Probability and Statistics

Lecture Notes

HAMZA BERAHMA

Student

December 6, 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.
$S_3 = \{2, 3, \dots, 12\}$

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

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.
$S_6 = \{1, 2, 3, \dots\} = \mathbb{N}^+$

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

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.

   $$A_1, A_2, \dots, A_n$$

   (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 $S = \{1, 2, \ldots, 12\}$. 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 $S' = \{(i, j) : i, j \in \{1, \ldots, 6\}\}$, so $|S'| = 36$. The sample space for the sum is $S = \{2, 3, \ldots, 12\}$. 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 $P(A \cap B) = P(B)P(A|B)$:

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

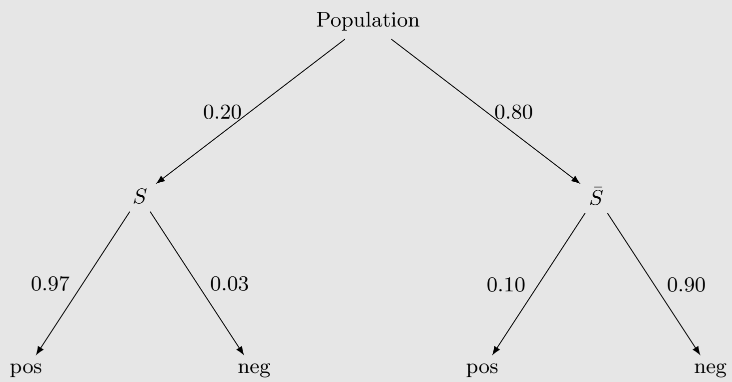


Figure 1.1: Tree diagram illustrating the relationships between disease status and test results.

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

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

Explanation. 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:

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

Explanation. 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:

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

Explanation. 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

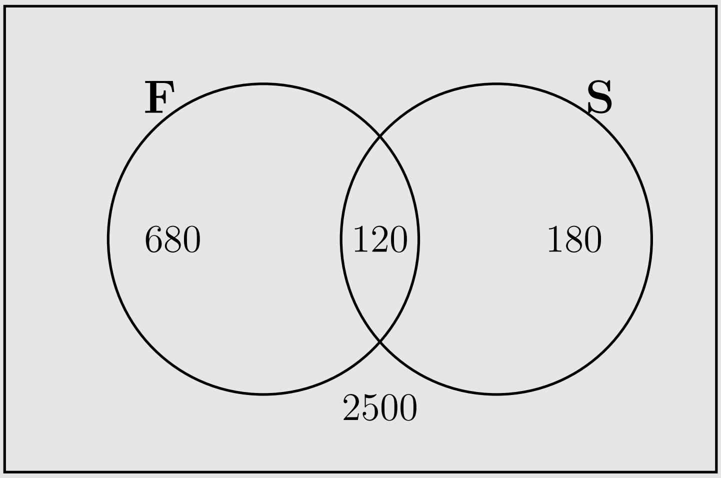


Figure 1.2: Venn diagram showing football (F) and swimming (S) students at AUI.

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.

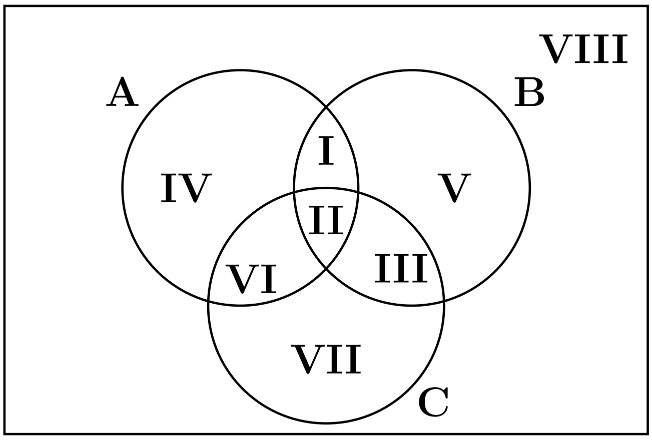


Figure 1.3: Three-set Venn diagram with labeled regions I–VIII.

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

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

Explanation. We sum all regions that include A.

(b) Probability of $A \cap C$

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

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

(c) Probability of $A \cup C$

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

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

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

(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:

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

□

Explanation. 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:

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

□

Explanation. 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

Explanation. 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:

$$4 \times 6 = 24$$

Explanation. 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

Explanation. 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)

Explanation. 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)

Explanation. 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 $\{a, b, c\}$?

$$|U| = 7^6$$

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

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

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

Explanation. 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

Explanation. 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 $\{a, b, c, d, e\}$:

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

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

3. Combinations / Selections

> **Explanation.** 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 $\{a, b, c, d, e\}$:
> $$\binom{5}{3} = 10$$
>
> > **Explanation.** Order does not matter; $\{a, b, c\} = \{c, b, a\}$, so divide permutations 3! to remove overcounting.

4. Permutations with Repetition

> **Explanation.** 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):
> $$\frac{5!}{2!} = 60$$

> **Explanation** (General Formula). 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

> **Explanation.** 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 $\{1, 2, \ldots, 6\}$ uniformly at random. Then the sample space is

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

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

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

(b) Define another random variable

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

Then

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

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:

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

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

Then the range of W is

$$\text{Range}(W) = \{1, 2, 3\},$$

since the first working camera could appear in the first, second, or third test at most (after testing 3 failed ones, you are guaranteed a working one next).

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,

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

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 mix them up and test them one by one until you find a working one. Let the random variable W be the number of cameras tested.

Then the range of W is

$$\text{Range}(W) = \{1, 2, 3\},$$

since the first working camera could appear in the first, second, or third test at most (after testing 3 failed ones, you are guaranteed a working one next).

We can compute the probability mass function $p_W(k) = P(W = k)$ as follows.

$$f_W(w) = \begin{cases} \frac{2}{5}, & w = 1, \\ \frac{3}{5} \cdot \frac{2}{4} = \frac{3}{10}, & w = 2, \\ \frac{3}{5} \cdot \frac{2}{4} = \frac{6}{20}, & w = 3, \\ 0, & \text{otherwise.} \end{cases}$$

Verification:

$$\frac{2}{5} + \frac{3}{10} + \frac{6}{20} = 1,$$

so the PMF is valid.

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 $p = \frac{1}{4}$, and the probability of failure is $q = 1 - p = \frac{3}{4}$.

(a) For $X = 8$:

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

(b) In general, for the geometric distribution:

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

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

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

3.2 More on discrete random variables

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

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

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

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

(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.

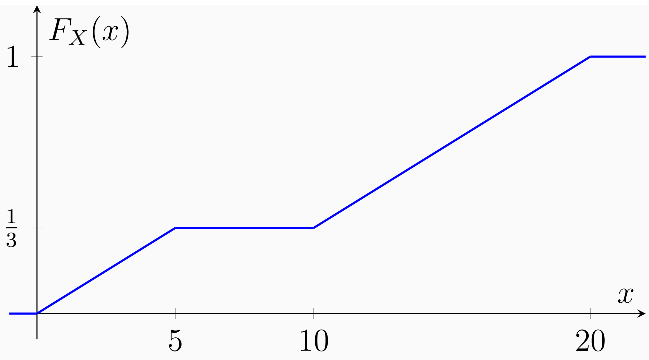
Aspect	**PMF**	**CDF**
Symbol	$p_X(x)$	$F_X(x)$
Definition	$p_X(x) = P(X = x)$	$F_X(x) = P(X \leq x)$
Domain	Discrete x	Real x
Range	$[0, 1]$	$[0, 1]$
Relation	$F_X(x) = \sum_{t \leq x} p_X(t)$	$p_X(x) = F_X(x) - F_X(x^-)$
Nature	Stepwise (non-continuous)	Non-decreasing, right-continuous

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

Proof. The total length of the support is $5 + 10 = 15$.

The cumulative distribution function is:

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

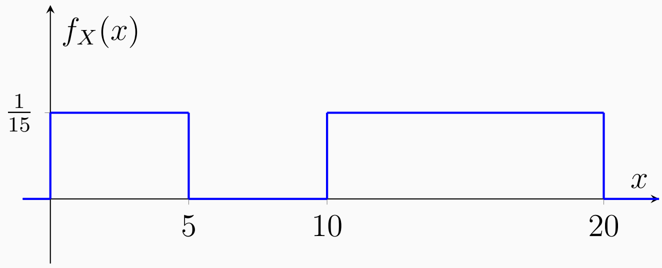


□

Exercise 7. Y has range $[0,10] \cup [20,25]$. It is uniform over $[0,10]$ and uniform over $[20,25]$ with $P(10 \leq Y \leq 25) = 3 \cdot P(0 \leq Y \leq 10)$. Find F_Y X is uniform over $[0,5] \cup [10,20]$. Find $f_X(x)$ and draw it.

Proof. The total length of the support is $(5-0)+(20-10) = 15$. For a uniform distribution, the pdf $f_X(x)$ is 1/(Total Length) over the support, and 0 elsewhere.

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



□

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

Proof. Let $P(0 \leq Y \leq 10) = p_1$ and $P(20 \leq Y \leq 25) = p_2$. Given: $p_1 + p_2 = 1$ and $p_2 = 3p_1$. Substituting gives $p_1 + 3p_1 = 1 \Rightarrow 4p_1 = 1 \Rightarrow p_1 = \frac{1}{4}$ and $p_2 = \frac{3}{4}$.

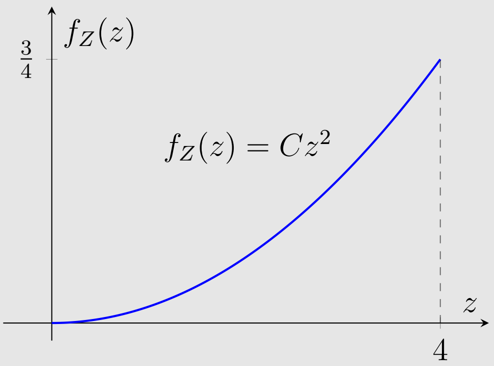
We find the pdf $f_Y(y)$ heights. For $0 \leq y \leq 10$ (length 10): $\int_0^{10} c_1\, dy = p_1 \Rightarrow 10c_1 = \frac{1}{4} \Rightarrow c_1 = \frac{1}{40}$. For $20 \leq y \leq 25$ (length 5): $\int_{20}^{25} c_2\, dy = p_2 \Rightarrow 5c_2 = \frac{3}{4} \Rightarrow c_2 = \frac{3}{20}$.

Now, we find the CDF $F_Y(y) = \int_{-\infty}^{y} f_Y(t)\,dt$:

$$F_Y(y) = \begin{cases} 0 & y < 0 \\ \int_0^y \frac{1}{40}\,dt = \frac{y}{40} & 0 \le y < 10 \\ F_Y(10) = \frac{10}{40} = \frac{1}{4} & 10 \le y < 20 \\ F_Y(20) + \int_{20}^{y} \frac{3}{20}\,dt = \frac{1}{4} + \frac{3(y-20)}{20} & 20 \le y < 25 \\ F_Y(25) = \frac{1}{4} + \frac{3(25-20)}{20} = \frac{1}{4} + \frac{15}{20} = 1 & y \ge 25 \end{cases}$$

□

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, its integral over its range $[0, 4]$ must be 1.

$$\int_0^4 Cz^2\,dz = 1$$

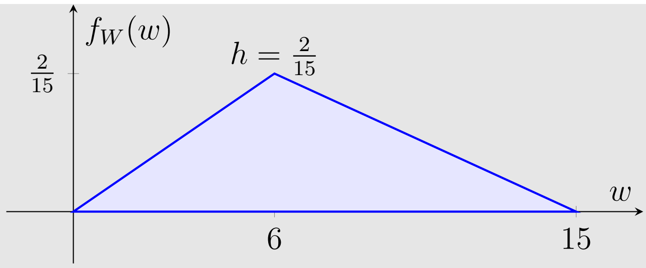
$$C\left[\frac{z^3}{3}\right]_0^4 = 1$$

$$C\left(\frac{4^3}{3} - \frac{0^3}{3}\right) = 1$$

$$C\left(\frac{64}{3}\right) = 1 \Rightarrow c = \frac{3}{64}$$

(The peak value at $z = 4$ is $f_Z(4) = \frac{3}{64}(4^2) = \frac{3 \cdot 16}{64} = \frac{48}{64} = \frac{3}{4}$.) □

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



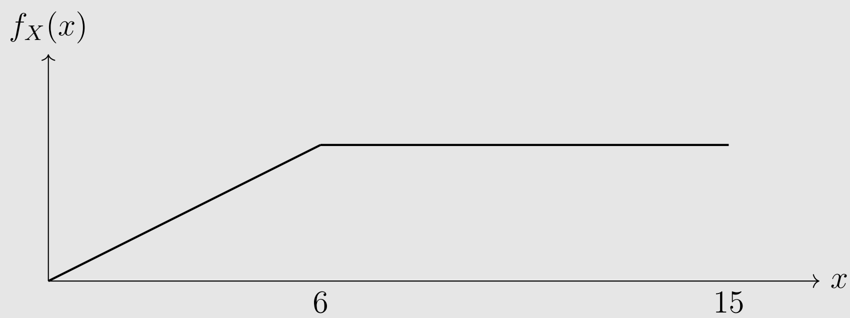
Proof. The pdf is a triangle with base $b = 15$. The area must be 1, so $\frac{1}{2} \cdot 15 \cdot h = 1$, which gives the height $h = \frac{2}{15}$ at the peak $w = 6$.

We find the equations for the two lines:

- **Line 1 ($0 \le w < 6$):** Passes through $(0, 0)$ and $(6, 2/15)$. Slope $m_1 = \frac{2/15}{6} = \frac{1}{45}$. Equation: $f_W(w) = \frac{1}{45}w$.
- **Line 2 ($6 \le w \le 15$):** Passes through $(6, 2/15)$ and $(15, 0)$. Slope $m_2 = \frac{0-2/15}{15-6} = -\frac{2}{135}$. Using point $(15, 0)$: $f_W(w) = -\frac{2}{135}(w - 15) = \frac{2}{9} - \frac{2}{135}w$.

$$f_W(w) = \begin{cases} \frac{1}{45}w & 0 \le w < 6 \\ \frac{2}{9} - \frac{2}{135}w & 6 \le w \le 15 \\ 0 & \text{otherwise} \end{cases}$$

$\square$



Exercise 11.

Write down $f_X(x)$.

(b) Calculate the following probabilities:

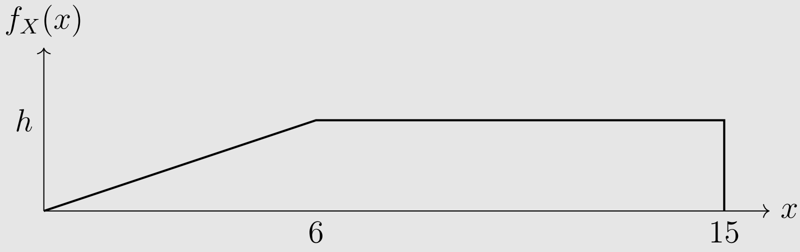
(i) $P(X \le 6)$

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

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

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

Exercise 12. 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) $P(X \leq 6)$

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

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

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

Solution 1. *1. find h and the piecewise form of f_X. The total area must equal 1. The area is the triangle on $[0, 6]$ plus the rectangle on $[6, 15]$:*

$$\frac{1}{2} \cdot 6 \cdot h \; + \; 9 \cdot h \; = \; 3h + 9h = 12h = 1 \quad \Longrightarrow \quad h = \frac{1}{12}.$$

On $[0, 6]$ the density is linear from 0 to h. Its slope is $h/6$, so for $0 \leq x \leq 6$

$$f_X(x) = \frac{h}{6}\, x = \frac{1}{12} \cdot \frac{x}{6} = \frac{x}{72}.$$

On $[6, 15]$, $f_X(x) = h = \frac{1}{12}$. Elsewhere $f_X(x) = 0$. Thus

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

2. probabilities. (i) $P(X \leq 6)$ is the area of the triangle:

$$P(X \leq 6) = \tfrac{1}{2} \cdot 6 \cdot h = 3h = 3 \cdot \frac{1}{12} = \frac{1}{4}.$$

(ii) $P(X \geq 6)$ is the area of the rectangle $[6, 15]$:

$$P(X \geq 6) = 9 \cdot h = 9 \cdot \frac{1}{12} = \frac{3}{4}.$$

(For a continuous distribution the inclusion of the endpoint does not matter.)

(iii) $P(4 \leq X \leq 12)$. Split at 6:

$$P(4 \leq X \leq 12) = \int_4^6 \frac{x}{72}\, dx + \int_6^{12} \frac{1}{12}\, dx.$$

Compute each term:

$$\int_4^6 \frac{x}{72}\,dx = \frac{x^2}{144}\bigg|_4^6 = \frac{36-16}{144} = \frac{20}{144} = \frac{5}{36},$$

$$\int_6^{12} \frac{1}{12}\,dx = \frac{1}{12}\cdot(12-6) = \frac{6}{12} = \frac{1}{2}.$$

So

$$P(4 \le X \le 12) = \frac{5}{36} + \frac{1}{2} = \frac{5+18}{36} = \frac{23}{36}.$$

(iv) $P(|X-6| \le 2) = P(4 \le X \le 8)$. *Split at* 6*:*

$$P(4 \le X \le 8) = \int_4^6 \frac{x}{72}\,dx + \int_6^8 \frac{1}{12}\,dx.$$

We already have $\int_4^6 \frac{x}{72}\,dx = \frac{5}{36}$. *The second integral is*

$$\int_6^8 \frac{1}{12}\,dx = \frac{1}{12}\cdot(8-6) = \frac{2}{12} = \frac{1}{6} = \frac{6}{36}.$$

Thus

$$P(|X-6| \le 2) = \frac{5}{36} + \frac{6}{36} = \frac{11}{36}.$$

3.3 Expectation of Random Variables

Definition 14 (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 the probability of rolling face k is proportional to $k+2$, for $k \in \{1,2,3,4\}$.

First, we find the probability mass function. The proportionality constants are:

$$1+2=3, \quad 2+2=4, \quad 3+2=5, \quad 4+2=6.$$

The sum is $3+4+5+6=18$.

Thus,

$$p_X(k) = \frac{k+2}{18}, \quad k \in \{1,2,3,4\}.$$

The expected value is

$$E(X) = \sum_{k=1}^{4} k \cdot p_X(k) = \sum_{k=1}^{4} k \cdot \frac{k+2}{18} = \frac{1}{18}\sum_{k=1}^{4} k(k+2).$$

Computing each term:

$$1(1+2)=3, \quad 2(2+2)=8, \quad 3(3+2)=15, \quad 4(4+2)=24.$$

Therefore,

$$E(X) = \frac{1}{18}(3+8+15+24) = \frac{50}{18} = \frac{25}{9} \approx 2.78.$$

Definition 15 (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 be a continuous random variable with pdf

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

The expected value is

$$E(X) = \int_0^6 x \cdot \frac{x}{72}\,dx + \int_6^{15} x \cdot \frac{1}{12}\,dx.$$

Computing the first integral:

$$\int_0^6 \frac{x^2}{72}\,dx = \frac{1}{72}\left[\frac{x^3}{3}\right]_0^6 = \frac{1}{72}\cdot\frac{216}{3} = \frac{216}{216} = 1.$$

Computing the second integral:

$$\int_6^{15} \frac{x}{12}\,dx = \frac{1}{12}\left[\frac{x^2}{2}\right]_6^{15} = \frac{1}{24}(225-36) = \frac{189}{24} = \frac{63}{8}.$$

Therefore,

$$E(X) = 1 + \frac{63}{8} = \frac{8+63}{8} = \frac{71}{8} = 8.875.$$

3.4 Variance and Standard Deviation

Definition 16 (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 17 (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:

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

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:**

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

(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 previously found $E(X) = 3.5$.

To find $\text{Var}(X)$, we first compute $E(X^2)$:

$$E(X^2) = \sum_{x=1}^{6} x^2 \cdot \frac{1}{6} = \frac{1}{6}(1+4+9+16+25+36) = \frac{91}{6}.$$

Then,

$$\text{Var}(X) = E(X^2) - [E(X)]^2 = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{91}{6} - \frac{49}{4} = \frac{182-147}{12} = \frac{35}{12} \approx 2.92.$$

The standard deviation is

$$\sigma_X = \sqrt{\frac{35}{12}} \approx 1.71.$$

3.5 Functions of Random Variables

Definition 18 (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. This case is unusual. If X is discrete, then $Y = g(X)$ is typically also discrete unless g involves randomization. In standard probability theory, a deterministic function of a discrete random variable remains discrete.

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

Example. Let X be uniformly distributed on $[0, 1]$ with pdf $f_X(x) = 1$ for $0 \le x \le 1$.

Define

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

Then Y is discrete with

$$P(Y = 0) = P(0 \le X < 0.5) = 0.5, \quad P(Y = 1) = P(0.5 \le X \le 1) = 0.5.$$

The pmf of Y is

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

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

Theorem 6 (Change of Variables Formula). Let X be a continuous random variable with pdf $f_X(x)$, and let $Y = g(X)$ where g is a differentiable, strictly monotonic function. Then the pdf of Y is

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

for y in the range of Y.

Example. Let X be uniformly distributed on $[0, 1]$, so $f_X(x) = 1$ for $0 \le x \le 1$.

Define $Y = 2X + 3$. Then Y ranges over $[3, 5]$.

Since $g(x) = 2x + 3$, we have $g^{-1}(y) = \frac{y-3}{2}$ and $\frac{d}{dy} g^{-1}(y) = \frac{1}{2}$.

Thus,

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

We can verify: $\int_3^5 \frac{1}{2}\, dy = \frac{1}{2}(5-3) = 1$.

Example. Let X have pdf $f_X(x) = 2x$ for $0 \le x \le 1$.

Define $Y = X^2$. Then Y ranges over $[0, 1]$.

Since $g(x) = x^2$, we have $g^{-1}(y) = \sqrt{y}$ (taking the positive root) and $\frac{d}{dy} g^{-1}(y) = \frac{1}{2\sqrt{y}}$.

Thus,

$$f_Y(y) = f_X(\sqrt{y}) \cdot \frac{1}{2\sqrt{y}} = 2\sqrt{y} \cdot \frac{1}{2\sqrt{y}} = 1, \quad 0 \le y \le 1.$$

So Y is uniformly distributed on $[0, 1]$.

3.6 Common Discrete Probability Distributions

3.6.1 Bernoulli Distribution

Definition 19 (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 20 (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 21 (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 22 (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 23 (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:** $\text{Var}(X) = n \cdot \dfrac{K}{N} \cdot \dfrac{N-K}{N} \cdot \dfrac{N-n}{N-1}$

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 24 (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