Statistics Pre-course
Probability

Ilaria Mozzetta

Tor Vergata University of Rome

September 2026

Part II Syllabus

1. Basic definitions and recap of Set Theory
2. Random Variables
3. Discrete Probability Distributions
4. Continuous Probability Distributions
5. Expected value and Variance of a Random Variable
6. Main Probability Distributions (Bernoulli, Binomial, Poisson, Uniform, Normal, Exponential,...)

Introduction to Probability

Why Probability?

- In descriptive statistics, we summarize observed data.
- In inferential statistics, we draw conclusions about a population from a sample.
- **Probability theory** serves as the mathematical bridge connecting observed samples to population characteristics under uncertainty.

Deterministic vs. Random Phenomena

- **Deterministic Process:** Outcome can be predicted with certainty under identical conditions (e.g., $E = mc^2$, free fall in vacuum).
- **Random Process:** Outcome cannot be predicted with certainty beforehand, even under identical conditions (e.g., coin toss, stock prices).

Random Experiments

Definition (Random Experiment)

A **random experiment** is an action or process that:

1. Has well-defined, identifiable possible outcomes.
2. Can be repeated under uniform conditions.
3. Has an outcome that is uncertain prior to execution.

Examples

- Tossing a fair coin 3 times.
- Rolling a six-sided die.
- Measuring the lifespan of a lightbulb (in hours).
- Recording daily returns of a financial asset.

Sample Space (Ω)

Definition (Sample Space)

The **sample space** (denoted by Ω or S) is the set of all possible basic outcomes of a random experiment.

Types of Sample Spaces

- **Finite Discrete:** $\Omega = \{H, T\}$ (Coin flip) or $\Omega = \{1, 2, 3, 4, 5, 6\}$ (Die roll).
- **Infinite Discrete:** $\Omega = \{0, 1, 2, 3, \dots\}$ (Number of phone calls received per hour).
- **Continuous:** $\Omega = \{t \in \mathbb{R} \mid t \geq 0\}$ (Lifespan of an electronic device).

Events

Definition (Event)

An **event** A is a subset of the sample space Ω ($A \subseteq \Omega$). An event occurs if the observed outcome belongs to A.

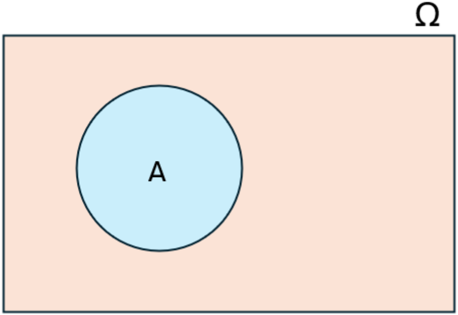
- **Elementary Event (ω):** A single outcome $\{\omega\} \in \Omega$.
- **Certain Event (Ω):** The outcome will always belong to Ω; $P(\Omega) = 1$.
- **Impossible Event ($\emptyset$):** The event containing no outcomes; $P(\emptyset) = 0$.

Example: Rolling a Die

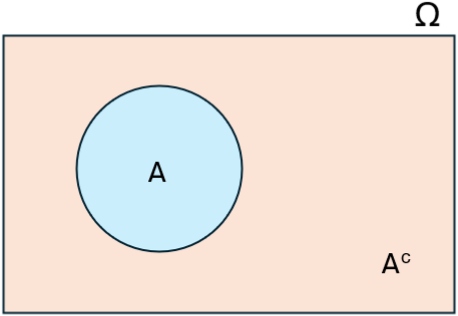
- Sample Space: $\Omega = \{1, 2, 3, 4, 5, 6\}$
- Event $A =$ "Even number": $A = \{2, 4, 6\}$
- Event $B =$ "Number greater than 4": $B = \{5, 6\}$

- Events are mathematically treated as sets.
- Sets can be finite (contain a finite number of objects) or infinite (consist of infinite elements).
- The cardinality of a given set is the number of objects that the set contains.
- E.g. if $E = \{1, 2, 3\}$ then the cardinality of E, denoted as $\#E = 3$.

Consider a generic set A included in an event space Ω:

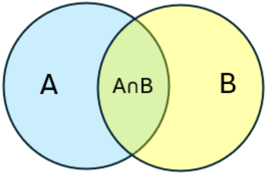


Complement: (A^c or $\bar{A}$) everything that is not in A



Example: $A =$ "the die returns an even number";
$A^c =$ "the die returns an odd number"

INTERSECTION: $(A \cap B)$ everything that is both in A and B

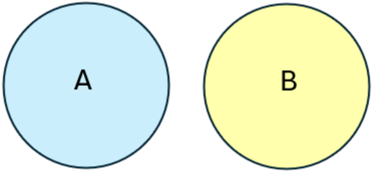


Example: $A =$ "the die returns an even number";
$B =$ "the die returns a number less than 5"
$\Rightarrow A \cap B = \{2, 4\}$

Recap of set theory: Disjoint sets

Disjoint sets: when A and B have no elements in common

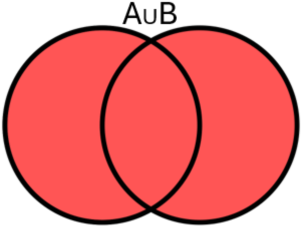
$$\Rightarrow A \cap B = \emptyset$$



Example: $A =$ "the die returns an even number";
$B =$ "the die returns a 5"
$\Rightarrow A$ and B are disjoint

Recap of set theory: Union

Union: $(A \cup B)$ everything that is either in A in B or both



Example: $A =$ "the die returns an even number";

$B =$ "the die returns a 5"

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

How do we define the probability?

Classical Approach

If all N outcomes in a finite sample space Ω are **equally likely**:

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

Frequentist Approach

The probability of event A is the long-run relative frequency of A in n identical, independent repetitions as $n \to \infty$:

$$P(A) = \lim_{n\to\infty} \frac{n_A}{n}$$

Subjective Interpretation

Probability measures an individual's degree of belief regarding the occurrence of an unrepeatable event (e.g., outcome of an election, economic growth rate).

Example

An urn contains 5 Red balls, 3 Blue balls, and 2 Green balls. A single ball is drawn at random.

- **Sample Space:** Total balls $\#\Omega = 5 + 3 + 2 = 10$. All balls are equally likely to be drawn.
- **Event R (Drawing a Red ball):** $\#R = 5$

$$\mathbb{P}(R) = \frac{\#R}{\#\Omega} = \frac{5}{10} = 0.50$$

- **Event $R \cup B$ (Drawing Red OR Blue):** $\#(R \cup B) = 5 + 3 = 8$

$$\mathbb{P}(R \cup B) = \frac{8}{10} = 0.80$$

- **Event G^c (Drawing NOT Green):**

$$\mathbb{P}(G^c) = 1 - \mathbb{P}(G) = 1 - \frac{2}{10} = 0.80$$

Probability Axioms (Kolmogorov)

Given a sample space Ω, a probability measure $\mathbb{P}$ satisfies:

1. **Non-negativity:** For any event A,

$$\mathbb{P}(A) \geq 0$$

2. **Normalization:**

$$\mathbb{P}(\Omega) = 1$$

3. **Countable Additivity:** For any sequence of mutually exclusive events $A_1, A_2, \ldots$ ($A_i \cap A_j = \emptyset$ for $i \neq j$):

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

Fundamental Properties of Probability

Direct consequences of Kolmogorov's axioms:

1. **Empty Set Rule:** $\mathbb{P}(\emptyset) = 0$
2. **Complement Rule:** $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$
3. **Boundedness:** $0 \leq \mathbb{P}(A) \leq 1$

The Addition Rule

For any two arbitrary events A and B:

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

If A and B are disjoint ($A \cap B = \emptyset$), then $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$.

Example

In a survey of 120 university students, 50 study French (F), 40 study Spanish (S), and 15 study both languages ($F \cap S$).

- Compute the probability that a student studies French but not Spanish.
- Compute the probability that a student studies either French or Spanish.
- Compute the probability that a student studies Spanish but not French.
- Compute the probability that a student studies neither language.

Solution

Total $\#\Omega = 120$, $\#F = 50$, $\#S = 40$, $\#(F \cap S) = 15$.

- **French but NOT Spanish ($\mathbb{P}(F \cap S^c)$):**

$$\mathbb{P}(F \cap S^c) = \frac{50 - 15}{120} = \frac{35}{120} \approx 0.292$$

- **French OR Spanish ($\mathbb{P}(F \cup S)$):**

$$\mathbb{P}(F \cup S) = \frac{50 + 40 - 15}{120} = \frac{75}{120} = 0.625$$

- **Spanish but NOT French ($\mathbb{P}(S \cap F^c)$):**

$$\mathbb{P}(S \cap F^c) = \frac{40 - 15}{120} = \frac{25}{120} \approx 0.208$$

- **Neither language ($\mathbb{P}((F \cup S)^c)$):**

$$\mathbb{P}((F \cup S)^c) = 1 - \mathbb{P}(F \cup S) = 1 - 0.625 = 0.375$$

Exercise

Draw a single card from a standard deck of 52 cards. Let $H =$ Heart and $F =$ Face card (Jack, Queen, King).

- Compute $\mathbb{P}(H \cap F^c)$ (Heart that is not a Face card).
- Compute $\mathbb{P}(H \cup F)$ (Heart or Face card).
- Compute $\mathbb{P}(F \cap H^c)$ (Face card that is not a Heart).
- Compute $\mathbb{P}(H \cap F)$ (Heart and Face card).

Conditional Probability

Sometimes we are interested in computing the probability of an event A given that B has already happened.

Conditional Probability

For any events A and B where $\mathbb{P}(B) > 0$:

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

Example

Out of 120 students, 50 study French (F) and 15 study both French and Spanish ($F \cap S$).

If we select a student and know they study French, what is the probability they also study Spanish?

$$\mathbb{P}(S \mid F) = \frac{\text{Students in both}}{\text{Total French students}} = \frac{15}{50} = 0.30$$

Law of total probability

Goal: Calculate the total probability of an event A by breaking it down into different scenarios.

Two-Scenario Formula

If an outcome A depends on whether scenario B or B^c happens:

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

Example

What is the overall probability that a student carries an umbrella (U)?

$$\mathbb{P}(U) = \mathbb{P}(U \mid \text{Rain})\mathbb{P}(\text{Rain}) + \mathbb{P}(U \mid \text{No Rain})\mathbb{P}(\text{No Rain})$$

- $\mathbb{P}(U)$ is simply a weighted average of carrying an umbrella across all possible weather conditions.

Bayes' Theorem

Goal: Reversing conditional probabilities. Working backward from an observed effect (A) to find the underlying cause (B).

Bayes' Formula

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

Why do we need it?

- We have seen the formula to compute how likely someone carries an umbrella when it rains: $\mathbb{P}(U \mid \text{Rain})$.
- But suppose you see a student walk into class with an umbrella (U). What is the probability that it is actually raining outside (Rain)? $\Rightarrow$ Bayes' Theorem

$\mathbb{P}(\text{Rain} \mid U)$

$$\mathbb{P}(\text{Rain} \mid U) = \frac{\mathbb{P}(U \mid \text{Rain})\mathbb{P}(\text{Rain})}{\mathbb{P}(U)}$$

Understanding the Parts

- $\mathbb{P}(\text{Rain})$: General chance of rain today.
- $\mathbb{P}(U \mid \text{Rain})$: How likely they grab an umbrella when it rains $\Rightarrow$ Conditional Probability.
- $\mathbb{P}(U)$: Total probability of seeing an umbrella $\Rightarrow$ TPL.

Example

Suppose we have two urns containing red (R) and black (B) balls:

- **Urn 1 (U_1):** Contains 8 red balls and 2 black balls (Total $= 10$).
- **Urn 2 (U_2):** Contains 4 red balls and 6 black balls (Total $= 10$).

We pick an urn at random with equal probability ($\mathbb{P}(U_1) = \mathbb{P}(U_2) = \frac{1}{2}$) and draw one ball. Compute:

1. The probability of drawing a red ball knowing that it comes from Urn 1: $\mathbb{P}(R \mid U_1)$
2. The overall probability of drawing a red ball: $\mathbb{P}(R)$
3. The probability that the ball came from Urn 1 knowing that it is red: $\mathbb{P}(U_1 \mid R)$

Example

Answer 1: Conditional Probability

$$\mathbb{P}(R \mid U_1) = \frac{8}{10} = \frac{4}{5} = 0.8$$

Answer 2: Law of Total Probability

$$\mathbb{P}(R) = \mathbb{P}(R \mid U_1)\mathbb{P}(U_1) + \mathbb{P}(R \mid U_2)\mathbb{P}(U_2)$$

$$\mathbb{P}(R) = \left(\frac{4}{5}\right) \cdot \left(\frac{1}{2}\right) + \left(\frac{2}{5}\right) \cdot \left(\frac{1}{2}\right) = \frac{4}{10} + \frac{2}{10} = \frac{6}{10} = \frac{3}{5} = 0.6$$

Answer 3: Bayes' Rule

$$\mathbb{P}(U_1 \mid R) = \frac{\mathbb{P}(R \mid U_1)\mathbb{P}(U_1)}{\mathbb{P}(R)}$$

$$\mathbb{P}(U_1 \mid R) = \frac{0.8 \cdot 0.5}{0.6} = \frac{0.4}{0.6} \approx 0.7$$

Independence

Two events A and B are **independent** if knowing that B occurred gives no information about the occurrence of A:

$$\mathbb{P}(A \mid B) = \mathbb{P}(A)$$

Implication

Substituting $\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$ into the condition above yields:

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

Warning: Disjoint $\neq$ Independent

Disjoint events with non-zero probability are **NEVER** independent, because if $A \cap B = \emptyset$, knowing B occurred guarantees A did **NOT** occur ($\mathbb{P}(A \mid B) = 0$).

Exercise

A factory uses two machines, Machine A (M_A) and Machine B (M_B), to manufacture parts:

- **Machine A (M_A):** Produces $\frac{2}{3}$ of all parts, with a 10% defect rate ($\frac{1}{10}$).
- **Machine B (M_B):** Produces $\frac{1}{3}$ of all parts, with a 40% defect rate ($\frac{4}{10}$).

An inspector selects one part at random from the total output. Compute:

Tasks:

1. The probability that the selected part is defective (D) knowing it came from Machine A: $\mathbb{P}(D \mid M_A)$.
2. The overall probability that a randomly chosen part is defective: $\mathbb{P}(D)$.
3. If a chosen part turns out to be defective, the probability that it came from Machine A: $\mathbb{P}(M_A \mid D)$.

Random Variables

Typically, we are not interested in a single outcome or events themselves but in a function of them.

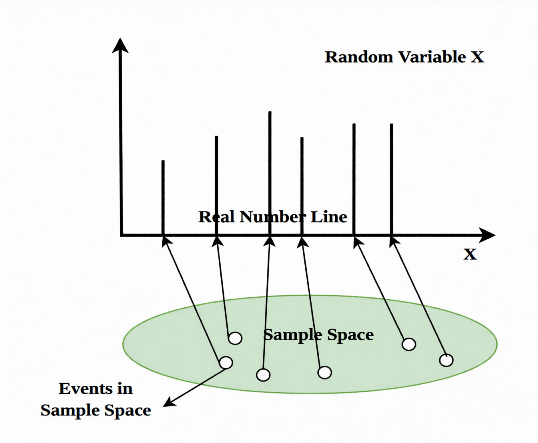
A random variable is any function from the event space to the real numbers.

Examples:

- Toss a coin three times and count the tails
- Roll two dice and sum the values on the faces

Random Variables

A random variable is any function from the event space to the real numbers.



X $\rightarrow$ the random variable: the random function before it is observed

x $\rightarrow$ a realization of the random variable: the number we observe

$\mathcal{X}$ $\rightarrow$ the support of the random variable: the set of the possible values that X can assume

Example: Toss a coin three times.

- Random variable $X =$ number of heads
- Support $\mathcal{X} = \{0, 1, 2, 3\}$
- Realization $x = 2$ (or 0, or 1 or 3)

Can we summarize the probabilities of all the outcomes? Yes, we use the distribution of the random variable.

Types of Random Variables

Depending on the nature of the state space (the set of possible values), RVs are classified into two main types:

1. **Discrete Random Variables:** Can take a finite or countably infinite set of values (e.g., number of defaults in a bank's portfolio).
2. **Continuous Random Variables:** Can take any value within a given interval on the real line (e.g., GDP growth rate, exact time until a financial crash).

Discrete Random Variables

- A random variable X is **discrete** if its sample space (the set of all possible values Ω) is **countable**, meaning it is either finite or countably infinite ($S_X = \{x_1, x_2, \dots\} \subset \mathbb{R}$).
- Examples:
  - Number of goals scored in a soccer match $(0, 1, 2, 3, \dots)$.
  - Number of heads when flipping 3 coins $(0, 1, 2,$ or $3)$.
  - Outcome of rolling a 6-sided die $(1, 2, 3, 4, 5,$ or $6)$.
  - Number of unread text messages on your phone $(0, 1, 2, \dots)$.

Probability distribution of discrete rvs

Idea

For a **discrete** random variable, its probability distribution is simply a list, table, or rule that tells us the exact probability of every single possible outcome.

- **In simple words:** Think of a total probability of 1. The distribution tells you exactly how that 1 is sliced up and handed out to each specific outcome.
- Imagine a simple table:
  - **Outcome (x):** 0 goal, 1 goal, 2 goal.
  - **Probability $P(X = x)$:** 0.5 chance, 0.4 chance, 0.1 chance.
- As long as the probabilities add up to 1 and none are negative, that table *is* the probability distribution!

In probability this is called Probability Mass Function (PMF)

Probability Mass Function (PMF)

Definition

The **probability mass function** (or PMF) of a discrete RV X describes the probability that X takes a specific value x.

$$p_x = P(X = x)$$

- It provides the distribution of the total probability across the different possible values of X.

x	x_1	x_2	$\dots$	x_k
$P(X = x)$	p_1	p_2	$\dots$	p_k

Properties of the PMF

To be a valid probability function, p_x must satisfy two strict conditions:

1. **Non-negativity:** $p_x \geq 0$ for all x in the state space. (Probabilities cannot be negative).
2. **Sum to One:** $\sum_x p_x = 1$. (Something must happen, so the sum of all mutually exclusive probabilities is 1).

Cumulative Distribution Function

Definition

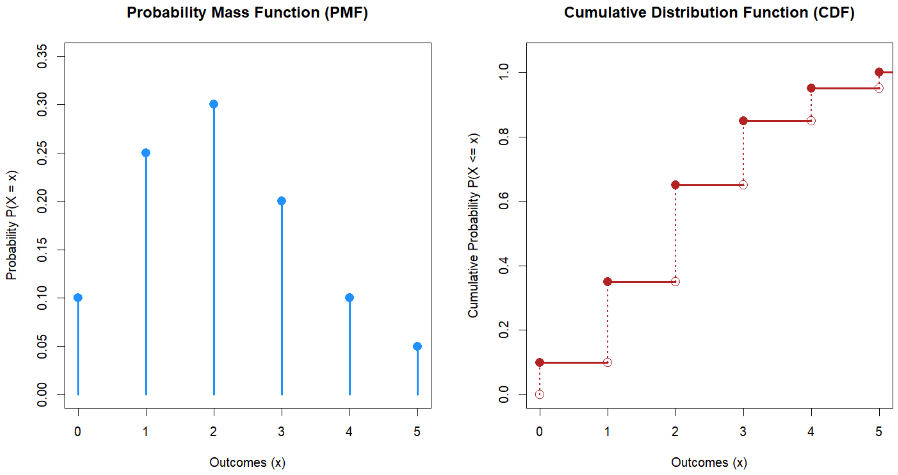
The **Cumulative Distribution Function (CDF)** $F(x)$ represents the probability that the random variable X takes a value less than or equal to x.

$$F(x) = P(X \leq x) = \sum_{t \leq x} p_t$$

- While the PMF gives the probability at a specific point, the CDF accumulates probability up to that point.

Properties of the CDF

- **Bounds:** $0 \leq F(x) \leq 1$ for any x.
- **Non-decreasing:** If $a < b$, then $F(a) \leq F(b)$.
- **Limits:**
  - $\lim_{x \to -\infty} F(x) = 0$
  - $\lim_{x \to \infty} F(x) = 1$
- For discrete RVs, the CDF is a **step function**.


Probability Mass Function (PMF)
Probability P(X = x)
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0
1
2
3
4
5
Outcomes (x)
Cumulative Distribution Function (CDF)
Cumulative Probability P(X <= x)
0.0
0.2
0.4
0.6
0.8
1.0
0
1
2
3
4
5
Outcomes (x)

Example

- A fair coin is tossed twice. The sample space of all equally likely outcomes is:
$$\Omega = \{\mathsf{HH}, \mathsf{HT}, \mathsf{TH}, \mathsf{TT}\}$$
- Let X denote the total number of Heads obtained. Construct the probability distribution of X.
- Find the probability of getting at least one Head.

Example

1. **Possible Values of X:**
   - 0 Heads (TT): $X = 0$
   - 1 Head (HT, TH): $X = 1$
   - 2 Heads (HH): $X = 2$

2. **Probability Distribution Table of X:**

x_i	0	1	2
$P(X = x_i)$	$\frac{1}{4} = 0.25$	$\frac{2}{4} = 0.50$	$\frac{1}{4} = 0.25$

3. **Probability of At Least One Head ($X \geq 1$):**

$$P(X \geq 1) = P(X = 1) + P(X = 2) = \frac{2}{4} + \frac{1}{4} = \frac{3}{4} = 0.75 \quad (75\%)$$

Exercise

Suppose a random variable X has the following probability distribution:

x	2	5	6	8	12	15	20
$P(X = x)$	0.15	0.08	0.22	0.18	0.14	0.17	?

- Fill in the missing value for $P(X = 20)$.
- Write down the cumulative distribution function $F_X(x)$.
- Evaluate the following probabilities:
  - X is at least 8 ($P(X \geq 8)$)
  - X is more than 8 ($P(X > 8)$)
  - X is less than 6 ($P(X < 6)$)

Continuous Random Variables

- **Uncountable Support:** A continuous random variable X can take any real value within an interval (e.g., $[a, b]$ or $\mathbb{R}$).
  - Examples: Exact weight, temperature, or elapsed waiting time.
- Since there are infinitely many points, the probability of taking any single exact point is zero: $P(X = x) = 0$ (Later we will see why).
- Therefore, we cannot use a PMF. Instead, we use a **Probability Density Function (PDF)**.

Probability Density Function (PDF)

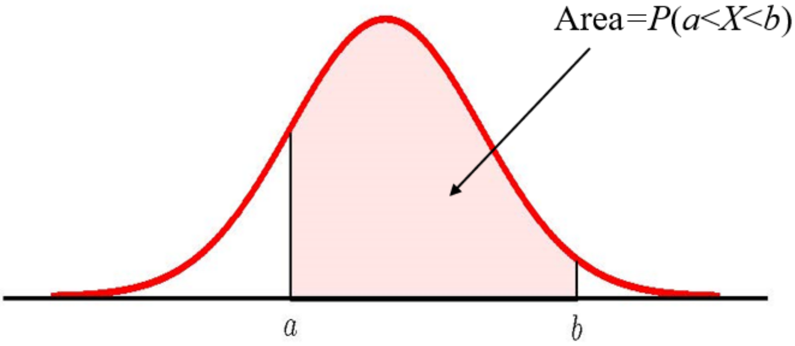
Definition

The function $f(x)$ is a PDF for a continuous RV X if:

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

- Probabilities are represented by **areas** under the density curve over an interval.
- $P(a \leq X \leq b) = \int_{a}^{b} f(x)dx$.

Probability as areas



Continuous CDF

- The Cumulative Distribution Function for a continuous RV is defined as:
$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt$$
- It accumulates the area under the PDF up to the point x.
- Fundamental Theorem of Calculus links them: $f(x) = \frac{d}{dx}F(x)$.

Exercise

Let X be a continuous random variable with the following probability density function

$$f_X(x) = \begin{cases} cx(1-x) & \text{if } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

- determine c such that this is a proper probability density function
- evaluate $\mathbb{P}(X = 0.5)$
- evaluate $\mathbb{P}(X \leq \frac{1}{2})$

Let Y be a continuous random variable with the following cumulative distribution function

$$F_Y(y) = \begin{cases} 1 & \text{if } y \geq 1 \\ 3y^2 - 2y^3 & \text{if } 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

evaluate $\mathbb{P}(Y \leq \frac{1}{2})$ using $F_Y(y)$

Discrete vs. Continuous Random Variables

Discrete RVs

PMF:

$$p_X(x) = P(X = x)$$

Probability over Set $A = \{x_1, \dots, x_k\}$**:**

$$P(X \in A) = \sum_{i=1}^{k} p_X(x_i)$$

Total Probability Axiom:

$$\sum_x p_X(x) = 1$$

Continuous RVs

PDF:

$$f_X(x) \quad \text{s.t } P(X \in A) = \int_A f_X(x)\, dx$$

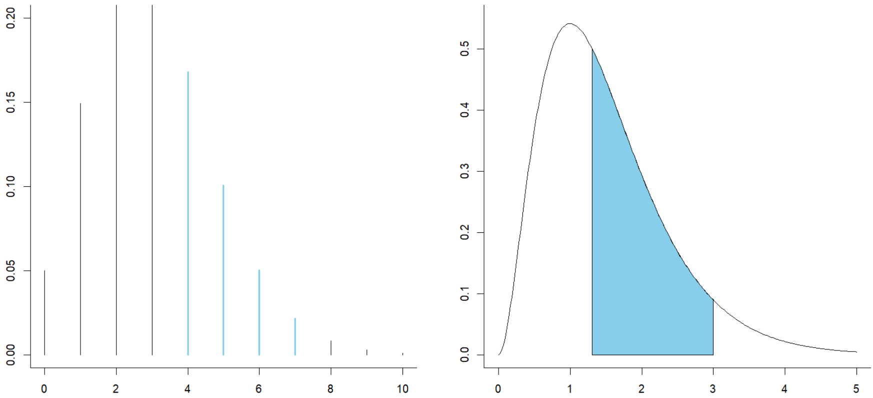
Probability over Interval $A = [a, b] : P(a \leq X \leq b) =$

$$\int_a^b f_X(x)\, dx = F_X(b) - F_X(a)$$

Total Probability Axiom:

$$\int_{-\infty}^{\infty} f_X(x)\, dx = 1$$

Discrete vs. Continuous Random Variables



Summary Measures: Mode and Median

While the probability distribution of a random variable X fully characterizes it, summary measures provide quick, intuitive insights into its central features.

Mode (x_m): The value that is "most likely" (maximizes probability mass or density).

Median (m): The value that divides the probability mass in half:

$$F_X(m) = P(X \leq m) = 0.5 \quad \text{(for continuous } X\text{)}$$

(General definition: $P(X \leq m) \geq 0.5$ and $P(X \geq m) \geq 0.5$)

Expected Value

Beyond mode and median, the single most fundamental summary measure of a distribution is its Expected Value.

What is the Expected Value (EV)?

The Expected Value, also called the **first moment**, is the average of the values in the support of X, weighted by the probability of each outcome.

It represents the long-run average of observed outcomes over many repeated trials.

EV for Discrete RVs

$$\mathbb{E}[X] = \sum_{x \in \mathcal{X}} x \, p_X(x)$$

EV for Continuous RVs

$$\mathbb{E}[X] = \int_{x \in \mathcal{X}} x \, f_X(x) \, dx$$

Properties of Expected Value

- $\mathbb{E}[c] = c$ for any constant c
- $\mathbb{E}[\mathbb{E}[X]] = \mathbb{E}[X]$ *(since $\mathbb{E}[X]$ is itself a constant)*
- $\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$ *(Linearity of expectation)*
- $\mathbb{E}[X - \mathbb{E}[X]] = 0$
- $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$

Expectation of a Function $g(X)$

For any real-valued function $g(X)$ whose expectation exists:

$$\text{Discrete: } \mathbb{E}[g(X)] = \sum_{x \in \mathcal{X}} g(x) p_X(x)$$

$$\text{Continuous: } \mathbb{E}[g(X)] = \int_{\mathcal{X}} g(x) f_X(x)\, dx$$

Measuring Variability

The Expected Value gives the center of a distribution, but tells us nothing about its dispersion (spread).

Example: Two investment plans with the exact same expected payout ($\mathbb{E}[X] = \$100$). We prefer the one with lower risk/variability.

Why do simple metrics fail?

- **Average deviation from the mean:**

$$\mathbb{E}[X - \mathbb{E}[X]] = 0 \quad \text{(Positive and negative errors cancel out!)}$$

- **Mean absolute deviation:**

$$\mathbb{E}[|X - \mathbb{E}[X]|] \quad \text{(Mathematically/computationally hard to differentiate)}$$

Variance

The **Variance** solves this by squaring the deviations, measuring spread around the mean:

$$\mathbb{V}[X] = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right]$$

Variance (Discrete RVs)

$$\mathbb{V}[X] = \sum_{x \in \mathcal{X}} (x - \mathbb{E}[X])^2 p_X(x)$$

Variance (Continuous RVs)

$$\mathbb{V}[X] = \int_{\mathcal{X}} (x - \mathbb{E}[X])^2 f_X(x)\, dx$$

Properties of Variance

- **Non-negativity:** $\mathbb{V}[X] \geq 0$, and $\mathbb{V}[X] = 0$ if and only if X is a constant.
- **Standard Deviation:** $SD(X) = \sqrt{\mathbb{V}[X]}$ roughly describes how far the values of the random variable fall, on average, from the expected value of the distribution.
- **Linear Transformation:** Variance is shift-invariant (insensitive to location shifts) but scales quadratically:

$$\mathbb{V}[aX + b] = a^2\mathbb{V}[X]$$

Note: The constant b disappears because $\mathbb{V}[b] = 0$—a fixed constant has zero variability.

A Computational Formula for Variance

In practice, computing variance via the definition could be difficult. We use:

$$\mathbb{V}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

where $\mathbb{E}[X^2]$ is the **Second Moment** of X, computed as:

Second Moment (Discrete)

$$\mathbb{E}[X^2] = \sum_{x \in \mathcal{X}} x^2 p_X(x)$$

Second Moment (Continuous)

$$\mathbb{E}[X^2] = \int_{\mathcal{X}} x^2 f_X(x)\, dx$$

Example: Computing Expected Value and Variance

Suppose X represents the number of service requests a tech support agent receives per hour, with the following probability distribution:

x	0	1	2	3
$P(X = x)$	0.10	0.40	0.30	0.20

Tasks:

1. Compute the Expected Value $\mathbb{E}[X]$.
2. Compute the Second Moment $\mathbb{E}[X^2]$.
3. Find the Variance $\mathbb{V}[X]$ and Standard Deviation $SD(X)$.

Example

1. Expected Value $\mathbb{E}[X]$:

$$\mathbb{E}[X] = \sum x \cdot P(X = x)$$

$$\mathbb{E}[X] = (0)(0.10) + (1)(0.40) + (2)(0.30) + (3)(0.20)$$

$$\mathbb{E}[X] = 0 + 0.40 + 0.60 + 0.60 = \mathbf{1.60} \text{ requests/hour}$$

2. Second Moment $\mathbb{E}[X^2]$:

$$\mathbb{E}[X^2] = \sum x^2 \cdot P(X = x)$$

$$\mathbb{E}[X^2] = (0^2)(0.10) + (1^2)(0.40) + (2^2)(0.30) + (3^2)(0.20)$$

$$\mathbb{E}[X^2] = 0 + 0.40 + 1.20 + 1.80 = \mathbf{3.40}$$

3. Variance $\mathbb{V}[X]$ and Standard Deviation $SD(X)$:

$$\mathbb{V}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = 3.40 - (1.60)^2 = 3.40 - 2.56 = \mathbf{0.84}$$

$$SD(X) = \sqrt{\mathbb{V}[X]} = \sqrt{0.84} \approx \mathbf{0.9165}$$

Exercise

Suppose the waiting time X (in minutes) for an elevator in an office building is a continuous random variable with probability density function (PDF):

$$f_X(x) = \begin{cases} \frac{3}{4}x(2-x) & \text{for } 0 \leq x \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

Tasks:

1. Verify that $f_X(x)$ is a valid Probability Density Function.
2. Calculate the Expected Value $\mathbb{E}[X]$.
3. Calculate the Second Moment $\mathbb{E}[X^2]$.
4. Find the Variance $\mathbb{V}[X]$ and Standard Deviation $SD(X)$.

What if we have more variables?

From One Variable to Two Variables

- **Univariate:** Analyzes a single random variable X in isolation.
- **Bivariate:** Analyzes two random variables (X, Y) recorded simultaneously on the same sample space.

The Joint Function $f(x, y)$

The function $f(x, y)$ specifies how X and Y behave together:

- **Discrete (Joint PMF):** $f(x, y) = P(X = x, Y = y)$
- **Continuous (Joint PDF):** A surface where probability equals the volume under $f(x, y)$ over a region.

Two Fundamental Rules

1. $f(x, y) \geq 0$ for all (x, y) (Probabilities are non-negative)
2. Total probability equals 1: $\sum_x \sum_y f(x, y) = 1$ or $\int \int f(x, y)\, dx\, dy = 1$

Breaking Down Joint Distributions

Once we have a joint distribution $f(x, y)$, we can analyze each variable individually or check if they affect one another:

Marginal Distribution

This is simply the probability distribution of a single variable on its own, completely ignoring the other variable $f(x), f(y)$.

Definition of Independence

X and Y are **independent** if knowing the value of one variable gives zero information about the other.
Mathematically, two variables are independent if their joint probability equals the product of their individual (marginal) probabilities:

$$f(x, y) = f_X(x) \cdot f_Y(y) \quad \text{for all } (x, y)$$

Example

Two-Way Table Representation

When X and Y take few discrete values, $f(x, y)$ is presented in a two-way table where each cell represents the probability of that joint outcome pair (x, y).

X	Y			$f_X(x)$
	70	85	100	
13	0.2	0.2	0.0	0.4
18	0.0	0.1	0.5	0.6
$f_Y(y)$	0.2	0.3	0.5	1.0

Measuring Relationship Between Variables

Covariance (σ_{XY})

Measures the **direction** of linear co-movement using expected values:

$$\sigma_{XY} = E[(X - E[X])(Y - E[Y])] = E[XY] - E[X]E[Y]$$

- $\sigma_{XY} > 0$: Variables tend to move in the same direction.
- $\sigma_{XY} < 0$: Variables tend to move in opposite directions.

Correlation Coefficient (ρ)

Measures the **strength** of the linear relationship on a standard scale $[-1, 1]$:

$$\rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

If X and Y are independent $\Rightarrow E[XY] = E[X]E[Y] \Rightarrow \sigma_{XY} = 0$ and $\rho = 0$.

Expected Value of the Product $E[XY]$

Definition

The expected value of $X \cdot Y$ is calculated by multiplying each outcome pair (x, y) by its joint probability $f(x, y)$:

- **Discrete Case:**

$$E[XY] = \sum_{x} \sum_{y} x \cdot y \cdot f(x, y)$$

- **Continuous Case:**

$$E[XY] = \iint x \cdot y \cdot f(x, y)\, dx\, dy$$

Example

Consider two discrete random variables X and Y with the following joint probability distribution $f(x, y)$:

$f(x,y)$	$Y=0$	$Y=1$
$X=0$	0.2	0.1
$X=1$	0.3	0.4

Task

Calculate the covariance σ_{XY} between X and Y using the formula:

$$\sigma_{XY} = E[XY] - E[X]E[Y]$$

Example

Find marginal probabilities and individual expected values

- $f_X(0) = 0.2 + 0.1 = 0.3$ and $f_X(1) = 0.3 + 0.4 = 0.7$

$$\Longrightarrow E[X] = (0 \cdot 0.3) + (1 \cdot 0.7) = 0.7$$

- $f_Y(0) = 0.2 + 0.3 = 0.5$ and $f_Y(1) = 0.1 + 0.4 = 0.5$

$$\Longrightarrow E[Y] = (0 \cdot 0.5) + (1 \cdot 0.5) = 0.5$$

Calculate $E[XY]$

$$E[XY] = \sum_x \sum_y x \cdot y \cdot f(x,y)$$

$$E[XY] = (0 \cdot 0)(0.2) + (0 \cdot 1)(0.1) + (1 \cdot 0)(0.3) + (1 \cdot 1)(0.4) = 0.4$$

Calculate Covariance σ_{XY}

$$\sigma_{XY} = E[XY] - E[X]E[Y] = 0.4 - (0.7 \cdot 0.5) = 0.4 - 0.35 = +0.05$$

Exercise

Consider two discrete random variables X and Y with the following joint probability distribution:

$f(x,y)$	$Y=0$	$Y=1$
$X=1$	0.1	0.3
$X=2$	0.4	0.2

Questions

1. Calculate the marginal probability distributions of X and Y.
2. Compute $E[X]$, $E[Y]$, and $E[XY]$.
3. Compute the covariance $\sigma_{XY} = E[XY] - E[X]E[Y]$.
4. Are X and Y independent? Briefly explain why or why not.