Probability

4. Basic Probability Theory

Ewen Gallic

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

amU FEG
Aix Marseille Université

Foreword

In this chapter, we move beyond single events and study how probabilities behave when events interact.

We introduce:

- joint events and joint probability tables,
- conditional probability and its interpretation,
- Bayes' rule and probability updating.

This chapter is based on:

- Mathieu Faure's (Aix Marseille Univ.) lecture slides, kindly shared,
- Chapter 1 of Hansen (2022),
- Chapters 3.1–3.4 of McClave et al. (2022).

1. Refresher on Outcomes and Events

Refresher

- In Chapter 2, we defined multiple notions:
  - a random experiment,
  - an outcome,
  - the sample space,
  - an event.
- Let us make a brief recap here.

Random Experiments and Outcomes

- A **random experiment** is a procedure that can be repeated under identical conditions and whose result cannot be predicted with certainty.
  - Example: rolling a fair six-sided die.
- An **outcome** is a single possible result of a random experiment.
  - Example: the number 6 appears.
- The set of all possible outcomes is called the **sample space**, denoted by Ω.
  - Example: $\Omega = \{1, 2, 3, 4, 5, 6\}$.
- An **event** A is a subset of outcomes Ω.
  - Example: $A = \{2, 4, 6\}$.
- An event containing exactly one outcome is called an **elementary event**.
  - Example: $A = \{2\}$.

2. Probability Function

Probability: Intuition and Notation

- The probability of an event A is denoted by $\mathbb{P}(A)$ (also written $P(A)$ or $\Pr(A)$).
- For an elementary event $\omega \in \Omega$, we write

$$\mathbb{P}(\omega) = \mathbb{P}(\{\omega\}).$$

- A **probability function** assigns a numerical value to each event:

$$\mathbb{P} : \mathcal{P}(\Omega) \longrightarrow [0, 1].$$

- Intuitively, $\mathbb{P}(A)$ measures how **likely** event A is to occur.
  - In simple settings, it can be interpreted as a **proportion**: the relative size of A compared to the whole sample space Ω.
  - Values close to 1 indicate very likely events, while values close to 0 indicate rare events.

Probability Function

Given a fundamental set Ω, equipped with a σ-algebra $\mathcal{A}$, a **probability** is a function that assigns to each event a real number between 0 and 1.

Definition 1. Probability Function (Three Axioms, Kolmogoroff, 1933)

*A **probability** on $(\Omega, \mathcal{A})$ is a function $\mathbb{P} : \mathcal{A} \to [0, 1]$ such that:*

(P1) $\mathbb{P}(A) \geq 0$,

(P2) $\mathbb{P}(\Omega) = 1$,

(P3) *For any sequence $(A_j)_{j \geq 1}$ of pairwise disjoint events (i.e. $A_i \cap A_j = \varnothing$ for $i \neq j$),*

$$\mathbb{P}\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} \mathbb{P}(A_j).$$

Some Remarks on the Definition

- A probability function $\mathbb{P}$ assigns a number to each event, not to individual outcomes directly.

1. The condition $\mathbb{P}(A) \geq 0$ means that probabilities are non-negative.
2. The condition $\mathbb{P}(\Omega) = 1$ means that
   - something in the sample space always happens (the probability that something happens is 1),
   - probabilities are measured relative to the whole universe.
3. The additivity property states that:
   - if events cannot occur simultaneously,
   - then the probability that one of them occurs is the sum of their individual probabilities.
   - That is, if A and B are disjoint, we have $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$.

Additivity Property: When Does It Apply?

- Consider a random experiment: rolling a fair six-sided die.
- The sample space is $\Omega = \{1, 2, 3, 4, 5, 6\}$.

Disjoint events

- Each outcome corresponds to an elementary event: $\{1\}, \{2\}, \dots, \{6\}$.
- These events are **mutually disjoint**: no two outcomes can occur at the same time.
- By the additivity property,

$$\mathbb{P}(\{1, 2\}) = \mathbb{P}(\{1\}) + \mathbb{P}(\{2\}).$$

Warning: applying the same reasoning to non-disjoint events is wrong!

Non-disjoint events

- Define the events:

$$A = \{\text{even number}\} = \{2, 4, 6\},$$
$$B = \{\text{number greater than } 3\} = \{4, 5, 6\}.$$

- The events A and B are **not disjoint**:

$$A \cap B = \{4, 6\} \neq \varnothing.$$

- Therefore,

$$\mathbb{P}(A \cup B) \neq \mathbb{P}(A) + \mathbb{P}(B).$$

- Adding probabilities would count outcomes in $A \cap B$ twice.

What Does a Probability Mean? Two interpretations

Frequentist interpretation

- Probability is the **long-run relative frequency** of an event in **repeated, identical experiments**.
  - Examples: the probability that the stock market increases, the probability that a soccer player scores a penalty.
- Repetition of the experiment allows frequencies to be observed.
- Some events cannot be repeated:
  - e.g. “global warming will exceed $2°$C”.
  - Although no empirical frequency is available, we may still reason using hypothetical or counterfactual worlds evolving randomly from similar initial conditions.

Subjective interpretation

- Probability represents a degree of belief.
- Even subjective probabilities must satisfy the axioms of probability.

Example 1: Accident Risk

- Experiment: randomly selecting an individual from a population.
- Define the events:

$$A = \{\text{the individual has a car accident within the year}\},$$
$$A^c = \{\text{no accident within the year}\}.$$

- The sample space is: $\Omega = \{A, A^c\}$.
- Using past data, we estimate: $\mathbb{P}(A) = 0.05, \qquad \mathbb{P}(A^c) = 0.95.$
- Interpretation: about 5% of randomly selected individuals experience a car accident within one year.

Example 2: Income Threshold

- Experiment: randomly selecting an individual from a population.
- Define the events:

$$A = \{\text{earns more than €12 per hour}\},$$
$$A^c = \{\text{earns at most €12 per hour}\}.$$

- Using a representative sample of $2{,}000$ individuals, we estimate:
  $\mathbb{P}(A) = 0.15, \qquad \mathbb{P}(A^c) = 0.85.$
- Interpretation:
  - Before observing the individual, their income category is unknown.
  - Probability reflects uncertainty about a randomly selected individual.

3. Properties of the Probability Function

Properties of the Probability Function

Proposition 1. Properties of the Probability Function

Let $\mathbb{P}$ be a probability on $(\Omega, \mathcal{A})$. Then the following properties hold:

1. ***Complement rule**: For any event $A \in \mathcal{A}$, $\mathbb{P}(A) = 1 - \mathbb{P}(A^c)$.*
2. ***Empty set**: $\mathbb{P}(\varnothing) = 0$.*
3. ***Upper bound**: $\mathbb{P}(A) \leq 1$*
4. ***Monotonicity**: If $A \subseteq B$, then $\mathbb{P}(A) \leq \mathbb{P}(B)$.*
5. ***Inclusion–exclusion**: For any events $A, B \in \mathcal{A}$,*

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

6. ***Boole's Inequality**: $\mathbb{P}(A \cup B) \leq \mathbb{P}(A) + \mathbb{P}(B)$.*
7. ***Bonferroni's Inequality**: $\mathbb{P}(A \cap B) \geq \mathbb{P}(A) + \mathbb{P}(B) - 1$*

Proofs of the Propositions (1/3)

- **Property 1 (Complement rule).** Since A and A^c are disjoint and $A \cup A^c = \Omega$, by Axiom **(P2)** and Axiom **(P3)**:

$$1 = \mathbb{P}(\Omega) = \mathbb{P}(A \cup A^c) = \mathbb{P}(A) + \mathbb{P}(A^c).$$

- **Property 2 (Empty set).** This follows immediately from Property 1 and Axiom **(P2)**, since

$$\varnothing = \Omega^c.$$

- **Property 3 (Upper bound).** Also from Property 1,
- **Property 4 (Monotonicity).** Let $A \subseteq B$. Then

$$B = A \cup (B \setminus A),$$

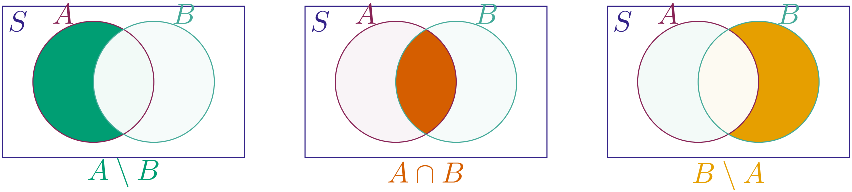
where A and $B \setminus A$ are disjoint. Hence,

$$\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A) \geq \mathbb{P}(A).$$

Proofs of the Propositions (2/3)

- **Property 5 (Inclusion–exclusion).** Note that:

$$A \setminus B = A - A \cap B$$



we obtain:

$$\begin{aligned}
\mathbb{P}(A \cup B) &= \mathbb{P}(A \setminus B) + \mathbb{P}(B \setminus A) + \mathbb{P}(A \cap B) \\
&= \mathbb{P}(A) - \mathbb{P}(A \cap B) + \mathbb{P}(B) - \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B) \\
&= \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).
\end{aligned}$$

Proofs of the Propositions (3/3)

- **Property 6 (Boole's Inequality).** From the Inclusion-Exclusion Principle (Proposition 5) and **(P1)**:

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \underbrace{\mathbb{P}(A \cap B)}_{\geq 0 \ (P1)} \leq \mathbb{P}(A) + \mathbb{P}(B).$$

- **Property 7 (Bonferroni's Inequality).** Rearranging the Inclusion-Exclusion Principle (Proposition 5) and using $\mathbb{P}(A \cup B) \leq 1$ (Proposition 3):

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cup B) \geq \mathbb{P}(A) + \mathbb{P}(B) - 1.$$

□

4. Equally Likely Outcomes, Equally Likely Events

Equally Likely Outcomes: Intuition

- In many situations, **symmetry** implies that some outcomes are equally likely.
- Typical examples:
  - flipping a fair coin,
  - rolling a fair die.
- A coin is **fair** if the event "head" is as likely as the event "tail".
- A die is **fair** if each face has the same probability.

Principle of Equally Likely Outcomes

Definition 2. Principle of Equally Likely Outcomes

*If a random experiment has n possible **outcomes** $a_1, a_2, \ldots, a_n$ which are **equally likely**, then:*

$$\mathbb{P}(a_i) = \frac{1}{n}, \qquad i = 1, \ldots, N.$$

Examples

- *Fair coin:*

$$\mathbb{P}(\textit{Head}) = \mathbb{P}(\textit{Tail}) = \frac{1}{2}.$$

- *Fair die:*

$$\mathbb{P}(1) = \mathbb{P}(2) = \cdots = \mathbb{P}(6) = \frac{1}{6}.$$

Finite Sample Space Ω

If $\Omega = \{1, 2, \ldots, n\}$, then $\mathcal{A} = \mathcal{P}(\Omega)$, the set of all subsets of Ω.

Definition 3. Elementary events, Elementary Probabilities

- ***Elementary events***: *events of the form $\{i\}$ (singletons).*
- ***Elementary probabilities***:

$$p_i := \mathbb{P}(\{i\}), \qquad i = 1, \ldots, n.$$

Since the elementary events form a partition of Ω,

$$\sum_{i=1}^{n} p_i = 1.$$

Finite Sample Space Ω

The probability of any event $A \subseteq \Omega$ is the sum of the probabilities of the elementary events it contains.

Example

Let $\Omega = \{1, 2, \ldots, 6\}$ *and* $A = \{2, 4, 6\}$. *Then*

$$\mathbb{P}(A) = p_2 + p_4 + p_6.$$

Definition 4. Equiprobability of elementary events

We say that the elementary events are ***equally likely*** *if*

$$p_1 = p_2 = \cdots = p_n = \frac{1}{n}.$$

Principle of Equally Likely Events

Definition 5. Principle of Equally Likely Events

If the elementary events of a finite sample space $\Omega = \{1, 2, \ldots, n\}$ are equally likely, then for any event $A \subseteq \Omega$,

$$\mathbb{P}(A) = \frac{\textit{card}(A)}{\textit{card}(\Omega)} = \frac{\textit{card}(A)}{n}.$$

Equally Likely Events: A Two-Coin Example

- Consider an experiment consisting in **tossing two fair coins**.
- To fully describe an outcome, we must specify the result of each coin.
- The **elementary outcomes** are therefore:
$$\Omega = \{HH, HT, TH, TT\}.$$
- Since the coins are fair, these elementary outcomes are **equally likely**:
$$\mathbb{P}(HH) = \mathbb{P}(HT) = \mathbb{P}(TH) = \mathbb{P}(TT) = \tfrac{1}{4}.$$
- Consider the event
$$A = \{\text{one head and one tail}\}.$$
- Then
$$A = \{HT, TH\}, \qquad \text{and} \qquad \mathbb{P}(A) = \frac{\text{card}(A)}{\text{card}(\Omega)} = \frac{2}{4} = \frac{1}{2}.$$

Why Not $\Omega = \{HH, TT, HT\}$?

- We might be tempted to define $\Omega_2 = \{HH, TT, HT\}$.
- This set does not describe elementary outcomes!
- The label HT actually describes an **event**, grouping two distinct elementary outcomes:

$$HT = \{(H, T), (T, H)\}.$$

- As a result, the elements of $\{HH, TT, HT\}$ are **not equally likely**.
- The principle of equally likely events must only be applied to **elementary outcomes**.

5. Joint Events

Setting: Joint Events

Consider a randomly selected individual from a population, and define the events:

- H: the individual's wage exceeds €12 per hour,
- C: the individual has a college degree.

We are interested in the probability of the **joint event**: $H \cap C$, (the individual both earns more than €12 per hour and has a college degree).

Using sample data, we estimate:

$$\mathbb{P}(H) = 0.15, \qquad \mathbb{P}(C) = 0.45.$$

With this information alone, the joint probability $\mathbb{P}(H \cap C)$ cannot be determined exactly.

Bounding the Joint Probability

The joint probability $\mathbb{P}(H \cap C)$ can, however, be bounded. Since

$$H \cap C \subseteq H \quad \text{and} \quad H \cap C \subseteq C,$$

Property 4 (Monotonicity) implies:

$$0 \leq \mathbb{P}(H \cap C) \leq \min\{\mathbb{P}(H), \mathbb{P}(C)\} = 0.15.$$

Here, however, assume now that, using sample data, we estimate:

$$\mathbb{P}(H \cap C) = 0.06.$$

Joint Probability Table

We are given: $\mathbb{P}(H) = 0.15, \qquad \mathbb{P}(C) = 0.45, \qquad \mathbb{P}(H \cap C) = 0.06.$

We can thus compute the probabilities of various intersection, as in the table below (more details in the next slide).

	C	C^c	Any Educ.
H	0.06	0.09	0.15
H^c	0.39	0.46	0.85
Any Wage	0.45	0.55	1

Table 1: Joint Probabilities: Wages and Education

- The coloured cells are the **joint probabilities**.
- They sum to 1, since the corresponding events form a partition of Ω.
- Column totals give the probabilities of C and C^c.
- Row totals give the probabilities of H and H^c.

Filling the Probability Table

- We start from the estimated joint probability: $\mathbb{P}(H \cap C) = 0.06$.
- Since
$$H = (H \cap C) \cup (H \cap C^c),$$
with disjoint events, additivity implies:
$$\mathbb{P}(H \cap C^c) = \mathbb{P}(H) - \mathbb{P}(H \cap C) = 0.15 - 0.06 = 0.09.$$
- Similarly, since
$$C = (H \cap C) \cup (H^c \cap C),$$
we obtain:
$$\mathbb{P}(H^c \cap C) = \mathbb{P}(C) - \mathbb{P}(H \cap C) = 0.45 - 0.06 = 0.39.$$
- Finally, the four joint events form a partition of Ω, so:
$$\mathbb{P}(H^c \cap C^c) = 1 - (0.06 + 0.09 + 0.39) = 0.46.$$

6. Conditional Probability

Motivation

- We now work on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$.
- We are interested in situations in which the probability that an event A occurs can be influenced by the occurrence of another event B.
- For example:
  - does receiving information about cancer screening affect the likelihood of dying from cancer?
  - does being Black affect the likelihood of being invited for a job interview?
- There are questions of **conditional probability**.

Abstract View

- Suppose that event B has occurred.
- Then the only way for A to occur is if the outcome lies in the intersection $A \cap B$.
- The relevant question becomes:
  What is the probability that $A \cap B$ occurs, given that B has occurred?
- This probability is *not* simply $\mathbb{P}(A \cap B)$.
- Intuitively, we restrict attention to B and **renormalize** probabilities so that the total probability of B becomes 1.

Conditional Probability

Definition 6. Conditional Probability

Let $B \in \mathcal{A}$ be an event such that $\mathbb{P}(B) > 0$. The ***conditional probability of A given B*** *is defined by*

$$\mathbb{P}(A \mid B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}, \qquad A \in \mathcal{A}.$$

- $\mathbb{P}(A \mid B)$ is read:
  - "the probability of A **given** B", or
  - "the probability of A **assuming** B has occurred".
- $\mathbb{P}(A)$ is sometimes called the **unconditional probability** of A, to distinguish it from $\mathbb{P}(A \mid B)$.
- In general, $\mathbb{P}(A \mid B) \neq \mathbb{P}(A)$.

Proposition

Proposition 2.

The mapping $\mathbb{P}(\cdot \mid B) : \mathcal{A} \to [0, 1]$ *is itself a probability function.*

Proof (1/2)

We verify that the mapping $\mathbb{P}(\cdot \mid B)$ satisfies the axioms of a probability function.

- **Axiom (P1) (Non-negativity).**

  For any event $A \in \mathcal{A}$, since $\mathbb{P}(A \cap B) \geq 0$ and $\mathbb{P}(B) > 0$:

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \geq 0,$$

- **Axiom (P2) (Total probability).**

$$\mathbb{P}(\Omega \mid B) = \frac{\mathbb{P}(\Omega \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B)}{\mathbb{P}(B)} = 1.$$

Proof (2/2)

- **Axiom (P3) (Additivity).**

  Let A and A' be two disjoint events. Then $(A \cap B)$ and $(A' \cap B)$ are also disjoint, and:

$$\begin{aligned}\mathbb{P}(A \cup A' \mid B) &= \frac{\mathbb{P}((A \cup A') \cap B)}{\mathbb{P}(B)}\\ &= \frac{\mathbb{P}((A \cap B) \cup (A' \cap B))}{\mathbb{P}(B)}\\ &= \frac{\mathbb{P}(A \cap B) + \mathbb{P}(A' \cap B)}{\mathbb{P}(B)}\\ &= \mathbb{P}(A \mid B) + \mathbb{P}(A' \mid B).\end{aligned}$$

□

Example

Two cards are drawn ***without replacement*** *from a standard 52-card deck. The order in which the cards are drawn matters.*

The sample space Ω *consists of all ordered two-card hands. There are* $A_{52}^2 = 52 \times 51 = 2652$ *possible outcomes.*

Let: A_1*: "the first card drawn is an Ace",* A_2*: "the second card drawn is an Ace".*

The (unconditional) probability that the second card is an Ace is:

$$\mathbb{P}(A_2) = \frac{4}{52} = \frac{1}{13}.$$

If we know that the first card drawn is an Ace, then one Ace has been removed from the deck. The probability that the second card is an Ace is therefore smaller. More precisely:

$$\mathbb{P}(A_2 \mid A_1) = \frac{\mathbb{P}(A_1 \cap A_2)}{\mathbb{P}(A_1)} = \frac{12/2652}{1/13} = \frac{1}{17}.$$

Note on the Example

The (unconditional) probability that the second card is an Ace is:

$$\mathbb{P}(A_2) = \frac{4}{52} = \frac{1}{13}.$$

- *Before observing anything, the second card is just a random card from the deck.*
- *Without additional information, its probability of being an Ace is the same as for any single draw.*
- *Consider that the sample space has* $52 \times 51 = 2652$ *different outcomes. To have* A_2 *occur, the second card must be an Ace (4 choices), and the first card can then be any of the remaining 51 cards. Hence, the number of favourable outcomes is* $4 \times 51 = 204$. *Hence* $P(A_2) = \frac{204}{2652} = \frac{1}{13}$.

7. Bayes' Rule

Law of Total Probability

Let $A_1, \ldots, A_n$ be a **partition** of the sample space Ω, that is:

$$\mathbb{P}(A_i) > 0, \qquad A_i \cap A_j = \varnothing \text{ for } i \neq j, \qquad \bigcup_{i=1}^{n} A_i = \Omega.$$

- The events $A_1, \ldots, A_n$ represent mutually exclusive and exhaustive cases.

Definition 7. Law of Total Probability

- *Let B be any event. Since $(A_1, \ldots, A_n)$ is a partition of Ω, we can decompose B as:*

$$B = \bigcup_{i=1}^{n} (B \cap A_i),$$

where the events $(B \cap A_i)$ are disjoint.

- *Using additivity, we obtain the **law of total probability**:*

$$\mathbb{P}(B) = \sum_{i=1}^{n} \mathbb{P}(B \cap A_i) = \sum_{i=1}^{n} \mathbb{P}(A_i)\,\mathbb{P}(B \mid A_i).$$

Bayes' Rule (Reverend Thomas Bayes)

Theorem 1. Bayes' Rule

Let $(A_1, \ldots, A_n)$ be a partition of the sample space Ω such that $\mathbb{P}(A_i) > 0$ for all i. Let B be an event with $\mathbb{P}(B) > 0$. Then, for each $i = 1, \ldots, n$,

$$\mathbb{P}(A_i \mid B) = \frac{\mathbb{P}(B \mid A_i)\,\mathbb{P}(A_i)}{\sum_{j=1}^{n} \mathbb{P}(A_j)\,\mathbb{P}(B \mid A_j)}.$$

Using the law of total probability, the denominator can be written as:

$$\mathbb{P}(B) = \sum_{i=1}^{n} \mathbb{P}(A_i)\,\mathbb{P}(B \mid A_i).$$

Note

This formula allows to calculate $\mathbb{P}(A_i|B)$ from $\mathbb{P}(B|A)$.

How to Read Bayes' Rule

- $\mathbb{P}(A)$: **prior probability** of A.
- $\mathbb{P}(B \mid A)$: how likely B is if A is true.
- $\mathbb{P}(A \mid B)$: **updated probability** of A after observing B.

Bayes' rule describes how probabilities are updated when new information becomes available.

Example: Medical Screening Test

A laboratory has developed a medical screening test for a certain disease. The test has the following properties:

- *it gives a positive result for a healthy person in 2% of cases;*
- *it gives a positive result for a person who has the disease in 96% of cases.*

In a given population, 3% of individuals have the disease. If a randomly selected individual tests positive, what is the probability that the individual is in fact healthy?

Let:

- D: the individual has the disease,
- D^c: the individual is healthy,
- $+$: the test result is positive.

Solution

From the data, we get:

$$\mathbb{P}(D) = 0.03, \quad \mathbb{P}(D^c) = 0.97,$$
$$\mathbb{P}(+ \mid D) = 0.96, \quad \mathbb{P}(+ \mid D^c) = 0.02.$$

We want to compute $\mathbb{P}(H \mid +)$. Using the law of total probability:

$$\begin{aligned}\mathbb{P}(+) &= \mathbb{P}(+ \mid D)\mathbb{P}(D) + \mathbb{P}(+ \mid D^c)\mathbb{P}(D^c)\\ &= 0.96 \times 0.03 + 0.02 \times 0.97\\ &= 0.0482.\end{aligned}$$

By **Bayes' rule**:

$$\mathbb{P}(H \mid +) = \frac{\mathbb{P}(+ \mid H)\mathbb{P}(H)}{\mathbb{P}(+)} = \frac{0.02 \times 0.97}{0.0482} \approx 0.40.$$

About 40% of individuals who test positive are in fact healthy.

- When the condition is rare (here, the prevalence is only 3%), false positives can represent a large share of positive test results.

Example: Urns

Three urns U_1, U_2, and U_3 have the following compositions:

- *U_1 contains 3 red balls, 4 blue balls, and 1 green ball;*
- *U_2 contains 1 red ball, 2 blue balls, and 3 green balls;*
- *U_3 contains 4 red balls, 3 blue balls, and 2 green balls.*

One urn is selected at random (each urn with equal probability), and then one ball is drawn at random from that urn.

Given that the ball drawn is red, what is the probability that it came from urn U_i, for $i = 1, 2, 3$?

Solution (1/2)

Let:

- A_i: the event "urn i is selected", for $i = 1, 2, 3$;
- R: the event "the ball drawn is red".

Since the urn is chosen uniformly:

$$\mathbb{P}(A_1) = \mathbb{P}(A_2) = \mathbb{P}(A_3) = \frac{1}{3}.$$

The conditional probabilities are:

$$\mathbb{P}(R \mid A_1) = \frac{3}{8}, \quad \mathbb{P}(R \mid A_2) = \frac{1}{6}, \quad \mathbb{P}(R \mid A_3) = \frac{4}{9}.$$

Using the law of total probability:

$$\mathbb{P}(R) = \sum_{i=1}^{3} \mathbb{P}(A_i)\mathbb{P}(R \mid A_i) = \frac{1}{3}\left(\frac{3}{8} + \frac{1}{6} + \frac{4}{9}\right).$$

Solution (2/2)

By Bayes' rule, for $i = 1, 2, 3$:

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

After simplification, we obtain:

$$\mathbb{P}(A_1 \mid R) = \frac{27}{71}, \qquad \mathbb{P}(A_2 \mid R) = \frac{12}{71}, \qquad \mathbb{P}(A_3 \mid R) = \frac{32}{71}.$$

Given that the ball drawn is red, it is most likely to come from urn A_3, even though all urns were initially equally likely.

Example: The Last House

The police are searching for a serial killer who may have taken refuge in a small village.

Intelligence services estimate that there is a 90% *chance that the suspect is indeed in the village.*
If he is in the village, he is assumed to be equally likely to be hiding in any of the 100 *houses.*

The local police search 99 *houses and find nothing.*

What is the probability that the suspect is in the last remaining house?

Solution (1/2)

Let:

- V: the suspect is in the village,
- H_i: the suspect is in house i, for $i = 1, \ldots, 100$.

We are given:

$$\mathbb{P}(V) = 0.9, \qquad \mathbb{P}(H_i \mid V) = \frac{1}{100}.$$

The police have searched houses 1 to 99 and did not find the suspect.

Let B be the event:

$$B = \{\text{houses } 1, \ldots, 99 \text{ are empty}\}.$$

We want to compute:

$$\mathbb{P}(H_{100} \mid B).$$

Solution (2/2)

- If V occurs, then B happens iff the suspect is in house 100:
$$\mathbb{P}(B \mid V) = \mathbb{P}(H_{100} \mid V) = \frac{1}{100}.$$
- If V^c occurs, the suspect is not in the village. Searching 99 houses cannot find him:
$$\mathbb{P}(B \mid V^c) = 1.$$
- Bayes' rule gives:
$$\mathbb{P}(V \mid B) = \frac{\mathbb{P}(V)\mathbb{P}(B \mid V)}{\mathbb{P}(V)\mathbb{P}(B \mid V) + \mathbb{P}(V^c)\mathbb{P}(B \mid V^c)} = \frac{0.9 \cdot \frac{1}{100}}{0.9 \cdot \frac{1}{100} + 0.1} \approx 0.083.$$
- Finally,
$$\mathbb{P}(H_{100} \mid B) = \mathbb{P}(V \mid B),$$
since if B occurs and the suspect is in the village, he must be in house 100.

Example: The Monty Hall Problem

The following problem is inspired by an American television programme. A contestant on a game show stands in front of three closed doors, in the presence of a presenter.

- *Behind one of the doors, there is a car, behind the other two are goats.*
- *The contestant chooses one door (say Door 1).*
- *The host then opens one of the other two doors and reveals a goat.*
- *The host **always** opens a door with a goat and **never** opens the contestant's chosen door.*
- *The contestant is offered a choice: **stay** with their original door or **switch** to the remaining closed door.*

Is it better to stay, switch, or does it not matter?

Events and the Key Question

Let:

- C_1: "the car is behind Door 1" (the initial choice),
- C_2: "the car is behind Door 2",
- C_3: "the car is behind Door 3",

Initially, by symmetry:

$$\mathbb{P}(C_1) = \mathbb{P}(C_2) = \mathbb{P}(C_3) = \frac{1}{3}.$$

Solution via Bayes' Rule

- Suppose the contestant's chose Door 1 and the host opens Door i (goat).
- Let O_i be the event "the host opens Door i $(i = 2, 3)$".
- Host behaviour:

$$\mathbb{P}(O_3 \mid C_1) = \frac{1}{2}, \qquad \mathbb{P}(O_3 \mid C_2) = 1, \qquad \mathbb{P}(O_3 \mid C_3) = 0.$$

- Hence, using Bayes' rule on the partition (C_1, C_2, C_3):

$$\mathbb{P}(C_i \mid O_3) = \frac{\mathbb{P}(O_3 \mid C_i)\,\mathbb{P}(C_i)}{\sum_{j=1}^{3} \mathbb{P}(O_3 \mid C_j)\,\mathbb{P}(C_j)} = \frac{\mathbb{P}(O_3 \mid C_i)\,\mathbb{P}(C_i)}{\frac{1}{3}\left(\frac{1}{2} + 1 + 0\right)} = \frac{\mathbb{P}(O_3 \mid C_i)\,\mathbb{P}(C_i)}{1/2}.$$

Hence:

$$\mathbb{P}(C_1 \mid O_3) = \frac{(1/2)(1/3)}{1/2} = \frac{1}{3}, \qquad \mathbb{P}(C_2 \mid O_3) = \frac{1 \cdot (1/3)}{1/2} = \frac{2}{3}.$$

- There is a two in three chance that the car is behind the door that the presenter did not open. $\Rightarrow$ always switch!

Why the Host's Action Changes Probabilities

We assume the contestant's initially chose Door 1.

- If C_1 is true (car behind Door 1), the host has **two** goat doors available to open.
- If C_2 is true, the host has only **one** choice: he must open Door 3.
- If C_3 is true, the host has only **one** choice: he must open Door 2.

So the host's action is **informative**! It is not independent of where the car is.

8. Independence

Independence

- Two events A and B are **independent** if learning that one of them occurred does **not change** the probability that the other occurs.
  - the occurrence of B provides no information about A (and vice versa).

Definition 8. Independence

*Let A and B be two events. We say that A and B are **independent** if*

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

Example: two coin tosses:

- Let A be "the first coin is Head".
- Let B be "the second coin is Head".
- Then A and B are independent if there is no mechanism linking the two coins.

Independence and Conditional Probability

Proposition 3. Independence and Conditional Probability

If A and B are independent and $\mathbb{P}(B) > 0$, then

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

Likewise, if $\mathbb{P}(A) > 0$, then $\mathbb{P}(B \mid A) = \mathbb{P}(B)$.

Proof:

- By definition of conditional probability, $\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$.
- By independence, $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$, hence

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

$\square$

Properties of Independence

Proposition 4. Properties of Independence

Let A and B be two independent events. Then:

1. A *and* B^c *are independent;*
2. A^c *and* B *are independent;*
3. A^c *and* B^c *are independent.*

Proof (1/2)

- **Property 1.**

$$
\begin{aligned}
\mathbb{P}(A \cap B^c) &= \mathbb{P}(A) - \mathbb{P}(A \cap B) \\
&= \mathbb{P}(A) - \mathbb{P}(A)\mathbb{P}(B) \\
&= \mathbb{P}(A)(1 - \mathbb{P}(B)) \\
&= \mathbb{P}(A)\mathbb{P}(B^c).
\end{aligned}
$$

  Therefore, A and B^c are independent.

- **Property 2.** The proof that A^c and B are independent is symmetric.

Proof (2/2)

- **Property 3.** Recall De Morgan's law:

$$A^c \cap B^c = (A \cup B)^c.$$

Hence,

$$\mathbb{P}(A^c \cap B^c) = 1 - \mathbb{P}(A \cup B).$$

Using the inclusion–exclusion formula and independence:

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

Therefore,

$$\begin{aligned}\mathbb{P}(A^c \cap B^c) &= 1 - \mathbb{P}(A) - \mathbb{P}(B) + \mathbb{P}(A)\mathbb{P}(B)\\ &= (1 - \mathbb{P}(A))(1 - \mathbb{P}(B))\\ &= \mathbb{P}(A^c)\mathbb{P}(B^c).\end{aligned}$$

Hence, A^c and B^c are independent.

$\square$

Example: Independence in an Urn

An urn contains balls numbered from 1 to 12. One ball is drawn at random. Consider the events:

- A: *"the number drawn is even",*
- B: *"the number drawn is a multiple of 3".*

1. *Are the events A and B independent?*
2. *Answer the same question for an urn containing balls numbered from 1 to 13.*

Solution: Urn with 12 Balls

The sample space is $\Omega = \{1, 2, \dots, 12\}$, with all outcomes equally likely.

- Even numbers:

$$A = \{2, 4, 6, 8, 10, 12\}, \qquad \mathbb{P}(A) = \frac{6}{12} = \frac{1}{2}.$$

- Multiples of 3:

$$B = \{3, 6, 9, 12\}, \qquad \mathbb{P}(B) = \frac{4}{12} = \frac{1}{3}.$$

- Then, for the intersection of A and B:

$$A \cap B = \{6, 12\}, \qquad \mathbb{P}(A \cap B) = \frac{2}{12} = \frac{1}{6}.$$

Since

$$\mathbb{P}(A \cap B) = \frac{1}{6} = \frac{1}{2} \times \frac{1}{3} = \mathbb{P}(A)\mathbb{P}(B),$$

the events A and B are **independent**.

Solution: Urn with 13 Balls

The sample space is now $\Omega = \{1, 2, \ldots, 13\}$.

- Even numbers:
$$A = \{2, 4, 6, 8, 10, 12\}, \qquad \mathbb{P}(A) = \frac{6}{13}.$$
- Multiples of 3:
$$B = \{3, 6, 9, 12\}, \qquad \mathbb{P}(B) = \frac{4}{13}.$$
- Intersection:
$$A \cap B = \{6, 12\}, \qquad \mathbb{P}(A \cap B) = \frac{2}{13}.$$

We compare:

$$\mathbb{P}(A \cap B) = \frac{2}{13}, \qquad \mathbb{P}(A)\mathbb{P}(B) = \frac{6}{13} \times \frac{4}{13} = \frac{24}{169}.$$

Since $\mathbb{P}(A \cap B) = \frac{2}{13} \neq \frac{24}{169} = \mathbb{P}(A)\mathbb{P}(B)$, the events A and B are **not independent**.

Mutual Independence

Definition 9. Mutual Independence

A family of events $A_1, \ldots, A_n$ is said to be ***mutually independent*** *if, for every subset of indices $\mathcal{I} \subseteq \{1, \ldots, n\}$,*

$$\mathbb{P}\left(\bigcap_{i \in \mathcal{I}} A_i\right) = \prod_{i \in \mathcal{I}} \mathbb{P}(A_i).$$

Pairwise vs Mutual Independence

Important distinction

- *Independence is often used to simplify probability calculations:*

$$\mathbb{P}(A_1 \cap \cdots \cap A_n) = \prod_{i=1}^{n} \mathbb{P}(A_i).$$

- *This simplification is **only valid** under **mutual independence**, which requires independence of **all subcollections**, not only pairs of events.*
  - ***Pairwise independence** guarantees this equality only for intersections of **two** events.*
- *When more than two events are involved, pairwise independence is **not sufficient**.*

*Pairwise independence **does not imply** mutual independence (see the example on the next slide).*

Example: Pairwise but Not Mutually Independent Events

Two Coins

Two fair coins are tossed. Consider the events:

- A_1: *"the first coin shows Head",*
- A_2: *"the second coin shows Head",*
- B: *"exactly one coin shows Head".*

It can be shown that the events are ***pairwise independent****:*

$$\mathbb{P}(A_1) = \mathbb{P}(A_2) = \mathbb{P}(B) = \frac{1}{2},$$

and for example,

$$\mathbb{P}(A_1 \cap B) = \mathbb{P}(A_1)\mathbb{P}(B) = \frac{1}{4}.$$

However, they are ***not mutually independent****, since:*

$$\mathbb{P}(A_1 \cap A_2 \cap B) = 0 \neq \frac{1}{8} = \mathbb{P}(A_1)\mathbb{P}(A_2)\mathbb{P}(B).$$

References I

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

Kolmogoroff, A. (1933). Grundbegriffe der wahrscheinlichkeitsrechnung .

McClave, J. T., Benson, P. G. and Sincich, T. (2022). *Statistics for business and economics, 14th Edition*. Pearson Education Limited.