = \sum_{x=1}^{6} x \cdot p_X(x) = \sum_{x=1}^{6} x \cdot \frac{1}{6} = \frac{1}{6}(1+2+3+4+5+6) = \frac{21}{6} = \frac{7}{2} = 3.5.$$

Example. Consider a 4-sided die where $P(X=k) \propto k+2$ for $k \in \{1,2,3,4\}$.

The normalization constant is $\sum_{k=1}^{4}(k+2) = 18$, so $p_X(k) = \frac{k+2}{18}$.
The expected value is:

$$E(X) = \sum_{k=1}^{4} k \cdot \frac{k+2}{18} = \frac{1}{18}\sum_{k=1}^{4} k(k+2) = \frac{1}{18}(3+8+15+24) = \frac{25}{9}.$$

Definition 16 (Expected Value - Continuous Case). For a continuous random variable X with probability density function $f_X(x)$, the *expected value* is defined as

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

provided this integral converges absolutely.

Example. Let X have pdf

$$f_X(x) = \begin{cases} \frac{x}{72}, & 0 \le x \le 6, \\ \frac{1}{12}, & 6 < x \le 15, \\ 0, & \text{otherwise.} \end{cases}$$

Then:

$$E(X) = \int_0^6 \frac{x^2}{72}\, dx + \int_6^{15} \frac{x}{12}\, dx = \frac{1}{72} \cdot \frac{216}{3} + \frac{1}{24}(225 - 36) = 1 + \frac{63}{8} = \frac{71}{8}.$$

3.4 Variance and Standard Deviation

Definition 17 (Variance). The *variance* of a random variable X is defined as

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

Definition 18 (Standard Deviation). The *standard deviation* of X is

$$\sigma_X = \sqrt{\mathrm{Var}(X)}.$$

Remark. For discrete random variables:

$$\mathrm{Var}(X) = \sum_{x \in R_X} (x - E(X))^2 \cdot p_X(x) = \sum_{x \in R_X} x^2 \cdot p_X(x) - [E(X)]^2.$$

For continuous random variables:

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

Theorem 3 (Properties of Expectation and Variance). Let X and Y be random variables, and let $a, b \in \mathbb{R}$ be constants.

(i) **Linearity of Expectation:**

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

(ii) **Variance of Linear Transformation:**

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

(iii) **Sum of Expectations:**

$$E(X + Y) = E(X) + E(Y).$$

(iv) **Variance is Non-negative:**

$$\text{Var}(X) \geq 0.$$

(v) **Constant Variance:**

$$\text{Var}(b) = 0 \text{ for any constant } b.$$

Example. Let X be the outcome of a fair 6-sided die. We have $E(X) = 3.5$. Then:

$$E(X^2) = \sum_{x=1}^{6} x^2 \cdot \frac{1}{6} = \frac{91}{6}, \quad \text{Var}(X) = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}, \quad \sigma_X = \sqrt{\frac{35}{12}}.$$

3.5 Functions of Random Variables

Definition 19 (Function of a Random Variable). Let X be a random variable and $g : \mathbb{R} \to \mathbb{R}$ be a function. Then $Y = g(X)$ is also a random variable.

Theorem 4 (Expectation of a Function - Discrete Case). If X is a discrete random variable with pmf $p_X(x)$ and $g : \mathbb{R} \to \mathbb{R}$, then

$$E[g(X)] = \sum_{x \in R_X} g(x) \cdot p_X(x).$$

Theorem 5 (Expectation of a Function - Continuous Case). If X is a continuous random variable with pdf $f_X(x)$ and $g : \mathbb{R} \to \mathbb{R}$, then

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

3.5.1 Case 1: Discrete X, Discrete $Y = g(X)$

Example. Let X be a discrete random variable with pmf

$$p_X(x) = \begin{cases} 0.2, & x = 1, \\ 0.3, & x = 2, \\ 0.5, & x = 3. \end{cases}$$

Define $Y = X^2$. Then Y takes values $1, 4, 9$ with probabilities:

$$p_Y(1) = P(X = 1) = 0.2, \quad p_Y(4) = P(X = 2) = 0.3, \quad p_Y(9) = P(X = 3) = 0.5.$$

The expected value of Y is

$$E(Y) = E(X^2) = \sum_x x^2 \cdot p_X(x) = 1(0.2) + 4(0.3) + 9(0.5) = 0.2 + 1.2 + 4.5 = 5.9.$$

3.5.2 Case 2: Discrete X, Continuous $Y = g(X)$

Remark. A deterministic function of a discrete random variable remains discrete. This case typically does not occur unless g involves randomization.

3.5.3 Case 3: Continuous X, Discrete $Y = g(X)$

When a continuous random variable X is transformed via a piecewise constant function, the resulting $Y = g(X)$ is discrete. We find its PMF by computing the probability that X falls into each constant region.

Example. Let $X \sim \text{Uniform}(0, 1)$ with pdf $f_X(x) = 1$ for $0 \le x \le 1$.

Define

$$Y = \begin{cases} 0, & 0 \le X < 0.5, \\ 1, & 0.5 \le X \le 1. \end{cases}$$

Since Y takes only values 0 and 1, we compute:

$$P(Y=0)=P(0\le X<0.5)=\int_0^{0.5} f_X(x)\,dx=\int_0^{0.5}1\,dx=0.5,$$

$$P(Y=1)=P(0.5\le X\le 1)=\int_{0.5}^{1} f_X(x)\,dx=\int_{0.5}^{1}1\,dx=0.5.$$

Therefore, Y has pmf:

$$p_Y(y)=\begin{cases}0.5, & y=0,\\ 0.5, & y=1,\\ 0, & \text{otherwise.}\end{cases}$$

3.5.4 Case 4: Continuous X, Continuous $Y=g(X)$

For continuous X and $Y=g(X)$, we use the *CDF method*: find $F_Y(y)=P(Y\le y)$ by expressing the event $\{Y\le y\}$ in terms of X, then differentiate to obtain the PDF. The key is to branch based on the monotonicity of g.

Theorem 6 (CDF Method for Transformations)**.** Let X be a continuous random variable with pdf $f_X(x)$ and CDF $F_X(x)$. For $Y=g(X)$:

1. If g is strictly increasing on the support of X, then for y in the range of Y:

$$F_Y(y)=P(Y\le y)=P(g(X)\le y)=P(X\le g^{-1}(y))=F_X(g^{-1}(y)),$$

and thus $f_Y(y)=f_X(g^{-1}(y))\cdot\frac{d}{dy}g^{-1}(y)$.

2. If g is strictly decreasing on the support of X, then:

$$F_Y(y)=P(Y\le y)=P(g(X)\le y)=P(X\ge g^{-1}(y))=1-F_X(g^{-1}(y)),$$

and thus $f_Y(y)=-f_X(g^{-1}(y))\cdot\frac{d}{dy}g^{-1}(y)=f_X(g^{-1}(y))\cdot\left|\frac{d}{dy}g^{-1}(y)\right|$.

3. If g is not monotonic, partition the domain and sum contributions from each monotonic piece.

Example (Linear Transformation)**.** Let $X\sim\text{Uniform}(0,1)$ with pdf $f_X(x)=1$ for $0\le x\le 1$ and CDF $F_X(x)=x$ for $0\le x\le 1$.

Define $Y=2X+3$. Since $g(x)=2x+3$ is strictly increasing, we use the increasing case.

For y in the range $[3,5]$, we have $g^{-1}(y)=\frac{y-3}{2}$. Using the CDF method:

$$F_Y(y)=P(Y\le y)=P(2X+3\le y)=P\left(X\le\frac{y-3}{2}\right)=F_X\left(\frac{y-3}{2}\right)=\frac{y-3}{2}.$$

Differentiating:

$$f_Y(y) = \frac{d}{dy}F_Y(y) = \frac{d}{dy}\left(\frac{y-3}{2}\right) = \frac{1}{2}, \quad 3 \le y \le 5.$$

We verify: $\int_3^5 \frac{1}{2}\,dy = 1$.

Example (Squared Transformation). Let X have pdf $f_X(x) = 2x$ for $0 \le x \le 1$, so $F_X(x) = x^2$.

Define $Y = X^2$. Since $g(x) = x^2$ is strictly increasing on $[0,1]$, for $y \in [0,1]$:

$$F_Y(y) = P(X^2 \le y) = P(X \le \sqrt{y}) = F_X(\sqrt{y}) = y.$$

Differentiating: $f_Y(y) = 1$ for $0 \le y \le 1$, so $Y \sim \text{Uniform}(0,1)$.

Example (Decreasing Transformation). Let X have pdf $f_X(x) = e^{-x}$ for $x > 0$, so $F_X(x) = 1 - e^{-x}$ for $x > 0$.

Define $Y = -X$. Since $g(x) = -x$ is strictly decreasing, we use the decreasing case.

For $y < 0$, we have $g^{-1}(y) = -y > 0$. Using the CDF method:

$$F_Y(y) = P(Y \le y) = P(-X \le y) = P(X \ge -y) = 1 - F_X(-y) = 1 - (1 - e^{-(-y)}) = e^y.$$

Differentiating:

$$f_Y(y) = \frac{d}{dy}F_Y(y) = \frac{d}{dy}(e^y) = e^y, \quad y < 0.$$

3.6 Common Discrete Probability Distributions

3.6.1 Bernoulli Distribution

Definition 20 (Bernoulli Distribution). A random variable X follows a *Bernoulli distribution* with parameter p (where $0 \le p \le 1$) if it represents a single trial with two possible outcomes: success (1) or failure (0).

We write $X \sim \text{Bernoulli}(p)$.

- **Range:** $R_X = \{0, 1\}$
- **PMF:**

$$p_X(x) = \begin{cases} p, & x = 1, \\ 1-p, & x = 0, \\ 0, & \text{otherwise.} \end{cases}$$

Alternatively: $p_X(x) = p^x(1-p)^{1-x}$ for $x \in \{0, 1\}$.

- **Expected Value:** $E(X) = p$
- **Variance:** $\text{Var}(X) = p(1-p)$

Example. A fair coin flip can be modeled as $X \sim \text{Bernoulli}(0.5)$, where $X = 1$ represents heads and $X = 0$ represents tails.

Then:
$$E(X) = 0.5, \quad \text{Var}(X) = 0.5(1-0.5) = 0.25.$$

3.6.2 Binomial Distribution

Definition 21 (Binomial Distribution). A random variable X follows a *binomial distribution* with parameters n (number of trials) and p (probability of success) if it represents the number of successes in n independent Bernoulli trials.

We write $X \sim \text{Binomial}(n, p)$ or $X \sim B(n, p)$.

- **Range:** $R_X = \{0, 1, 2, \ldots, n\}$
- **PMF:**
$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, 2, \ldots, n.$$
- **Expected Value:** $E(X) = np$
- **Variance:** $\text{Var}(X) = np(1-p)$

Example. A fair coin is flipped 10 times. Let X be the number of heads. Then $X \sim \text{Binomial}(10, 0.5)$.

The probability of getting exactly 6 heads is:
$$P(X = 6) = \binom{10}{6} (0.5)^6 (0.5)^4 = \frac{210}{1024} = \frac{105}{512} \approx 0.205.$$

The expected number of heads is:
$$E(X) = 10 \cdot 0.5 = 5.$$

The variance is:
$$\text{Var}(X) = 10 \cdot 0.5 \cdot 0.5 = 2.5.$$

Example. A multiple-choice exam has 20 questions, each with 4 choices. If a student guesses randomly on all questions, what is the probability they get at least 8 correct?

Let X be the number of correct answers. Then $X \sim \text{Binomial}(20, 0.25)$.

We want:

$$P(X \geq 8) = \sum_{k=8}^{20} \binom{20}{k} (0.25)^k (0.75)^{20-k}.$$

The expected number of correct answers is:

$$E(X) = 20 \cdot 0.25 = 5.$$

3.6.3 Geometric Distribution

Definition 22 (Geometric Distribution). A random variable X follows a *geometric distribution* with parameter p if it represents the number of trials needed to get the first success in a sequence of independent Bernoulli trials.

We write $X \sim \text{Geometric}(p)$.

- **Range:** $R_X = \{1, 2, 3, \ldots\}$
- **PMF:**
$$p_X(k) = (1-p)^{k-1} p, \quad k = 1, 2, 3, \ldots$$
- **Expected Value:** $E(X) = \dfrac{1}{p}$
- **Variance:** $\text{Var}(X) = \dfrac{1-p}{p^2}$
- **Memoryless Property:** $P(X > n + m \mid X > n) = P(X > m)$

Example. Roll a fair 6-sided die until you get a 4. Let X be the number of rolls needed. Then $X \sim \text{Geometric}(1/6)$.

The probability that exactly 3 rolls are needed:

$$P(X = 3) = \left(\frac{5}{6}\right)^2 \cdot \frac{1}{6} = \frac{25}{216} \approx 0.116.$$

The expected number of rolls is:

$$E(X) = \frac{1}{1/6} = 6.$$

Remark. Some textbooks define the geometric distribution as the number of *failures* before the first success, in which case the range is $\{0, 1, 2, \ldots\}$ and the PMF is $p_X(k) = (1-p)^k p$. Always check the definition being used.

3.6.4 Negative Binomial (Pascal) Distribution

Definition 23 (Negative Binomial Distribution). A random variable X follows a *negative binomial distribution* (also called *Pascal distribution*) with parameters r and p if it represents the number of trials needed to achieve exactly r successes in a sequence of independent Bernoulli trials.

We write $X \sim \text{NegBin}(r, p)$.

- **Range:** $R_X = \{r, r+1, r+2, \ldots\}$
- **PMF:**
$$p_X(k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}, \quad k = r, r+1, r+2, \ldots$$
- **Expected Value:** $E(X) = \dfrac{r}{p}$
- **Variance:** $\text{Var}(X) = \dfrac{r(1-p)}{p^2}$

Remark. The negative binomial distribution generalizes the geometric distribution. When $r = 1$, we get $X \sim \text{Geometric}(p)$.

Example. A basketball player has a free throw success rate of 70%. How many shots must she take to make exactly 5 baskets?

Let X be the number of shots needed. Then $X \sim \text{NegBin}(5, 0.7)$.

The probability she needs exactly 8 shots:

$$P(X = 8) = \binom{7}{4} (0.7)^5 (0.3)^3 = 35 \cdot (0.7)^5 \cdot (0.3)^3 \approx 0.124.$$

The expected number of shots is:

$$E(X) = \frac{5}{0.7} \approx 7.14.$$

Remark. An alternative parameterization defines the negative binomial as the number of *failures* before the r-th success, in which case the range is $\{0, 1, 2, \ldots\}$ and the PMF is adjusted accordingly.

3.6.5 Hypergeometric Distribution

Definition 24 (Hypergeometric Distribution). A random variable X follows a *hypergeometric distribution* if it represents the number of successes in n draws without replacement from a finite population of size N containing exactly K success states.

We write $X \sim \text{Hypergeometric}(N, K, n)$.

- **Parameters:**
  - N: total population size
  - K: number of success states in the population
  - n: number of draws
- **Range:** $R_X = \{\max(0, n - N + K), \ldots, \min(n, K)\}$
- **PMF:**

$$p_X(k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}, \quad k \in R_X.$$

- **Expected Value:** $E(X) = n \cdot \dfrac{K}{N}$
- **Variance:** $\mathrm{Var}(X) = n \cdot \dfrac{K}{N} \cdot \dfrac{N-K}{N} \cdot \dfrac{N-n}{N-1}$

Example. A deck of 52 cards contains 13 hearts. If you draw 5 cards without replacement, what is the probability of getting exactly 2 hearts?

Let X be the number of hearts drawn. Then $X \sim \text{Hypergeometric}(52, 13, 5)$.

$$P(X = 2) = \frac{\binom{13}{2}\binom{39}{3}}{\binom{52}{5}} = \frac{78 \cdot 9139}{2598960} \approx 0.274.$$

The expected number of hearts is:

$$E(X) = 5 \cdot \frac{13}{52} = 1.25.$$

Example. A box contains 10 balls: 6 red and 4 blue. If 3 balls are drawn without replacement, find the probability distribution of X, the number of red balls drawn.

Here $N = 10$, $K = 6$, $n = 3$, so $X \sim \text{Hypergeometric}(10, 6, 3)$.

The possible values are $k \in \{0, 1, 2, 3\}$:

$$P(X = 0) = \frac{\binom{6}{0}\binom{4}{3}}{\binom{10}{3}} = \frac{1 \cdot 4}{120} = \frac{1}{30},$$

$$P(X = 1) = \frac{\binom{6}{1}\binom{4}{2}}{\binom{10}{3}} = \frac{6 \cdot 6}{120} = \frac{3}{10},$$

$$P(X = 2) = \frac{\binom{6}{2}\binom{4}{1}}{\binom{10}{3}} = \frac{15 \cdot 4}{120} = \frac{1}{2},$$

$$P(X=3) = \frac{\binom{6}{3}\binom{4}{0}}{\binom{10}{3}} = \frac{20 \cdot 1}{120} = \frac{1}{6}.$$

Verification: $\frac{1}{30} + \frac{3}{10} + \frac{1}{2} + \frac{1}{6} = \frac{2+18+30+10}{60} = 1.$ ✓

Remark. The hypergeometric distribution is used for sampling *without replacement*. When the population size N is very large compared to the sample size n, the hypergeometric distribution can be approximated by the binomial distribution with $p = K/N$.

3.6.6 Poisson Distribution

Definition 25 (Poisson Distribution). A random variable X follows a *Poisson distribution* with parameter $\lambda > 0$ if it represents the number of events occurring in a fixed interval of time or space, where events occur independently at a constant average rate λ.

We write $X \sim \text{Poisson}(\lambda)$.

- **Range:** $R_X = \{0, 1, 2, 3, \ldots\}$
- **PMF:**
$$p_X(k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, 2, \ldots$$
- **Expected Value:** $E(X) = \lambda$
- **Variance:** $\text{Var}(X) = \lambda$
- **Note:** For the Poisson distribution, the mean equals the variance.

Example. A call center receives an average of 3 calls per minute. Let X be the number of calls in a given minute. Then $X \sim \text{Poisson}(3)$.

The probability of receiving exactly 5 calls in a minute:

$$P(X=5) = \frac{3^5 e^{-3}}{5!} = \frac{243e^{-3}}{120} \approx 0.101.$$

The probability of receiving no calls:

$$P(X=0) = \frac{3^0 e^{-3}}{0!} = e^{-3} \approx 0.050.$$

Example. The number of typos on a page follows a Poisson distribution with mean 2. What is the probability that a randomly selected page has at most 1 typo?

Let $X \sim \text{Poisson}(2)$. We want $P(X \leq 1) = P(X=0) + P(X=1)$:

$$P(X=0) = \frac{2^0 e^{-2}}{0!} = e^{-2},$$

$$P(X=1) = \frac{2^1 e^{-2}}{1!} = 2e^{-2}.$$

Therefore:

$$P(X \leq 1) = e^{-2} + 2e^{-2} = 3e^{-2} \approx 0.406.$$

Theorem 7 (Poisson Approximation to Binomial). If $X \sim \text{Binomial}(n,p)$ with n large, p small, and $\lambda = np$ moderate, then X is approximately $\text{Poisson}(\lambda)$.

A common rule of thumb: use the Poisson approximation when $n \geq 20$ and $p \leq 0.05$.

Example. A factory produces 1000 items per day, and the probability that any item is defective is 0.002. What is the probability that there are exactly 3 defective items in a day?

Let X be the number of defective items. Exactly, $X \sim \text{Binomial}(1000, 0.002)$.

Using the Poisson approximation with $\lambda = 1000 \cdot 0.002 = 2$:

$$P(X=3) \approx \frac{2^3 e^{-2}}{3!} = \frac{8e^{-2}}{6} = \frac{4e^{-2}}{3} \approx 0.180.$$

3.6.7 Summary Table

Distribution	PMF	Mean	Variance
Bernoulli(p)	$p^x(1-p)^{1-x}$, $x \in \{0,1\}$	p	$p(1-p)$
Binomial(n,p)	$\binom{n}{k}p^k(1-p)^{n-k}$	np	$np(1-p)$
Geometric(p)	$(1-p)^{k-1}p$, $k \geq 1$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
NegBin(r,p)	$\binom{k-1}{r-1}p^r(1-p)^{k-r}$	$\frac{r}{p}$	$\frac{r(1-p)}{p^2}$
Hypergeom(N,K,n)	$\frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$	$n\frac{K}{N}$	$n\frac{K}{N}\frac{N-K}{N}\frac{N-n}{N-1}$
Poisson(λ)	$\frac{\lambda^k e^{-\lambda}}{k!}$, $k \geq 0$	λ	λ

3.6.8 Deriving Expectations and Variances

In this section, we'll derive the expected values and variances for each discrete distribution. These derivations help build intuition and understanding.

Bernoulli Distribution

Exercise 12 (DIY: Bernoulli Expectation and Variance). Let $X \sim \text{Bernoulli}(p)$.

(a) Show that $E(X) = p$ using the definition $E(X) = \sum_x x \cdot p_X(x)$.

(b) Show that $E(X^2) = p$ (hint: $X^2 = X$ for a Bernoulli random variable).

(c) Use the formula $\text{Var}(X) = E(X^2) - [E(X)]^2$ to show that $\text{Var}(X) = p(1-p)$.

Proof. (a) By definition:

$$E(X) = \sum_{x \in \{0,1\}} x \cdot p_X(x) = 0 \cdot (1-p) + 1 \cdot p = p$$

(b) Since $X \in \{0, 1\}$, we have $X^2 = X$ (because $0^2 = 0$ and $1^2 = 1$). Therefore:

$$E(X^2) = E(X) = p$$

(c) Using the variance formula:

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

□

Binomial Distribution

Exercise 13 (DIY: Binomial Expectation and Variance). Let $X \sim \text{Binomial}(n, p)$.

(a) Use the fact that $X = \sum_{i=1}^n X_i$ where each $X_i \sim \text{Bernoulli}(p)$ to show that $E(X) = np$.

(b) Use the independence of the X_i's and the fact that $\text{Var}(X_1 + \cdots + X_n) = \text{Var}(X_1) + \cdots + \text{Var}(X_n)$ for independent random variables to show that $\text{Var}(X) = np(1-p)$.

(c) **Challenge:** Derive $E(X)$ directly using the definition and the identity $\sum_{k=0}^n k\binom{n}{k}p^k(1-p)^{n-k} = np$.

Proof. (a) Let $X_i \sim \text{Bernoulli}(p)$ for $i = 1, \ldots, n$, where $X_i = 1$ if trial i is a success. Then $X = \sum_{i=1}^n X_i$.

By linearity of expectation:

$$E(X) = E\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n E(X_i) = \sum_{i=1}^n p = np$$

(b) Since the X_i's are independent:

$$\text{Var}(X) = \text{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \text{Var}(X_i) = \sum_{i=1}^{n} p(1-p) = np(1-p)$$

(c) **Direct derivation:** Using $k\binom{n}{k} = n\binom{n-1}{k-1}$ and letting $j = k-1$:

$$E(X) = \sum_{k=1}^{n} n\binom{n-1}{k-1} p^k (1-p)^{n-k} = np \sum_{j=0}^{n-1} \binom{n-1}{j} p^j (1-p)^{n-1-j} = np.$$

□

Geometric Distribution

Exercise 14 (DIY: Geometric Expectation and Variance). Let $X \sim \text{Geometric}(p)$.

(a) Show that $E(X) = \frac{1}{p}$ using the formula for the sum of a geometric series: $\sum_{k=1}^{\infty} kq^{k-1} = \frac{1}{(1-q)^2}$ for $|q| < 1$.

(b) Show that $E(X^2) = \frac{2-p}{p^2}$ using $\sum_{k=1}^{\infty} k^2 q^{k-1} = \frac{1+q}{(1-q)^3}$.

(c) Use the variance formula to show that $\text{Var}(X) = \frac{1-p}{p^2}$.

Proof. (a) By definition:

$$E(X) = \sum_{k=1}^{\infty} k \cdot (1-p)^{k-1} p = p \sum_{k=1}^{\infty} k(1-p)^{k-1}$$

Using the formula with $q = 1-p$:

$$E(X) = p \cdot \frac{1}{(1-(1-p))^2} = p \cdot \frac{1}{p^2} = \frac{1}{p}$$

(b)

$$E(X^2) = \sum_{k=1}^{\infty} k^2 \cdot (1-p)^{k-1} p = p \sum_{k=1}^{\infty} k^2 (1-p)^{k-1}$$

Using the formula with $q = 1-p$:

$$E(X^2) = p \cdot \frac{1+(1-p)}{(1-(1-p))^3} = p \cdot \frac{2-p}{p^3} = \frac{2-p}{p^2}$$

(c)

$$\mathrm{Var}(X) = E(X^2) - [E(X)]^2 = \frac{2-p}{p^2} - \left(\frac{1}{p}\right)^2 = \frac{2-p-1}{p^2} = \frac{1-p}{p^2}$$

□

Negative Binomial Distribution

Exercise 15 (DIY: Negative Binomial Expectation and Variance). Let $X \sim \mathrm{NegBin}(r, p)$.

(a) Use the fact that $X = X_1 + X_2 + \cdots + X_r$ where each $X_i \sim \mathrm{Geometric}(p)$ are independent to show that $E(X) = \frac{r}{p}$.

(b) Use the independence of the X_i's to show that $\mathrm{Var}(X) = \frac{r(1-p)}{p^2}$.

Proof. (a) Let X_i be the number of additional trials needed to get the i-th success after the $(i-1)$-th success. Then $X_i \sim \mathrm{Geometric}(p)$ and $X = \sum_{i=1}^{r} X_i$.

By linearity of expectation:

$$E(X) = \sum_{i=1}^{r} E(X_i) = \sum_{i=1}^{r} \frac{1}{p} = \frac{r}{p}$$

(b) Since the X_i's are independent:

$$\mathrm{Var}(X) = \sum_{i=1}^{r} \mathrm{Var}(X_i) = \sum_{i=1}^{r} \frac{1-p}{p^2} = \frac{r(1-p)}{p^2}$$

□

Hypergeometric Distribution

Exercise 16 (DIY: Hypergeometric Expectation). Let $X \sim \mathrm{Hypergeometric}(N, K, n)$.

Show that $E(X) = n \cdot \frac{K}{N}$ using indicator variables.

Hint: Let $X_i = 1$ if the i-th draw is a success, 0 otherwise. Then $X = \sum_{i=1}^{n} X_i$. Note that $P(X_i = 1) = \frac{K}{N}$ for each i (think about symmetry).

Proof. Let $X_i = 1$ if the i-th item drawn is a success (one of the K good items), and $X_i = 0$ otherwise.

By symmetry, each draw is equally likely to be any of the N items, so:

$$P(X_i = 1) = \frac{K}{N}$$

Therefore:

$$E(X_i) = \frac{K}{N}$$

Since $X = \sum_{i=1}^{n} X_i$, by linearity of expectation:

$$E(X) = \sum_{i=1}^{n} E(X_i) = n \cdot \frac{K}{N}$$

Note: This is true even though the X_i's are not independent! Linearity of expectation doesn't require independence. □

Remark. The variance of the hypergeometric distribution is more complex because the draws are not independent. The formula is:

$$\text{Var}(X) = n \cdot \frac{K}{N} \cdot \frac{N-K}{N} \cdot \frac{N-n}{N-1}$$

Notice that this is similar to the binomial variance $np(1-p)$ but with an extra factor $\frac{N-n}{N-1}$, which accounts for the fact that we're sampling without replacement. When N is large compared to n, this factor is close to 1, and the hypergeometric variance approaches the binomial variance.

Poisson Distribution

Exercise 17 (DIY: Poisson Expectation and Variance). Let $X \sim \text{Poisson}(\lambda)$.

(a) Show that $E(X) = \lambda$ using the Taylor series $e^\lambda = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!}$.

(b) Show that $E(X^2) = \lambda^2 + \lambda$ using the identity $k^2 = k(k-1) + k$.

(c) Show that $\text{Var}(X) = \lambda$.

Proof. (a) By definition:

$$E(X) = \sum_{k=0}^{\infty} k \cdot \frac{\lambda^k e^{-\lambda}}{k!} = e^{-\lambda} \sum_{k=1}^{\infty} k \cdot \frac{\lambda^k}{k!}$$

For $k \geq 1$, we have $\frac{k}{k!} = \frac{1}{(k-1)!}$, so:

$$E(X) = e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} = e^{-\lambda}\lambda \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} = e^{-\lambda}\lambda e^{\lambda} = \lambda$$

where we substituted $j = k - 1$.

(b) Using $k^2 = k(k-1) + k$:

$$E(X^2) = \sum_{k=0}^{\infty} k^2 \cdot \frac{\lambda^k e^{-\lambda}}{k!} = \sum_{k=0}^{\infty} [k(k-1) + k] \cdot \frac{\lambda^k e^{-\lambda}}{k!}$$

For $k \geq 2$, we have $\frac{k(k-1)}{k!} = \frac{1}{(k-2)!}$, so:

$$E(X^2) = e^{-\lambda} \sum_{k=2}^{\infty} \frac{\lambda^k}{(k-2)!} + e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!}$$

Let $j = k-2$ in the first sum and $j = k-1$ in the second:

$$E(X^2) = e^{-\lambda} \lambda^2 \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} + e^{-\lambda} \lambda \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} = \lambda^2 + \lambda$$

(c)

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

This shows that for the Poisson distribution, the mean equals the variance!

□

3.7 More Examples and Applications

3.7.1 Advanced Expectation Problems

Example (Expected Value of a Function). Let $X \sim \text{Binomial}(n,p)$. Find $E\left(\frac{1}{X+1}\right)$.

Using $\frac{1}{k+1}\binom{n}{k} = \frac{1}{n+1}\binom{n+1}{k+1}$ and letting $j = k+1$:

$$E\left(\frac{1}{X+1}\right) = \frac{1}{n+1}\sum_{j=1}^{n+1}\binom{n+1}{j}p^{j-1}(1-p)^{n+1-j} = \frac{1}{p(n+1)}[1-(1-p)^{n+1}].$$

Example (Conditional Expectation). A fair coin is flipped until heads appears. Let $X \sim \text{Geometric}(1/2)$ be the number of flips. Given $X > 3$, what is $E(X)$?

By the memoryless property: $E(X \mid X > 3) = 3 + E(X) = 5$.

3.7.2 Real-World Applications

Example (Quality Control). Items are defective with probability 0.01, independently. Inspect until 5 defectives are found. Let $X \sim \text{NegBin}(5, 0.01)$ be the number inspected.

Then $E(X) = \frac{5}{0.01} = 500$ items on average.

Example (Network Reliability). A network has L links, each failing independently with probability 0.05. Let X be the number of failed links.

Then $E(X) = 0.05L$. For large L, we approximate $X \sim \text{Poisson}(\lambda = 0.05L)$, so:

$$P(X \leq 3) \approx \sum_{k=0}^{3}\frac{\lambda^k e^{-\lambda}}{k!}.$$

Example (Birthday Problem Variant). In a class of 30 students, what is the expected number of pairs sharing the same birthday? (Assume 365 days)

For each of the $\binom{30}{2} = 435$ pairs, let $X_{ij} = 1$ if they share a birthday. Then $P(X_{ij} = 1) = \frac{1}{365}$, so by linearity:

$$E(\text{number of pairs}) = 435 \cdot \frac{1}{365} \approx 1.19.$$

3.7.3 Challenging Problems

Exercise 18. Let $X \sim \text{Poisson}(\lambda)$. Find $E(X(X-1)(X-2))$.

Hint: Use the fact that for $k \geq 3$, we have $\frac{k(k-1)(k-2)}{k!} = \frac{1}{(k-3)!}$.

Proof.

$$E(X(X-1)(X-2)) = \sum_{k=0}^{\infty} k(k-1)(k-2) \cdot \frac{\lambda^k e^{-\lambda}}{k!}$$

For $k < 3$, the terms are 0. For $k \geq 3$:

$$\frac{k(k-1)(k-2)}{k!} = \frac{1}{(k-3)!}$$

Therefore:

$$E(X(X-1)(X-2)) = e^{-\lambda} \sum_{k=3}^{\infty} \frac{\lambda^k}{(k-3)!} = e^{-\lambda}\lambda^3 \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} = \lambda^3$$

where we substituted $j = k - 3$. □

Exercise 19. A fair die is rolled n times. Let X be the number of times a 6 appears, and let Y be the number of times an odd number appears. Find:

(a) $E(X)$ and $E(Y)$

(b) $E(XY)$

(c) $\mathrm{Cov}(X, Y) = E(XY) - E(X)E(Y)$

Proof. (a) $X \sim \text{Binomial}(n, 1/6)$, so $E(X) = n/6$.

$Y \sim \text{Binomial}(n, 1/2)$ (since 3 out of 6 outcomes are odd), so $E(Y) = n/2$.

(b) To find $E(XY)$, we use indicator variables. Let $X_i = 1$ if roll i is a 6, and $Y_i = 1$ if roll i is odd.

Then $X = \sum_{i=1}^{n} X_i$ and $Y = \sum_{i=1}^{n} Y_i$.

$$XY = \left(\sum_{i=1}^{n} X_i\right)\left(\sum_{j=1}^{n} Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} X_i Y_j$$

$$E(XY) = \sum_{i=1}^{n}\sum_{j=1}^{n} E(X_i Y_j)$$

When $i \neq j$: $E(X_i Y_j) = E(X_i)E(Y_j) = \frac{1}{6} \cdot \frac{1}{2} = \frac{1}{12}$ (by independence).

When $i = j$: $E(X_i Y_i) = P(\text{roll } i \text{ is both 6 and odd}) = 0$ (since 6 is even).

Therefore:

$$E(XY) = \sum_{i \neq j} \frac{1}{12} + \sum_{i=1}^{n} 0 = n(n-1) \cdot \frac{1}{12} = \frac{n(n-1)}{12}$$

(c)

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y) = \frac{n(n-1)}{12} - \frac{n}{6} \cdot \frac{n}{2} = \frac{n(n-1)}{12} - \frac{n^2}{12} = -\frac{n}{12}$$

This makes sense: X and Y are negatively correlated because a 6 cannot be odd!

□

Exercise 20. Let $X \sim \text{Geometric}(p)$. Show that for any positive integers m and n:

$$P(X > m + n \mid X > m) = P(X > n)$$

This is the *memoryless property* of the geometric distribution.

Proof.

$$P(X > m + n \mid X > m) = \frac{P(X > m + n \text{ and } X > m)}{P(X > m)} = \frac{P(X > m + n)}{P(X > m)}$$

Now, $P(X > k) = \sum_{j=k+1}^{\infty} (1-p)^{j-1} p = p(1-p)^k \sum_{j=0}^{\infty} (1-p)^j = p(1-p)^k \cdot \frac{1}{p} = (1-p)^k$

Therefore:

$$P(X > m + n \mid X > m) = \frac{(1-p)^{m+n}}{(1-p)^m} = (1-p)^n = P(X > n)$$

This shows that the geometric distribution "forgets" how many failures have already occurred! □

3.8 Summary and Key Takeaways

3.8.1 Key Formulas to Remember

Concept	Formula
$E(X)$ (discrete)	$\sum_x x \cdot p_X(x)$
$E(X)$ (continuous)	$\int_{-\infty}^{\infty} x \cdot f_X(x)\, dx$
$E[g(X)]$ (discrete)	$\sum_x g(x) \cdot p_X(x)$
$E[g(X)]$ (continuous)	$\int_{-\infty}^{\infty} g(x) \cdot f_X(x)\, dx$
$\text{Var}(X)$	$E(X^2) - [E(X)]^2$
σ_X	$\sqrt{\text{Var}(X)}$
$E(aX + b)$	$aE(X) + b$
$\text{Var}(aX + b)$	$a^2\text{Var}(X)$
$E(X + Y)$	$E(X) + E(Y)$ (always)

3.8.2 Distribution Summary

All discrete distributions covered:

- **Bernoulli:** Single trial, $E(X) = p$, $\text{Var}(X) = p(1 - p)$
- **Binomial:** n independent trials, $E(X) = np$, $\text{Var}(X) = np(1 - p)$
- **Geometric:** Trials until first success, $E(X) = 1/p$, $\text{Var}(X) = (1 - p)/p^2$, memoryless
- **Negative Binomial:** Trials until r successes, $E(X) = r/p$, $\text{Var}(X) = r(1-p)/p^2$
- **Hypergeometric:** Sampling without replacement, $E(X) = nK/N$
- **Poisson:** Rare events, $E(X) = \lambda$, $\text{Var}(X) = \lambda$ (mean = variance!)

3.8.3 Problem-Solving Strategies

1. **For expectation problems:**
   - Use linearity of expectation whenever possible
   - Consider indicator variables for counting problems
   - Use the definition $E[g(X)] = \sum g(x)p_X(x)$ or the integral version
2. **For variance problems:**
   - Always use $\text{Var}(X) = E(X^2) - [E(X)]^2$ (usually easier than the definition)
   - Remember $\text{Var}(aX + b) = a^2\text{Var}(X)$

- For independent X and Y: $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y)$

3. **For distribution identification:**

   - Look for keywords: "until first success" → Geometric
   - "number of successes in n trials" → Binomial (with replacement) or Hypergeometric (without replacement)
   - "rare events" or "average rate" → Poisson

Chapter 4

Discrete Joint Distributions

4.1 Basic Concepts

Definition 26 (Joint Probability Mass Function). Let X and Y be discrete random variables. The *joint probability mass function (joint PMF)* of X and Y is defined as:

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

The joint PMF satisfies:

1. $p_{X,Y}(x, y) \geq 0$ for all x, y
2. $\sum_x \sum_y p_{X,Y}(x, y) = 1$

Definition 27 (Marginal PMF). The *marginal PMF* of X is:

$$p_X(x) = \sum_y p_{X,Y}(x, y)$$

Similarly, the marginal PMF of Y is:

$$p_Y(y) = \sum_x p_{X,Y}(x, y)$$

Exercise 21 (DIY: Marginals Sum to 1). Prove that if $p_{X,Y}(x, y)$ is a valid joint PMF, then:

$$\sum_x p_X(x) = 1 \quad \text{and} \quad \sum_y p_Y(y) = 1$$

Proof.

$$\sum_x p_X(x) = \sum_x \sum_y p_{X,Y}(x, y) = 1$$

since the joint PMF sums to 1. Similarly:

$$\sum_y p_Y(y) = \sum_y \sum_x p_{X,Y}(x,y) = 1$$

□

Example (Simple Joint Distribution). Consider two random variables X and Y with the following joint PMF:

$p_{X,Y}(x,y)$	$X=1$	$X=2$	$X=3$
$Y=2$	0.1	0.2	0.1
$Y=3$	0.15	0.15	0.1
$Y=4$	0.05	0.1	0.05

Find:

(a) The marginal PMFs of X and Y

(b) $P(X=2)$

(c) $P(Y \geq 3)$

(d) $P(X=2, Y=3)$

Solution. (a) **Marginal PMF of X:**

$$p_X(1) = 0.1+0.15+0.05 = 0.3, \quad p_X(2) = 0.2+0.15+0.1 = 0.45, \quad p_X(3) = 0.1+0.1+0.05 = 0.$$

Marginal PMF of Y:

$$p_Y(2) = 0.1+0.2+0.1 = 0.4, \quad p_Y(3) = 0.15+0.15+0.1 = 0.4, \quad p_Y(4) = 0.05+0.1+0.05 = 0.2$$

(b) $P(X=2) = p_X(2) = 0.45$

(c) $P(Y \geq 3) = p_Y(3) + p_Y(4) = 0.4 + 0.2 = 0.6$

(d) $P(X=2, Y=3) = p_{X,Y}(2,3) = 0.15$

□

4.2 Conditional Distributions

Definition 28 (Conditional PMF). The *conditional PMF* of Y given $X = x$ is:

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

Similarly, the conditional PMF of X given $Y = y$ is:

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

Exercise 22 (DIY: Conditional PMF is Valid PMF). Prove that for a fixed x with $p_X(x) > 0$, the conditional PMF $p_{Y|X}(y \mid x)$ satisfies:

(a) $p_{Y|X}(y \mid x) \geq 0$ for all y

(b) $\sum_y p_{Y|X}(y \mid x) = 1$

Proof. (a) Since $p_{X,Y}(x,y) \geq 0$ and $p_X(x) > 0$, we have $p_{Y|X}(y \mid x) = \frac{p_{X,Y}(x,y)}{p_X(x)} \geq 0$.

(b)

$$\sum_y p_{Y|X}(y \mid x) = \sum_y \frac{p_{X,Y}(x,y)}{p_X(x)} = \frac{1}{p_X(x)} \sum_y p_{X,Y}(x,y) = \frac{p_X(x)}{p_X(x)} = 1$$

□

Example (Conditional Distributions). Using the joint distribution from the previous example, find:

(a) $p_{Y|X}(y \mid 2)$ (the conditional PMF of Y given $X = 2$)

(b) $p_{X|Y}(x \mid 3)$ (the conditional PMF of X given $Y = 3$)

(c) $P(Y = 4 \mid X = 1)$

Solution. (a) Given $X = 2$, we have $p_X(2) = 0.45$:

$$p_{Y|X}(2 \mid 2) = \frac{0.2}{0.45} = \frac{4}{9}, \quad p_{Y|X}(3 \mid 2) = \frac{0.15}{0.45} = \frac{1}{3}, \quad p_{Y|X}(4 \mid 2) = \frac{0.1}{0.45} = \frac{2}{9}$$

(b) Given $Y = 3$, we have $p_Y(3) = 0.4$:

$$p_{X|Y}(1 \mid 3) = \frac{0.15}{0.4} = \frac{3}{8}, \quad p_{X|Y}(2 \mid 3) = \frac{0.15}{0.4} = \frac{3}{8}, \quad p_{X|Y}(3 \mid 3) = \frac{0.1}{0.4} = \frac{1}{4}$$

(c) $P(Y = 4 \mid X = 1) = p_{Y|X}(4 \mid 1) = \frac{p_{X,Y}(1,4)}{p_X(1)} = \frac{0.05}{0.3} = \frac{1}{6}$

□

4.3 Independence

Definition 29 (Independence). Two discrete random variables X and Y are *independent* if and only if:

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

Equivalently, X and Y are independent if:

$$p_{Y|X}(y \mid x) = p_Y(y) \quad \text{for all } x, y \text{ with } p_X(x) > 0$$

Exercise 23 (DIY: Equivalence of Independence Definitions). Prove that the two definitions of independence are equivalent:

(a) If $p_{X,Y}(x, y) = p_X(x) \cdot p_Y(y)$ for all x, y, then $p_{Y|X}(y \mid x) = p_Y(y)$ for all x, y with $p_X(x) > 0$.

(b) If $p_{Y|X}(y \mid x) = p_Y(y)$ for all x, y with $p_X(x) > 0$, then $p_{X,Y}(x, y) = p_X(x) \cdot p_Y(y)$ for all x, y.

Proof. (a) If $p_{X,Y}(x, y) = p_X(x) \cdot p_Y(y)$, then:

$$p_{Y|X}(y \mid x) = \frac{p_{X,Y}(x, y)}{p_X(x)} = \frac{p_X(x) \cdot p_Y(y)}{p_X(x)} = p_Y(y)$$

provided $p_X(x) > 0$.

(b) If $p_{Y|X}(y \mid x) = p_Y(y)$ for all x, y with $p_X(x) > 0$, then:

$$p_{X,Y}(x, y) = p_{Y|X}(y \mid x) \cdot p_X(x) = p_Y(y) \cdot p_X(x) = p_X(x) \cdot p_Y(y)$$

For x with $p_X(x) = 0$, we have $p_{X,Y}(x, y) = 0 = p_X(x) \cdot p_Y(y)$ as well.

□

Example (Checking Independence). Determine if X and Y are independent in the previous example.

Solution. We check if $p_{X,Y}(x,y) = p_X(x) \cdot p_Y(y)$ for all x, y.
For example, check $(x,y) = (1,2)$:

$$p_{X,Y}(1,2) = 0.1$$

$$p_X(1) \cdot p_Y(2) = 0.3 \cdot 0.4 = 0.12$$

Since $0.1 \neq 0.12$, X and Y are **not independent**. □

Example (Independent Random Variables). Consider two independent random variables X and Y with:

$$p_X(1) = 0.4, \quad p_X(2) = 0.6$$

$$p_Y(2) = 0.3, \quad p_Y(3) = 0.5, \quad p_Y(4) = 0.2$$

Find the joint PMF $p_{X,Y}(x,y)$.

Solution. Since X and Y are independent:

$$p_{X,Y}(x,y) = p_X(x) \cdot p_Y(y)$$

$p_{X,Y}(x,y)$	$X = 1$	$X = 2$
$Y = 2$	$0.4 \cdot 0.3 = 0.12$	$0.6 \cdot 0.3 = 0.18$
$Y = 3$	$0.4 \cdot 0.5 = 0.20$	$0.6 \cdot 0.5 = 0.30$
$Y = 4$	$0.4 \cdot 0.2 = 0.08$	$0.6 \cdot 0.2 = 0.12$

□

4.4 Expectation and Functions of Two Random Variables

Exercise 24 (DIY: Linearity of Expectation). Prove that for any constants a, b, c:

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

Hint: Use the definition $E[g(X,Y)] = \sum_x \sum_y g(x,y) \cdot p_{X,Y}(x,y)$.

Proof.

$$E(aX + bY + c) = \sum_x \sum_y (ax + by + c) \cdot p_{X,Y}(x, y)$$
$$= a \sum_x \sum_y x \cdot p_{X,Y}(x, y) + b \sum_x \sum_y y \cdot p_{X,Y}(x, y) + c \sum_x \sum_y p_{X,Y}(x, y)$$
$$= a \sum_x x \sum_y p_{X,Y}(x, y) + b \sum_y y \sum_x p_{X,Y}(x, y) + c \cdot 1$$
$$= a \sum_x x \cdot p_X(x) + b \sum_y y \cdot p_Y(y) + c = aE(X) + bE(Y) + c$$

This holds **regardless of whether X and Y are independent**! □

Example (Expectation Calculations). Using the joint distribution from the first example, find:

(a) $E(X)$ and $E(Y)$

(b) $E(XY)$

(c) $E(2X + 3Y - 1)$

(d) $E(X^2)$

Solution. (a)

$$E(X) = 1(0.3) + 2(0.45) + 3(0.25) = 0.3 + 0.9 + 0.75 = 1.95$$

$$E(Y) = 2(0.4) + 3(0.4) + 4(0.2) = 0.8 + 1.2 + 0.8 = 2.8$$

(b)

$$E(XY) = \sum_x \sum_y xy \cdot p_{X,Y}(x, y)$$

$$= 1 \cdot 2(0.1) + 1 \cdot 3(0.15) + 1 \cdot 4(0.05) + 2 \cdot 2(0.2) + 2 \cdot 3(0.15)$$
$$+2 \cdot 4(0.1) + 3 \cdot 2(0.1) + 3 \cdot 3(0.1) + 3 \cdot 4(0.05)$$
$$= 0.2 + 0.45 + 0.2 + 0.8 + 0.9 + 0.8 + 0.6 + 0.9 + 0.6 = 5.45$$

(c)

$$E(2X+3Y-1) = 2E(X)+3E(Y)-1 = 2(1.95)+3(2.8)-1 = 3.9+8.4-1 = 11.3$$

(d)

$$E(X^2) = \sum_x x^2 \cdot p_X(x) = 1^2(0.3)+2^2(0.45)+3^2(0.25) = 0.3+1.8+2.25 = 4.35$$

□

Exercise 25 (DIY: Expectation of Product for Independent RVs)**.** Prove that if X and Y are independent discrete random variables, then:

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

Hint: Use the fact that $p_{X,Y}(x, y) = p_X(x) \cdot p_Y(y)$ when X and Y are independent.

Proof.

$$E(XY) = \sum_x \sum_y xy \cdot p_{X,Y}(x, y) = \sum_x \sum_y xy \cdot p_X(x) \cdot p_Y(y)$$

$$= \sum_x x \cdot p_X(x) \sum_y y \cdot p_Y(y) = E(X) \cdot E(Y)$$

□

Exercise 26 (DIY: Law of Total Expectation)**.** Prove the law of total expectation for discrete random variables:

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

where $E(Y \mid X = x) = \sum_y y \cdot p_{Y|X}(y \mid x)$ is the conditional expectation.

Proof.

$$\sum_x E(Y \mid X = x) \cdot p_X(x) = \sum_x \left[\sum_y y \cdot p_{Y|X}(y \mid x) \right] \cdot p_X(x)$$

$$= \sum_x \sum_y y \cdot \frac{p_{X,Y}(x, y)}{p_X(x)} \cdot p_X(x) = \sum_x \sum_y y \cdot p_{X,Y}(x, y) = E(Y)$$

□

Exercise 27 (DIY: Expectation of Function of Independent RVs)**.** Prove that if X and Y are independent discrete random variables and $g : \mathbb{R} \to \mathbb{R}$ and $h : \mathbb{R} \to \mathbb{R}$ are functions, then:

$$E[g(X)h(Y)] = E[g(X)] \cdot E[h(Y)]$$

Hint: This generalizes the result $E(XY) = E(X)E(Y)$ for independent random variables.

Proof.

$$E[g(X)h(Y)] = \sum_x \sum_y g(x)h(y) \cdot p_{X,Y}(x, y) = \sum_x \sum_y g(x)h(y) \cdot p_X(x) \cdot p_Y(y)$$

$$= \sum_x g(x) \cdot p_X(x) \sum_y h(y) \cdot p_Y(y) = E[g(X)] \cdot E[h(Y)]$$

□

4.5 Exam-Style Practice Problems

Exercise 28. Let X and Y be discrete random variables with joint PMF:

$p_{X,Y}(x,y)$	$X=1$	$X=2$	$X=3$
$Y=2$	0.1	0.15	0.05
$Y=3$	0.2	0.15	0.1
$Y=4$	0.1	0.05	0.1

(a) Find the marginal PMFs of X and Y.

(b) Find $P(X = 2 \mid Y = 3)$.

(c) Are X and Y independent? Justify your answer.

(d) Find $E(X)$, $E(Y)$, and $E(XY)$.

(e) **DIY:** Verify that $E(XY) = E(X)E(Y)$ using the definition of expectation. Does this contradict independence? Explain.

Proof. (a)

$$p_X(1) = 0.1{+}0.2{+}0.1 = 0.4, \quad p_X(2) = 0.15{+}0.15{+}0.05 = 0.35, \quad p_X(3) = 0.05{+}0.1{+}0.1 = 0$$

$$p_Y(2) = 0.1{+}0.15{+}0.05 = 0.3, \quad p_Y(3) = 0.2{+}0.15{+}0.1 = 0.45, \quad p_Y(4) = 0.1{+}0.05{+}0.1 = 0.$$

(b)

$$P(X = 2 \mid Y = 3) = \frac{0.15}{0.45} = \frac{1}{3}$$

(c) Check: $p_{X,Y}(1,2) = 0.1$ and $p_X(1) \cdot p_Y(2) = 0.4 \cdot 0.3 = 0.12$. Since $0.1 \neq 0.12$, X and Y are **not independent**.

(d)

$$\begin{aligned}
E(X) &= 1(0.4) + 2(0.35) + 3(0.25) = 0.4 + 0.7 + 0.75 = 1.85 \\
E(Y) &= 2(0.3) + 3(0.45) + 4(0.25) = 0.6 + 1.35 + 1.0 = 2.95 \\
E(XY) &= 1 \cdot 2(0.1) + 1 \cdot 3(0.2) + 1 \cdot 4(0.1) + 2 \cdot 2(0.15) + 2 \cdot 3(0.15) \\
&\quad + 2 \cdot 4(0.05) + 3 \cdot 2(0.05) + 3 \cdot 3(0.1) + 3 \cdot 4(0.1) \\
&= 0.2 + 0.6 + 0.4 + 0.6 + 0.9 + 0.4 + 0.3 + 0.9 + 1.2 = 5.5
\end{aligned}$$

(e) **DIY Solution:** We already calculated $E(X) = 1.85$, $E(Y) = 2.95$, and $E(XY) = 5.5$.

Now $E(X)E(Y) = 1.85 \cdot 2.95 = 5.4575 \neq 5.5 = E(XY)$.

This does not contradict independence. In fact, since $E(XY) \neq E(X)E(Y)$, this confirms that X and Y are **not independent**. If they were independent, we would have $E(XY) = E(X)E(Y)$.

□

Exercise 29. Let X and Y be independent discrete random variables with:

$$p_X(1) = 0.3, \quad p_X(2) = 0.5, \quad p_X(3) = 0.2$$

$$p_Y(2) = 0.4, \quad p_Y(3) = 0.3, \quad p_Y(4) = 0.2, \quad p_Y(5) = 0.1$$

(a) Find the joint PMF $p_{X,Y}(x, y)$.

(b) Find $E(X)$, $E(Y)$, and $E(XY)$.

(c) Find $E(2X - 3Y + 1)$.

Proof. (a) Since X and Y are independent, $p_{X,Y}(x, y) = p_X(x) \cdot p_Y(y)$:

$p_{X,Y}(x,y)$	$X = 1$	$X = 2$	$X = 3$
$Y = 2$	0.12	0.20	0.08
$Y = 3$	0.09	0.15	0.06
$Y = 4$	0.06	0.10	0.04
$Y = 5$	0.03	0.05	0.02

(b)

$$E(X) = 1(0.3) + 2(0.5) + 3(0.2) = 0.3 + 1.0 + 0.6 = 1.9$$
$$E(Y) = 2(0.4) + 3(0.3) + 4(0.2) + 5(0.1) = 0.8 + 0.9 + 0.8 + 0.5 = 3.0$$

Since X and Y are independent, $E(XY) = E(X)E(Y) = 1.9 \cdot 3.0 = 5.7$.

(c)

$$E(2X-3Y+1) = 2E(X)-3E(Y)+1 = 2(1.9)-3(3.0)+1 = 3.8-9.0+1 = -4.2$$

□

Exercise 30. Let X and Y have joint PMF:

$p_{X,Y}(x,y)$	$X = 1$	$X = 2$
$Y = 2$	a	b
$Y = 3$	c	d

where $a+b+c+d=1$ and all probabilities are non-negative.

(a) Express the marginal PMFs in terms of a, b, c, d.

(b) Find conditions on a, b, c, d such that X and Y are independent.

(c) If $E(X) = 1.6$ and $E(Y) = 2.4$, find a, b, c, d assuming independence.

(d) **DIY:** Show that if X and Y are independent, then the conditions in part (b) reduce to a single constraint. What is this constraint?

Proof. (a)

$$p_X(1) = a + c, \quad p_X(2) = b + d$$
$$p_Y(2) = a + b, \quad p_Y(3) = c + d$$

(b) For independence, we need $p_{X,Y}(x, y) = p_X(x) \cdot p_Y(y)$ for all x, y:

$$a = (a+c)(a+b), \quad b = (b+d)(a+b), \quad c = (a+c)(c+d), \quad d = (b+d)(c+d)$$

(c) If independent and $E(X) = 1.6$, $E(Y) = 2.4$:

$$1 \cdot (a + c) + 2 \cdot (b + d) = 1.6 \Rightarrow a + c + 2b + 2d = 1.6$$

$$2 \cdot (a + b) + 3 \cdot (c + d) = 2.4 \Rightarrow 2a + 2b + 3c + 3d = 2.4$$

Also, $a + b + c + d = 1$.

Using independence: $a = (a+c)(a+b)$, $b = (b+d)(a+b)$, $c = (a+c)(c+d)$, $d = (b+d)(c+d)$.

Let $p = a + c$ and $q = a + b$. Then $p_X(1) = p$, $p_X(2) = 1 - p$, $p_Y(2) = q$, $p_Y(3) = 1 - q$.

From $E(X) = 1.6$: $p + 2(1 - p) = 1.6 \Rightarrow 2 - p = 1.6 \Rightarrow p = 0.4$.

From $E(Y) = 2.4$: $2q + 3(1 - q) = 2.4 \Rightarrow 3 - q = 2.4 \Rightarrow q = 0.6$.

Therefore:

$$a = p \cdot q = 0.4 \cdot 0.6 = 0.24$$
$$b = (1 - p) \cdot q = 0.6 \cdot 0.6 = 0.36$$
$$c = p \cdot (1 - q) = 0.4 \cdot 0.4 = 0.16$$
$$d = (1 - p) \cdot (1 - q) = 0.6 \cdot 0.4 = 0.24$$

Verification: $0.24 + 0.36 + 0.16 + 0.24 = 1.00$

(d) **DIY Solution:** If X and Y are independent, then $p_{X,Y}(x, y) = p_X(x) \cdot p_Y(y)$ for all x, y.

Let $p = p_X(1) = a + c$ and $q = p_Y(2) = a + b$. Then:

$$a = p \cdot q, \quad b = (1-p) \cdot q, \quad c = p \cdot (1-q), \quad d = (1-p) \cdot (1-q)$$

The single constraint is that $a+b+c+d = 1$, which is automatically satisfied:

$$pq + (1-p)q + p(1-q) + (1-p)(1-q) = q + (1-q) = 1$$

So independence reduces the four parameters (a, b, c, d) to just two parameters (p, q).

□

Exercise 31. Let X and Y be random variables with:

$$E(X) = 2, \quad E(Y) = 3$$

Find:

(a) $E(3X + 2Y - 5)$

(b) $E(X^2)$ if $\mathrm{Var}(X) = 4$

(c) $E(Y^2)$ if $\mathrm{Var}(Y) = 9$

Proof. (a)

$$E(3X + 2Y - 5) = 3E(X) + 2E(Y) - 5 = 3(2) + 2(3) - 5 = 6 + 6 - 5 = 7$$

(b)

$$E(X^2) = \mathrm{Var}(X) + [E(X)]^2 = 4 + 2^2 = 8$$

(c)

$$E(Y^2) = \mathrm{Var}(Y) + [E(Y)]^2 = 9 + 3^2 = 18$$

□

Exercise 32. Let X and Y have joint PMF:

$p_{X,Y}(x,y)$	$X=1$	$X=2$	$X=3$
$Y=2$	0.06	0.12	0.12
$Y=3$	0.08	0.16	0.16
$Y=4$	0.06	0.12	0.12

(a) Find the marginal PMFs.

(b) Show that X and Y are independent.

(c) Find $E(XY)$ and verify that $E(XY) = E(X)E(Y)$.

Proof. (a)

$$p_X(1) = 0.20, \quad p_X(2) = 0.40, \quad p_X(3) = 0.40$$
$$p_Y(2) = 0.30, \quad p_Y(3) = 0.40, \quad p_Y(4) = 0.30$$

(b) Check: $p_{X,Y}(1,2) = 0.06 = p_X(1) \cdot p_Y(2) = 0.20 \cdot 0.30 = 0.06$

$p_{X,Y}(2,3) = 0.16 = p_X(2) \cdot p_Y(3) = 0.40 \cdot 0.40 = 0.16$

$p_{X,Y}(3,4) = 0.12 = p_X(3) \cdot p_Y(4) = 0.40 \cdot 0.30 = 0.12$

All checks pass, so X and Y are **independent**.

(c)

$$E(X) = 1(0.20) + 2(0.40) + 3(0.40) = 0.20 + 0.80 + 1.20 = 2.20$$
$$E(Y) = 2(0.30) + 3(0.40) + 4(0.30) = 0.60 + 1.20 + 1.20 = 3.00$$
$$E(XY) = \sum_x \sum_y xy \cdot p_{X,Y}(x,y) = 1 \cdot 2(0.06) + 1 \cdot 3(0.08) + 1 \cdot 4(0.06)$$

$+2\cdot2(0.12)+2\cdot3(0.16)+2\cdot4(0.12)+3\cdot2(0.12)+3\cdot3(0.16)+3\cdot4(0.12) = 6.60$

Verification: $E(X)E(Y) = 2.20 \cdot 3.00 = 6.60$

$\square$

4.6 Summary and Key Formulas

4.6.1 Joint Distribution Formulas

Concept	Formula
Joint PMF	$p_{X,Y}(x,y) = P(X=x, Y=y)$
Marginal PMF of X	$p_X(x) = \sum_y p_{X,Y}(x,y)$
Marginal PMF of Y	$p_Y(y) = \sum_x p_{X,Y}(x,y)$
Conditional PMF	$p_{Y\|X}(y\|x) = \frac{p_{X,Y}(x,y)}{p_X(x)}$
Independence	$p_{X,Y}(x,y) = p_X(x) \cdot p_Y(y)$ for all x, y
$E[g(X,Y)]$	$\sum_x \sum_y g(x,y) \cdot p_{X,Y}(x,y)$
$E(aX+bY+c)$	$aE(X) + bE(Y) + c$ (always)
$E(XY)$	$\sum_x \sum_y xy \cdot p_{X,Y}(x,y)$

4.6.2 Key Properties

1. **Linearity of Expectation:** $E(aX+bY+c) = aE(X)+bE(Y)+c$ holds **always**, even if X and Y are dependent.
2. **Independence:** If X and Y are independent:
   - $E(XY) = E(X)E(Y)$

4.6.3 Problem-Solving Strategy

1. **Always start by finding marginals** - they're needed for everything else.
2. **For independence checks:** Verify $p_{X,Y}(x,y) = p_X(x) \cdot p_Y(y)$ for all (x,y).
3. **For expectations:** Use linearity whenever possible - it's the easiest approach.
4. **For conditional probabilities:** Use $P(A \mid B) = \frac{P(A \cap B)}{P(B)}$.

Key Terms and Concepts

Joint PMF
The probability mass function $p_{X,Y}(x,y) = P(X = x, Y = y)$ for discrete random variables X and Y. It satisfies: (1) $p_{X,Y}(x,y) \geq 0$ for all x, y, and (2) $\sum_x \sum_y p_{X,Y}(x,y) = 1$.

Marginal PMF
The probability mass function of a single random variable obtained by summing the joint PMF over all values of the other variable: $p_X(x) = \sum_y p_{X,Y}(x,y)$ and $p_Y(y) = \sum_x p_{X,Y}(x,y)$.

Conditional PMF
The probability mass function of one random variable given the value of another: $p_{Y|X}(y|x) = \frac{p_{X,Y}(x,y)}{p_X(x)}$ (provided $p_X(x) > 0$). Conditional PMFs are valid PMFs.

Independence
Two discrete random variables X and Y are independent if $p_{X,Y}(x,y) = p_X(x) \cdot p_Y(y)$ for all x, y. Equivalently, $p_{Y|X}(y|x) = p_Y(y)$ for all x, y with $p_X(x) > 0$.

Linearity of Expectation
For any constants a, b, c: $E(aX + bY + c) = aE(X) + bE(Y) + c$. This holds **always**, regardless of whether X and Y are independent.

Expectation of Product
For independent random variables: $E(XY) = E(X)E(Y)$. More generally, $E[g(X)h(Y)] = E[g(X)]E[h(Y)]$ for independent X and Y. Note: $E(XY) = E(X)E(Y)$ is necessary but not sufficient for independence.

Law of Total Expectation
$E(Y) = \sum_x E(Y|X = x) \cdot p_X(x)$, where $E(Y|X = x) = \sum_y y \cdot p_{Y|X}(y|x)$ is the conditional expectation.

Problem-Solving Strategy
Always start by finding marginals. For independence checks, verify $p_{X,Y}(x,y) = p_X(x) \cdot p_Y(y)$ for all (x,y). For expectations, use linearity whenever possible.

Keywords and Important Concepts

Joint Distributions
Joint PMF, Marginal PMF, Conditional PMF, Independence, Joint probability mass function

Expectation Linearity of Expectation, $E(XY)$, $E[g(X,Y)]$, Law of Total Expectation, Conditional Expectation

Key Properties
Independence criteria ($p_{X,Y} = p_X \cdot p_Y$), Expectation properties, Conditional distributions, Product rule for independent RVs

Problem-Solving Techniques
Finding marginals, Checking independence, Computing expectations, Using linearity, Conditional probabilities

Important Formulas
$p_X(x) = \sum_y p_{X,Y}(x,y)$, $p_{Y|X}(y|x) = \frac{p_{X,Y}(x,y)}{p_X(x)}$, $E(aX + bY + c) = aE(X) + bE(Y) + c$, $E(XY) = E(X)E(Y)$ (if independent)

For detailed definitions and a complete index, see the previous pages.