Probability

5. Random Variables

Ewen Gallic

Bachelor in Economics
International Program in Economics and Management
Academic Year 2026

amU FEG
Aix Marseille Université

Foreword

- Random experiments generate uncertain outcomes.
- Outcomes themselves are often complex and difficult to analyse.
- In this chapter, we will define **random variables** which provide a numerical description of an outcome.
- Random variables allow us to:
  - describe uncertainty quantitatively,
  - assign probabilities to numerical events,
  - summarize random phenomena relevant for decision-making.
- This chapter is based on:
  - Mathieu Faure's (Aix Marseille Univ.) lecture slides, kindly shared,
  - Chapter 2 of Hansen (2022),
  - Chapter 3 of Blitzstein and Hwang (2019).

1. Random Variables

1.1. Definition of a Random Variable

From Random Experiments to Numbers

- Consider a **random experiment**.
  - For example, a store opens for the day and customers arrive.
  - Each customer may or may not buy the latest volume of *One Piece*.
- An outcome $\omega \in \Omega$ occurs.
  - ω describes the entire sequence of customers and their purchases.
  - Different days correspond to different outcomes.
- Outcomes themselves are often not convenient to work with.
  - The store does not need the full description of the day.
  - What matters is a numerical summary.
- We therefore associate a **numerical value** to each outcome.
  - Number of copies of the latest *One Piece* volume demanded that day.
  - This number helps decide how many books to keep in stock.
- The numerical description of the outcome of a random experiment will be what we call a **random variable**.

Definition

Definition 1. Random Variable

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. A ***random variable*** *is a function*

$$X : \Omega \longrightarrow \mathbb{R},$$

which assigns a real number to each outcome $\omega \in \Omega$.

- For each outcome ω, the value $X(\omega)$ is called a **realization** of X.
- The set of all possible realizations is

$$\mathcal{D}_X = \{X(\omega) : \omega \in \Omega\} \subset \mathbb{R}.$$

- $\mathcal{D}_X$ is called the **support** (or **set of possible values**) of the random variable X.
- Before the realization of X, the result is unknown, it is **random**.

Example

Example: Bookstore

In the example of the bookstore: We define a random variable

$$X(\omega) = \textit{number of copies of the latest volume sold during the day.}$$

- *The supports of X is*
$$\mathcal{D}_X = \{0, 1, 2, \ldots\}$$

Tossing a Coin Three Times

- *A coin is tossed three times. Before tossing the coin, the exact outcome is unknown.*
- *The natural sample space associated with this experiment is:*

$$\Omega = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}.$$

- Each element of Ω fully describes the outcome of the experiment.
- We associate a numerical value with each outcome of the experiment.
- Define the random variable. For example:

$$X(\omega) = \text{number of times Head appears.}$$

  - Hence,
  $$X(HTH) = 2, \qquad X(TTT) = 0, \qquad X(HHH) = 3.$$
  - The possible values of X are $\{0, 1, 2, 3\}$.

Constant Random Variables and Indicator Random Variables

- The simplest example of a random variable is a **constant** random variable:
  - $X(\omega)$ is identically equal to some fixed value $x_0 \in \mathbb{R}$,
  - for all $\omega \in \Omega$.
- Another fundamental example is an **indicator random variable**, which can only take the values 0 or 1.

Definition 2. Indicator Random Variable

*Let $A \in \mathcal{A}$ be an event. The **indicator** of A, denoted by $\mathbf{1}_A$ (or $\mathbb{I}_A$), is the random variable defined by*

$$\mathbf{1}_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A, \\ 0 & \text{if } \omega \notin A. \end{cases}$$

The indicator of A takes the value 1 if the event A occurs, and 0 otherwise.

1.2. Distribution of a Random Variable

Distribution of a Random Variable

- A random variable X is characterized by the probabilities with which it takes different values.
- The collection of these probabilities is called the **distribution of the random variable X**.

Definition 3. Distribution of a Random Variable

Let X be a random variable with set of possible values $\mathcal{D}_X$. The ***distribution*** *of X is defined by the probabilities*

$$\mathbb{P}(X \in B), \qquad B \subset \mathcal{D}_X.$$

We denote by $\mathbb{P}_X$ the distribution of X, defined by

$$\mathbb{P}_X(B) = \mathbb{P}(X \in B).$$

Example: Tossing a Coin Three Times

Recall the example where a coin is tossed three times, and $X =$ number of times Head appears.

- The possible values of X are $\mathcal{D}_X = \{0, 1, 2, 3\}$. To determine the distribution of X, we compute the probability that X takes each of these values:
$$\mathbb{P}_X(\{0\}) = \mathbb{P}(X = 0) = \mathbb{P}(TTT) = \frac{1}{8}.$$

  Similarly, we obtain:
$$\mathbb{P}_X(\{1\}) = \mathbb{P}(\text{exactly one Head}) = \mathbb{P}(\{(HTT), (THT), (TTH)\}) = \frac{3}{8},$$
$$\mathbb{P}_X(\{2\}) = \mathbb{P}(\text{exactly two Heads}) = \mathbb{P}(\{(THH), (HTH), (HHT)\}) = \frac{3}{8},$$
$$\mathbb{P}_X(\{3\}) = \mathbb{P}(HHH) = \frac{1}{8}.$$

Technical Remark

- *In general, the definition of $\mathbb{P}(X \in B)$ requires a technical condition.*
- *The set*

$$\{X \in B\}$$

  must be an ***event****, that is, an element of the sigma-algebra $\mathcal{A}$ (as stated in the definition).*
- *This is important, since the probability $\mathbb{P}$ is only defined for events in $\mathcal{A}$.*
- *Throughout this course, we will assume that this condition always holds.*
- *In the language of measure theory, this means that a random variable is* ***measurable****.*

Discrete and Continuous Random Variables

- Random variables are classified according to their **set of possible values** (the support $\mathcal{D}_X$).
- A random variable X is called **discrete** if its support $\mathcal{D}_X$ is **finite or countable**.
  - Examples:

$$\mathcal{D}_X \text{ finite}, \quad \mathcal{D}_X = \mathbb{N}, \quad \mathcal{D}_X = \mathbb{Z}.$$

- A random variable X is called **continuous** if it can take values in an interval of $\mathbb{R}$.
  - Examples:

$$\mathcal{D}_X = \mathbb{R}, \quad \mathcal{D}_X = [0, +\infty), \quad \mathcal{D}_X = [a, b],\ a < b.$$

2. Discrete Random Variables

Discrete Random Variables

Definition 4. Discrete Random Variable

Let X be a random variable taking values in a ***discrete subset*** *$\mathcal{D}_X \subset \mathbb{R}$ (that is, $\mathcal{D}_X$ is finite or countable).*
Since $\mathcal{D}_X$ is discrete, the distribution of X is fully characterized by the collection $\{p_x\}_{x \in \mathcal{D}_X}$, where

$$p_x = \mathbb{P}(X = x).$$

- By definition of a probability distribution, we necessarily have:

$$\sum_{x \in \mathcal{D}_X} p_x = 1.$$

- In the case $\mathcal{D}_X = \mathbb{N}$, this means that the series with general term p_n is convergent.

This characterization does not apply to continuous random variables.

Remark: What Does $\mathbb{P}(X = x)$ Mean?

*When we write $\mathbb{P}(X = x)$, the expression $X = x$ denotes an **event**.*
Formally, the event $\{X = x\}$ is defined as

$$\{X = x\} = \{\omega \in \Omega : X(\omega) = x\}.$$

This set contains all outcomes ω for which the r.v. X takes the value x.

Example: Three Coin Tosses

Let X be the number of Heads obtained when tossing a coin three times.

- The event $\{X = 1\}$ consists of the outcomes

$$\{HTT, THT, TTH\}.$$

- This set is a subset of the sample space Ω.
- Therefore, $\{X = 1\}$ is an event, and $\mathbb{P}(X = 1)$ is well-defined.

Example of a Random Variable: Sum of Two Dice

Two fair six-sided dice, numbered from 1 to 6, are rolled. Define the random variable X as the ***sum of the two dice****.*

- The sample space is $\Omega = \{(1,1),(1,2),\ldots,(6,5),(6,6)\}$.
- The cardinality of the sample space is $|\Omega| = 36$.
- The support of X is $\mathcal{D}_X = \{2,3,\ldots,12\}$.
- The distribution of X is given by the probabilities $\{\mathbb{P}(X=i)\}_{i=2,\ldots,12}$.
- Since the dice are fair, all outcomes in Ω are equally likely.
- For example:
$$\mathbb{P}(X=2) = \mathbb{P}(\text{both dice show } 1) = \frac{1}{36}.$$
- Similarly:
$$\mathbb{P}(X=7) = \mathbb{P}(\{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}) = \frac{6}{36} = \frac{1}{6}.$$

2.1. Probability Mass Function

Probability Mass Function

Definition 5. Probability Mass Function

*Let X be a discrete random variable with support $\mathcal{D}_X$. The **probability mass function** (PMF) of X is the function*

$$p_X : \mathbb{R} \longrightarrow [0, 1], \qquad p_X(x) = \mathbb{P}(X = x).$$

Probability Mass Function: One Die

A fair six-sided die is rolled once. Define the random variable X as the ***number shown on the die****.*

- The sample space is $\Omega = \{1, 2, 3, 4, 5, 6\}$.
- The support of X is

$$\mathcal{D}_X = \{1, 2, 3, 4, 5, 6\}.$$

- Since the die is fair, each outcome is equally likely.
- The PMF: for each $x \in \mathcal{D}_X$,

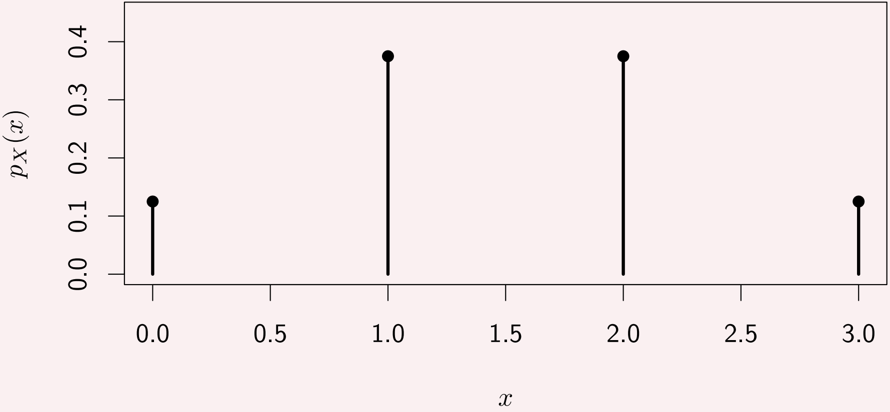
$$p_x = \mathbb{P}(X = x) = \frac{1}{6}.$$

Probability Mass Function: Three Coins

Three fair coins are tossed. Define the random variable X as the ***number of Heads****.*

- The sample space is $\Omega = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$.
- The support of X is
$$\mathcal{D}_X = \{0, 1, 2, 3\}.$$
- The PMF:
$$\begin{aligned}
p_0 &= \mathbb{P}(X = 0) = \mathbb{P}(\{TTT\}) = 1/8,\\
p_1 &= \mathbb{P}(X = 1) = \mathbb{P}(\{HTT, THT, TTH\}) = 3/8,\\
p_2 &= \mathbb{P}(X = 2) = \mathbb{P}(\{HHT, HTH, THH\}) = 3/8,\\
p_3 &= \mathbb{P}(X = 3) = \mathbb{P}(\{HHH\}) = 1/8.
\end{aligned}$$

Plot of the PMF



Probability Mass Function of X

Properties of the PMF

Proposition 1. Basic Properties of the PMF

1. ***Nonnegativity.*** *For all* $x \in \mathbb{R}$, $p_X(x) \geq 0$. *Moreover,*

$$p_X(x) > 0 \text{ if } x \in \mathcal{D}_X, \qquad p_X(x) = 0 \text{ if } x \notin \mathcal{D}_X.$$

2. ***Normalization.*** *The PMF satisfies*

$$\sum_{x \in \mathcal{D}_X} p_X(x) = 1.$$

Proof of the Properties

1. **Nonnegativity.** By definition, $p_X(x) = \mathbb{P}(X = x)$. Since probabilities are always nonnegative, we have $p_X(x) \geq 0$ for all x. Moreover, if $x \notin \mathcal{D}_X$, the event $\{X = x\}$ is empty, hence $p_X(x) = 0$.

2. **Normalization.** Since X is discrete, its support can be written as

$$\mathcal{D}_X = \{x_1, x_2, x_3, \ldots\}.$$

The events $\{X = x_i\}$ are pairwise disjoint and their union is the whole sample space:

$$\bigcup_{i=1}^{\infty} \{X = x_i\} = \Omega.$$

Therefore, by countable additivity of probability,

$$\sum_{i=1}^{\infty} \mathbb{P}(X = x_i) = \mathbb{P}(\Omega) = 1.$$

2.2. Distribution Function

Distribution Function

Definition 6. Distribution Function

*The **distribution function** of a random variable X is the function*

$$F_X : \mathbb{R} \longrightarrow [0,1], \qquad F_X(x) = \mathbb{P}(X \leq x),$$

that is, the probability of the event $\{X \leq x\}$.

- The function F_X is also called the **cumulative distribution function** (CDF).
- The subscript X indicates the random variable considered.
- When there is no ambiguity, we may simply write F instead of F_X.

Remark

Assume for a moment that $\mathcal{D}_X = \mathbb{N}$. In this case, the fact that the distribution function fully characterizes the distribution of X follows immediately from the identity

$$\mathbb{P}(X = x) = \mathbb{P}(X \leq x) - \mathbb{P}(X \leq x - 1) = F_X(x) - F_X(x - 1).$$

- *Indeed, the event $\{X \leq x\}$ can be written as:*

$$\{X \leq x\} = \{X = 0\} \cup \cdots \cup \{X = x - 1\} \cup \{X = x\}.$$

- *Similarly,*

$$\{X \leq x - 1\} = \{X = 0\} \cup \cdots \cup \{X = x - 1\}.$$

- *Since the events $\{X = k\}$ are disjoint,*

$$\mathbb{P}(X \leq x) - \mathbb{P}(X \leq x - 1) = \mathbb{P}(X = x).$$

The same principle applies for any discrete support $\mathcal{D}_X$.

Properties of the Distribution Function

Proposition 2. Properties of the Distribution Function

Let F_X be the distribution function (CDF) of a random variable X. Then:

1. ***Monotonicity:*** *The function F_X is non-decreasing, that is,*

$$x \leq y \;\Rightarrow\; F_X(x) \leq F_X(y).$$

2. ***Limit at $-\infty$:***

$$\lim_{x \to -\infty} F_X(x) = 0.$$

3. ***Limit at $+\infty$:***

$$\lim_{x \to +\infty} F_X(x) = 1.$$

Proof: Monotonicity

- Let $F_X(x) = \mathbb{P}(X \leq x)$.
- If $x \leq y$, then
$$\{X \leq x\} \subset \{X \leq y\}.$$
- By monotonicity of probability,
$$F_X(x) = \mathbb{P}(X \leq x) \leq \mathbb{P}(X \leq y) = F_X(y).$$

Intuition of the Proof: Limit at $-\infty$

- Recall that $F_X(x) = \mathbb{P}(X \leq x)$.
- Since X is discrete, it takes values in a countable set $\{x_1, x_2, x_3, \ldots\}$.
- By definition,
$$F_X(x) = \sum_{x_i \leq x} \mathbb{P}(X = x_i).$$
- As x becomes very negative:
  - fewer and fewer values x_i satisfy $x_i \leq x$,
  - the sum contains fewer (or no) terms,
  - since all probabilities are nonnegative, the sum becomes smaller and smaller.
- Recall that the total probability mass is finite: $\sum_{i=1}^{\infty} \mathbb{P}(X = x_i) = 1$.
- Hence, as $x \to -\infty$,
$$F_X(x) \longrightarrow 0.$$

Intuition of the Proof: Limit at $+\infty$

- Since X is discrete, it takes values in a countable set $\{x_1, x_2, x_3, \ldots\}$.
- By definition,

$$F_X(x) = \sum_{x_i \leq x} \mathbb{P}(X = x_i).$$

- As x increases:
  - more and more values x_i satisfy $x_i \leq x$,
  - the sum keeps adding nonnegative terms,
  - more and more probability mass is accumulated.
- In the limit, the sum includes all possible values of X. Since

$$\sum_{i=1}^{\infty} \mathbb{P}(X = x_i) = 1,$$

we obtain

$$F_X(x) \longrightarrow 1 \quad \text{as } x \to +\infty.$$

Example: CDF of a Discrete Random Variable

Consider again the experiment where three fair coins are tossed, and let X denote the number of times Heads appears.

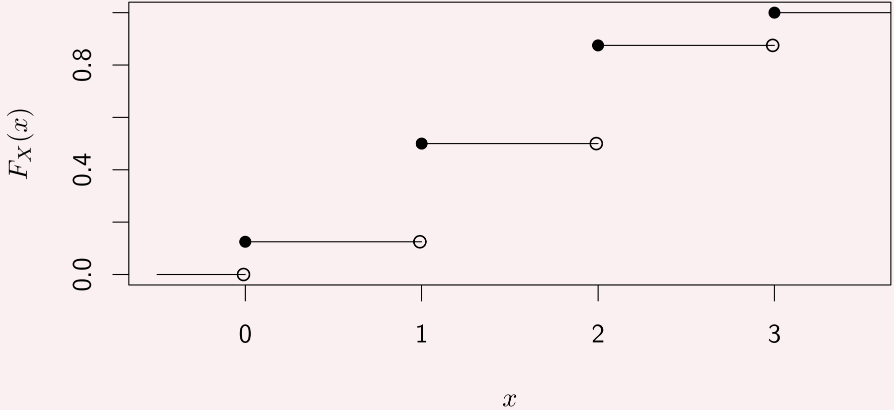
- The probability mass function of X is given by:

$$\mathbb{P}(X = 0) = \frac{1}{8}, \quad \mathbb{P}(X = 1) = \frac{3}{8}, \quad \mathbb{P}(X = 2) = \frac{3}{8}, \quad \mathbb{P}(X = 3) = \frac{1}{8}.$$

- The cumulative distribution function $F_X(x) = \mathbb{P}(X \leq x)$ is therefore:

$$F_X(x) = \begin{cases} 0, & \text{if } x < 0, \\ 1/8, & \text{if } 0 \leq x < 1, \\ 1/2, & \text{if } 1 \leq x < 2, \\ 7/8, & \text{if } 2 \leq x < 3, \\ 1, & \text{if } x \geq 3. \end{cases}$$

CDF of a Discrete Random Variable



CDF of X (Number of Heads in 3 Coin Tosses)

2.3. Expectation

Expectation

- The expectation describes the **central tendency** of a distribution.
- It is a **weighted average** of the possible values of X, where the weights are the probabilities.
- Other common names: **expected value**, **mean**, or **average** of the distribution.
- Notation: the expectation of X is written $\mathbb{E}(X)$ (or $\mathbb{E}[X]$).

Definition 7. Expectation (Discrete Case)

*Let X be a discrete random variable with support $\mathcal{D}_X = \{x_1, x_2, \ldots\}$ and PMF p_X. The **expectation** of X (when it exists) is defined by*

$$\mu_X = \mathbb{E}(X) = \sum_{x \in \mathcal{D}_X} x\, p_X(x).$$

Note: if X takes values in a **finite** set, then $\mathbb{E}(X)$ always exists.

Expectation of a Fair Die

Let X denote the face shown when rolling a fair six-sided die.

- Sample space: $\Omega = \{1, 2, 3, 4, 5, 6\}$.
- Support of X: $\mathcal{D}_X = \{1, 2, 3, 4, 5, 6\}$.
- Probability mass function: $p_X(x) = \frac{1}{6}, \qquad x \in \mathcal{D}_X$.

$$\mathbb{E}(X) = \sum_{x \in \mathcal{D}_X} x\, p_X(x) = \frac{1}{6}(1 + 2 + 3 + 4 + 5 + 6) = \frac{21}{6} = \frac{7}{2} = 3.5.$$

Exercise

Let X be a discrete random variable with support $\mathcal{D}_X = \{-2, -1, 0, 1, 2\}$, and probability mass function given by:

$$\begin{aligned}
\mathbb{P}(X = -2) &= \frac{1}{4},\\
\mathbb{P}(X = -1) &= \frac{1}{4},\\
\mathbb{P}(X = 0) &= \frac{1}{6},\\
\mathbb{P}(X = 1) &= \frac{1}{6},\\
\mathbb{P}(X = 2) &= \frac{1}{6}.
\end{aligned}$$

Compute the expectation of X.

Solution

The expectation of X is:

$$\mathbb{E}(X) = \sum_{x \in \mathcal{D}_X} x \, \mathbb{P}(X = x) = (-2) \cdot \frac{1}{4} + (-1) \cdot \frac{1}{4} + 0 \cdot \frac{1}{6} + 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} = \frac{-1}{4}.$$

3. Continuous Random Variables

Continuous Random Variables

When a random variable takes values in a **continuous subset** of $\mathbb{R}$, it becomes difficult to describe probabilities of the form

$$\mathbb{P}(X \in B)$$

for arbitrary subsets $B \subset \mathcal{D}_X$, especially since such sets may be very complicated.

As in the discrete case, the distribution of X can be characterized through its **distribution function**.

Definition 8. Continuous Random Variable

Let X be a random variable taking values in a subset of $\mathbb{R}$. If its distribution function

$$F_X(x) = \mathbb{P}(X \leq x)$$

*is a **continuous function** of x, then X is called a **continuous random variable**.*

Why $\mathbb{P}(X = x) = 0$ for a Continuous Random Variable

Let X be a continuous random variable.

- X takes values on a continuous scale (e.g., height, income, time).
- A single value x corresponds to a **single point** on the real line.
- In contrast, probabilities are assigned to **intervals**.

For example, if X denotes height (in cm):

$$\mathbb{P}(X = 170) = 0, \qquad \mathbb{P}(169.9 \leq X \leq 170.1) > 0.$$

Other example: suppose you pick a number uniformly at random in $[0, 1]$.

- The probability it lies in $[0.4, 0.5]$ is 0.1.
- But the probability is equals exactly 0.4 is 0: you would need infinite precision to hit exactly 0.4.

Remark: Properties of the CDF (Continuous Case)

When the random variable X is **continuous**, its distribution function $F_X(x)$ satisfies the same global properties as in the discrete case, with an important difference.

Proposition 3. Properties of the CDF for a Continuous Random Variable

Let X be a continuous random variable with distribution function

$$F_X(x) = \mathbb{P}(X \leq x).$$

Then:

1. F_X *is continuous on* $\mathbb{R}$*;*
2. $\lim_{x \to -\infty} F_X(x) = 0$*;*
3. $\lim_{x \to +\infty} F_X(x) = 1$*.*

In the continuous case, $\mathbb{P}(X = x) = 0 \quad$ for all $x \in \mathbb{R}$.

Right-Continuity of the Distribution Function (1/3)

Proposition 4. Right-Continuity of the CDF

For any random variable X and any $x \in \mathbb{R}$,

$$\lim_{h \to 0^+} F_X(x+h) = F_X(x).$$

Let us give an intuition of the proof. First, let us show that

$$0 \leq F_X(x+h) - F_X(x) = \mathbb{P}(X \in (x, x+h]).$$

Recall that $F_X(x) = \mathbb{P}(X \leq x)$. So, for $h > 0$:

$$F_X(x+h) - F_X(x) = \mathbb{P}(X \leq x+h) - \mathbb{P}(X \leq x).$$

Since $x < x + h$, we have the inclusion:

$$\{X \leq x\} \subset \{X \leq x+h\}.$$

Right-Continuity of the Distribution Function (2/3)

Hence, the event $\{X \leq x + h\}$ can be decomposed as a disjoint union:

$$\{X \leq x + h\} = \{X \leq x\} \cup \{x < X \leq x + h\}.$$

Since the two events are disjoint:

$$\mathbb{P}(X \leq x + h) = \mathbb{P}(X \leq x) + \mathbb{P}(x < X \leq x + h).$$

Rearranging gives:

$$\mathbb{P}(X \leq x + h) - \mathbb{P}(X \leq x) = \mathbb{P}(x < X \leq x + h)$$

The event $\{x < X \leq x + h\}$ is exactly

$$\{X \in \{x, x + h\}\}.$$

Hence:

$$F_X(x + h) - F_X(x) = \mathbb{P}(X \in (x, x + h]).$$

Right-Continuity of the Distribution Function (3/3)

Recall that probabilities are always nonnegative, so:

$$0 \leq F_X(x+h) - F_X(x) = \mathbb{P}(X \in (x, x+h]).$$

As $h \to 0^+$:

- the interval $(x, x+h]$ shrinks,
- it approaches the empty set:
  - in the limit, there is no number strictly larger than x and at mots equal to x, so there is no point left in that interval),
- therefore, $\mathbb{P}(X \in (x, x+h]) \to 0$.

Hence, $F_X(x+h) \to F_X(x)$ as $h \to 0^+$: the CDF is always **right-continuous**.

Left Limits and Continuity of the CDF

Let us now consider the **left limit** (and see why it is usually not left continuous).

For $h > 0$, we have:

$$0 \leq F_X(x) - F_X(x-h) = \mathbb{P}(X \in (x-h, x]).$$

As $h \to 0^+$:

- the interval $(x-h, x]$ shrinks to the singleton $\{x\}$:
  - as $h \to 0^+$, all points strictly smaller than x are eventually excluded,
- hence,

$$\mathbb{P}(X \in (x-h, x]) \longrightarrow \mathbb{P}(X = x).$$

Proposition 5. Continuity Points of the CDF

The distribution function F_X is continuous at $x \in \mathbb{R}$ if and only if

$$\mathbb{P}(X = x) = 0.$$

Example: Verifying the Properties of a CDF

Let the function F_X, defined on $\mathbb{R}$ by $x \mapsto \frac{1}{2}\left(1 + \frac{x}{\sqrt{x^2+1}}\right)$. Let us verify that it satisfies the properties of a CDF.

1. **Monotonicity.** The function is differentiable on $\mathbb{R}$:

$$F'_X(x) = \frac{1}{(x^2+1)^{3/2}} > 0.$$

   Hence, F_X is strictly increasing on $\mathbb{R}$.

2. **Right-continuity.** F_X continuous on $\mathbb{R}$, so in particular, right-continuous.

Example: Verifying the Properties of a CDF (continued)

3. **Limits at infinity.**

$$\lim_{x\to-\infty} F_X(x) = 0, \qquad \lim_{x\to+\infty} F_X(x) = 1,$$

Indeed:

$$\left(\frac{x}{\sqrt{x^2+1}}\right)^2 = \frac{x^2}{x^2+1} \xrightarrow[x\to\pm\infty]{} 1.$$

Since $\frac{x}{\sqrt{x^2+1}}$ has the same sign as x, we conclude that

$$\lim_{x\to+\infty} \frac{x}{\sqrt{x^2+1}} = 1, \qquad \lim_{x\to-\infty} \frac{x}{\sqrt{x^2+1}} = -1.$$

Details on $F_X'(x)$ (1/2)

$$F_X'(x) = \frac{d}{dx} F_X(x) = \frac{d}{dx}\frac{1}{2}\left(1 + \frac{x}{\sqrt{x^2+1}}\right) = \frac{1}{2}\frac{d}{dx}\frac{x}{(x^2+1)^{1/2}}.$$

Let $u = x$ and $v = (x^2+1)^{1/2}$.

Let us calculate the derivative of $v = g \circ h$. This function is a composite function with:

$$\begin{cases} h(x) = x^2 + 1 \\ g(x) = x^{1/2}. \end{cases}$$

Recall the chain rule: $\frac{d}{dx}(g \circ h) = \frac{dg}{dh}\frac{dh}{dx}$.

The derivative of each:

$$\begin{cases} \frac{d}{dx}(x^2+1) = 2x \\ \frac{d}{dx}x^{1/2} = \frac{1}{2}x^{-1/2}. \end{cases}$$

Hence,

$$\frac{d}{dx}(x^2+1)^{1/2} = \frac{1}{2}(x^2+1)^{-1/2} \cdot 2x = x(x^2+1)^{-1/2}$$

Details on $F'_X(x)$ (2/2)

$$F'_X(x) = \frac{d}{dx} F_X(x) = \frac{d}{dx}\frac{1}{2}\left(1 + \frac{x}{\sqrt{x^2+1}}\right) = \frac{1}{2}\frac{d}{dx}\frac{x}{(x^2+1)^{1/2}}.$$

We defined $u = x$ and $v = (x^2+1)^{1/2}$.

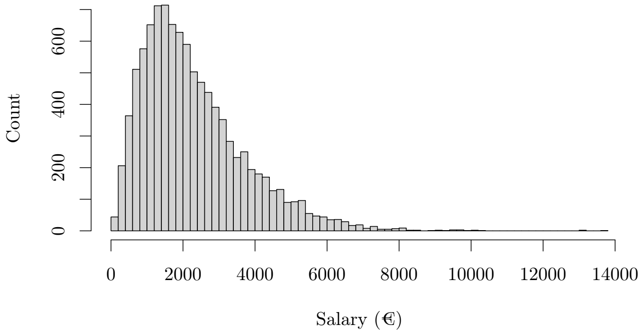
The derivative of each: $u' = \frac{d}{dx}u = 1$ and $v' = \frac{d}{dx}v = x(x^2+1)^{-1/2}$. Hence:

$$\begin{aligned}
\frac{d}{dx}\frac{x}{(x^2+1)^{1/2}} &= \frac{u'v - uv'}{u^2}\\
&= \frac{(x^2+1)^{1/2} - x \cdot x(x^2+1)^{-1/2}}{((x^2+1)^{1/2})^2}\\
&= \frac{(x^2+1)^{-1/2}\left[(x^2+1) - x^2\right]}{(x^2+1)}\\
&= (x^2+1)^{-\frac{1}{2}-1} = (x^2+1)^{-3/2} = \frac{1}{(x^2+1)^{3/2}}
\end{aligned}$$

Visual Illustration: Distribution of Monthly Salaries

Consider a population of $10,000$ individuals. Let X denote the **monthly salary** (in euros) of an individual randomly selected from this population.

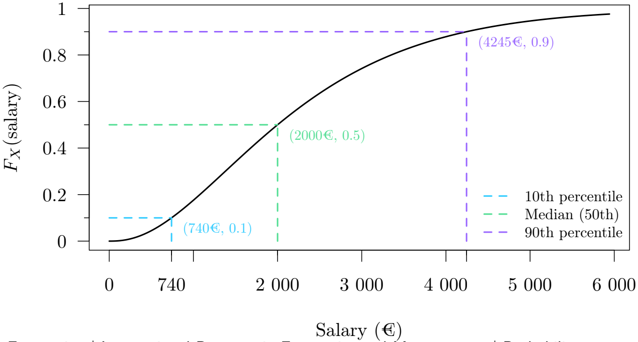
The distribution of monthly salaries in the population is shown in the histogram below. We observe that the distribution is **right-skewed**: a small number of individuals earns very high salaries.



Visual Illustration: Cumulative Distribution Function

The graph below represents the **cumulative distribution function** (CDF) of monthly salaries. From the graph, we can read that if an individual is randomly selected from the population, there is a 10% probability that their monthly salary is less than or equal to €750. Formally,

$$F_X(750) = \mathbb{P}(X \leq 750) = 0.10.$$



Reading Probabilities from the CDF

For any real numbers $a < b$, the CDF allows us to compute interval probabilities:

$$\mathbb{P}(a < X \leq b) = F_X(b) - F_X(a) .$$

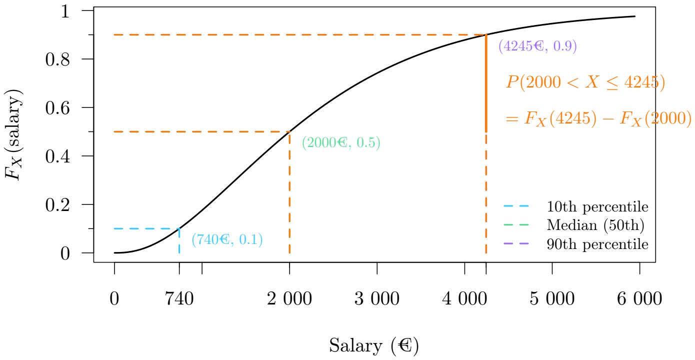


Figure 1: The probability $\mathbb{P}(a < X \leq b)$ is the vertical difference $F_X(b) - F_X(a)$.

Inequalities and the CDF

Proposition 6.

Let X be a ***continuous*** *random variable with distribution function*

$$F_X(x) = \mathbb{P}(X \leq x).$$

For any $x \in \mathbb{R}$, we have:

$$\mathbb{P}(X < x) = \mathbb{P}(X \leq x) = F_X(x),$$

and

$$\mathbb{P}(X \geq x) = \mathbb{P}(X > x) = 1 - F_X(x).$$

Since X is continuous,

$$\mathbb{P}(X = x) = \lim_{\varepsilon \to 0^+} \mathbb{P}(x \leq X \leq x + \varepsilon) = \lim_{\varepsilon \to 0^+} \left(F_X(x+\varepsilon) - F_X(x)\right) = 0.$$

Therefore, including or excluding the endpoint x does not change the probability. □.

3.1. Density Functions

Density of a Continuous Random Variable

- A **continuous random variable** does not admit a probability mass function:

$$\mathbb{P}(X = x) = 0 \quad \text{for all } x \in \mathbb{R}.$$

- Instead, probabilities are described through the **distribution function** F_X.
- When F_X is differentiable, its derivative plays a role analogous to the PMF in the discrete case.

Definition 9. Density Function

Let X be a continuous random variable with distribution function F_X. If F_X is differentiable on $\mathbb{R}$, its ***density function*** *is defined by*

$$f_X(x) = \frac{d}{dx} F_X(x).$$

- The density function is also called the **probability density function (PDF)**.

Properties of a Probability Density Function

Definition 10. Properties of a PDF

When it exists, the density f_X of a continuous random variable X satisfies the following properties:

1. ***Nonnegativity:*** $f_X(x) \geq 0$ *for all* $x \in \mathbb{R}$.
2. ***Normalization:***

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

3. ***Probability of an interval:*** *for any* $a < b$,

$$\mathbb{P}(X \in [a, b]) = \int_a^b f_X(x)\,dx.$$

Proofs of the Properties (1/2)

1. **Nonnegativity.** The distribution function F_X is non-decreasing. Since f_X is its derivative, we must have
$$f_X(x) \geq 0 \quad \text{for all } x.$$

2. **Integration to 1.** By the Fundamental Theorem of Calculus,
$$\int_a^b f_X(x)\,dx = F_X(b) - F_X(a).$$
Taking the limits $a \to -\infty$ and $b \to +\infty$, and using
$$\lim_{x \to -\infty} F_X(x) = 0, \qquad \lim_{x \to +\infty} F_X(x) = 1,$$
we obtain
$$\int_{-\infty}^{+\infty} f_X(x)\,dx = 1.$$

Proofs of the Properties (2/2)

3. **Probability of an interval.** For $a < b$,

$$\int_a^b f_X(x)\,dx = F_X(b) - F_X(a) = \mathbb{P}(X \leq b) - \mathbb{P}(X \leq a) = \mathbb{P}(X \in]a, b]).$$

Since X is continuous, $\mathbb{P}(X = a) = 0$, hence

$$\mathbb{P}(X \leq a) = \mathbb{P}(X < a) + \mathbb{P}(X = a) = \mathbb{P}(X < a).$$

Therefore,

$$\mathbb{P}(X \in [a, b]) = \int_a^b f_X(x)\,dx.$$

□

Remark

With this third property, we understand that the area underneath the density function are probabilities!

Example: From CDF to Density

An Absolutely Continuous Distribution

Consider (again) the function F_X, defined on $\mathbb{R}$ by

$$F_X(x) = \frac{1}{2}\left(1 + \frac{x}{\sqrt{x^2+1}}\right).$$

The random variable X is **absolutely continuous** since F_X is differentiable on $\mathbb{R}$. Its density f_X is given by the derivative of F_X:

$$f_X(x) = \frac{1}{(x^2+1)^{3/2}}, \qquad x \in \mathbb{R}.$$

Conversely, the distribution function F_X can be recovered from the density f_X via

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\, dt.$$

Another Example

Let f_X be the function defined by

$$f_X(x) = \begin{cases} \dfrac{a}{x} & \textit{if } -e \leq x < -1, \\ x + 1 - a & \textit{if } -1 \leq x \leq 0, \\ 0 & \textit{otherwise.} \end{cases}$$

1. *For which value of a is f_X a probability density function?*
2. *Compute the CDF F_X of the corresponding random variable.*

Solution (1/4): Finding a

For f_X to be a probability density function, we need:

1. $f_X(x) \geq 0$ for all x,
2. $\displaystyle\int_{-\infty}^{+\infty} f_X(x)\,dx = 1.$

Nonnegativity.

- On $[-e,-1)$, we have $x<0$, so $\dfrac{a}{x} \geq 0 \iff a \leq 0.$
- On $[-1,0]$, the minimum of $x+1-a$ occurs at $x=-1$, hence

$$x+1-a \geq 0 \;\forall x \in [-1,0] \iff -a \geq 0 \iff a \leq 0.$$

Solution (2/4): Finding a

Normalization. We now compute:

$$\int_{-e}^{-1} \frac{a}{x}\,dx + \int_{-1}^{0} (x+1-a)\,dx = 1.$$

Now,

$$\int_{-e}^{-1} \frac{a}{x}\,dx = a\big[\ln|x|\big]_{-e}^{-1} = a(0-1) = -a,$$

and

$$\int_{-1}^{0} (x+1-a)\,dx = \left[\frac{x^2}{2} + (1-a)x\right]_{-1}^{0} = 0 - \left(\frac{1}{2} - 1 + a\right) = \frac{1}{2} - a.$$

Therefore,

$$-a + \left(\frac{1}{2} - a\right) = 1 \quad \Longleftrightarrow \quad \frac{1}{2} - 2a = 1 \quad \Longleftrightarrow \quad a = -\frac{1}{4}.$$

Solution (3/4): The CDF F_X

With $a = -\frac{1}{4}$, we compute $F_X(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f_X(t)\, dt$.

- If $x < -e$, then $F_X(x) = 0$.
- If $-e \leq x < -1$, then

$$F_X(x) = \int_{-e}^{x} \frac{a}{t}\, dt = a\big(\ln|x| - \ln(e)\big) = a(\ln|x| - 1) = \frac{1}{4}\big(1 - \ln|x|\big).$$

- If $-1 \leq x \leq 0$, then

$$F_X(x) = \underbrace{\int_{-e}^{-1} f_X(t)\, dt}_{=F_X(-1)=-a} + \int_{-1}^{x} (t + 1 - a)\, dt.$$

  Since $F_X(-1) = \frac{1}{4}$ and $1 - a = \frac{5}{4}$, we get

$$F_X(x) = \frac{1}{4} + \left[\frac{t^2}{2} + \frac{5}{4}t\right]_{-1}^{x} = \frac{x^2}{2} + \frac{5}{4}x + 1.$$

- If $x \geq 0$, then $F_X(x) = 1$.

Solution (4/4): The CDF F_X

Hence,

$$F_X(x) = \begin{cases} 0 & \text{if } x < -e, \\ \frac{1}{4}(1 - \ln|x|) & \text{if } -e \leq x < -1, \\ \frac{x^2}{2} + \frac{5}{4}x + 1 & \text{if } -1 \leq x \leq 0, \\ 1 & \text{if } x \geq 0. \end{cases}$$

3.2. Expectation

Expectation

Definition 11. Expectation of a Continuous Random Variable

*Let X be a continuous random variable with density function f_X. The **expectation** (or **mean**) of X is defined by*

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

provided the integral is finite.

- The expectation represents the **average value** of X.
- It can be interpreted as the **center of mass** of the distribution.

Example (1/2)

Let $\theta > 0$ and let X be a continuous random variable with density function

$$f_X(x) = \begin{cases} \frac{2x}{\theta^2} & \text{if } 0 \leq x \leq \theta, \\ 0 & \text{otherwise.} \end{cases}$$

Let us calculate the expectation of X.

Solution

The expectation of X is

$$
\begin{aligned}
\mathbb{E}(X) &= \int_{-\infty}^{+\infty} x\, f_X(x)\, dx \\
&= \int_0^{\theta} x\, \frac{2x}{\theta^2}\, dx \\
&= \frac{2}{\theta^2} \int_0^{\theta} x^2\, dx \\
&= \frac{2}{\theta^2} \left[\frac{1}{3} x^3\right]_0^{\theta} \\
&= \frac{2}{3}\, \theta.
\end{aligned}
$$

4. Independent Random Variables

Independent Random Variables

Definition 12. Independent Random Variables

Let $X_1, X_2, \dots, X_n$ be random variables defined on the same probability space. They are said to be ***(mutually) independent*** *if, for all events $A_1, A_2, \dots, A_n \subset \mathbb{R}$,*

$$\mathbb{P}(X_1 \in A_1, \dots, X_n \in A_n) = \prod_{i=1}^{n} \mathbb{P}(X_i \in A_i).$$

Notation

If X and Y are independent random variables, we write

$$X \perp\!\!\!\perp Y.$$

Independence and Indicator Functions

Proposition 7. (Admitted) Indicator Functions and Independence

Let $E_1, \ldots, E_n$ be events. Then the random variables

$$\mathbf{1}_{E_1}, \mathbf{1}_{E_2}, \ldots, \mathbf{1}_{E_n}$$

are mutually independent if and only if the events $E_1, \ldots, E_n$ are mutually independent.

Joint Density under Independence

Proposition 8. Joint Density of Independent Random Variables

Let $X_1, X_2, \ldots, X_n$ be independent continuous random variables with densities $f_{X_1}, f_{X_2}, \ldots, f_{X_n}$.
Then the random vector $(X_1, X_2, \ldots, X_n)$ admits a joint density given by

$$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i).$$

Remark

*The product $\prod_{i=1}^{n} f_{X_i}(x_i)$ is the **joint density** of the random vector $(X_1, \ldots, X_n)$.*

Independence of Discrete Random Variables

Three independent fair coins are tossed.

- *Let* X_1 *be the indicator random variable of the event* $E_1 =$ *"the first coin shows Heads".*
- *Let* X_2 *be the indicator random variable of the event* $E_2 =$ *"the second coin shows Heads".*
- *Let* X_3 *be the indicator random variable of the event* $E_3 =$ *"the third coin shows Heads".*

1. *Compute* $\mathbb{P}(X_1 = 1)$, $\mathbb{P}(X_2 = 1)$ *and* $\mathbb{P}(X_3 = 1)$.
2. *Compute* $\mathbb{P}(X_1 = 1,\, X_2 = 1,\, X_3 = 1)$.
3. *Are the random variables* X_1, X_2 *and* X_3 *mutually independent?*

Solution

- Each outcome having probability $\frac{1}{8}$.
- The values taken by the random variables:

Outcome	X_1	X_2	X_3
(H, H, H)	1	1	1
(H, H, T)	1	1	0
(H, T, H)	1	0	1
(H, T, T)	1	0	0
(T, H, H)	0	1	1
(T, H, T)	0	1	0
(T, T, H)	0	0	1
(T, T, T)	0	0	0

- Since each coin is fair,

$$\mathbb{P}(X_1 = 1) = \mathbb{P}(X_2 = 1) = \mathbb{P}(X_3 = 1) = \frac{1}{2}.$$

- The event $\{X_1 = 1, X_2 = 1, X_3 = 1\}$ corresponds to the outcome (H, H, H), hence

$$\mathbb{P}(X_1 = 1, X_2 = 1, X_3 = 1) = \frac{1}{8}.$$

- Since

$$\mathbb{P}(X_1 = 1, X_2 = 1, X_3 = 1) = \mathbb{P}(X_1 = 1)\,\mathbb{P}(X_2 = 1)\,\mathbb{P}(X_3 = 1),$$

the random variables X_1, X_2, X_3 are mutually independent.

Independence of Continuous Random Variables

Let X_1, X_2 and X_3 be continuous random variables whose joint density is given by

$$f_{X_1,X_2,X_3}(x_1,x_2,x_3) = \begin{cases} 1 & \textit{if } (x_1,x_2,x_3) \in [0,1]^3, \\ 0 & \textit{otherwise.} \end{cases}$$

1. *Show that f_{X_1,X_2,X_3} is a valid joint density.*
2. *Compute the marginal densities f_{X_1}, f_{X_2}, and f_{X_3}.*
3. *Prove that X_1, X_2, and X_3 are mutually independent.*

Solution: Valid Joint Density

First, we verify that the function $f_{X_1,X_2,X_3}(x_1, x_2, x_3) = 1$ on $[0,1]^3$ integrates to 1.

By definition: $\int_{[0,1]^3} 1\, dx_1\, dx_2\, dx_3 = \int_0^1 \int_0^1 \int_0^1 1\, dx_1\, dx_2\, dx_3$.

Let us compute the integrals step by step:

$$\int_0^1 1\, dx_1 = [x_1]_0^1 = 1,$$

$$\int_0^1 \left(\int_0^1 1\, dx_1\right) dx_2 = \int_0^1 1\, dx_2 = 1,$$

$$\int_0^1 \left(\int_0^1 \int_0^1 1\, dx_1\, dx_2\right) dx_3 = \int_0^1 1\, dx_3 = 1.$$

Therefore, $\int_{[0,1]^3} 1\, dx_1\, dx_2\, dx_3 = 1$.

The function f_{X_1,X_2,X_3} is nonnegative and $\int_{[0,1]^3} 1\, dx_1\, dx_2\, dx_3 = 1$, hence it is a valid joint density.

Solution: Marginal Distributions

For $x_1 \in [0,1]$,

$$f_{X_1}(x_1) = \int_{\mathbb{R}} \int_{\mathbb{R}} f_{X_1,X_2,X_3}(x_1,x_2,x_3)\, dx_2 dx_3$$
$$= \int_0^1 \int_0^1 1\, dx_2 dx_3 = 1,$$

and for $x_1 \notin [0,1]$, then $f_{X_1}(x_1) = 0$.

Therefore,

$$f_{X_1}(x) = \begin{cases} 1 & \text{if } x \in [0,1], \\ 0 & \text{otherwise.} \end{cases}$$

By symmetry, the same holds for f_{X_2} and f_{X_3}.

Solution: Independence

For any $(x_1, x_2, x_3) \in \mathbb{R}^3$,

$$f_{X_1}(x_1)\, f_{X_2}(x_2)\, f_{X_3}(x_3) = \begin{cases} 1 & \text{if } (x_1, x_2, x_3) \in [0,1]^3, \\ 0 & \text{otherwise,} \end{cases}$$

which is exactly $f_{X_1,X_2,X_3}(x_1, x_2, x_3)$.

Thus,

$$f_{X_1,X_2,X_3}(x_1, x_2, x_3) = f_{X_1}(x_1)\, f_{X_2}(x_2)\, f_{X_3}(x_3).$$

By the characterization of independence for continuous random variables, X_1, X_2, and X_3 are mutually independent.

5. Numerical Characteristics of a Random Variable

5.1. Transformations of a Random Variable

Proposition 9. Transformations

Let X be a random variable and let

$$Y = g(X),$$

where $g : \mathcal{D}_X \to \mathbb{R}$ is a (deterministic) function.
Then Y is also a random variable. The support of Y is $\mathcal{D}_Y = g(\mathcal{D}_X)$.

Intuitively: A random variable assigns a number to each outcome. Applying a function to this number produces another number, and therefore defines a new random variable.

More formally:

- X is a mapping from the sample space Ω to $\mathcal{D}_X \subset \mathbb{R}$,
- g maps $\mathcal{D}_X$ to $\mathcal{D}_Y \subset \mathbb{R}$,
- Hence Y is also a mapping from Ω to $\mathbb{R}$.

Transformations of a Discrete Random Variable

- Let X be a discrete random variable with support $\mathcal{D}_X = \{x_1, x_2, \ldots\}$ and probability mass function $p_X(x_j)$.
- Let $Y = g(X)$, where $g : \mathcal{D}_X \to \mathcal{D}_Y \subset \mathbb{R}$ is a function.
- Applying the transformation to each support point of X gives $y_j = g(x_j)$.

Case 1: the values y_j are unique (no redundancies).

- The support of Y is $\mathcal{D}_Y = \{y_1, y_2, \ldots\}$
- The probabilities are preserved: $p_Y(y_j) = p_X(x_j)$.
- The transformation moves the support points, but keeps their probabilities unchanged.

Case 2: the values y_j are not unique.

- Several support points of X are mapped to the same value of Y.
- The corresponding probabilities are combined.
- The transformation reduces the number of support points.

Transformations of a Discrete Random Variable

Example

Let $\mathcal{D}_X = \{-1, 0, 1\}$ and define $Y = X^2$.

- The support of Y is $\mathcal{D}_Y = \{0, 1\}$.
- Since $g(-1) = g(1) = 1$, we have:

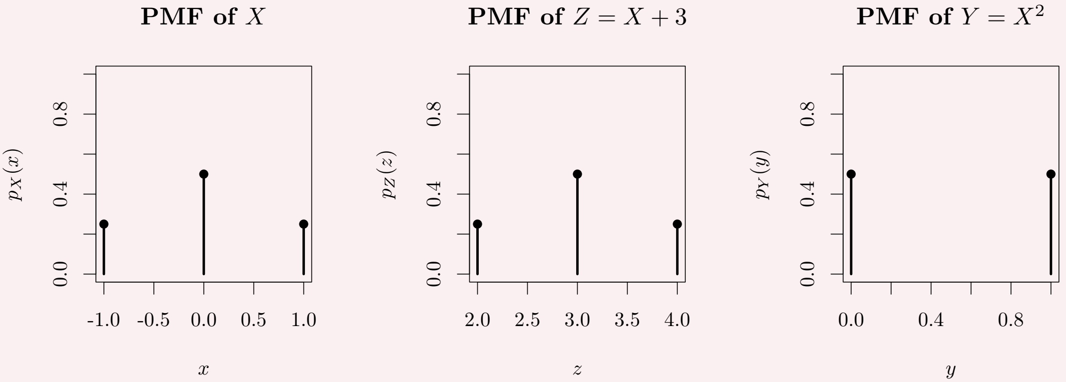
$$p_Y(0) = p_X(0),$$
$$p_Y(1) = p_X(-1) + p_X(1).$$

Remark

More generally, for any $y_i \in \mathcal{D}_Y$,

$$p_Y(y_i) = \sum_{j:\, g(x_j) = y_i} p_X(x_j).$$

Graphical Illustration



PMF before (left) and after transformation: $Z = X + 3$ (middle) and $Y = X^2$ (right)

5.2. Expectation

Expectation of Transformations

Definition 13. Expectation of Transformations

Let X be a discrete random variable with support $\mathcal{D}_X = \{x_1, x_2, \ldots\}$ and PMF p_X. Let g be a deterministic function.
The expectation of the transformed random variable $g(X)$ is defined by

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

Note that to compute $\mathbb{E}[g(X)]$, we do not need the distribution of $g(X)$. We only need the distribution of X.

Example: Expectation of X^2

Let X be a random variable with support $\mathcal{D}_X = \{-1, 0, 1\}$, and probability mass function

$$\mathbb{P}(X = -1) = \frac{1}{4}, \quad \mathbb{P}(X = 0) = \frac{1}{2}, \quad \mathbb{P}(X = 1) = \frac{1}{4}.$$

Let $g(x) = x^2$.

Then

$$\mathbb{E}[X^2] = \sum_{x \in \mathcal{D}_X} x^2 \, \mathbb{P}(X = x) = (-1)^2 \cdot \frac{1}{4} + 0^2 \cdot \frac{1}{2} + 1^2 \cdot \frac{1}{4} = \frac{1}{2}.$$

Expectation of an Indicator Function.

Proposition 10. Expectation of an Indicator Function.

Let A be an event. Then

$$\mathbb{E}(\mathbf{1}_A) = \mathbb{P}(A).$$

In particular, $\mathbb{E}(\mathbf{1}_\Omega) = 1$.

Proof. $\mathbf{1}_A$ is a discrete variable that only takes two values: 0 or 1. Hence,

$$\mathbb{E}(\mathbf{1}_A) = 0 \times \mathbb{P}(\mathbf{1}_A = 0) + 1 \times \mathbb{P}(\mathbf{1}_A = 1) = \mathbb{P}(\mathbf{1}_A = 1) = \mathbb{P}(A)$$

$\square$

Additivity of Expectation

Proposition 11. Additivity of Expectation

Let X and Y be random variables with finite expectations. Then,

$$\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y).$$

Remark

The identity

$$\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y)$$

does not require X and Y to be independent.
Linearity of expectation holds for any random variables with finite expectations, regardless of their dependence structure.

Proof (1/2)

We prove the result in the discrete case, assuming that X and Y take values in $\mathbb{N}$.

By definition,

$$\mathbb{E}(X+Y) = \sum_{n\in\mathbb{N}} n\,\mathbb{P}(X+Y=n).$$

For any $n \in \mathbb{N}$,

$$\mathbb{P}(X+Y=n) = \sum_{i=0}^{n} \mathbb{P}(X=i, Y=n-i)$$

Substituting into the expectation, we obtain

$$\mathbb{E}(X+Y) = \sum_{n\in\mathbb{N}} \sum_{i=0}^{n} n\,\mathbb{P}(X=i,\, Y=n-i).$$

Proof (2/2)

Since $n = i + (n - i)$, we can write

$$\mathbb{E}(X+Y) = \sum_{n\in\mathbb{N}} \sum_{i=0}^{n} \big(i + (n-i)\big)\, \mathbb{P}(X = i,\, Y = n-i)$$
$$= \sum_{n\in\mathbb{N}} \sum_{i=0}^{n} i\, \mathbb{P}(X = i,\, Y = n-i)$$
$$+ \sum_{n\in\mathbb{N}} \sum_{i=0}^{n} (n-i)\, \mathbb{P}(X = i,\, Y = n-i).$$

Change of variable $j = n - i$ (see details next slides).

First term :

$$\sum_{i\in\mathbb{N}} i \sum_{j\in\mathbb{N}} \mathbb{P}(X = i,\, Y = j) = \sum_{i\in\mathbb{N}} i\, \mathbb{P}(X = i) = \mathbb{E}(X).$$

Second term:

$$\sum_{j\in\mathbb{N}} j \sum_{i\in\mathbb{N}} \mathbb{P}(X = i,\, Y = j) = \sum_{j\in\mathbb{N}} j\, \mathbb{P}(Y = j) = \mathbb{E}(Y).$$

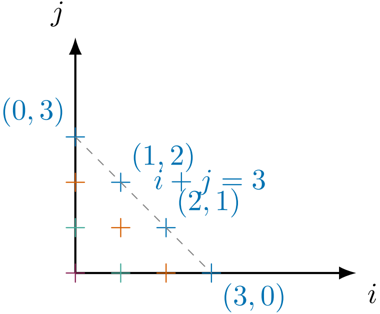
Therefore,

$$\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y). \qquad \square$$

More Explanation (First Term)

- Let $j = n - i$.
- For each fixed n, the **inner sum** runs over all possible pairs
$$(i, n-i) = (0, n), (1, n-1), \dots, (n, 0),$$
i.e., all integer points $(i, j) \in \mathbb{N}^2$ such that $i + j = n$ (one diagonal).
- When n increases:
  - If $n = 0$: $(0, 0)$
  - If $n = 1$: $(0, 1), (1, 0)$
  - If $n = 2$: $(0, 2), (1, 1), (2, 0)$
  - If $n = 3$: $(3, 0), (1, 2), (2, 1), (0, 3)$
  - Each time n increases, we move to the next diagonal $\{(i, j) \in \mathbb{N}^2 : i + j = n\}$
  - So as $n = 0, 1, 2, \dots$, we systematically cover all pairs $(i, j) \in \mathbb{N}^2$ exactly once.

$$\sum_{n\in\mathbb{N}} \sum_{i=0}^n i\mathbb{P}(X = i, Y = n - i)$$
$$= \sum_i \sum_j i\mathbb{P}(X = i, Y = j)$$



More Explanation (Second Term)

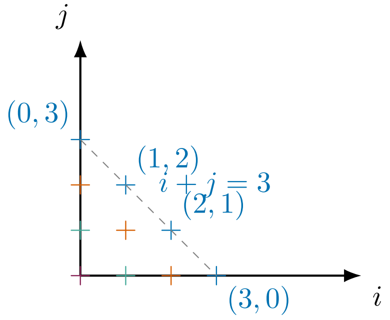
- Let $j = n - i$.
- For each fixed n, the **inner sum** runs over the pairs

$$(i, n-i) = (0, n), (1, n-1), \ldots, (n, 0).$$

- When n increases, we move to the next diagonal:
  - $n = 0$: $(0,0)$
  - $n = 1$: $(0,1), (1,0)$
  - $n = 2$: $(0,2), (1,1), (2,0)$
  - $n = 3$: $(0,3), (1,2), (2,1), (3,0)$
  - Each diagonal is $\{(i,j) \in \mathbb{N}^2 : i + j = n\}$.
  - Hence, as $n = 0, 1, 2, \ldots$, we enumerate all pairs $(i,j) \in \mathbb{N}^2$ exactly once.
- Under this change of variables, the coefficient becomes $n - i = j$.

$$\sum_{n\in\mathbb{N}} \boxed{\sum_{i=0}^{n}(n-i)\mathbb{P}(X=i, Y=n-i)}$$

$$= \sum_{i\in\mathbb{N}} \sum_{j\in\mathbb{N}} j\mathbb{P}(X=i, Y=j).$$



Linearity of Expectation

Proposition 12. Linearity of Expectation

Let X and Y be random variables such that $\mathbb{E}(X)$ and $\mathbb{E}(Y)$ exist, and let $a, b \in \mathbb{R}$. Then

$$\mathbb{E}(aX + bY) = a\,\mathbb{E}(X) + b\,\mathbb{E}(Y).$$

Remark

As a particular case, for any constant $a \in \mathbb{R}$,

$$\mathbb{E}(a + X) = a + \mathbb{E}(X).$$

Proof (1/2)

To prove the proposition, it is enough to show that

$$\mathbb{E}(aX) = a\,\mathbb{E}(X).$$

Discrete case. Assume that X is a discrete random variable with support $\mathcal{D}_X = E$, and define $Y = aX$.

Then Y is discrete with support $\mathcal{D}_Y = \{ax : x \in E\}$, and

$$\begin{aligned}
\mathbb{E}(aX) &= \sum_{y \in \mathcal{D}_Y} y\,\mathbb{P}(Y = y) \\
&= \sum_{x \in E} (ax)\,\mathbb{P}(X = x) \\
&= a \sum_{x \in E} x\,\mathbb{P}(X = x) \\
&= a\,\mathbb{E}(X).
\end{aligned}$$

Proof (2/2)

Continuous case. Assume that X is a continuous random variable with density f_X, and let $Y = aX$.

If $a > 0$, then Y is continuous with density

$$f_Y(y) = \frac{1}{a} f_X\left(\frac{y}{a}\right).$$

Therefore,

$$\begin{aligned}\mathbb{E}(aX) &= \int_{\mathbb{R}} y\, f_Y(y)\, dy \\ &= \int_{\mathbb{R}} y\, \frac{1}{a} f_X\left(\frac{y}{a}\right)\, dy.\end{aligned}$$

Using the change of variable $y = ax$ (so $dy = a\, dx$), we obtain

$$\mathbb{E}(aX) = a \int_{\mathbb{R}} x\, f_X(x)\, dx = a\, \mathbb{E}(X).$$

The case $a < 0$ is handled similarly. If $a = 0$, $Y = aX = 0$ a.s., hence $\mathbb{E}(aX) = \mathbb{E}(0) = 0$ and the identity holds. □

Extension: Linearity for Linear Combinations

Proposition 13. Linearity of Expectation

Let $X_1, \ldots, X_n$ be random variables with finite expectations, and let $a_1, \ldots, a_n \in \mathbb{R}$. Then

$$\mathbb{E}\left(\sum_{k=1}^{n} a_k X_k\right) = \sum_{k=1}^{n} a_k \, \mathbb{E}(X_k).$$

- This result follows directly from the linearity of sums.
- No independence assumption is required.

Expectation of a Symmetric Continuous Random Variable

Proposition 14. Expectation of a Symmetric Distribution

Let X be an absolutely continuous random variable with finite expectation. If its density f_X is even, that is,

$$f_X(-x) = f_X(x) \quad \textit{for all } x \in \mathbb{R},$$

then

$$\mathbb{E}(X) = 0.$$

Proof

By definition,

$$\mathbb{E}(X) = \int_{\mathbb{R}} x\, f_X(x)\, dx.$$

Using the change of variable

$$u = -x \quad \Longleftrightarrow \quad x = -u \quad \Longleftrightarrow \quad dx = -du$$

We obtain

$$\mathbb{E}(X) = \int_{x=-\infty}^{x=+\infty} x f_X(x)\, dx = \int_{u=+\infty}^{u=-\infty} (-u)\, f_X(-u)\, (-du) = \int_{u=-\infty}^{u=+\infty} (-u)\, f_X(-u)\, (du).$$

Since f_X is even, $f_X(-u) = f_X(u)$, hence

$$\mathbb{E}(X) = -\int_{\mathbb{R}} u\, f_X(u)\, du = -\mathbb{E}(X).$$

Therefore, $\mathbb{E}(X) = 0$. □

Nonnegativity of the Expectation

Proposition 15. Nonnegativity of the Expectation

Let X be a nonnegative random variable with a finite expectation. Then

$$\mathbb{E}(X) \geq 0.$$

Proof (discrete case). If $X \geq 0$, then $\mathbb{P}(X = x) = 0$ for all $x < 0$. Hence, every term in the sum $\mathbb{E}(X) = \sum_{x \in \mathcal{D}_X} x \, \mathbb{P}(X = x)$ is nonnegative, which implies $\mathbb{E}(X) \geq 0$.

Proof (continuous case). If X is a nonnegative continuous random variable, then

$$\mathbb{P}(X < 0) = 0.$$

Therefore, its density f_X satisfies $f_X(x) = 0$ for all $x < 0$, and

$$\mathbb{E}(X) = \int_0^{+\infty} x \, f_X(x) \, dx \geq 0.$$

□

Proposition 16. Existence of the Expectation

Let $\varphi : \mathbb{R} \to \mathbb{R}$ *be a deterministic func. and let* X *be a real-valued r.v. Define* $Y = \varphi(X)$.

1. ***Discrete case.*** *Assume that* X *is discrete with support* $\mathcal{D}_X$ *and PMF* $p_X(x) = \mathbb{P}(X = x)$. *Then* Y *has a finite expectation if and only if*
$$\sum_{x \in \mathcal{D}_X} \left|\varphi(x)\right| p_X(x) < \infty.$$
In that case,
$$\mathbb{E}(Y) = \mathbb{E}[\varphi(X)] = \sum_{x \in \mathcal{D}_X} \varphi(x)\, p_X(x).$$

2. ***Continuous case.*** *Assume that* X *is continuous with density* f_X. *Then* Y *has a finite expectation if and only if*
$$\int_{\mathbb{R}} \left|\varphi(x)\right| f_X(x)\, dx < \infty.$$
In that case,
$$\mathbb{E}(Y) = \mathbb{E}[\varphi(X)] = \int_{\mathbb{R}} \varphi(x)\, f_X(x)\, dx.$$

Proof (Discrete Case, 1/3)

Let X be discrete with support $\mathcal{D}_X$ and PMF $p_X(x) = \mathbb{P}(X = x)$. Let $Y = \varphi(X)$.

For any $x \in \mathcal{D}_X$, on the event $\{X = x\}$ we have $Y = \varphi(x)$.

First, let us show that if the expectation exists, we have $\mathbb{E}(Y) = \sum_{x \in \mathcal{D}_X} \varphi(x)\, p_X(x)$.

By definition of expectation for a discrete random variable,

$$\mathbb{E}(Y) = \sum_{y \in \mathcal{D}_Y} y\, \mathbb{P}(Y = y).$$

We group together all $x \in \mathcal{D}_X$ that produce the same value $y = \varphi(x)$:

$$\mathbb{P}(Y = y) = \sum_{x \in \mathcal{D}_X : \, \varphi(x) = y} \mathbb{P}(X = x).$$

Proof (Discrete Case, 2/3)

Hence, we have:

$$\begin{aligned}
\mathbb{E}(Y) &= \sum_{y \in \mathcal{D}_Y} y \sum_{x \colon \varphi(x) = y} \mathbb{P}(X = x) \\
&= \sum_{x \in \mathcal{D}_X} \varphi(x)\, \mathbb{P}(X = x) \\
&= \sum_{x \in \mathcal{D}_X} \varphi(x)\, p_X(x).
\end{aligned}$$

Now, let us turn to the **second part**. To justify the existence condition, write for every $x \in \mathcal{D}_X$:

$$\varphi(x) = \varphi^+(x) - \varphi^-(x), \qquad |\varphi(x)| = \varphi^+(x) + \varphi^-(x),$$

where

$$\varphi^+(x) = \max\{\varphi(x), 0\}, \qquad \varphi^-(x) = \max\{-\varphi(x), 0\}.$$

Proof (Discrete Case, 3/3)

Then,

$$\sum_x \varphi(x)p_X(x) = \sum_x \varphi^+(x)p_X(x) - \sum_x \varphi^-(x)p_X(x).$$

- If $\sum_x |\varphi(x)|p_X(x) < \infty$, then both sums above are finite, hence $\mathbb{E}[\varphi(X)]$ is finite.
- Conversely, if $\mathbb{E}[\varphi(X)]$ is finite, neither the positive nor the negative part can diverge, which implies

$$\sum_x |\varphi(x)|p_X(x) < \infty.$$

Therefore,

$$\mathbb{E}[\varphi(X)] \text{ is finite} \iff \sum_{x\in\mathcal{D}_X} |\varphi(x)|\, p_X(x) < \infty. \qquad \square$$

Remark

Why do we have:

$$\sum_{y\in\mathcal{D}_Y} y \sum_{x:\,\varphi(x)=y} \mathbb{P}(X=x) = \sum_{x\in\mathcal{D}_X} \varphi(x)\,\mathbb{P}(X=x)$$

Each $x\in\mathcal{D}_X$ appears only once in the sum. Consequently, the collection of sets $\{\{x:\varphi(x)=y\}: y\in\mathcal{D}_Y\}$ forms a partition of $\mathcal{D}_X$.
As an example, consider $\mathcal{D}_X=\{-1,0,1\}$, $\quad \varphi(x)=x^2$. Then:

$$\begin{aligned}
\mathbb{E}(Y) &= \sum_{y\in\{0,1\}} y\,\mathbb{P}(Y=y) = \sum_{y\in\{0,1\}} y \sum_{x\in\mathcal{D}_X:\,\varphi(x)=y} \mathbb{P}(X=x)\\
&= 0\cdot\mathbb{P}(X=0) + 1\cdot\left[\mathbb{P}(X=-1)+\mathbb{P}(X=1)\right]\\
&= \sum_{x\in\{-1,0,1\}} x^2\,\mathbb{P}(X=x) = \sum_{x\in\mathcal{D}_X} \varphi(x)\,\mathbb{P}(X=x).
\end{aligned}$$

Finiteness of Expectations

The expectation of a discrete random variable may be **infinite** (or undefined), even when its probability mass function is perfectly valid.

A Discrete Random Variable with Infinite Expectation

Let X be a discrete random variable with support $\mathcal{D}_X = \{2^k : k = 1, 2, 3, \ldots\}$, and probability mass function $\mathbb{P}(X = 2^k) = 2^{-k}, \qquad k = 1, 2, 3, \ldots$

We can verify that this defines a valid probability mass function:

- Nonnegativity: $\mathbb{P}(X = 2^k) = 2^{-k} \geq 0$.
- Normalization: $\sum_{k=1}^{\infty} \mathbb{P}(X = 2^k) = \sum_{k=1}^{\infty} 2^{-k} = 1$. (see next slide)

The expectation of X is

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

Normalization of the Probability Mass Function

In the previous slide, we wrote that $\sum_{k=1}^{\infty} \mathbb{P}(X = 2^k) = 1$.

By definition,

$$\sum_{k=1}^{\infty} \mathbb{P}(X = 2^k) = \sum_{k=1}^{\infty} 2^{-k} = \sum_{k=1}^{\infty} \frac{1}{2^k} = \sum_{k=1}^{\infty} \left(\frac{1}{2}\right)^k = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots$$

- This is a geometric series with common ratio $r = \frac{1}{2}$.
- For any geometric series with $|r| < 1$, we have:

$$\sum_{k=0}^{\infty} r^k = \frac{1}{1-r}.$$

Applying this result,

$$\sum_{k=1}^{\infty} 2^{-k} = \sum_{k=0}^{\infty} 2^{-k} - 1 = 2 - 1 = 1.$$

The St. Petersburg Paradox

The previous example is closely related to the **St. Petersburg paradox**.

- A fair coin is tossed repeatedly until the first Head appears.
- If the first Head appears on toss k, the payoff is €2^k.
  - if $K = 1$, the player is paid €2; if $K = 2$, they are paid €4, etc.
- The probability of this event is 2^{-k}.
- As shown in the previous slide, the expected payoff is infinite.
- In practice, only a few individuals are willing to pay a high price to receive this random payout (hence the paradox).

Expectation That Is Not Defined (1/3)

A discrete random variable may have an expectation that is **not defined**, even though it has a valid probability mass function.

Example: A Symmetric Heavy-Tailed Distribution

Let X be a discrete random variable with support $\mathcal{D}_X = \mathbb{Z} \setminus \{0\}$ and probability mass function

$$\mathbb{P}(X = k) = \frac{C}{k^2}, \qquad k \in \mathbb{Z} \setminus \{0\},$$

where $C > 0$ is a normalizing constant.

Expectation That Is Not Defined (2/3)

First, we show that the series $\sum_{k\neq 0} \frac{1}{k^2}$ converges so that it defines a valus probability mass function. Since the distribution is symmetric,

$$\sum_{k\neq 0} \frac{1}{k^2} = \sum_{k=-\infty}^{-1} \frac{1}{k^2} + \sum_{k=1}^{+\infty} \frac{1}{k^2} = 2\sum_{k=1}^{\infty} \frac{1}{k^2}.$$

We use integral comparison. For $x \geq 1$, the function $f(x) = 1/x^2$ is positive and decreasing. Hence,

$$\sum_{k=1}^{\infty} \frac{1}{k^2} \leq 1 + \int_1^{\infty} \frac{1}{x^2}\, dx = \left[-\frac{1}{x}\right]_1^{\infty} = 1.$$

Therefore,

$$\sum_{k=1}^{\infty} \frac{1}{k^2} \quad \text{converges.}$$

C can be chosen such that $\sum_k \mathbb{P}(X = k) = 1$ (this is the case for $C = 3/\pi^2$).

Expectation That Is Not Defined (3/3)

We have shown that $\frac{C}{k^2}$ (for a suitable C) defines a valid probability mass function. The expectation would be

$$\mathbb{E}(X) = \sum_{k\neq 0} k \cdot \frac{C}{k^2} = C \sum_{k\neq 0} \frac{1}{k}.$$

For $x \geq 1$, $f(x) = 1/x$ is positive, continuous, and decreasing. But

$$\int_1^{\infty} \frac{1}{x}\, dx = [\ln x]_1^{\infty} = +\infty.$$

Therefore, $\sum_{k=1}^{\infty} \frac{1}{k} = +\infty$.

The positive part of $\sum_{k\neq 0} \frac{1}{k}$ diverges to $+\infty$ and the negative part diverges to $-\infty$, so the two-sided series is not well-defined.

Therefore, $\mathbb{E}(X)$ is **not defined**.

5.3. Moments

Definition 14. Moments of Order p

Let $p \in \mathbb{N}^*$ and let X be a real-valued random variable.
We say that X ***admits a (finite) moment of order*** *p if*

$$\mathbb{E}(|X|^p) < \infty.$$

Equivalently:

- ***Discrete case:*** *if X has support $\mathcal{D}_X$ and PMF $p_X(x) = \mathbb{P}(X = x)$, then*

$$\sum_{x \in \mathcal{D}_X} |x|^p \, p_X(x) < \infty.$$

- ***Continuous case:*** *if X has density f_X, then*

$$\int_{\mathbb{R}} |x|^p \, f_X(x) \, dx < \infty.$$

Moments: Uncentered and Central

Uncentered moments describe the magnitude of a random variable, while central moments describe deviations from its mean.

Definition 15. Uncentered Moment of Order p

Let $p \in \mathbb{N}^*$ and let X be a random variable. If $\mathbb{E}(|X|^p) < \infty$, the ***uncentered moment of order*** p *of X is defined by*

$$\mathbb{E}(X^p),$$

which is then finite.

Definition 16. Central Moment of Order p

Let $p \in \mathbb{N}^*$ and let X be a random variable with finite expectation $\mathbb{E}(X)$. If $\mathbb{E}(|X - \mathbb{E}(X)|^p) < \infty$, the ***central moment of order*** p *of X is defined by*

$$\mathbb{E}[(X - \mathbb{E}(X))^p] .$$

5.4. Variance

Variance

Definition 17. Variance of a Random Variable

*Let X be a random variable admitting a finite moment of order 2 (i.e. $\mathbb{E}(X^2) < \infty$). The **variance** of X is defined by*

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

Remark

*The variance is the **central moment of order** 2.*

Example with a Discrete Random Variable

Let X be a random variable with support $\mathcal{D}_X = \{0, 1\}$ and probability distribution

$$\mathbb{P}(X = 1) = p, \qquad \mathbb{P}(X = 0) = 1 - p, \qquad p \in [0, 1].$$

The expectation of X is

$$\mathbb{E}(X) = \sum_{x \in \mathcal{D}_X} x\, \mathbb{P}(X = x) = 0 \cdot (1 - p) + 1 \cdot p = p.$$

The variance of X is

$$\begin{aligned} \mathrm{Var}(X) &= \mathbb{E}\left[(X - \mathbb{E}(X))^2\right] \\ &= (0 - p)^2(1 - p) + (1 - p)^2 p \\ &= p(1 - p). \end{aligned}$$

Example with a Continuous Variable

Let $\theta > 0$ and let X be a continuous random variable with density function

$$f_X(x) = \begin{cases} \dfrac{2x}{\theta^2} & \textit{if } 0 \leq x \leq \theta, \\ 0 & \textit{otherwise.} \end{cases}$$

Recall that (see earlier example)

$$\mathbb{E}(X) = \frac{2\theta}{3}.$$

The variance of X is

$$\mathrm{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}(X))^2\right] = \int_0^\theta \left(x - \frac{2\theta}{3}\right)^2 \frac{2x}{\theta^2}\, dx.$$

Expanding the square,

$$\left(x - \frac{2\theta}{3}\right)^2 = x^2 - \frac{4\theta}{3}x + \frac{4\theta^2}{9}.$$

Therefore,

$$\begin{aligned}
\mathrm{Var}(X) &= \frac{2}{\theta^2}\int_0^\theta \left(x^3 + \frac{4\theta^2}{9}x - \frac{4\theta}{3}x^2\right)\, dx \\
&= \frac{2}{\theta^2}\left(\frac{\theta^4}{4} - \frac{2\theta^4}{9} - \frac{4\theta^4}{9}\right). \\
&= \frac{2}{\theta^2}\left(\frac{9\theta^4 - 8\theta^4}{36}\right) \\
&= \frac{\theta^2}{18}.
\end{aligned}$$

Alternative Formula for the Variance

Proposition 17. Variance Formula

Let X be a random variable with a finite moment of order 2. Then,

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

Proof

By definition,

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

Expanding the square,

$$(X - \mathbb{E}(X))^2 = X^2 + \mathbb{E}(X)^2 - 2X\,\mathbb{E}(X).$$

Taking expectations and using linearity,

$$\begin{aligned}
\mathrm{Var}(X) &= \mathbb{E}(X^2) + \mathbb{E}(\mathbb{E}(X)^2) - \mathbb{E}(2X\,\mathbb{E}(X)) \\
&= \mathbb{E}(X^2) + \mathbb{E}(X)^2 - 2\mathbb{E}(X\,\mathbb{E}(X)) \\
&= \mathbb{E}(X^2) + \mathbb{E}(X)^2 - 2\,\mathbb{E}(X)\,\mathbb{E}(X) \\
&= \mathbb{E}(X^2) - \mathbb{E}(X)^2.
\end{aligned}$$

□

Example with a Discrete Random Variable (revisited)

Let X be a random variable with support $\mathcal{D}_X = \{0, 1\}$ and probability distribution

$$\mathbb{P}(X = 1) = p, \qquad \mathbb{P}(X = 0) = 1 - p, \qquad p \in [0, 1].$$

We have

$$\mathbb{E}(X) = 0 \cdot (1 - p) + 1 \cdot p = p.$$

In addition, since $X^2 = X$ when $X \in \{0, 1\}$,

$$\mathbb{E}(X^2) = \mathbb{E}(X) = p.$$

Therefore,

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

Example with a Continuous Variable (revisited)

Let $\theta > 0$ and let X have density

$$f_X(x) = \begin{cases} \frac{2x}{\theta^2} & \text{if } 0 \leq x \leq \theta, \\ 0 & \text{otherwise.} \end{cases}$$

Recall that $\mathbb{E}(X) = \frac{2\theta}{3}$.

We first compute $\mathbb{E}(X^2)$:

$$\mathbb{E}(X^2) = \int_0^\theta x^2 \frac{2x}{\theta^2}\, dx = \frac{2}{\theta^2} \int_0^\theta x^3\, dx = \frac{2}{\theta^2} \cdot \frac{\theta^4}{4} = \frac{\theta^2}{2}.$$

Using $\mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$, we obtain

$$\mathrm{Var}(X) = \frac{\theta^2}{2} - \left(\frac{2\theta}{3}\right)^2 = \frac{\theta^2}{2} - \frac{4\theta^2}{9} = \theta^2 \left(\frac{9-8}{18}\right) = \frac{\theta^2}{18}.$$

Proposition 18. Properties of the Variance of a Random Variable

Let X be a random variable admitting a finite moment of order 2, and let $a \in \mathbb{R}$.

1. ***Translation invariance.***
$$\mathrm{Var}(a + X) = \mathrm{Var}(X).$$
2. ***Scaling by a constant.***
$$\mathrm{Var}(aX) = a^2 \, \mathrm{Var}(X).$$
3. ***Nonnegativity.***
$$\mathrm{Var}(X) \geq 0.$$
In addition,
   - $\mathrm{Var}(X) = 0$ *if and only if*
$$\mathbb{P}\big(X = \mathbb{E}(X)\big) = 1,$$
that is, X is almost surely constant.
   - *In particular, if X is a non-degenerate continuous random variable, then*
$$\mathrm{Var}(X) > 0.$$

Proof: Translation Invariance

We prove that $\mathrm{Var}(a+X) = \mathrm{Var}(X)$. Recall that

$$\mathbb{E}(a+X) = a + \mathbb{E}(X).$$

Then,

$$\begin{aligned}
\mathrm{Var}(a+X) &= \mathbb{E}\left[(a+X-\mathbb{E}(a+X))^2\right] \\
&= \mathbb{E}\left[(a+X-(a+\mathbb{E}(X)))^2\right] \\
&= \mathbb{E}[(X-\mathbb{E}(X))^2] \\
&= \mathrm{Var}(X).
\end{aligned}$$

$\square$

Proof: Scaling by a Constant

We prove that $\text{Var}(aX) = a^2\ \text{Var}(X)$.

Using the alternative formula for the variance,

$$\text{Var}(aX) = \mathbb{E}\left[(aX)^2\right] - \left[\mathbb{E}(aX)\right]^2.$$

Recall that, by linearity of expectation,

$$\mathbb{E}\left[(aX)^2\right] = a^2\,\mathbb{E}(X^2), \qquad \mathbb{E}(aX) = a\,\mathbb{E}(X).$$

Substituting into the expression of the variance,

$$\begin{aligned}\text{Var}(aX) &= a^2\,\mathbb{E}(X^2) - a^2\,\mathbb{E}(X)^2 \\ &= a^2\left(\mathbb{E}(X^2) - \mathbb{E}(X)^2\right) \\ &= a^2\ \text{Var}(X).\end{aligned}$$

□

Proof: Nonnegativity (1/)

We prove that $\mathrm{Var}(X) \geq 0$.

Recall that

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

For every outcome ω, we have $(X(\omega) - \mathbb{E}(X))^2 \geq 0$. Taking expectations preserves inequalities (see proof), hence

$$\mathrm{Var}(X) = \mathbb{E}\Big[(X - \mathbb{E}(X))^2\Big] \geq 0.$$

□

Proof: Nonnegativity (2/)

Now, we prove that $\mathrm{Var}(X) = 0 \Leftrightarrow \mathbb{P}(X = \mathbb{E}(X)) = 1$.

- First, we prove that $\mathrm{Var}(X) = 0 \Rightarrow \mathbb{P}(X = \mathbb{E}(X)) = 1.$ If $\mathrm{Var}(X) = 0$, then $\mathbb{E}(Y) = 0$ with $Y = (X - \mathbb{E}(X))^2 \geq 0$. A nonnegative random variable has zero expectation only if it is zero almost surely (an event is said to happen almost surely if it happens with probability 1), so

$$(X - \mathbb{E}(X))^2 = 0 \quad \text{a.s.}$$

Since a square is zero if and only if the quantity itself is zero, this implies

$$X - \mathbb{E}(X) = 0 \quad \text{a.s.},$$

or equivalently,

$$\mathbb{P}(X = \mathbb{E}(X)) = 1.$$

That is, X is almost surely constant.

Proof: Nonnegativity (3/)

- Second, we prove that $\mathrm{Var}(X) = 0$ $\Leftarrow$ $\mathbb{P}(X = \mathbb{E}(X)) = 1.$

  If $\mathbb{P}(X = \mathbb{E}(X)) = 1$, then, for almost every outcome ω,

$$X(\omega) = \mathbb{E}(X).$$

  Then $(X - \mathbb{E}(X))^2 = 0$ a.s.. Then,

$$\mathrm{Var}(X) = \mathbb{E}[\underbrace{(X - \mathbb{E}(X))^2}_{=0 \text{ a.s.}}] = 0.$$

□

Proof: Nonnegativity (4/4)

We prove that if X is a non-degenerate continuous random variable, then $\mathrm{Var}(X) > 0$.

Definition 18. (Non-)degenerate Random Variable

*A random variable is **degenerate** if it is almost surely constant, i.e.*

$$\exists c \in \mathbb{R} \quad \textit{s.t.} \quad \mathbb{P}(X = c) = 1.$$

*It is **non-degenerate** if it is not almost surely constant:*

$$\forall c \in \mathbb{R}, \quad \mathbb{P}(X = c) < 1.$$

From the previous steps, we know that $\mathrm{Var}(X) \geq 0$. We show that if X is non-degenerate continuous, we cannot have $\mathrm{Var}(X) = 0$: if X were continuous and $\mathbb{P}(X = \mathbb{E}(X)) = 1$, then all the probability mass would be concentrated at a single point, which contradicts continuity. Therefore, X cannot be almost surely constant, and $\mathrm{Var}(X) > 0$. □

Exercise: Standardizing Income

Assume we are interested in studying the monthly income of individuals in France. Let X denote an individual's monthly income, measured in euros. The estimated values of the mean and variance of X are

$$\mathbb{E}(X) = 2{,}600 \quad \textit{and} \quad \mathrm{Var}(X) = 1{,}000^2.$$

To simplify the analysis, we define a new random variable

$$Y = \frac{X - 2{,}600}{1\,000}.$$

1. *Compute the expectation of Y.*
2. *Compute the variance of Y.*

Solution

1. Using linearity of expectation,

$$\mathbb{E}(Y) = \mathbb{E}\left(\frac{X - 2{,}600}{1{,}000}\right) = \frac{\mathbb{E}(X) - 2{,}600}{1{,}000} = 0.$$

2. Using the properties of the variance,

$$\mathrm{Var}(Y) = \mathrm{Var}\left(\frac{X - 2{,}600}{1{,}000}\right) = \left(\frac{1}{1{,}000}\right)^2 \mathrm{Var}(X) = 1.$$

Symmetric Continuous Random Variable

Proposition 19. Variance of a Symmetric Continuous Random Variable

Let X be an absolutely continuous random variable with density f_X. Assume that f_X is even, that is,

$$f_X(-x) = f_X(x) \quad \text{for all } x \in \mathbb{R}.$$

If $\mathbb{E}(X^2) < \infty$, then

$$\mathrm{Var}(X) = \mathbb{E}(X^2) = 2 \int_0^{+\infty} x^2 \, f_X(x) \, dx.$$

Proof (1/3)

Since f_X is even, X has a symmetric distribution, and we have already shown that $\mathbb{E}(X) = 0$.

By definition, $\mathrm{Var}(X) = \mathbb{E}\big[(X - \mathbb{E}(X))^2\big] = \mathbb{E}(X^2)$.

In addition, since $x^2 f_X(x)$ is an even function, we can write

$$\mathbb{E}(X^2) = \underbrace{\int_{-\infty}^{0} x^2 f_X(x)\, dx}_{:=I_-} + \int_{0}^{+\infty} x^2 f_X(x)\, dx.$$

Consider the **first integral, I_-**.

Make the change of variable $u = -x$ (so $x = -u$ and $dx = -du$). When $x \to -\infty$, we have $u \to +\infty$, and when $x \to 0$, we have $u \to 0$. Therefore,

$$I_- = \int_{u=+\infty}^{u=0} (-u)^2 f_X(-u)\, (-du).$$

Proof (2/3)

Since $(-u)^2 = u^2$, we get

$$\boxed{I_-} = \int_{+\infty}^{0} u^2 f_X(-u)\,(-du)$$

Reminder

Recall that for any integrable function g,

$$\int_a^b g(u)\,(-du) = \int_b^a g(u)\,du.$$

Integrating "backwards" from $-\infty$ to 0 with a negative step is the same as integrating "forwards" from 0 to $+\infty$.

Applying this property,

$$\boxed{I_-} = \int_{+\infty}^{0} u^2 f_X(-u)\,(-du) = \int_0^{+\infty} u^2 f_X(-u)\,du.$$

Proof (3/3)

Since f_X is even, $f_X(-u) = f_X(u)$, and therefore

$$\int_{-\infty}^{0} x^2 f_X(x)\, dx = \int_0^{+\infty} u^2 f_X(u)\, du.$$

Hence,

$$\mathbb{E}(X^2) = 2 \int_0^{+\infty} x^2 f_X(x)\, dx.$$

□

Example: Symmetric Continuous Random Variable

Let X be a random variable with density $f_X(x) = \frac{1}{2}e^{-|x|}$ for $x \in \mathbb{R}$.

The density f_X is symmetric, so $\mathbb{E}(X) = 0$ and $\mathrm{Var}(X) = \mathbb{E}(X^2)$.

Using symmetry, we obtain

$$\begin{aligned}
\mathrm{Var}(X) &= 2\int_0^{+\infty} x^2\, f_X(x)\, dx \\
&= 2\int_0^{+\infty} x^2\, \frac{1}{2}e^{-x}\, dx \\
&= \int_0^{+\infty} x^2 e^{-x}\, dx.
\end{aligned}$$

Integration by parts

The integration by part formula states:

$$\begin{aligned}\int_a^b u(x)v'(x)\,dx &= \left[u(x)v(x)\right]_a^b - \int_a^b u'(x)v(x)\,dx \\ &= u(b)v(b) - u(a)v(a) - \int_a^b u'(x)v(x)\,dx.\end{aligned}$$

$$\mathrm{Var}(X) = \int_0^{+\infty} x^2 e^{-x}\,dx$$

Let us choose:

- $u(x) = x^2$ which gives $u'(x) = 2x$,
- $v'(x) = e^{-x}$ which gives $v(x) = \int e^{-x}\,dx = -e^{-x}$.

We compute

$$\mathrm{Var}(X) = \int_0^{+\infty} x^2 e^{-x} = \left[x^2(-e^{-x})\right]_0^{+\infty} - \int_0^{+\infty} 2x(-e^{-x})\,dx.$$

Since

$$\lim_{x\to+\infty} x^2(-e^{-x}) = 0 \quad \text{(the exp. term dominates any polynomial)},$$

We have:

$$\mathrm{Var}(X) = (0-0) + \int_0^{+\infty} 2x\,e^{-x}\,dx = 2\int_0^{+\infty} x\,e^{-x}\,dx$$

A second integration by parts is required for $J := \int_0^{+\infty} x\, e^{-x}\, dx$. We choose:

- $u(x) = x$ which gives $u'(x) = 1$,
- $v'(x) = e^{-x}$ which gives $v(x) = \int e^{-x}\, dx = -e^{-x}$.

Applying the integration by parts formula:

$$J = \int_0^{+\infty} x\, e^{-x}\, dx = \left[-xe^{-x}\right]_0^{+\infty} - \int_0^{+\infty} (-e^{-x})\, dx.$$

Since

$$\lim_{x\to\infty} -xe^{-x} = 0, \qquad \lim_{x\to 0} -xe^{-x} = 0,$$

We have

$$\begin{aligned} J &= (0 - 0) + \int_0^{+\infty} e^{-x}\, dx \\ &= \left[-e^{-x}\right]_0^{+\infty} = \left(\lim_{x\to+\infty} -e^{-x}\right) - \left(\lim_{x\to 0} -e^{-x}\right) \\ &= 0 - (-1) = 1 \end{aligned}$$

Hence, $\mathrm{Var}(X) = 2J = 2 \times 1 = 2$.

Standard Deviation

Definition 19. Standard Deviation

*The **standard deviation** of a random variable X is defined as the positive square root of its variance:*

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

- The standard deviation provides a measure of **dispersion** that is easier to interpret than the variance.
- Indeed, the variance is based on squared deviations from the expectation, which results in a quantity expressed in squared units. This makes direct comparison with the expectation less intuitive.
- By taking the square root, the standard deviation is expressed in the **same unit** as the random variable itself.
- For example, if X measures a quantity in liters, then $\mathrm{Var}(X)$ is expressed in liters squared, whereas the standard deviation is expressed in liters.

5.5. Covariance

Covariance

Covariance measures how two variables vary together. While variance measures individual uncertainty, covariance measures whether uncertainties move together.

Definition 20. Covariance

Let X and Y be random variables with finite second moments. The covariance between X and Y is defined by

$$\mathrm{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))]$$

Covariance

Proposition 20. Covariance

Let X and Y be random variables with finite second moments. Then

$$\mathrm{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].$$

Proof.

$$\begin{aligned}
\mathrm{Cov}(X, Y) &= \mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))] \\
&= \mathbb{E}[XY - X\mathbb{E}(Y) - Y\mathbb{E}(X) + \mathbb{E}(X)\mathbb{E}(Y)] \\
&= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] - \mathbb{E}[Y]\mathbb{E}[X] + \mathbb{E}[X]\mathbb{E}[Y] \\
&= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].
\end{aligned}$$

$\square$

Variance of a Sum

Proposition 21. Variance of the Sum of Two Random Variables

Let X and Y be random variables with finite second moments. Then,

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X,Y).$$

Proof.

$$\begin{aligned}
\mathrm{Var}(X+Y) &= \mathbb{E}[(X+Y)^2] - (\mathbb{E}(X+Y))^2 \\
&= \mathbb{E}[X^2 + Y^2 + 2XY] - (\mathbb{E}(X) + \mathbb{E}(Y))^2 \\
&= \mathbb{E}(X^2) + \mathbb{E}(Y^2) + 2\mathbb{E}(XY) - \mathbb{E}(X)^2 - \mathbb{E}(Y)^2 - 2\mathbb{E}(X)\mathbb{E}(Y) \\
&= \mathbb{E}(X^2) - \mathbb{E}(X)^2 + \mathbb{E}(Y^2) - \mathbb{E}(Y)^2 + 2(\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)) \\
&= \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X,Y).
\end{aligned}$$

□

Independence

Proposition 22. Consequences of Independence

*Let X and Y be **independent** random variables.*

1. *If X and Y both have a finite expectation, then*

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y).$$

2. *If X and Y both have finite second moments, then*

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y).$$

We will admit the first proposition. Its proof is long and technically involved.

Proof

We prove that $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y)$.

Recall the definition of covariance:

$$\text{Cov}(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y).$$

If X and Y are independent, we admit that (proposition 1)

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y).$$

Therefore,

$$\text{Cov}(X,Y) = 0.$$

Using the variance decomposition formula,

$$\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) + 2\,\text{Cov}(X,Y),$$

we conclude that

$$\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y). \qquad \square$$

Covariance and Independence

We have shown that if X and Y are independent, then $\mathrm{Cov}(X, Y) = 0$.

The converse is not true.

$\mathrm{Cov}(X, Y) = 0 \not\Longrightarrow X \perp\!\!\!\perp Y$

Let Z be a random variable with support $\{-1, 1\}$ such that

$$\mathbb{P}(Z = 1) = \mathbb{P}(Z = -1) = \tfrac{1}{2}.$$

Let X be a random variable, independent of Z, with a symmetric density f_X and a finite second-order moment.
Define $Y = ZX$.

Then X and Y are not independent: conditional on $X = x$, the variable Y can only take the values x or $-x$.

$\mathrm{Cov}(X, Y) = 0 \not\Longrightarrow X \perp\!\!\!\perp Y$

We compute the covariance between X and $Y = ZX$.

Since X has a symmetric density, $\mathbb{E}(X) = 0$. In addition, since Z is symmetric, $\mathbb{E}(Z) = 0$.

First, since Z and X are independent:

$$\mathbb{E}(Y) = \mathbb{E}(ZX) = \mathbb{E}(Z)\mathbb{E}(X) = 0,$$

Next, $$\mathrm{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\,\mathbb{E}(Y) = \mathbb{E}(X \cdot ZX) = \mathbb{E}(ZX^2).$$

Using independence again,

$$\mathbb{E}(ZX^2) = \mathbb{E}(Z)\,\mathbb{E}(X^2) = 0.$$

Therefore, even though $\mathrm{Cov}(X, Y) = 0$, X and Y are not independent.

5.6. Quantiles

Quantiles of a Discrete Random Variable

Definition 21. Quantile of a Discrete Random Variable

Let X be a discrete random variable with distribution function F_X and let $\alpha \in [0, 1]$.
*The **quantile of order** α of X is defined as*

$$q_X(\alpha) = \inf\{x \in \mathbb{R} : F_X(x) \geq \alpha\}.$$

- In the discrete case, the distribution function F_X is **non-decreasing and stepwise**.
- As a result, F_X is **not invertible** in the usual sense.
- Quantiles are therefore defined using a **generalized inverse**.
- A quantile of order α is a value below which at least $\alpha \times 100\%$ of the probability mass lies.

Note on the Infimum

Meaning of the Notation

The expression

$$\inf\{x \in \mathbb{R} \ : \ F_X(x) \geq \alpha\}$$

is read as: "the smallest real number x such that the distribution function $F_X(x)$ is greater than or equal to α"

- The set $\{x \in \mathbb{R} : F_X(x) \geq \alpha\}$ contains all values of x for which the cumulative probability has reached at least α.
- The symbol $\inf$ (infimum) means the **smallest** value in this set.
- In the discrete case, this corresponds to the **first jump** of F_X that reaches or exceeds the level α.

Example: Quantiles of a Discrete Random Variable

Let X be a discrete random variable with support $\mathcal{D}_X = \{1, 2, 3\}$, and probability mass function

$$\mathbb{P}(X = 1) = 0.2, \qquad \mathbb{P}(X = 2) = 0.5, \qquad \mathbb{P}(X = 3) = 0.3.$$

The distribution function F_X is given by:

$$F_X(1) = 0.2, \quad F_X(2) = 0.7, \quad F_X(3) = 1.$$

- $q_X(0.25) = 2$, since $F_X(1) < 0.25 \leq F_X(2)$,
- $q_X(0.5) = 2$,
- $q_X(0.8) = 3$.

Median

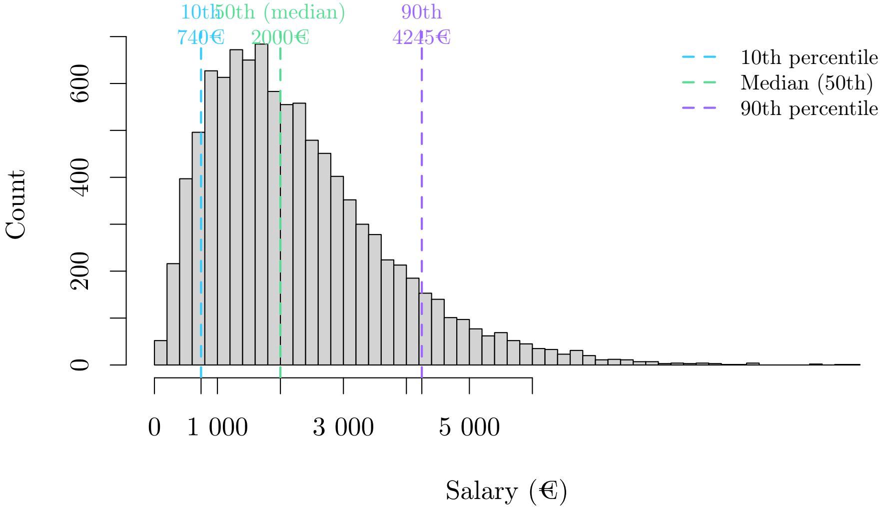
Median

*The quantile of order $\alpha = \frac{1}{2}$ is called a **median**.*
In the discrete case, the median may not be unique, but the definition

$$q_X(\tfrac{1}{2}) = \inf\{x : F_X(x) \geq \tfrac{1}{2}\}$$

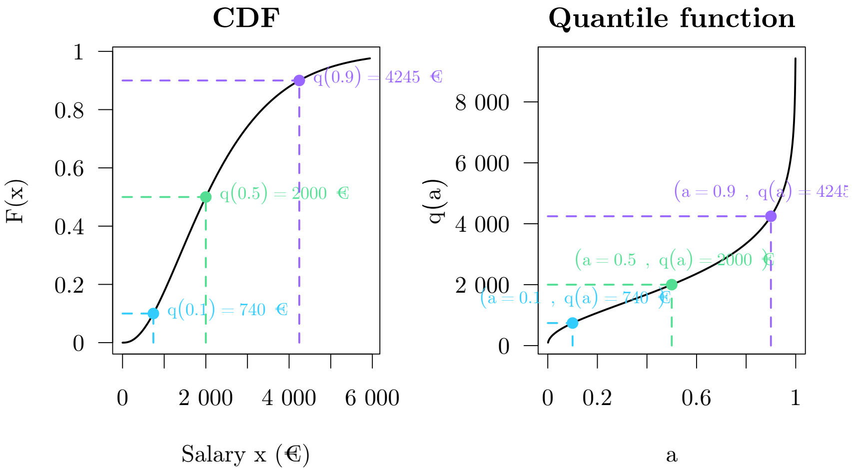
selects a specific value.

Distribution of Salaries



Three Specific Quantiles of on the Distribution of Salaries.

CDF and Quantiles



CDF of X (left) and "inverse CDF" of X (right).

6. Exercises

Exercise

Let X be a real-valued random variable with density f_X defined by

$$f_X(x) = \begin{cases} 0 & \textit{if } x < 0, \\ 4xe^{-2x} & \textit{if } x \geq 0. \end{cases}$$

1. *Verify that f_X is a probability density function.*
2. *Compute the expectation and the variance of X.*

Solution: Density (1/3)

To be a probability density function, f_X must satisfy two conditions:

1. **Nonnegativity.** For $x \geq 0$, we have $4xe^{-2x} \geq 0$, and for $x < 0$, $f_X(x) = 0$. Hence, $f_X(x) \geq 0$ for all $x \in \mathbb{R}$.
2. **Normalization.** We must check that $\int_{\mathbb{R}} f_X(x)\, dx = 1$.

   Since $f_X(x) = 0$ for $x < 0$,

$$\int_{\mathbb{R}} f_X(x)\, dx = \int_0^{+\infty} 4xe^{-2x}\, dx.$$

Solution: Density (2/3)

Let $I = \int_0^{+\infty} xe^{-2x}\, dx$. Let us define:

- $u(x) = x$, and $u'(x) = 1$,
- $v'(x) = e^{-2x}$, which gives $v(x) = -\frac{1}{2}e^{-2x}$

Using integration by parts, we obtain

$$
\begin{aligned}
I &= \int_0^{+\infty} u(x)v'(x)\, dx = [u(x)v(x)]_0^{+\infty} - \int_0^{+\infty} u'(x)v(x)\, dx \\
&= \left[-\frac{x^2}{2}e^{-2x}\right]_0^{+\infty} + \int_0^{+\infty} \frac{1}{2}e^{-2x}\, dx \\
&= \left(\lim_{x\to+\infty} -\frac{x^2}{2}e^{-2x} - 0\right) + \left[-\frac{1}{4}e^{-2x}\right]_0^{+\infty} \\
&= (0-0) + \left(\lim_{x\to+\infty}\left(-\frac{1}{4}e^{-2x}\right) - \left(-\frac{1}{4}e^{0}\right)\right) \\
&= 0 + \left(0 + \frac{1}{4}\right) = \frac{1}{4}.
\end{aligned}
$$

Solution: Density (3/3)

We calculated

$$I = \int_0^{+\infty} xe^{-2x}\,dx = \frac{1}{4}.$$

Therefore,

$$\int_{\mathbb{R}} f_X(x)\,dx = 4\int_0^{+\infty} xe^{-2x}\,dx = 4I = 1.$$

Hence, f_X is a valid probability density function.

Solution: Expectation (1/3)

The expectation of X is given by

$$\mathbb{E}(X) = \int_0^{+\infty} x\, f_X(x)\, dx = \int_0^{+\infty} x \cdot 4xe^{-2x}\, dx = 4\int_0^{+\infty} x^2 e^{-2x}\, dx.$$

Let $J = \int_0^{+\infty} x^2 e^{-2x}\, dx$.

Solution: Expectation (2/3)

We have $J = \int_0^{+\infty} x^2 e^{-2x}\, dx$. Again, we use integration by parts:

- $u(x) = x^2 \qquad u'(x) = 2x, \qquad v'(x) = e^{-2x} \qquad v(x) = -\frac{1}{2}e^{-2x}$.

Then,

$$
\begin{aligned}
J &= [u(x)v(x)]_0^{+\infty} - \int_0^{+\infty} u'(x)v(x)\, dx \\
&= \left[-\frac{x^2}{2}e^{-2x}\right]_0^{+\infty} + \int_0^{+\infty} \frac{2x}{2}e^{-2x}\, dx \\
&= (0-0) + \int_0^{+\infty} xe^{-2x}\, dx \\
&= \int_0^{+\infty} xe^{-2x}\, dx.
\end{aligned}
$$

Let $K = \int_0^{+\infty} xe^{-2x}\, dx$.

Solution: Expectation (3/3)

We have $K = \int_0^{+\infty} xe^{-2x}\,dx$. Consider a second integration by parts:

- $u(x) = x \qquad u'(x) = 1 \qquad v'(x) = e^{-2x} \qquad v(x) = -\frac{1}{2}e^{-2x}$.

Then,

$$
\begin{aligned}
K &= \left[-\frac{x}{2}e^{-2x}\right]_0^{+\infty} + \int_0^{+\infty} \frac{1}{2}e^{-2x}\,dx \\
&= (0-0) + \left[-\frac{1}{4}e^{-2x}\right]_0^{+\infty} \\
&= 0 + \frac{1}{4} = \frac{1}{4}.
\end{aligned}
$$

Hence,

$$\mathbb{E}(X) = 4J = 4K = 1.$$

Solution: Variance

We have: $\text{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$.

We first compute the second moment:

$$\mathbb{E}(X^2) = \int_0^{+\infty} x^2 f_X(x)\, dx = \int_0^{+\infty} x^2 \cdot 4xe^{-2x}\, dx = \int_0^{+\infty} 4x^3 e^{-2x}\, dx.$$

Using integration by parts:

- $u(x) = 4x^3$ $\quad u'(x) = 12x$ $\quad v'(x) = e^{-2x}$ $\quad v(x) = -\frac{1}{2}e^{-2x}$.

Hence,

$$\begin{aligned}\mathbb{E}(X^2) &= \left[-2x^3 e^{-2x}\right]_0^{+\infty} + \int_0^{+\infty} \frac{12x}{2} e^{-2x}\, dx \\ &= 6\int_0^{+\infty} x^2 e^{-2x}\, dx = 6J = 6 \times \frac{1}{4} = \frac{3}{2}.\end{aligned}$$

Finally,

$$\text{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \frac{3}{2} - 1 = \frac{1}{2}.$$

Exercise

Let X be a real-valued random variable with density $f_X(x)$ defined by

$$f_X(x) = \begin{cases} 0 & \textit{if} \quad x < 1, \\ \frac{\alpha \ln x}{x^3} & \textit{if} \quad x \geq 1. \end{cases}$$

1. *Calculate $\alpha > 0$ such that f_X is a probability density function.*
2. *Compute the expectation and the variance of X.*

Solution: Value for α (1/3)

To be a probability density function, f_X must satisfy two conditions:

1. **Nonnegativity. (1/)** For $x < 1$, we have $f_X(x) = 0$. For $x \geq 1$,

$$\frac{\alpha \ln x}{x^3} \geq 0 \iff \alpha \ln x \geq 0 \quad (\text{since } x^3 \geq 0 \quad \text{for } x \geq 1)$$
$$\iff \alpha \geq 0 \quad (\text{since } \ln x \geq 0 \quad \text{for } x \geq 1).$$

Hence, for $x \geq 1$, $\ln(x) \geq 0$, hence $\alpha > 0 \implies f_X(x) \geq 0$.

Solution: Value for α (2/3)

1. **Normalization.** We must check that $\int_{\mathbb{R}} f_X(x)\,dx = 1$.

$$\int_{\mathbb{R}} f_X(x)\,dx = \int_1^{+\infty} \frac{\alpha \ln x}{x^3}\,dx = \alpha \int_1^{+\infty} \ln(x)\,x^{-3}\,dx$$

Let $I = \int_1^{+\infty} \ln(x)\,x^{-3}\,dx$. We integrate by parts. Let:

- $u(x) = \ln(x)$ $\quad u'(x) = 1/x$ $\quad v'(x) = x^{-3}$ $\quad v(x) = -\frac{1}{2}x^{-2}$.

Hence,

$$\begin{aligned}
I &= \left[-\frac{1}{2}x^{-2}\,\ln(x)\right]_1^{+\infty} + \int_1^{+\infty} \frac{1}{2x}x^{-2}\,dx \\
&= \left[-\frac{\ln(x)}{2x^2}\right]_1^{+\infty} + \int_1^{+\infty} \frac{1}{2}x^{-3}\,dx \\
&= \frac{1}{2}\int_1^{+\infty} x^{-3}\,dx.
\end{aligned}$$

$\lim_{x\to+\infty} -\frac{\ln(x)}{2x^2} = 0$, since $x^2 \to +\infty$ much faster than $\ln(x)$ when $x \to +\infty$.

Solution: Value for α (3/3)

Hence,

$$
\begin{aligned}
I &= \frac{1}{2}\int_1^{+\infty} x^{-3}\,dx \\
&= \frac{1}{2}\left[-\frac{1}{2}x^{-2}\right]_1^{+\infty} \\
&= \frac{1}{2}\left(0-(-\frac{1}{2})\right) = \frac{1}{4}
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\int_{\mathbb{R}} f_X(x)\,dx = 1 &\iff \alpha I = 1 \\
&\iff \alpha\frac{1}{4} = 1 \\
&\iff \alpha = 4.
\end{aligned}
$$

The function f_X is a density function iff $\alpha = 4$.

Solution: Expectation

We compute

$$\mathbb{E}(X) = \int_{\mathbb{R}} x\, f_X(x)\, dx = \int_1^{+\infty} \frac{4x\ln(x)}{x^3}\, dx = 4\int_1^{+\infty} \ln(x)x^{-2}\, dx.$$

Let $J = \int_1^{+\infty} \ln(x)x^{-2}\, dx$. We integrate by parts:

- $u(x) = \ln(x) \qquad u'(x) = 1/x \qquad v'(x) = x^{-2} \qquad v(x) = -x^{-1}$.

Hence,

$$J = \left[-\frac{\ln(x)}{x}\right]_1^{+\infty} + \int_1^{+\infty} \frac{1}{x^2}\, dx = \left[-x^{-1}\right]_1^{+\infty} = 0 - (-1) = 1.$$

Therefore, $\mathbb{E}(X) = 4J = 4$.

Solution: Variance

Recall that $\mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$, provided that $\mathbb{E}(X^2)$ is finite.

We compute the second moment:

$$\mathbb{E}(X^2) = \int_{\mathbb{R}} x^2 f_X(x)\,dx = \int_1^{+\infty} x^2 \cdot \frac{4\ln(x)}{x^3}\,dx = 4\int_1^{+\infty} \frac{\ln(x)}{x}\,dx.$$

Let

$$L = \int_1^{+\infty} \frac{\ln(x)}{x}\,dx.$$

Using the change of variable $u = \ln(x)$ (so $du/dx = 1/x$, hence $du = dx/x$), we obtain

$$L = \int_0^{+\infty} u\,du = +\infty.$$

Therefore, $\mathbb{E}(X^2) = +\infty$.

Since the second moment does not exist, the variance of X is not finite: $\mathrm{Var}(X) = +\infty$

7. Transformations of Continuous Random Variables

Foreword

- In many applications, we are interested in the distribution of a random variable Y defined as
$$Y = g(X),$$
where X is an absolutely continuous random variable and g is a function with suitable regularity properties.
- Recall from earlier that under such conditions, Y is itself a random variable.
- Our goal in this section is to derive the distribution (and in particular the density) of Y from the density of X.
- Throughout this section, we assume that:
  - X is a continuous random variable with density f_X,
  - g is a real-valued function defined on an interval (α, β),
  - the support of X satisfies
$$D_X \subset (\alpha, \beta),$$
where α and β may take the values $-\infty$ and $+\infty$.

Why Are Monotone Transformations So Important?

- In this section, we will distinguish between:
  - **monotone transformations**, and
  - **non-monotone transformations**, which require special treatment.
- The cases where g is **strictly increasing or strictly decreasing** cover most transformations used in practice. They appear systematically when:
  - rescaling or shifting a random variable (standardization),
  - changing units (e.g. euros to thousands of euros),
  - transforming distributions to simplify calculations,
  - constructing new random variables with a prescribed support.
- In these situations, the transformation is one-to-one (i.e. bijective on its domain), so:
  - the cumulative distribution function of Y is easy to compute,
  - the density formula involves a single inverse function,
  - no summation over multiple preimages is required.

When g Is Strictly Increasing (1/2)

- Assume that g **is strictly increasing** on (α, β).
- Then g is one-to-one (bijective) and admits an inverse function

$$g^{-1} : (g(\alpha), g(\beta)) \to (\alpha, \beta).$$

- Let us compute the cumulative distribution function of $Y = g(X)$.

For $y \in (g(\alpha), g(\beta))$,

$$\begin{aligned} F_Y(y) &= \mathbb{P}(Y \leq y) \\ &= \mathbb{P}(g(X) \leq y) \\ &= \mathbb{P}(X \leq g^{-1}(y)) \quad (g \text{ is strictly increasing}) \\ &= F_X(g^{-1}(y)). \end{aligned}$$

And if $y \notin (g(\alpha), g(\beta))$,

$$F_Y(y) = \begin{cases} 0, & y < g(\alpha), \\ 1, & y > g(\beta). \end{cases}$$

When g Is Strictly Increasing (2/2)

- Assume in addition that g is **differentiable**.
- Then $Y = g(X)$ admits a density.

Proposition 23. Density under a Strictly Increasing Transformation

Under the previous assumptions, the random variable $Y = g(X)$ has a density f_Y given by

$$f_Y(y) = f_X(g^{-1}(y))\,(g^{-1})'(y),$$

for $y \in (g(\alpha), g(\beta))$, and

$$f_Y(y) = 0 \quad \textit{otherwise.}$$

Proof (1/2)

Let $y \in (g(\alpha), g(\beta))$. From the previous slide, the cumulative distribution function of $Y = g(X)$ is

$$F_Y(y) = F_X(g^{-1}(y)).$$

Define two functions:

$$u(y) := g^{-1}(y), \qquad h(x) := F_X(x).$$

Then

$$F_Y(y) = (h \circ u)(y).$$

- For the outer function, since X has density f_X, the CDF F_X is differentiable and
$$h'(x) = F_X'(x) = f_X(x).$$
- Because g is differentiable and strictly increasing, its inverse is differentiable on $(g(\alpha), g(\beta))$, so $u'(y) = (g^{-1})'(y)$ exists.

Proof (2/2)

We have:

$$F_Y(y) = (h \circ u)(y), \quad h'(x) = f_X(x), \quad \text{and } u'(y) = (g^{-1})'(y)$$

By the chain rule,

$$F_Y'(y) = (h \circ u)'(y) = h'(u(y))\, u'(y).$$

Hence,

$$f_Y(y) = F_Y'(y) = f_X(g^{-1}(y))\, (g^{-1})'(y).$$

For $y \notin (g(\alpha), g(\beta))$, F_Y is constant, hence $f_Y(y) = 0$. $\square$.

Example

Let X be a random variable with cumulative distribution function

$$F_X(x) = \begin{cases} 1 - e^{-x}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

The support of X is $\mathcal{D}_X = (0, +\infty)$. The CDF F_X is continuously differentiable on $\mathbb{R}$, and the density of X is

$$f_X(x) = \begin{cases} e^{-x}, & x > 0, \\ 0, & x \leq 0. \end{cases}$$

Let $g(x) = e^x$ and define $Y = g(X)$.
What is the density of Y?

g is strictly increasing and differentiable on $\mathbb{R}$. Its inverse is

$$g^{-1}(y) = \ln(y), \qquad y > 0.$$

Since $X \in (0, +\infty)$, the support of Y is $\mathcal{D}_Y = (1, +\infty)$.

Using the density transformation formula,

$$f_Y(y) = f_X(g^{-1}(y))\,(g^{-1})'(y), \qquad y > 1.$$

We compute

$$f_X(\ln y) = e^{-\ln y} = \frac{1}{e^{\ln y}} = \frac{1}{y}, \qquad (g^{-1})'(y) = \frac{1}{y}.$$

Therefore,

$$f_Y(y) = \begin{cases} \dfrac{1}{y^2}, & y > 1, \\ 0, & y \leq 1. \end{cases}$$

When g Is Strictly Decreasing (1/2)

- Assume that g **is strictly decreasing** on (α, β).
- Then g is one-to-one (bijective) and admits an inverse function

$$g^{-1} : (g(\beta), g(\alpha)) \to (\alpha, \beta).$$

- Let us compute the cumulative distribution function of $Y = g(X)$.

For $y \in (g(\beta), g(\alpha))$,

$$\begin{aligned} F_Y(y) &= \mathbb{P}(Y \leq y) \\ &= \mathbb{P}(g(X) \leq y) \\ &= \mathbb{P}(X \geq g^{-1}(y)) \quad (g \text{ is strictly decreasing}) \\ &= 1 - F_X(g^{-1}(y)). \end{aligned}$$

And if $y \notin (g(\beta), g(\alpha))$,

$$F_Y(y) = \begin{cases} 0, & y < g(\beta), \\ 1, & y > g(\alpha). \end{cases}$$

When g Is Strictly Decreasing (2/2)

- Assume in addition that g is **differentiable**.
- Then $Y = g(X)$ admits a density.

Proposition 24. Density under a Strictly Decreasing Transformation

Under the previous assumptions, the random variable $Y = g(X)$ has a density f_Y given by

$$f_Y(y) = -\, f_X(g^{-1}(y))\,(g^{-1})'(y),$$

for $y \in (g(\beta), g(\alpha))$, and

$$f_Y(y) = 0 \quad \textit{otherwise.}$$

Proof (1/2) (read it at home)

Let $y \in (g(\beta), g(\alpha))$. From the previous slide,

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(g(X) \leq y) = \mathbb{P}\big(X \geq g^{-1}(y)\big) = 1 - F_X\big(g^{-1}(y)\big).$$

Define

$$u(y) := g^{-1}(y), \qquad h(x) := 1 - F_X(x).$$

Then

$$F_Y(y) = (h \circ u)(y).$$

- Since X has density f_X, the CDF F_X is differentiable and

$$h'(x) = \big(1 - F_X(x)\big)' = -F'_X(x) = -f_X(x).$$

- Because g is differentiable and strictly decreasing, its inverse is differentiable on $(g(\beta), g(\alpha))$, so $u'(y) = (g^{-1})'(y)$ exists.

Proof (2/2) (read it at home)

We have:

$$F_Y(y) = (h \circ u)(y), \quad h'(x) = -f_X(x), \quad u'(y) = (g^{-1})'(y).$$

By the chain rule,

$$F_Y'(y) = (h \circ u)'(y) = h'(u(y))\, u'(y).$$

Hence,

$$f_Y(y) = F_Y'(y) = -\, f_X(g^{-1}(y))\, (g^{-1})'(y), \qquad y \in (g(\beta), g(\alpha)).$$

For $y \notin (g(\beta), g(\alpha))$, F_Y is constant, hence $f_Y(y) = 0$. □

Unified Density Formula for Monotone Transformations

Let X be a continuous random variable with density f_X, and let $g : (\alpha, \beta) \to \mathbb{R}$ be a ***monotone and differentiable*** *function.*
Define $Y = g(X)$. Then Y admits a density f_Y given by

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

for all y in the image of (α, β) by g, and

$$f_Y(y) = 0 \quad \text{otherwise.}$$

In particular, we have:

- *if g is strictly increasing, $(g^{-1})'(y) > 0$;*
- *if g is strictly decreasing, $(g^{-1})'(y) < 0$.*

Special Case: g Is an Affine Function

- Consider the affine transformation $g(x) = ax + b$, where $a \neq 0$ and $b \in \mathbb{R}$.
- Let $Y = aX + b$ (this will be useful when standardizing random variables).

Proposition 25. Affine Transformation of a Continuous Random Variable

Let X be a continuous random variable with density f_X.
Then the random variable $Y = aX + b$ admits a density f_Y given by

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right), \qquad y \in \mathbb{R}.$$

In particular:

- *if $a > 0$, the transformation is strictly increasing;*
- *if $a < 0$, the transformation is strictly decreasing.*

Example: An Affine Transformation

Let X be a continuous random variable with density $f_X(x) = e^{-x}\,\mathbf{1}_{\{x>0\}}$.
Let $Y = 2X + 1$.

Since $g(x) = 2x + 1$ is strictly increasing and differentiable, the affine transformation formula applies:

$$f_Y(y) = \frac{1}{|2|}\, f_X\left(\frac{y-1}{2}\right).$$

Therefore,

$$f_Y(y) = \begin{cases} \dfrac{1}{2}\, e^{-(y-1)/2}, & y > 1, \\ 0, & y \leq 1. \end{cases}$$

What Happens for a Non-Monotone Transformation?

- In the previous results, the transformation g was assumed to be **monotone** (hence one-to-one).
- In that case, each value y has a *unique preimage*

$$g(x) = y \quad \Longleftrightarrow \quad x = g^{-1}(y),$$

  and the density transformation formula involves a single term.
- When g is **not one-to-one**, several values of X may lead to the same value of Y.
- The probability mass associated with all these values must be **added**.

If $g(x) = y$ has several solutions $x_1, \ldots, x_k$, then the density of Y at y is obtained by summing the contributions of all these points.

Let us illustrate this phenomenon with an example.

Example: A Non-Monotone Transformation

Let

$$g(x) = x^2, \qquad Y = g(X) = X^2.$$

The function g is **not injective** on $\mathbb{R}$, since

$$g(x) = g(-x) = x^2 \quad \text{for all } x \in \mathbb{R}.$$

1. In a first step, we compute the cumulative distribution function. For $y \geq 0$,

$$\begin{aligned} F_Y(y) &= \mathbb{P}(Y \leq y) = \mathbb{P}(X^2 \leq y) \\ &= \mathbb{P}(-\sqrt{y} \leq X \leq \sqrt{y}) \\ &= F_X(\sqrt{y}) - F_X(-\sqrt{y}). \end{aligned}$$

   For $y < 0$, $F_Y(y) = 0$.

2 In a second step, we calculate the density of Y. Assume that X admits a density f_X. For $y > 0$, differentiating $F_Y(y)$ gives

$$\begin{aligned} f_Y(y) &= \frac{d}{dy}[F_X(\sqrt{y}) - F_X(-\sqrt{y})] \\ &= \frac{1}{2\sqrt{y}} f_X(\sqrt{y}) + \frac{1}{2\sqrt{y}} f_X(-\sqrt{y}). \end{aligned}$$

Therefore,

$$f_Y(y) = \begin{cases} \frac{1}{2\sqrt{y}}\Big(f_X(\sqrt{y}) + f_X(-\sqrt{y})\Big), & y > 0, \\ 0, & y \leq 0. \end{cases}$$

References I

Blitzstein, J. K. and Hwang, J. (2019). *Introduction to probability*. Chapman and Hall/CRC.

Hansen, B. (2022). *Probability and statistics for economists*. Princeton University Press.

A. Additional Proof

Expectation Preserves Inequalities

Proposition 26. Expectation Preserves Inequalities

Let X and Y be integrable random variables such that $X \leq Y$ almost surely. Then,

$$\mathbb{E}(X) \leq \mathrm{Var}(X)$$

Proof. Since $X \leq Y$ a.s., then $Y - X \geq 0$ a.s. The expectation of a nonnegative random variable is nonnegative, hence $\mathbb{E}(Y - X) \geq 0$.

By linearity of expectation,

$$\mathbb{E}(Y - X) = \mathbb{E}(Y) - \mathbb{E}(X),$$

which implies

$$\mathbb{E}(X) \leq \mathbb{E}(Y).$$

Back to the proof of nonnegativity of the variance.