



People's Democratic Republic of Algeria
Ministry of Higher Education and Scientific Research
SETIF 1 University — Ferhat Abbas
Faculty of Economic, Commercial and Management Science
Department of Basic Education



PEDAGOGICAL HANDOUT

STATISTICS 2

مطبوعة موجهة لطلبة السنة الأولى جذع مشترك للعلوم الإقتصادية و التجارية و علوم التسيير

Prepared for:

First-Year LMD

Second Semester

Prepared by:

Dr. Manel REBBOUH

Associate Professor

Manel.rebbouh@univ-setif.dz

Manelrebbouh23@gmail.com

Academic Year 2025 / 2026

Preface

This handout is intended as a complete pedagogical resource for the probability-theory component of the Statistics 2 module, taught to first-year LMD students in Economics, Commercial Sciences, and Management at the SETIF 1 University.

The content is organised into three progressively structured chapters. Chapter 1 introduces the foundational concepts of probability theory: sample spaces, events, the three definitions of probability, and the main rules of probability calculation, including Bayes' theorem and combinatorial analysis. Chapter 2 develops the theory of random variables, both discrete and continuous, together with their numerical summaries (expectation, variance, standard deviation, and standardisation). Chapter 3 completes the course with the seven most important probability distributions used in economics and management, along with their applications, approximations, and reference tables.

Each chapter follows the same pedagogical structure: learning objectives aligned with Bloom's taxonomy, prerequisite checks, clear notation, worked examples in economic contexts, common-mistake warnings, visual mind maps, summary cards, and a set of practice exercises designed to reflect the style of final examinations. All bibliographic references follow the APA 7 format and are drawn from recent academic sources (2019–2024) in English and French.

Students are encouraged to study each chapter in sequence, work through every example with pen and paper, and attempt all practice exercises before moving on. The mind maps and summary cards are specifically designed to aid revision and should be consulted repeatedly throughout the semester.

I thank the Department of Basic Education and the Faculty of Economic, Commercial and Management Sciences for their continuing support, and my students for the questions and feedback that shaped much of this material.

Dr. Manel REBBOUH

Associate Professor

SETIF 1 University — Ferhat Abbas

Table of Contents

Part	Title	Approx. pages
Front Matter	Cover · Preface · TOC · Abstract · Notation	i – viii
Chapter 1	Introduction to Probability Theory	1 – 41
Chapter 2	Random Variables	42 – 86
Chapter 3	Special Probability Laws	87 – 139
Appendices	Z-Table · Binomial · Poisson Tables (in Ch. 3)	inside Ch. 3
Back Matter	Unified References (APA 7)	140 – 144

Abstract of The Handout

ملخص

تقدّم هذه المطبوعة البيداغوجية مادة متكاملة في نظرية الاحتمالات، موجهة لطلبة السنة الأولى LMD - جذع مشترك لكلية العلوم الاقتصادية والتجارية وعلوم التسيير بجامعة سطيف 1 فرحات عباس - حيث تتطرق هذه المطبوعة الى ثلاثة فصول متدرّجة: الأول يتناول الأسس الرياضية للاحتمال (فضاء العينة، الأحداث، قواعد الحساب، التحليل التوافقي، ونظرية بايز)؛ الثاني يتناول المتغيرات العشوائية المنفصلة والمستمرة مع مقاييس التمرکز والتشتت (الأمل الرياضي، التباين، الانحراف المعياري، والتقييس)؛ الثالث يعرض سبعة توزيعات احتمالية أساسية (برنولي، ثنائي الحدين، فوق الهندسي، بواسون، المنتظم، الأسّي، والطبيعي) مع تطبيقاتها الاقتصادية وجداولها المرجعية.

تتميز هذه المطبوعة بـ 11 خريطة ذهنية ملوّنة تلخّص المفاهيم الأساسية، وأكثر من 55 مثالاً محلّولاً في سياقات اقتصادية وإدارية، و58 تمريناً غير محلّول على ثلاثة مستويات (أسس، تطبيقات، توسيع). كل فصل يتضمن أهدافاً تعليمية، وقائمة تحذيرات من الأخطاء الشائعة، وبطاقات ملخّصة قابلة للمراجعة قبل الامتحان.

الكلمات المفتاحية: نظرية الاحتمال؛ المتغيرات العشوائية المنفصلة؛ المتغيرات العشوائية المتصلة؛ التوزيعات الاحتمالية.

Abstract

This pedagogical booklet provides a complete course in probability theory for first-year LMD students of the Basic Education Department in Economic, Commercial and Management Sciences Faculty at SETIF 1 University Ferhat Abbas. The material is organised into three progressively structured chapters. Chapter 1 presents the mathematical foundations of probability: sample spaces and events, the three definitions of probability (classical, empirical, axiomatic), the addition and multiplication rules, conditional probability, total probability, Bayes' theorem, and combinatorial analysis. Chapter 2 develops the theory of random variables, both discrete (PMF, CDF) and continuous (PDF, CDF), together with their numerical summaries, including expected value, variance, standard deviation, and the Z-score used for standardization. Chapter 3 completes the course with the seven most important named probability distributions used in economics and management: Bernoulli, Binomial, Hypergeometric, and Poisson (discrete families); Uniform, Exponential, and Normal (continuous families). Each family is introduced through its generating mechanism, its PMF or PDF, its moments, and representative economic applications. Approximations between distributions (Binomial $\rightarrow$ Poisson, Binomial $\rightarrow$ Normal, Poisson $\rightarrow$ Normal) are discussed with their validity conditions. Three reference tables (standard normal, selected binomial, selected Poisson) are provided as appendices to Chapter 3.

The handout features 11 full-page mind maps that visually summarise the key concepts, more than 55 worked examples set in economic and managerial contexts, and 58 unsolved practice exercises graded on three levels (foundations, applications, extension). Each chapter includes learning objectives, common-mistake warnings, and revision-oriented summary cards.

Keywords: Probability theory; discrete random variables; continuous random variables; probability distributions.

List of Symbols and Notation

The table below gathers all mathematical symbols used throughout the booklet. Symbols specific to a single chapter are noted accordingly; others recur throughout.

Set-theoretic and probability notation (Chapter 1)

Symbol	Reading	Meaning
E, Ω	sample space	Set of all possible outcomes of a random experiment.
ω	omega	A single outcome, element of E .
$A, B, C, \dots$	events	Subsets of E , i.e. sets of outcomes.
$A \cup B$	A union B	Event "A or B (or both) occurs".
$A \cap B$	A intersection B	Event "both A and B occur".
$\bar{A}, A^c$	complement of A	Event "A does not occur".
$\emptyset$	empty set	Impossible event: contains no outcomes.
$P(A)$	probability of A	Probability that event A occurs; $P(A) \in [0, 1]$.
$P(A B)$	P of A given B	Conditional probability of A given that B has occurred.
$n!$	n factorial	Product $1 \cdot 2 \cdot \dots \cdot n$, with $0! = 1$.
$C(n, k), C_n^k$	n choose k	Binomial coefficient: $n! / [k! (n - k)!]$.
$A(n, k), A_n^k$	arrangements	Ordered arrangements: $n! / (n - k)!$.
P_n	permutations	Orderings of n distinct elements: $n!$.

Random variables and moments (Chapter 2)

Symbol	Reading	Meaning
X, Y, Z	random variable	Function from E to $\mathbb{R}$.
x, y, z	realisation	Specific value that a random variable can take.
$p(x), p_X(x)$	PMF	Probability mass function: $P(X = x)$ for discrete X .
$f(x), f_X(x)$	PDF	Probability density function for continuous X .
$F(x), F_X(x)$	CDF	Cumulative distribution function: $P(X \leq x)$.
$E(X), \mu$	expected value	Mean of X , the long-run average.
$\text{Var}(X), \sigma^2$	variance	$E[(X - \mu)^2]$: average squared deviation from the mean.
σ, σ_X	standard deviation	$\sqrt{\text{Var}(X)}$, in the same units as X .
Z	Z-score	$(X - \mu)/\sigma$: standardised value.

$\mathbf{1_A, I_A}$	indicator of A	1 if A occurs, 0 otherwise.
$\mathbb{R}$	real numbers	Set of all real numbers.

Named distributions (Chapter 3)

Symbol	Reading	Meaning
$\mathbf{X \sim B(p)}$	Bernoulli	$X \in \{0, 1\}$ with $P(X = 1) = p$.
$\mathbf{X \sim B(n, p)}$	Binomial	Count of successes in n i.i.d. Bernoulli(p) trials.
$\mathbf{X \sim H(N, K, n)}$	Hypergeometric	Count of successes in n draws without replacement.
$\mathbf{X \sim P(\lambda)}$	Poisson	Count of rare events at constant rate $\lambda > 0$.
$\mathbf{X \sim U(a, b)}$	Uniform	Uniformly distributed on $[a, b]$.
$\mathbf{X \sim Exp(\lambda)}$	Exponential	Waiting time in a Poisson process of rate λ .
$\mathbf{X \sim N(\mu, \sigma^2)}$	Normal	Bell-shaped distribution with mean μ and variance σ^2 .
λ	lambda	Rate parameter (Poisson, Exponential).
e	Euler's number	$e \approx 2.71828$; base of the natural exponential.
π	pi	$\pi \approx 3.14159$; used in the normal density.
$\Phi(z)$	Phi of z	CDF of the standard normal $N(0, 1)$.
$\phi(z)$	phi of z	PDF of the standard normal $N(0, 1)$.

CHAPTER 1

Introduction to Probability Theory

Sample spaces · Events · Probability rules · Bayes' theorem · Combinatorics

Before a firm commits capital to a new market, before a bank approves a loan, before a government adjusts subsidies on a staple good, someone must answer a question that has no certain answer: *what is likely to happen?* Economics and management science live with this question permanently. Probability theory is the formal language that lets us work with it. The goal is not to eliminate uncertainty but to measure it, compare alternatives, and defend our choices with numerical evidence.

This opening chapter develops the vocabulary and basic rules that every subsequent chapter will assume. The material here is not difficult, but it is unforgiving: a small confusion about what an event is, or about when two events can be treated as independent, will propagate through every calculation that follows. Read slowly, work the examples with pen and paper, and do not skip the exercises.

i In This Chapter

- §1.1 Introduction
- §1.2 Learning Objectives (Bloom's taxonomy)
- §1.3 Prerequisites Check
- §1.4 Notation and Symbols
- §1.5 Random Experiments and Events
- §1.6 The Three Definitions of Probability
- §1.7 Rules of Probability Calculation
- §1.8 Combinatorial Analysis
- §1.9 Worked Examples in an Economic Context
- §1.10 Common Mistakes and Warnings
- §1.11 Key Takeaways
- §1.12 Chapter Summary Cards
- §1.13 Practice Exercises
- §1.14 References

1.1 Introduction

Most phenomena that concern an economist cannot be predicted with certainty. The monthly inflation rate, the number of defaulting borrowers at a regional bank, the quantity of wheat imported in a given quarter, the duration of a job search: each of these varies in ways that no deterministic rule can capture in advance. Yet the variation is not random in the ordinary sense of the word. It follows patterns, and those patterns can be described mathematically.

Probability theory provides that description. It assigns a number between 0 and 1 to every outcome we might observe, and it supplies rules for combining those numbers when outcomes interact. The framework was built over three centuries of work beginning with Pascal and Fermat in 1654 and reaching a rigorous axiomatic form with Kolmogorov in 1933 (Blitzstein & Hwang, 2019). What began as a correspondence about gambling problems now underpins insurance pricing, portfolio theory, statistical quality control, econometric inference, and every modern application of data science.

For the first-year LMD student in economics, the subject serves three distinct purposes. It trains quantitative reasoning under uncertainty, which is the core competency of any applied economist. It provides the mathematical foundation required for inferential statistics and econometrics in later semesters. And it builds the habit of separating what is known from what is assumed, a habit that matters as much in policy analysis as in academic research (Hansen, 2022).

The present chapter stays at the foundational level. We define what a random experiment is, how events are built from outcomes, and how probabilities are assigned and manipulated. The formalism may feel abstract at first, but each concept is anchored to at least one example drawn from economic or managerial practice. By the end of the chapter, you should be able to read a short problem description and translate it into the correct probability notation without hesitation.

1.1.1 A brief historical note

Probability theory emerged from practical questions about gambling, insurance, and demographic records. The correspondence between Blaise Pascal and Pierre de Fermat in 1654, triggered by the Chevalier de Méré's questions about dice games, is traditionally regarded as the birth of the field. Jakob Bernoulli's posthumous *Ars Conjectandi* (1713) introduced the first version of the law of large numbers. Pierre-Simon Laplace, in his *Théorie analytique des probabilités* (1812), systematised the classical definition and applied it to errors of measurement in astronomy (Blitzstein & Hwang, 2019).

The modern axiomatic foundation is due to Andrey Kolmogorov, whose *Grundbegriffe der Wahrscheinlichkeitsrechnung* (1933) resolved long-standing paradoxes by grounding probability in measure theory. The twentieth century saw probability become an autonomous

mathematical discipline and, through the work of economists such as Frank Ramsey, Bruno de Finetti, and Leonard Savage, the formal basis of rational choice under uncertainty (Hansen, 2022). Today, the same formalism supports portfolio theory, quality control, econometric inference, Bayesian forecasting, and machine learning.

For the student entering the subject for the first time, this long history matters less than the underlying unity of the framework. The three definitions of probability we study in §1.6 (classical, empirical, and axiomatic) are not competing theories but complementary viewpoints that arose historically and that apply in different situations today.

1.2 Learning Objectives

The table below lists the concrete abilities you should possess after completing this chapter. Each row maps to one level of Bloom's cognitive taxonomy (Devore, Berk, & Carlton, 2021); mastering the lower levels is a prerequisite for the higher ones.

Bloom level	Action verb	Specific competency
Remember	Define · Identify	Random experiment, sample space, event, axioms of probability.
Understand	Explain · Distinguish	Difference between classical, empirical, and axiomatic approaches; mutually exclusive vs independent events.
Apply	Compute · Use	Addition and multiplication rules; conditional probabilities; counting formulas.
Analyse	Decompose · Compare	Partition of the sample space; structure of a probability tree.
Evaluate	Justify · Interpret	Use Bayes' theorem to update prior beliefs; comment on posterior probabilities in a managerial context.

The competencies above map directly to outcomes **C1–C4** of the official course specification (fiche pédagogique).

How to study this chapter effectively

The material in this chapter is cumulative: each definition is used by the next, and every rule of §1.7 depends on the vocabulary introduced in §1.5. A suggested study path follows.

- **Step 1.** Read the prerequisite check (§1.3) and make sure you can answer every item. If not, revise the relevant first-semester material before continuing.
- **Step 2.** Read §1.5 and §1.6 slowly, noting each definition in your own words. A concept you cannot rephrase is a concept you do not yet own.

- **Step 3.** Work through Examples 1.1 to 1.14 with pen and paper. Do not read the solutions passively; rewrite each calculation yourself.
- **Step 4.** Study §1.7 (the four rules) and §1.8 (combinatorics). These two sections contain all the machinery used in every exercise.
- **Step 5.** Go through Examples 1.15 to 1.23 to see how the rules combine in realistic settings.
- **Step 6.** Read §1.10 (common mistakes) carefully. These are the traps that cost the most marks in the final exam.
- **Step 7.** Attempt the exercises in §1.13 in order of difficulty. If you get stuck on one, return to the relevant worked example rather than consulting a solution immediately.
- **Step 8.** Before moving to Chapter 2, make sure you can reproduce from memory the statement of Bayes' theorem and of the total probability theorem, along with one worked example of each.

A realistic total workload for this chapter is approximately twelve hours: two hours for the reading, four hours for the worked examples, and six hours for the exercises. Students who spread this effort over two to three weeks typically retain the material far better than those who attempt to absorb it in a single session.

1.3 Prerequisites Check

Before starting this chapter, you should be comfortable with the notions listed below, all of which come from your first-semester Mathematics 1 and Statistics 1 courses. If any item seems unfamiliar, review the corresponding section of your first-semester notes before proceeding.

- **From set theory:** the meaning of $\in$, $\subset$, $\subseteq$, $\cup$, $\cap$, and the complement operation; how to read a Venn diagram; the idea of a partition.
- **From algebra:** factorials (including the convention $0! = 1$); binomial coefficients; basic manipulation of fractions and proportions; elementary limits.
- **From descriptive statistics:** the difference between a population and a sample; the notion of relative frequency; how a histogram represents a distribution of observed values.

The short diagnostic below lets you test yourself. If you can answer all five items without hesitation, you have the required background. Answers are given in §1.13.6.

i Diagnostic 1.1 — Self-check before starting

- Let $E = \{1, 2, 3, 4, 5, 6\}$, $A = \{2, 4, 6\}$, $B = \{1, 2, 3\}$. Write $A \cup B$ and $A \cap B$ as explicit sets.
- What is the complement of A in E ?

- Compute $5! / (3! \times 2!)$ and $C(6, 2)$.
- A die is rolled 60 times; the face "4" appears 12 times. What is the relative frequency of "4"?
- Are the sets $\{1, 3\}$ and $\{2, 4\}$ disjoint? Justify briefly.

1.4 Notation and Symbols Used in This Chapter

The table below lists the symbols used throughout this chapter. A complete booklet-wide notation list appears in the front matter.

Symbol	Reading	Meaning
E, Ω	"omega"	Sample space: set of all possible outcomes of the random experiment.
$ E , n(E)$	cardinality of E	Number of elements (outcomes) in the sample space.
A, B, C	—	Events (subsets of E).
$ A $	cardinality of A	Number of outcomes that compose event A .
$\bar{A}, A^c$	"A bar"	Complement of A : every outcome in E that is not in A .
$A \cup B$	"A union B"	The event "A occurs, or B occurs, or both".
$A \cap B$	"A intersection B"	The event "A and B both occur".
$\emptyset$	"empty set"	The impossible event.
$A \subset B$	"A included in B"	Every outcome of A is also in B .
$P(A)$	"probability of A"	A number in $[0, 1]$ measuring the chance that A occurs.
$P(B A)$	"probability of B given A"	Conditional probability of B once A is known to have occurred.
$n!$	"factorial of n"	$n \times (n-1) \times \dots \times 2 \times 1$, with $0! = 1$.
$C(n, k)$	"n choose k"	Number of unordered selections of k items from n .
$A(n, k)$	"arrangements"	Number of ordered selections of k items from n without repetition.

1.5 Random Experiments and Events

1.5.1 The notion of a random experiment

An **experiment** is any procedure whose execution produces an observable result. When the result is fixed in advance by known laws and initial conditions (a free-falling object covering a predictable distance, for example), we speak of a **deterministic experiment**. When the result cannot be known in advance even if the procedure is repeated under identical conditions, the experiment is **random** (Hurlin & Mignon, 2022).

The notion of "identical conditions" deserves attention. Tossing a coin twice does not mean producing the same outcome twice; it means performing the same physical procedure twice. The outcome may differ, and that difference is the object of interest. Three features characterise a random experiment useful for our purposes:

- the set of all possible outcomes is known before the experiment is performed;
- no particular outcome can be predicted with certainty;
- the experiment is at least conceptually repeatable.

Example 1.1 — *Four random experiments from economic life*

- (a) Observing whether a given loan applicant repays on time or defaults.
- (b) Recording the closing price of a listed share on the next trading day.
- (c) Counting the number of customers who walk into a retail branch during one hour.
- (d) Measuring the time (in minutes) a client waits before being served at a post-office counter.

In (a), the two possible outcomes are discrete and easy to list. In (b), the price could in principle take any positive real value; the outcome set is uncountable. In (c), outcomes are non-negative integers. In (d), outcomes lie in $[0, \infty)$. These distinctions will matter once we start building probability models in Chapter 2.

1.5.2 The sample space

The **sample space**, written E or Ω , is the set of all possible distinct outcomes of a random experiment. Every valid outcome must belong to E , and the listed outcomes must be mutually exclusive: no outcome can be simultaneously two different elements of E .

Example 1.2 — *Constructing sample spaces*

- Tossing a fair coin once: $E = \{H, T\}$, so $|E| = 2$.
- Rolling a fair six-sided die: $E = \{1, 2, 3, 4, 5, 6\}$, $|E| = 6$.
- Tossing two distinguishable coins: $E = \{(H,H), (H,T), (T,H), (T,T)\}$, $|E| = 4$.
- Drawing one card from a standard deck: $|E| = 52$.
- Choosing one of 500 employees at random: $|E| = 500$.

When outcomes are equally likely and E is finite, counting its elements is often enough to compute probabilities, as we shall see in §1.6.1. When E is infinite or outcomes are not equally likely, we need additional machinery that we introduce in Chapter 2.

1.5.3 Events and their types

An **event** is any subset of the sample space. After the experiment is carried out, we say the event A has **occurred** if the observed outcome is one of the elements of A ; otherwise A has **not occurred**. Events are denoted by capital Latin letters, and the number of outcomes composing event A is written $|A|$.

Several categories of events are named because they recur constantly.

An **elementary event** (or singleton) contains exactly one outcome. On a die, $A = \{5\}$ is elementary.

A **compound event** groups several outcomes under a common description. On a die, "the result is even" gives $A = \{2, 4, 6\}$, which has three elementary outcomes.

The **certain event** is E itself. It occurs no matter what outcome is observed, so $P(E) = 1$.

The **impossible event** is $\emptyset$. It contains no outcomes, so $P(\emptyset) = 0$.

The **complement** of A , written $\bar{A}$ or A^c , consists of all outcomes of E that are not in A :

$$\bar{A} = \{\omega \in E : \omega \notin A\}$$

By construction, $A \cup \bar{A} = E$ and $A \cap \bar{A} = \emptyset$.

Two events are **mutually exclusive** (or incompatible) when $A \cap B = \emptyset$: they share no outcomes and cannot both occur on the same trial. Rolling a die, "result is 2" and "result is 5" are mutually exclusive; "result is even" and "result is prime" are not, because 2 belongs to both.

Two events A and B are **independent** when the occurrence of one leaves the probability of the other unchanged. We formalise this condition in §1.7.4 once we have conditional probability at our disposal. For now, the intuition is that sampling *with replacement* produces independence across draws, while sampling *without replacement* introduces dependence.

i Intuition to remember

- Independence captures the idea that information about A is irrelevant to B .
- Disjointness (mutual exclusivity) captures the opposite: learning A tells you B cannot have occurred.
- These two properties are mathematically distinct and almost always mutually incompatible (except when one event has zero probability).

1.5.4 Operations on events and De Morgan's laws

Events are sets, and set algebra therefore applies to them without modification. The three operations needed in this course are union, intersection, and complement. Their intuitive meanings are given in Table 1.1.

Table 1.1: Set operations on events

Operation	Notation	Meaning in words
Union	$A \cup B$	A occurs, or B occurs, or both
Intersection	$A \cap B$	A and B both occur
Complement	$\bar{A}$	A does not occur

Source: Elaborated by the Author

Two identities, known as **De Morgan's laws**, link these operations:

$$(A \cup B)^c = \bar{A} \cap \bar{B}$$

$$(A \cap B)^c = \bar{A} \cup \bar{B}$$

The first law says that the event "neither A nor B" is the same as " $\bar{A}$ and $\bar{B}$ ". The second says that "A and B do not both occur" is the same as " $\bar{A}$ or $\bar{B}$ (or both)". Both identities extend to any finite number of events and are frequently used to simplify compound statements. A visual verification with a Venn diagram is a useful exercise.

Example 1.3 — Translating a managerial statement

A supplier classifies each delivery as "on time" (T) or "complete" (C). Express in set notation: "the delivery is on time but not complete".

The answer is $T \cap \bar{C}$. The full event space has four compound possibilities: $T \cap C$, $T \cap \bar{C}$, $\bar{T} \cap C$, $\bar{T} \cap \bar{C}$. If the manager later asks for "deliveries that failed on at least one dimension", the answer is $\bar{T} \cup \bar{C}$, which by De Morgan's second law equals the complement of $T \cap C$, a useful simplification when computing probabilities.

⚠ Watch out — common confusion

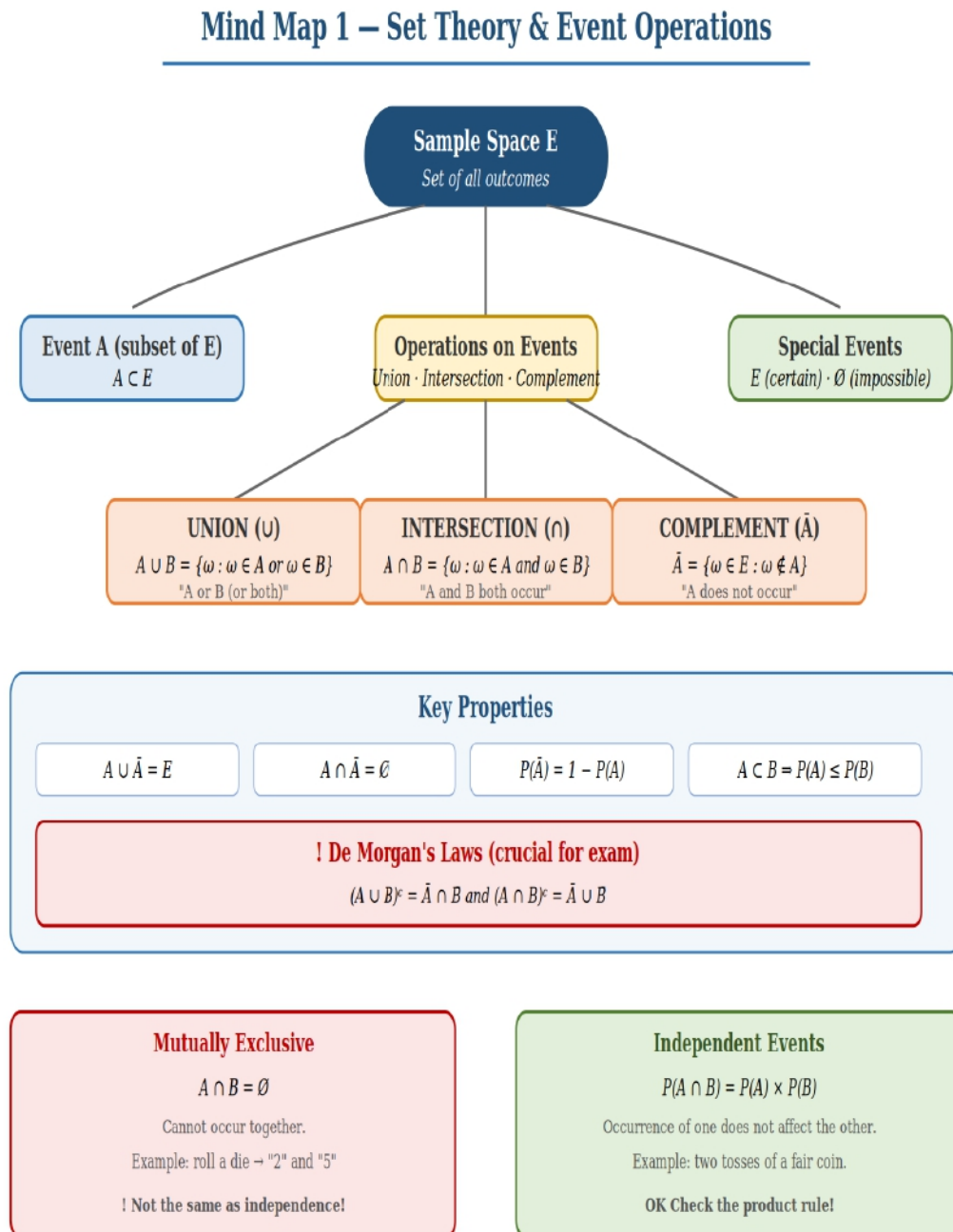
Mutually exclusive $\neq$ independent. Two incompatible events (with positive probabilities) cannot be independent, because knowing that one occurred rules out the other entirely. Students confuse these two notions almost every year.

"Not A" is $\bar{A}$, not $1 - A$. $1 - A$ has no meaning as a set; the correct object is the complementary event, whose probability is $1 - P(A)$.

Visual Summary — Mind Map of §1.5

The diagram below summarises the structure of §1.5: the sample space, the three operations on events (union, intersection, complement), De Morgan's laws, and the essential distinction between mutually exclusive and independent events.

Figure 1.1: Set theory and event operations: a visual map



Source: Elaborated by the Author

1.6 The Three Definitions of Probability

Three complementary definitions of probability are used in this course. Each answers a slightly different question, and each has a domain in which it applies most naturally.

1.6.1 The classical (Laplace) definition

When the sample space is finite and the outcomes are equally likely, the probability of an event A is the ratio:

$$P(A) = \frac{|A|}{|E|} = \frac{(\text{number of favourable outcomes})}{(\text{total number of outcomes})};$$

This definition, attributed to Laplace in 1812, is the one encountered in every introductory problem involving dice, coins, cards, or balanced urns. Its strength is operational clarity: the calculation reduces to counting. Its limitation is the equiprobability assumption, which fails for many real-world experiments. The daily return on an asset, for instance, is certainly not uniformly distributed over all possible values (Ross, 2024).

Example 1.4 — A simple classical calculation

A fair coin is tossed twice. Let A be the event "both tosses show the same face".

Sample space: $E = \{(H,H), (H,T), (T,H), (T,T)\}$, $|E| = 4$.

Favourable outcomes: $A = \{(H,H), (T,T)\}$, $|A| = 2$.

$$P(A) = \frac{2}{4} = 0.50$$

In the long run, half of the trials of this two-toss experiment will produce two identical faces.

1.6.2 The empirical (frequentist) definition

When the equiprobability assumption is questionable but the experiment can be repeated many times, we estimate probabilities from observed frequencies. If A has occurred x times in n independent trials, the **relative frequency** of A is:

$$f_n(A) = \frac{x}{n}$$

The relative frequency fluctuates from one set of trials to the next, but the law of large numbers (which we revisit in Chapter 3) guarantees that:

$$\lim_{n \rightarrow \infty} \frac{x}{n} = P(A) \quad (\text{as } n \rightarrow \infty)$$

The empirical definition is how actuaries determine life-insurance premiums from mortality tables, how quality engineers estimate defect rates from production samples, and how election forecasters convert poll data into probability statements (Lecoutre, 2023).

Example 1.5 — Estimating a default probability

A bank observes that 42 out of 500 loans granted last year defaulted. Under the assumption that current conditions are similar, the estimated probability of default is:

$$f_{500} = \frac{42}{500} = 0.084 \text{ (or) } 8.4\%$$

The estimate would become more reliable if the bank could examine a larger historical sample, a direct consequence of the law of large numbers.

1.6.3 The axiomatic (Kolmogorov) definition

Kolmogorov's 1933 axiomatisation placed probability on a rigorous mathematical footing. A probability is any function P that assigns a real number to each event in E and satisfies three axioms:

Axiom 1. $P(A) \geq 0$ for every event A .

Axiom 2. $P(E) = 1$.

Axiom 3. If $A_1, A_2, \dots$ are mutually exclusive events, then

$$P(A_1 \cup A_2 \cup \dots) = P(A_1) + P(A_2) + \dots$$

Every property we use in this course (the complement rule, the general addition rule, the conditional probability formula) can be derived from these three axioms (Blitzstein & Hwang, 2019). The axiomatic approach does not tell us how to compute a particular $P(A)$ in a specific problem; it tells us what properties any such assignment must respect to deserve the name "probability".

1.6.4 Properties derived from the axioms

The following consequences are used so often that we state them explicitly and encourage you to memorise them.

- **(a)** For any event A , $0 \leq P(A) \leq 1$.
- **(b)** $P(\emptyset) = 0$.
- **(c)** $P(\bar{A}) = 1 - P(A)$. (*Complement rule*)
- **(d)** If $A \subset B$, then $P(A) \leq P(B)$. (*Monotonicity*)
- **(e)** $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. (*General addition rule*)

Property (c) is often the fastest route to a solution: when an "at least one" question appears, computing the probability of the complementary event and subtracting from 1 is usually simpler than enumerating the favorable outcomes directly.

★ **Exam tip**

- When you see "at least one", think "complement": $P(\text{at least one}) = 1 - P(\text{none})$.
- When you see "exactly k" or "at most k", direct counting is often clearer.
- Always check that your final probability lies in $[0, 1]$. A negative or >1 answer signals an arithmetic mistake.

1.7 Rules of Probability Calculation

1.7.1 The addition rule

Mutually exclusive events ($A \cap B = \emptyset$):

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

Non-mutually exclusive events ($A \cap B \neq \emptyset$):

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

The second formula subtracts the probability of the intersection because $P(A)$ and $P(B)$ each already include the overlap; adding them without correction would double-count it.

Example 1.6 — Addition rule on a die

A fair six-sided die is rolled once. Let A = "result is even" = $\{2, 4, 6\}$ and C = "result is prime" = $\{2, 3, 5\}$. Then $P(A) = P(C) = 3/6 = 1/2$, and $A \cap C = \{2\}$ so $P(A \cap C) = 1/6$. Therefore:

$$P(A \cup C) = \frac{1}{2} + \frac{1}{2} - \frac{1}{6} = \frac{5}{6}$$

The five favourable outcomes are $\{2, 3, 4, 5, 6\}$, which confirms the result by direct enumeration.

1.7.2 The multiplication rule

The multiplication rule governs the probability that two events both occur. Its form depends on whether the events are independent.

Independent events:

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

Dependent events (general case):

$$P(A \cap B) = P(A) * P(B | A) = P(B) * P(A | B)$$

The second formulation introduces **conditional probability**, which we study in §1.7.3.

Example 1.7 — Drawing balls with and without replacement

A box contains 10 balls: 4 white (W) and 6 black (B). Two balls are drawn one after the other. What is the probability that the first is white and the second is black?

With replacement: the two draws are independent.

$$P(W_1 \cap B_2) = \frac{4}{10} * \frac{6}{10} = 0.24$$

Without replacement: the two draws are dependent. After removing one white ball, 9 balls remain (3 white, 6 black). Hence $P(B_2 | W_1) = 6/9$, and:

$$P(W_1 \cap B_2) = \frac{4}{10} * \frac{6}{9} = \frac{24}{90} \approx 0.267$$

The two results differ because replacing the first ball resets the composition of the box, while not replacing it shifts the second-draw probabilities.

1.7.3 Conditional probability

The **conditional probability** of B given that A has occurred (with $P(A) > 0$) is defined as:

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

The idea is that once A is known to have occurred, the effective sample space shrinks from E to A. Within this restricted space, we ask how likely B is. Conditional probability formalises how information revealed during an experiment alters our assessment of what remains uncertain, a mechanism that recurs throughout decision theory, credit scoring, and Bayesian econometrics (Hansen, 2022).

Example 1.8 — Conditional probability with two dice

Two fair dice are rolled. Let A = "the sum equals 6" and B = "at least one die shows 2".

- $A = \{(1,5), (2,4), (3,3), (4,2), (5,1)\}$, $|A| = 5$, $P(A) = 5/36$.
- B contains all 11 pairs with at least one 2; $P(B) = 11/36$.
- $A \cap B = \{(2,4), (4,2)\}$, $P(A \cap B) = 2/36$.

The conditional probability of B given A is:

$$P(B | A) = \frac{2/36}{5/36} = \frac{2}{5} = 0.40$$

Knowing that the sum is 6 raises the chance of seeing at least one 2 from $11/36 \approx 0.31$ to 0.40. The information "sum = 6" is therefore relevant to B.

1.7.4 Independence — the formal definition

Two events A and B are **independent** when:

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

equivalently, when $P(B | A) = P(B)$, provided $P(A) > 0$. Independence means that knowing A occurred does not change the probability of B, and vice versa.

⚠ Watch out — the two most common errors

- Assuming independence because "the events look unrelated". Independence is a numerical property, so always verify $P(A \cap B) = P(A) \times P(B)$ before using it.

- Confusing independence with mutual exclusivity. If A and B are mutually exclusive and have positive probabilities, then they are dependent: knowing A occurred forces $P(B | A) = 0 \neq P(B)$.

Example 1.9 — Checking independence

A retail firm tracks two product categories: food (F) and electronics (X). In a survey of 1 000 clients, 300 bought food, 200 bought electronics, and 60 bought both. Are the two purchases independent?

Compute the probabilities:

- $P(F) = 0.30$
- $P(X) = 0.20$
- $P(F \cap X) = 0.06$

Under independence we would expect $P(F) \times P(X) = 0.30 \times 0.20 = 0.06$, which matches the observed intersection. The two purchases can be treated as independent in this sample.

Had the observed intersection been 0.10 instead, we would conclude that F and X are positively associated: buying food raises the chance of buying electronics.

Table 1.2 summarises the two notions that students most often confuse.

Table 1.2: Independence versus mutual exclusivity

Property	Mutually exclusive events	Independent events
Definition	$A \cap B = \emptyset$	$P(A \cap B) = P(A) \times P(B)$
Can both occur?	No, never	Yes, and with positive probability
$P(A \cap B)$	0	$P(A) \times P(B) > 0$ (if both have positive probability)
$P(A B)$	0	$P(A)$, unchanged by B
Typical example	Rolling a 2 and a 5 on a single die	Outcomes of two successive coin tosses
Relation to each other	Incompatible with independence (when both have positive probability)	Compatible with any numerical overlap, as long as the product rule holds

Source: Elaborated by the Author

The bottom row deserves emphasis: two events with $P(A) > 0$ and $P(B) > 0$ cannot be both mutually exclusive and independent at the same time. The only way for both conditions to hold simultaneously is for one of the events to have probability zero.

1.7.5 The total probability theorem

Consider a collection of events $A_1, A_2, \dots, A_k$ that form a **partition** of the sample space:

- the events are mutually exclusive: $A_i \cap A_j = \emptyset$ whenever $i \neq j$;
- they are exhaustive: $A_1 \cup A_2 \cup \dots \cup A_k = E$.

The partition represents alternative "causes" or "origins" of any observed event. For any event B ,

$$P(B) = \sum_{i=1}^k P(A_i) \cdot P(B | A_i)$$

The total probability of B is the sum of its probabilities under each cause, weighted by the prior probability of that cause. The theorem is the workhorse of quality control, medical screening, credit analytics, and any problem where an observation can arise from distinguishable subpopulations (Camm et al., 2024).

Example 1.10 — Total probability — two production lines

A factory operates two assembly lines. Line L_1 produces 60% of daily output with a 3% defect rate; line L_2 produces the remaining 40% with a 5% defect rate. An item is selected at random from the day's output. What is the probability that it is defective?

Let D = "defective". The causes L_1 and L_2 partition the production. Applying the total probability theorem:

$$P(D) = P(L_1) \cdot P(D | L_1) + P(L_2) \cdot P(D | L_2)$$

$$P(D) = 0.60 \cdot 0.03 + 0.40 \cdot 0.05 = 0.018 + 0.020 = 0.038$$

Roughly 3.8% of daily output is defective on average, a figure that could not be recovered from either line's rate in isolation.

1.7.6 Bayes' theorem

Bayes' theorem answers the inverse question: once we observe event B , how likely is it that a particular cause A_i was responsible? Using the conditional probability definition twice and substituting the total probability expression in the denominator:

$$P(A_i | B) = \frac{P(B | A_i) \cdot P(A_i)}{\sum_j P(B | A_j) \cdot P(A_j)}$$

The Bayesian vocabulary distinguishes two quantities:

- $P(A_i)$ is the *prior probability* — our belief about A_i before observing B ;
- $P(A_i | B)$ is the *posterior probability* — our updated belief after observing B .

Bayes' theorem makes the updating explicit: the posterior is proportional to the prior times the likelihood $P(B | A_i)$. The framework has become central in modern econometrics, machine learning, medical diagnostics, and decision analysis (Hansen, 2022).

Example 1.11 — Bayes' theorem applied to Example 1.10

A randomly chosen item has been found defective. What is the probability that it came from line L_1 ?

From Example 1.10 we already have $P(D) = 0.038$. Then:

$$P(L_1 | D) = \frac{0.60 * 0.03}{0.038} = \frac{0.018}{0.038} \approx 0.474$$

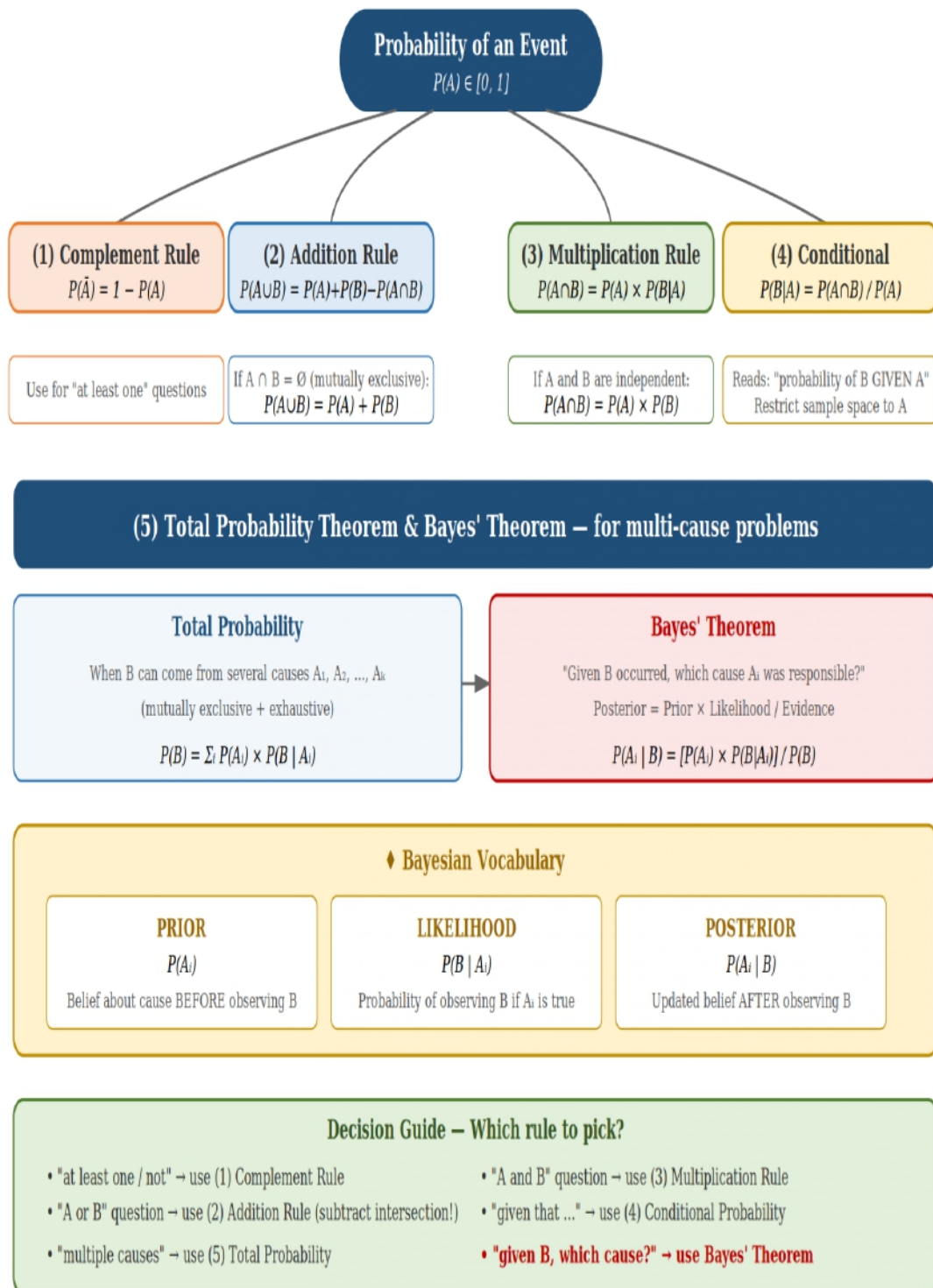
Although L_1 produces the larger share of total output (60%), it accounts for only 47.4% of defects because its defect rate is lower. The complementary posterior $P(L_2 | D) \approx 0.526$ shows that L_2 , despite its smaller output share, contributes the majority of defective items, a finding with clear managerial implications for where to concentrate quality-improvement effort.

Visual Summary — Mind Map of §1.7

The diagram below gathers the four basic rules of probability calculation (complement, addition, multiplication, conditional), the total probability theorem, and Bayes' theorem into a single visual decision guide. Use it as a quick-reference aid while solving exercises.

Figure 1.2 — The rules of probability calculation and Bayes' theorem

Mind Map 2 – Probability Rules & Bayes' Theorem



Source: Elaborated by the Author

1.8 Combinatorial Analysis

When a sample space contains many outcomes, listing them one by one is impractical. Combinatorics provides the counting tools needed to apply the classical definition of probability in such cases. Three structures are sufficient for this course.

1.8.1 Permutations

A **permutation** is an ordered arrangement of n distinct objects. The number of permutations of n objects is:

$$P(n) = n! = n * (n - 1) * \dots * 2 * 1, \quad (\text{with } 0! = 1)$$

Example 1.12 — Delivery routing

A logistics manager must schedule daily visits to five client sites. In how many distinct orders can the route be planned?

Any of the 5 sites can be visited first, then any of the remaining 4 second, and so on. Therefore the number of routes is $5! = 120$. If the vehicle returns to a single depot and the starting point is fixed, the relevant count is the number of circular permutations, $(5 - 1)! = 24$.

1.8.2 Arrangements

An **arrangement** selects k objects from n , where order matters and no object is repeated:

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

Example 1.13 — Allocating distinct roles

From 8 candidates, a management board selects three to fill the positions of chairperson, vice-chairperson, and secretary. Because the three positions are distinct, order matters: swapping two candidates between the chair and secretary seats produces a different outcome. The number of possible assignments is:

$$A(8, 3) = \frac{8!}{5!} = 8 * 7 * 6 = 336$$

1.8.3 Combinations

A **combination** selects k objects from n where order does not matter and no object is repeated. The count is the binomial coefficient:

$$C(n, k) = \frac{n!}{k! * (n - k)!}$$

Combinations are the right tool whenever the selected objects form a *group* with no internal hierarchy: a committee, a sample, a hand of cards.

Example 1.14 — Auditing a sample

A regional tax office must audit 5 firms chosen at random from 20 firms registered in a given sector. How many distinct samples are possible?

The order in which the 5 firms are selected has no operational meaning, since the audit applies to the group. Therefore:

$$C(20, 5) = \frac{20!}{5! * 15!} = 15,504$$

1.8.4 Summary of counting formulas

The four situations most encountered in this course are collected in Table 1.3.

Table 1.3: Summary of counting methods

Selection type	Order matters?	Repetition?	Formula
Combination	No	No	$C(n, k) = n! / [k!(n-k)!]$
Arrangement	Yes	No	$A(n, k) = n! / (n-k)!$
Permutation (all)	Yes	No	$n!$
Ordered with repetition	Yes	Yes	n^k

Source: Elaborated by the Author

★ Mnemonic

- If the objects keep their identity and their position matters → arrangement (or permutation).
- If only the group matters → combination.
- With repetition → use n^k (ordered) or $C(n + k - 1, k)$ (unordered).

1.8.5 Probability trees — a visual computation device

A probability tree is a diagram that organises a sequential experiment into branches. Each branch represents one possible stage of the experiment, and each node carries a conditional probability. Reading the tree makes the multiplication rule (for the probability of a path) and the total probability theorem (for the probability of an event across several paths) mechanical rather than abstract.

The construction is always the same: at each level of the tree, the branch labels are conditional probabilities given the events along the path from the root to that node. The

probability of any path is the product of the labels along that path. The probability of a compound event is the sum of the probabilities of all paths leading to it.

Example 1.20 — Probability tree for a quality inspection

A factory has two production lines L_1 and L_2 contributing 60% and 40% of daily output. Defect rates are 3% and 5% respectively. We can summarise the situation in a tree with two levels: the line (L_1 or L_2) and the quality status (defective D or good G).

The first-level probabilities are $P(L_1) = 0.60$ and $P(L_2) = 0.40$.

The second-level branches carry the conditional probabilities

$$P(D | L_1) = 0.03, P(G | L_1) = 0.97, P(D | L_2) = 0.05, P(G | L_2) = 0.95.$$

The path probabilities are:

$$P(L_1 \cap D) = 0.60 * 0.03 = 0.018$$

$$P(L_1 \cap G) = 0.60 * 0.97 = 0.582$$

$$P(L_2 \cap D) = 0.40 * 0.05 = 0.020$$

$$P(L_2 \cap G) = 0.40 * 0.95 = 0.380$$

These four probabilities sum to 1, which is a useful consistency check. The marginal probability of a defective item is $P(D) = 0.018 + 0.020 = 0.038$, the same result we obtained from the total probability theorem in Example 1.10.

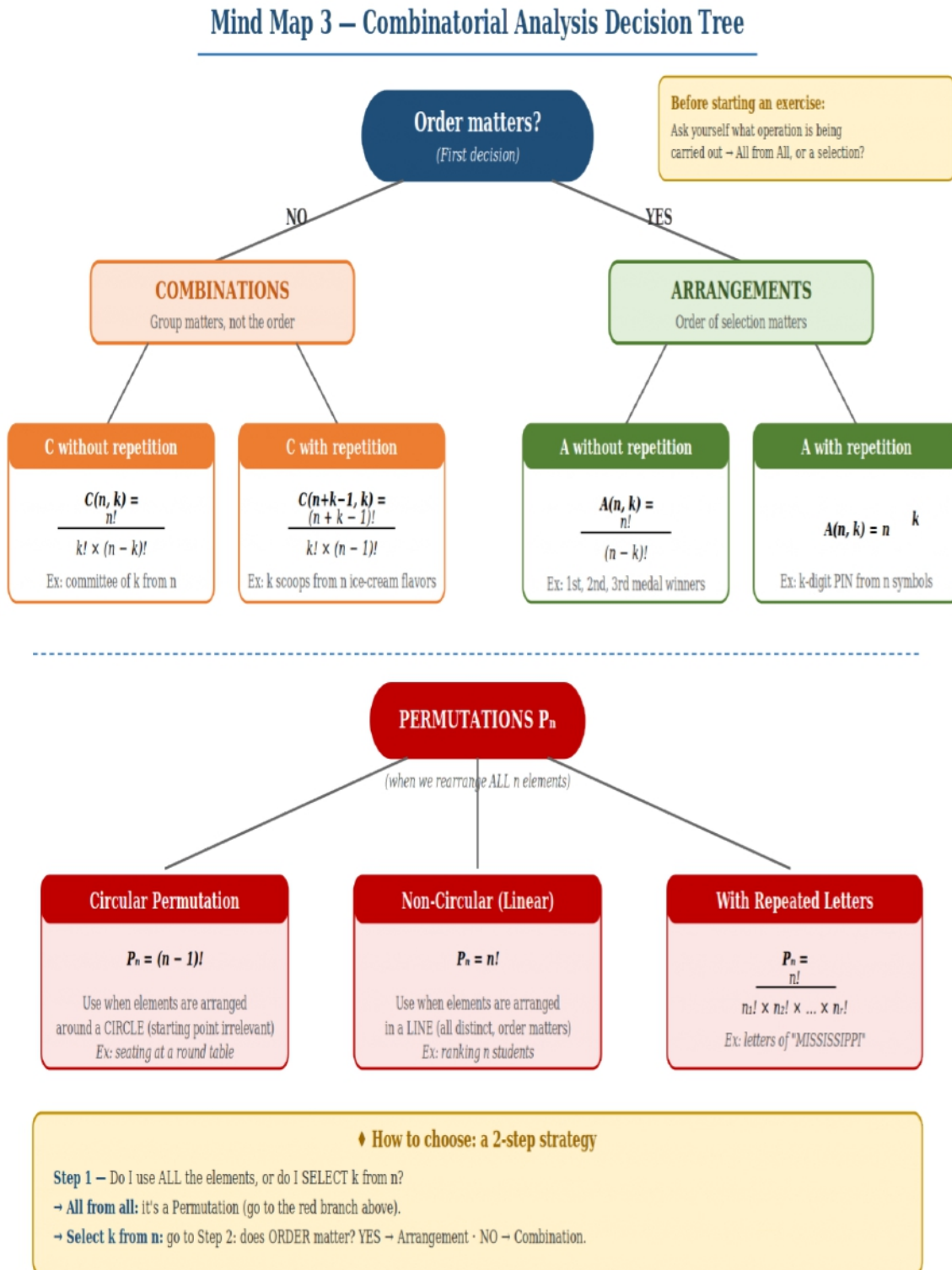
★ **When to draw a probability tree**

- Whenever the experiment has two or more sequential stages (e.g. first a cause is realised, then an observation is made).
- Whenever the statement contains several conditional probabilities.
- Whenever you need to apply Bayes' theorem, the tree makes the numerator and denominator of Bayes' formula visible at a glance.

Visual Summary — Mind Map of §1.8

The decision tree below is the single most useful tool for choosing the correct combinatorial formula on an exam. Start at the root ("Order matters?") and follow the branch that matches your problem. The Permutations sub-tree below covers the special case where ALL n elements are rearranged.

Figure 1.3: Combinatorial analysis: a decision tree for selecting the correct formula



Source: Elaborated by the Author

1.9 Worked Examples in an Economic Context

Example 1.15 Screening loan applications

A bank records that 80% of loan applicants have a regular monthly income (R) and 20% do not ($\bar{R}$). Among applicants with regular income, 95% repay their loan on schedule (S); among those without, only 60% do.

- (a) What is the probability that a randomly chosen applicant will repay on schedule?
- (b) A loan has been repaid on schedule. What is the probability that the borrower had regular income?

Solution (a). The events R and $\bar{R}$ partition the applicant population. By the total probability theorem:

$$P(S) = P(R) * P(S | R) + P(\bar{R}) * P(S | \bar{R})$$

$$P(S) = 0.80 * 0.95 + 0.20 * 0.60 = 0.760 + 0.120 = 0.880$$

On average, 88% of loans are repaid on schedule.

Solution (b). Using Bayes' theorem:

$$P(R | S) = \frac{P(R)}{P(S)} * P(S | R); = \frac{0.80 * 0.95}{0.880} = \frac{0.760}{0.880} \approx 0.864$$

Among loans repaid on time, approximately 86.4% were taken by applicants with regular income, a figure the bank can use when calibrating its credit-scoring model.

Example 1.16 Selecting a committee

An association counts 7 economists and 5 managers. A working committee of 4 members is drawn at random from the 12.

- (a) How many different committees are possible?
- (b) What is the probability that the committee contains exactly 2 economists?
- (c) What is the probability that it contains at least one manager?

Solution (a). $C(12, 4) = 12! / (4! \times 8!) = 495$.

Solution (b). Favourable committees: 2 economists from 7, and 2 managers from 5.

$$C(7, 2) * C(5, 2) = 21 * 10 = 210$$

$$P(\text{exactly 2 economists}) = \frac{210}{495} \approx 0.424$$

Solution (c). The complement of "at least one manager" is "no manager", i.e. all 4 drawn from the economists. There are $C(7, 4) = 35$ such committees, so:

$$P((no\ manager)) = \frac{35}{495} \approx 0.071$$

$$P((at\ least\ one\ manager)) = 1 - 0.071 \approx 0.929$$

Using the complement rule saved us from enumerating the four cases "exactly 1 manager", "exactly 2 managers", and so on.

Example 1.17 Sampling without replacement

A warehouse receives 50 electronic components, among which 5 are defective. The quality inspector randomly tests 3 of them (without replacement). What is the probability that none of the three tested components is defective?

Favourable selections draw all 3 from the 45 non-defective items: $C(45, 3)$. Total selections: $C(50, 3)$. Therefore:

$$P(\text{no defective}) = \frac{\binom{45}{3}}{\binom{50}{3}} = \frac{14,190}{19,600} \approx 0.724$$

About 72.4% of random inspections of size 3 will fail to reveal the defect. That would not be acceptable if defect detection were the main goal, and it argues for a larger sample size.

Example 1.18 A three-cause medical screening

A regional health authority tracks three sources of a respiratory infection: seasonal flu (F_1), bronchitis (F_2), and environmental allergies (F_3), with prevalences 0.55, 0.30, and 0.15 respectively among all consulting patients. Among patients with each condition, the probability of presenting "high fever" (H) is 0.70, 0.50, and 0.05.

A new patient presents high fever. What is the posterior probability of each cause?

Step 1 — Total probability.

$$P(H) = 0.55 * 0.70 + 0.30 * 0.50 + 0.15 * 0.05$$

$$P(H) = 0.385 + 0.150 + 0.0075 = 0.5425$$

Step 2 — Bayes' theorem for each cause.

$$P(F_1 | H) = \frac{0.385}{0.5425} \approx 0.710$$

$$P(F_2 | H) = \frac{0.150}{0.5425} \approx 0.277$$

$$P(F_3 | H) = \frac{0.0075}{0.5425} \approx 0.014$$

Given the symptom "high fever", the most likely diagnosis is seasonal flu (71%), followed by bronchitis (27.7%). Allergies become almost negligible once the fever is observed, even though

they accounted for 15% of the prior distribution. A three-cause Bayesian update therefore reorders the diagnoses dramatically.

Example 1.19 Urn problem with conditional structure

Two urns U_1 and U_2 contain balls as follows: U_1 has 5 white and 3 red; U_2 has 2 white and 6 red. A fair coin is tossed; if it shows heads, a ball is drawn from U_1 , otherwise from U_2 . The drawn ball is white. What is the probability that it came from U_1 ?

Let W = "white ball drawn".

$$P(W | U_1) = \frac{5}{8} \quad , \quad P(W | U_2) = \frac{2}{8} = \frac{1}{4}$$

$$P(U_1) = P(U_2) = \frac{1}{2}$$

$$P(W) = \frac{1}{2} * \frac{5}{8} + \frac{1}{2} * \frac{2}{8} = \frac{7}{16}$$

$$P(U_1 | W) = \frac{P(U_1)}{P(W)} * P(W | U_1); = \frac{(1/2)}{7/16} * (5/8); = \frac{5}{7} \approx 0.714$$

Observing a white ball nearly doubles the prior probability 0.5 to approximately 0.714, reflecting the richer white composition of urn U_1 .

Example 1.20 Tree diagram for a three-stage experiment

A small consulting firm receives project proposals through three channels: direct (D) with prior probability 0.40, referrals (R) with 0.35, and open calls (C) with 0.25. The probability of proposal acceptance given each channel is 0.70, 0.50, and 0.30 respectively. A randomly selected proposal has been accepted. From which channel is it most likely to have come?

Step 1. Build the probability tree. The first stage is the channel (D, R, C); the second stage is the outcome (A = accepted, $\bar{A}$ = rejected). Each path has a product probability.

$$P(D \cap A) = 0.40 * 0.70 = 0.280$$

$$P(R \cap A) = 0.35 * 0.50 = 0.175$$

$$P(C \cap A) = 0.25 * 0.30 = 0.075$$

Step 2. Total probability of acceptance:

$$P(A) = 0.280 + 0.175 + 0.075 = 0.530$$

Step 3. Posterior probabilities via Bayes' theorem:

$$P(D | A) = \frac{0.280}{0.530} \approx 0.528$$

$$P(R | A) = \frac{0.175}{0.530} \approx 0.330$$

$$P(C | A) = \frac{0.075}{0.530} \approx 0.142$$

The direct channel D accounts for the majority of accepted proposals (52.8%), even though it represented only 40% of incoming proposals *ex ante*. The open-call channel C, despite contributing a quarter of proposals, accounts for less than 15% of accepted ones. A marketing decision about where to allocate future effort can be made on the basis of these posterior probabilities.

Verification: the three posteriors sum to 1, which they must since D, R, C form a partition.

Example 1.21 Counting sequences under constraints

A training programme requires each participant to follow five modules labelled M_1, M_2, M_3, M_4, M_5 , but the order in which modules are taken is constrained: M_1 must be taken first, and M_5 must be taken last. How many ordered programmes satisfy these constraints?

The first and last positions are fixed. The three middle modules M_2, M_3, M_4 can be taken in any order. The number of arrangements of three distinct objects is $3! = 6$. Therefore:

$$(\text{Number of admissible programmes}) = 3! = 6$$

If the constraint were relaxed to " M_1 must be first" alone (with any order for the last four), the count would rise to $4! = 24$. And with no constraints at all, the count is $5! = 120$. The reduction from 120 to 6 shows how strongly an ordering constraint can shrink the admissible set, a fact with practical implications in scheduling problems.

Example 1.22 A mixed selection problem

From a pool of 10 candidates (6 with a background in economics and 4 in management), a jury selects 4 to interview, then chooses 2 of the 4 for final hiring. What is the probability that both final hires have an economics background?

We solve this in two steps, using the multiplication rule.

Step 1. Probability of the interview composition. Let X be the number of economists among the four interviewed. Under random selection:

$$P(X = 4) = \frac{\binom{6}{4}\binom{4}{0}}{\binom{10}{4}} = \frac{15}{210} \approx 0.071$$

$$P(X = 3) = \frac{\binom{6}{3}\binom{4}{1}}{\binom{10}{4}} = \frac{80}{210} \approx 0.381$$

$$P(X = 2) = \frac{\binom{6}{2}\binom{4}{2}}{\binom{10}{4}} = \frac{90}{210} \approx 0.429$$

Step 2. Given X economists among the four, the probability that both final hires are economists is $C(X, 2) / C(4, 2) = C(X, 2) / 6$. Summing over the partition $X \in \{2, 3, 4\}$:

$$\begin{aligned} P(\text{both hired economists}) &= \sum_{X=2}^4 P(X) * \frac{C(X, 2)}{6} \\ &= 0.429 * \frac{1}{6} + 0.381 * \frac{3}{6} + 0.071 * \frac{6}{6} \\ &= 0.0715 + 0.1905 + 0.0710 \approx 0.333 \end{aligned}$$

One chance in three. The symmetry is not accidental: under random selection throughout the process, the final-hire composition must reflect the pool composition, and $6/10 \times 5/9 = 1/3$ is exactly what uniform sampling of two economists from the original pool would give.

Example 1.23 Mini-case: Loan portfolio analysis

A small commercial bank holds a portfolio of 300 corporate loans distributed across three sectors: agriculture (90 loans), manufacturing (120 loans), and services (90 loans). Historical data show one-year default rates of 6%, 4%, and 3% respectively. The bank's risk officer is preparing the monthly report.

Question (a). What is the probability that a randomly selected loan in the portfolio defaults within one year?

The sectors partition the portfolio. Prior probabilities:

$$P(A) = \frac{90}{300} = 0.30, \quad P(M) = \frac{120}{300} = 0.40, \quad P(S) = \frac{90}{300} = 0.30$$

Applying the total probability theorem:

$$\begin{aligned} P(D) &= 0.30 * 0.06 + 0.40 * 0.04 + 0.30 * 0.03 = 0.018 + 0.016 + 0.009 \\ &= 0.043 \end{aligned}$$

The expected default rate for the whole portfolio is 4.3%, a figure useful for loan-loss provisioning.

Question (b). A loan has defaulted. From which sector is it most likely to have come?

Bayes' theorem gives the posterior probability of each sector given default:

$$\begin{aligned} P(A | D) &= \frac{0.30 * 0.06}{0.043} \approx 0.419 \\ P(M | D) &= \frac{0.40 * 0.04}{0.043} \approx 0.372 \\ P(S | D) &= \frac{0.30 * 0.03}{0.043} \approx 0.209 \end{aligned}$$

Agriculture accounts for roughly 42% of defaults despite representing only 30% of the portfolio, because its default rate is the highest. The risk officer will flag this sector as a priority for portfolio monitoring. Manufacturing, with 40% of the portfolio but a moderate default rate, contributes 37% of defaults. Services contribute proportionally the least.

Managerial takeaway. Even a low prior weight combined with a high conditional probability can produce a large posterior, and vice versa. Decisions about capital allocation, provisioning, or relationship management should rely on posteriors rather than priors whenever the triggering event (default) is what concerns the decision-maker.

1.10 Common Mistakes — Things to Watch For

The mistakes described below recur in almost every cohort of first-year students. Reading them carefully before tackling the exercises will save you time and marks. Each mistake is followed by a short diagnostic question that will help you identify whether you are about to commit it.

⚠ Mistake 1 — Treating mutually exclusive events as independent

- If A and B are mutually exclusive with $P(A) > 0$ and $P(B) > 0$, then $P(A \cap B) = 0 \neq P(A) \times P(B)$, so A and B are dependent.
- Always test the multiplicative condition $P(A \cap B) = P(A) \times P(B)$ numerically before asserting independence.

Self-check: boldcolor

Quick test: if two events cannot occur together ($A \cap B = \emptyset$), then knowing A occurred immediately tells you B did not. That is dependence, not independence.

⚠ Mistake 2 — Wrong denominator in conditional probability

- $P(B | A)$ has denominator $P(A)$, not $P(B)$ or $P(A \cap B)$.
- A useful memory aid: the bar | is read "given"; what comes after the bar goes to the denominator.
- Equivalent formulations: $P(A \cap B) = P(A) \times P(B | A) = P(B) \times P(A | B)$.

In a tree diagram, the conditional probability labels always point from a parent node to its children; their denominator is the probability of the parent. Confusing this structure produces wrong probabilities even when the correct values are all on the tree.

⚠ Mistake 3 — Forgetting to subtract the intersection

- $P(A \cup B) = P(A) + P(B)$ holds only when A and B are mutually exclusive.
- In every other case, subtract $P(A \cap B)$; forgetting this subtraction gives probabilities greater than 1.
- For three events: $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$. (Inclusion-exclusion.)

⚠ Mistake 4 — Confusing "permutations" and "combinations"

- Ask yourself: would swapping two chosen objects change the outcome?
- If yes → arrangement or permutation ($n! / (n-k)!$ or $n!$).

- If no $\rightarrow$ combination $C(n, k)$.
- Poker hands: combinations, because the order in which you receive cards does not matter.
- Olympic medal podium: arrangements, because gold $\neq$ silver $\neq$ bronze.

⚠ Mistake 5 — Missing the partition condition in total probability

- The total probability theorem requires the events A_i to be both mutually exclusive and exhaustive.
- If the A_i overlap or do not cover E , the formula $P(B) = \sum P(A_i) P(B | A_i)$ no longer holds.
- Sanity check: $P(A_1) + P(A_2) + \dots + P(A_k)$ must equal 1.

⚠ Mistake 6 — Mixing up prior and posterior in Bayes' theorem

- $P(A_i)$ is the prior; $P(A_i | B)$ is the posterior.
- The denominator $P(B)$ always equals the weighted sum $\sum P(A_j) P(B | A_j)$. Never forget to compute it.
- After applying Bayes, the posteriors $P(A_i | B)$ must sum to 1 over the partition, a useful consistency check.

⚠ Mistake 7 — Arithmetic and sanity checks

- Final probability must lie in $[0, 1]$. Any negative value or value > 1 indicates a mistake, usually a misplaced sign or a double-count.
- All posterior probabilities $P(A_i | B)$ should sum to 1 over a partition.
- Reasonableness: if a posterior probability is drastically different from the prior, re-examine the likelihoods to make sure the evidence is as informative as the computation suggests.

1.11 Key Takeaways

The most important ideas of this chapter, condensed.

- ✓ A random experiment produces outcomes that cannot be predicted individually; the sample space E lists every possible outcome.
- ✓ An event is a subset of E ; set algebra (union, intersection, complement) and De Morgan's laws apply without modification.
- ✓ Probability is a function P with three axioms: $P(A) \geq 0$, $P(E) = 1$, and countable additivity for mutually exclusive events.
- ✓ Three computational definitions coexist: classical (equally likely outcomes), empirical (observed frequency), and axiomatic (Kolmogorov).
- ✓ Complement rule: $P(\bar{A}) = 1 - P(A)$. Use it for "at least one" questions.
- ✓ Addition rule: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$; the subtraction disappears when events are mutually exclusive.
- ✓ Multiplication rule: $P(A \cap B) = P(A) P(B | A)$; if A and B are independent, this reduces to $P(A) P(B)$.
- ✓ Conditional probability: $P(B | A) = P(A \cap B) / P(A)$; reading a scenario sequentially usually makes the conditioning structure obvious.
- ✓ Total probability: $P(B) = \sum_i P(A_i) P(B | A_i)$ over a partition, the key tool for problems with multiple causes.
- ✓ Bayes' theorem: $P(A_i | B) \propto P(A_i) P(B | A_i)$. It inverts the direction of conditioning and updates beliefs after evidence.
- ✓ Counting: $n!$ for full permutations; $A(n, k)$ for ordered selections; $C(n, k)$ for unordered selections.
- ✓ Mutually exclusive $\neq$ independent: this single confusion generates the largest number of errors in introductory probability.

Bridging to Chapter 2

The concepts of this chapter pave the way for the notion of a random variable, which is introduced formally in Chapter 2. A random variable is a numerical summary of a random experiment: rather than describing every outcome, we encode each outcome as a number (for example, a cost, a count, a duration), and we study the probability distribution of that number.

Three bridges in particular will be essential in what follows:

- **The sample space E** becomes the domain on which a random variable is defined.
- **The addition and multiplication rules** will reappear as the rules for computing the probability distribution of a sum or a product of random variables.
- **The Bayes–total probability framework** carries over verbatim to conditional expectations and to the law of total expectation.

Mastering the present chapter therefore pays returns not only in the rest of this booklet but in every quantitative economics course you will take subsequently.

1.12 Chapter Summary Cards

The cards below condense the chapter into four visual summaries, designed for quick revision before an assessment.

Card 1 — Essential Concepts at a Glance

This card gathers the definitions and formulae that every student must know cold before any examination.

✓ ESSENTIAL CONCEPTS

Random experiment: procedure with uncertain outcome; outcomes listed in the sample space E .

Event: subset of E ; denoted by capital letters A, B, C .

Probability: function P : events $\rightarrow [0, 1]$ satisfying Kolmogorov's three axioms.

Complement: $P(\bar{A}) = 1 - P(A)$.

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

Multiplication: $P(A \cap B) = P(A) P(B | A)$; $P(A) P(B)$ if independent.

Conditional: $P(B | A) = P(A \cap B) / P(A)$.

Total probability: $P(B) = \sum_i P(A_i) P(B | A_i)$ over a partition.

Bayes: $P(A_i | B) = P(A_i) P(B | A_i) / P(B)$.

Counting: $n!$; $A(n, k) = n!/(n-k)!$; $C(n, k) = n!/[k!(n-k)!]$.

Card 2 — Terminology Table

This table is a quick-reference glossary. Use it during revision to check your fluency with symbols and vocabulary.

Term	Symbol	Meaning
Sample space	E, Ω	Set of all outcomes
Event	A, B, C	Subset of E
Probability	$P(A)$	Chance that A occurs
Complement	$\bar{A}$	A does not occur
Union	$A \cup B$	A or B (or both)
Intersection	$A \cap B$	A and B both
Conditional	$P(B A)$	B given A has occurred
Factorial	$n!$	$n \times (n-1) \times \dots \times 1$
Combination	$C(n, k)$	Unordered selection
Arrangement	$A(n, k)$	Ordered selection

Card 3 — Decision Guide: Which Rule to Apply?

Most exam problems reduce to identifying which formula applies. This card gives you the decision rule to use when you read the problem statement.

⚠ DECISION GUIDE

"At least one" or "not" → use the complement rule $P(\bar{A}) = 1 - P(A)$.

Two events, "and" question → multiplication rule; check whether they are independent.

Two events, "or" question → addition rule; subtract the intersection unless mutually exclusive.

"Given that ..." → conditional probability $P(B | A) = P(A \cap B) / P(A)$.

Multiple causes / origins → total probability over a partition.

"Observed B , probability of cause A_i " → Bayes' theorem.

Large sample space → combinatorics ($C, A, n!$) to count efficiently.

Sequential stages → draw a probability tree and label each branch with its conditional probability.

Card 4 — Typical Exam Questions

The exam typically tests five core competencies, listed below. Make sure you can solve each type before sitting the final test.

⚠ TYPICAL EXAM PROBLEMS

- Q1.** Construct the sample space and compute $P(A)$ for a structured experiment.
- Q2.** Use the addition and multiplication rules in a compound-event problem.
- Q3.** Apply the total probability theorem to a 2- or 3-cause situation.
- Q4.** Apply Bayes' theorem and comment on the posterior probabilities.
- Q5.** Combinatorial selection (committee, sample, team).

★ EXAM TIPS

- Read the full statement twice before writing anything.
- Translate each information sentence into a probability statement on paper.
- Draw a tree diagram whenever the problem involves sequential events.
- Always check that $\sum P(A_i | B) = 1$ over a partition after applying Bayes.
- Interpret every numerical result in the context of the original problem (a percentage, a proportion, a frequency).
- Budget your time: spend roughly 1 minute per mark, and keep 10 minutes at the end for verification.

1.13 Practice Exercises

The exercises below form a complete set for revision and exam preparation. They are ordered by increasing difficulty. Solutions are not provided, and that is deliberate: attempt every problem with pen, paper, and the tools developed in this chapter.

1.13.1 Level 1 — Foundations

Exercise 1.1. A box contains 8 red balls and 4 blue balls. One ball is drawn at random. Compute (a) $P(\text{red})$, (b) $P(\text{blue})$, (c) $P(\text{not red})$.

Exercise 1.2. A fair six-sided die is rolled once. Let $A = \text{"result greater than 4"}$ and $B = \text{"result is even"}$. (a) Find $P(A)$ and $P(B)$. (b) Are A and B mutually exclusive? (c) Compute $P(A \cup B)$.

Exercise 1.3. A student's probability of passing Statistics is 0.65; of passing Microeconomics, 0.70. Assuming the two results are independent, find the probability that the student passes both subjects.

Exercise 1.4. In a certain city, 60% of households own a car (C) and 30% own a motorcycle (M). 15% own both. (a) What is the probability that a randomly chosen household owns a car or a motorcycle? (b) What is the probability that it owns neither?

Exercise 1.5. Compute: (a) $7!$; (b) $C(8, 3)$; (c) $A(10, 4)$; (d) $C(n, 0)$ and $C(n, n)$.

1.13.2 Level 2 — Applications

Exercise 1.6. A factory uses two machines M_1 and M_2 , producing respectively 60% and 40% of daily output. Defect rates are 3% for M_1 and 5% for M_2 . An item is drawn at random and found defective. Using Bayes' theorem, find the probability that it was produced by M_1 .

Exercise 1.7. A public park has two entrances G_1 and G_2 . Of all visitors, 70% use G_1 and 30% use G_2 . Of visitors entering by G_1 , 40% also exit by G_1 ; of those entering by G_2 , 50% exit by G_2 . A visitor is selected at random. (a) What is the probability that this visitor exits by G_1 ? (b) Given that the visitor exited by G_2 , what is the probability that they entered by G_1 ?

Exercise 1.8. A committee of 3 is drawn at random from 9 men and 3 women. Compute the probability that the committee contains exactly one woman.

Exercise 1.9. An insurance firm classifies drivers as low-risk (L), medium-risk (M), or high-risk (H) with proportions 0.50, 0.35, and 0.15. Accident rates are 2%, 6%, and 15%. A randomly chosen policyholder had an accident last year. What is the probability that they are high-risk?

Exercise 1.10. A box contains 15 electronic components, 3 of which are defective. 4 components are drawn simultaneously. (a) Compute the probability that none of them is defective. (b) Compute the probability that exactly one is defective. (c) Compute the probability that at least one is defective.

1.13.3 Level 3 — Extension

Exercise 1.11. A firm operates three plants P_1 , P_2 , and P_3 , producing 3 000, 5 000, and 2 000 units per day respectively. Defect rates are 5%, 2%, and 5%. A randomly selected unit is inspected and found defective. (a) Represent the situation with a probability tree. (b) Compute the probability that the unit came from each plant. (c) Compute the total probability that a randomly chosen unit is defective.

Exercise 1.12. Twenty-five numbered balls (1 to 25) are placed in an urn. Three are drawn simultaneously. Compute the probability that (a) all three are even; (b) all three are multiples of 5; (c) the sum of two of them equals 10.

Exercise 1.13. Nine friends attend a football match. In how many ways can they sit (a) in a row of 9 seats; (b) in a row of 7 seats (only 7 sit, 2 stand); (c) around a circular table with 9 seats?

Exercise 1.14. A bank has three branches with respective daily numbers of customers 200, 300, and 500. The proportion of customers who open an account during the day is 10% at branch 1, 15% at branch 2, and 12% at branch 3. (a) What is the probability that a random customer (selected from across the whole bank) opened an account today? (b) If a customer opened an account today, what is the probability that they visited branch 2?

Exercise 1.15. A rapid test for a disease has a sensitivity of 95% ($P(\text{positive} \mid \text{disease}) = 0.95$) and a specificity of 90% ($P(\text{negative} \mid \text{no disease}) = 0.90$). In a given population, 2% of individuals have the disease. A randomly tested person has a positive result. What is the probability that they actually have the disease? Comment on the unexpected numerical outcome.

1.13.4 Supplementary Exam-Style Problems

Exercise 1.16 — Monty Hall variant. A game show has three identical doors. Behind one door is a prize; behind the other two are empty rooms. A contestant picks one door at random. The host then opens one of the remaining two doors to reveal an empty room. The contestant is offered the choice to switch to the third (unopened) door or to keep the original choice. Using conditional probability, compute the probability of winning under each strategy. Which strategy is optimal and by how much?

Exercise 1.17 — Credit scoring with Bayes. A credit-scoring model classifies applicants as low-risk (70% of applicants) or high-risk (30%). The model correctly flags a high-risk applicant 85% of the time and correctly clears a low-risk applicant 90% of the time. An applicant has been flagged as high-risk by the model. What is the probability that this applicant is genuinely high-risk? Interpret the result in light of the base-rate effect.

Exercise 1.18 — Committee formation with constraints. A university department must form an evaluation committee of 5 members from a pool of 8 professors, 6 researchers, and 4 administrators. Compute: (a) the total number of possible committees; (b) the number of

committees with at least 2 professors; (c) the number of committees with exactly 1 administrator and at least 2 researchers.

Exercise 1.19 — Multi-stage urn problem. Two urns contain: urn A has 4 white and 6 black balls; urn B has 7 white and 3 black balls. An urn is chosen at random (each with probability $1/2$), then two balls are drawn without replacement from the selected urn. (a) Compute the probability that both drawn balls are white. (b) Given that both drawn balls are white, what is the probability that urn B was selected?

Exercise 1.20 — Independence test in a survey. A market-research survey of 1 000 consumers reports: 450 purchased product X, 380 purchased product Y, and 170 purchased both. (a) Compute $P(X)$, $P(Y)$, and $P(X \cap Y)$. (b) Are purchases of X and Y independent? (c) If they are not independent, compute the conditional probabilities $P(X | Y)$ and $P(Y | X)$ and comment on the direction of the association.

1.13.6 Answers to the Prerequisites Diagnostic (§1.3)

- (a) $A \cup B = \{1, 2, 3, 4, 6\}$; $A \cap B = \{2\}$.
- (b) Complement of A in E = $\{1, 3, 5\}$.
- (c) $5! / (3! \times 2!) = 120 / 12 = 10$; $C(6, 2) = 15$.
- (d) Relative frequency = $12 / 60 = 0.20$.
- (e) Yes, they are disjoint: their intersection is empty.

1.14 References

The references below are the primary academic sources consulted for the preparation of this chapter. The booklet-wide bibliography appears at the end of the volume.

1. Blitzstein, J. K., & Hwang, J. (2019). *Introduction to probability* (2nd ed.). Chapman & Hall/CRC. <https://doi.org/10.1201/9780429428357>
2. Camm, J. D., Cochran, J. J., Fry, M. J., Ohlmann, J. W., Anderson, D. R., Sweeney, D. J., & Williams, T. A. (2024). *Statistics for business and economics* (15th ed.). Cengage Learning.
3. Devore, J. L., Berk, K. N., & Carlton, M. A. (2021). *Modern mathematical statistics with applications* (3rd ed.). Springer. <https://doi.org/10.1007/978-3-030-55156-8>
4. Hansen, B. E. (2022). *Probability and statistics for economists*. Princeton University Press.
5. Hurlin, C., & Mignon, V. (2022). *Statistique et probabilités en économie-gestion* (2e éd.). Dunod.
6. Lecoutre, J.-P. (2023). *Statistique et probabilités : Cours et exercices corrigés* (7e éd.). Dunod.
7. Ross, S. M. (2024). *A first course in probability* (10th ed., Global Edition). Pearson Education.

End of Chapter 1

You have now mastered the foundations of probability theory.

Next: Chapter 2 will introduce random variables
and their probability distributions.

✓ **Before moving on, make sure you can:**

- Build the sample space of any random experiment described in words.
- Compute unconditional and conditional probabilities using the four basic rules.
- Apply the total probability theorem over any partition of the sample space.
- Use Bayes' theorem to update prior beliefs after observing new evidence.
- Count outcomes efficiently using combinations, arrangements, and permutations.
- Spot and avoid the seven common mistakes listed in §1.10.

CHAPTER 2

Random Variables

Discrete · Continuous · PMF · PDF · CDF · Expectation · Variance

Chapter 1 taught us how to assign probabilities to events. In practice, though, we rarely care about the outcomes themselves. We care about some number attached to each one: a profit, a cost, a duration, a count, a return. Random variables are the machinery that turns probabilistic descriptions into numbers we can compute with.

This chapter develops that machinery in both the discrete setting (where outcomes are countable, such as defects in a batch or customers arriving per hour) and the continuous setting (where outcomes form an interval, such as waiting times, stock returns, or measurement errors). By the end of the chapter you will be able to describe a random variable by its probability mass or density function, compute its expected value and variance, apply simple transformations, and interpret the results in an economic context.

i In This Chapter

- §2.1 Introduction — From outcomes to numbers
- §2.2 Learning Objectives (Bloom's taxonomy)
- §2.3 Prerequisites Check
- §2.4 Notation and Symbols
- §2.5 Definition of a Random Variable
- §2.6 Discrete Random Variables (PMF, CDF)
- §2.7 Continuous Random Variables (PDF, CDF)
- §2.8 Comparison: Discrete vs Continuous
- §2.9 Expectation of a Random Variable
- §2.10 Variance and Standard Deviation
- §2.11 Transformations of Random Variables
- §2.12 Worked Examples in Economic Context
- §2.13 Common Mistakes — Watch Out
- §2.14 Key Takeaways
- §2.15 Chapter Summary Cards
- §2.16 Practice Exercises
- §2.17 References

2.1 Introduction

2.1.1 Why we need random variables

Consider a simple economic experiment: a consumer is asked whether they will purchase a new product. The sample space $E = \{\text{Yes}, \text{No}\}$ contains only two outcomes, but the manager's interest is rarely in the outcomes themselves. What matters is the associated numerical quantity: revenue generated, profit margin, marketing cost saved. These numbers are the objects on which managerial decisions are based, and they are exactly what random variables formalise.

A **random variable** is a rule that attaches a number to each outcome of a random experiment. The qualifier "random" does not mean "arbitrary"; the rule itself is fully determined in advance. It means instead that the number we will observe is unknown until the experiment is performed. In the consumer example, we could define $X = 1$ if Yes and $X = 0$ if No; this is the simplest random variable, known as an indicator variable, and it already carries enough information to compute average purchase rates and their variability (Blitzstein & Hwang, 2019).

The power of this formalism is that it replaces the often complicated object "sample space E " with a numerical object "the distribution of X ", which can be visualised, summarised by descriptive statistics, and combined with other variables through familiar algebraic operations. Once this translation is made, the entire apparatus of calculus and algebra becomes available for reasoning about uncertainty.

2.1.2 A brief historical note

The concept of a random variable emerged gradually. Jakob Bernoulli (1713) implicitly worked with what we would now call a binomial random variable when counting successes in repeated trials. Abraham de Moivre (1738) and Pierre-Simon Laplace (1812) used continuous analogues, namely density functions, to approximate binomial probabilities by what became the normal distribution. The modern abstract definition, in which a random variable is a measurable function from the sample space to the real line, was formalised by Andrey Kolmogorov (1933) together with the axiomatic foundation of probability theory (Ross, 2024).

For economics and management, the decisive step was the incorporation of random variables into decision-making under uncertainty. Frank Ramsey (1926), Bruno de Finetti (1937), and Leonard Savage (1954) showed that rational choice among risky alternatives reduces to the comparison of expected utilities, a computation that presupposes the random-variable framework developed in this chapter (Hansen, 2022). Today, nearly every quantitative tool used in economics, finance, actuarial science, or business analytics starts from a specification of one or more random variables and their joint distribution.

2.1.3 How this chapter extends Chapter 1

Chapter 1 developed the probability of events A, B, C subsets of the sample space. The present chapter lifts that framework to numerical quantities:

- **From events to numbers.** An event tells us whether "something happened". A random variable tells us "how much". The transition is achieved by assigning a real number to each outcome.
- **From probabilities to distributions.** Rather than studying individual $P(A), P(B), P(C)$, we now study the entire probabilistic behaviour of X at once, captured by its PMF, PDF, or CDF.
- **From counting to summation / integration.** The rules $\sum p(x_i) = 1$ and $\int f(x) dx = 1$ are direct continuations of axiom 2 of Kolmogorov ($P(E) = 1$) applied to the numerical range of X .
- **From $P(A \cap B)$ to $E(XY)$.** Independence and conditional probability, introduced in Chapter 1, reappear here as properties of expectations and variances of sums and products.

The reward for this abstraction is considerable: once the random-variable framework is in place, the machinery of calculus, algebra, and optimisation becomes directly applicable to problems of uncertainty. The single most important outcome of this chapter, the pair of summary statistics $(E(X), \text{Var}(X))$, will appear in every subsequent chapter of this booklet and in every quantitative course you will take thereafter.

2.2 Learning Objectives

After working through this chapter you should be able to perform the tasks listed below. Each row maps to one level of Bloom's cognitive taxonomy.

Bloom level	Action verb	Specific competency
Remember	Define · Identify	Random variable, PMF, PDF, CDF, expected value, variance.
Understand	Explain · Distinguish	Difference between discrete and continuous random variables; relation between PDF and CDF.
Apply	Compute · Use	Calculate $E(X)$, $\text{Var}(X)$, $P(a \leq X \leq b)$ for both types; apply linear transformations.
Analyse	Decompose · Compare	Decide whether a given problem calls for a discrete or continuous model; compare the dispersion of two random variables.
Evaluate	Justify · Interpret	Interpret $E(X)$, $\text{Var}(X)$, and standardised scores in economic and managerial contexts.

These competencies map to outcomes **C2–C5** of the official course specification (fiche pédagogique) and build directly on Chapter 1.

How to study this chapter effectively

This chapter is more computational than Chapter 1: you will be integrating, differentiating, summing, and interpreting numerical summaries. A suggested study path follows.

- **Step 1.** Make sure you can perform the prerequisites in §2.3, especially basic integration (for the continuous part) and summation with series (for the discrete part).
- **Step 2.** Read §2.5 carefully. The definition of a random variable is abstract, but everything else in the chapter depends on it.
- **Step 3.** Study §2.6 and §2.7 in parallel, noting the close correspondence between discrete and continuous formulas (sum $\leftrightarrow$ integral, PMF $\leftrightarrow$ PDF).
- **Step 4.** Work through every example with pen and paper. Do not read the solutions passively.
- **Step 5.** Focus §2.9 and §2.10: expectation and variance are the two most important summary measures and they will appear in every subsequent chapter.
- **Step 6.** After the worked examples in §2.12, attempt the practice exercises in §2.16 without looking at the examples.

A realistic total workload for this chapter is approximately fifteen hours: three hours for the reading, five hours for the worked examples, and seven hours for the exercises.

2.3 Prerequisites Check

Before starting this chapter, you should be comfortable with the concepts listed below. If any item seems unfamiliar, review it before continuing.

- **From Chapter 1:** sample space, event, probability function, the three axioms of Kolmogorov, conditional probability, and the total probability theorem.
- **From calculus:** the meaning of a derivative and a definite integral; integration of simple functions (polynomials, exponentials); limits at infinity; the fundamental theorem of calculus.
- **From algebra:** summation notation Σ , geometric series, factorials, manipulation of fractions and powers.
- **From descriptive statistics:** the arithmetic mean, the variance and standard deviation of a data set, the idea of a frequency distribution.

The short diagnostic below lets you test yourself. If you can answer all six items without hesitation, you have the required background.

i Diagnostic 2.1 — Self-check before starting

- Compute $\int$ from 0 to 2 of $3x^2 dx$.
- Compute the derivative of $F(x) = x^3/3 + 2x$ at $x = 1$.
- Compute the sum Σ from $k=0$ to 3 of $(1/2)^k$.
- State the total probability theorem and give a one-sentence interpretation.
- For the data set $\{2, 4, 4, 6, 8, 10\}$, compute the arithmetic mean and the variance.
- Explain why $0 \leq P(A) \leq 1$ is a direct consequence of Kolmogorov's axioms.

Answers are given at the end of §2.16.

2.4 Notation and Symbols

The table below lists the symbols introduced in this chapter. The booklet-wide notation list is in the front matter.

Symbol	Reading	Meaning
X, Y, Z	capital letters	Random variables.
x, y, z	lowercase letters	Particular values (realisations) that a random variable can take.
X(E)	"range of X"	Set of all possible values of X.
P(X = x)	"P of X equals x"	Probability that X takes the value x (discrete).
p(x) or p_X(x)	"p of x"	Probability mass function (PMF) — discrete case.
f(x) or f_X(x)	"f of x"	Probability density function (PDF) — continuous case.
F(x) or F_X(x)	"big F of x"	Cumulative distribution function (CDF): $F(x) = P(X \leq x)$.
E(X), μ_X	"expected value of X"	Mean of X (also called mathematical expectation).
Var(X), σ^2	"variance of X"	Dispersion of X around its mean.
σ or σ_X	"sigma of X"	Standard deviation of X: $\sigma = \sqrt{\text{Var}(X)}$.
Z	"Z-score"	Standardised variable $Z = (X - \mu)/\sigma$.
$\mathbb{R}$	"R"	Set of real numbers.

2.5 Definition of a Random Variable

2.5.1 Formal definition

Let E be the sample space of a random experiment. A **random variable** X is a function that assigns a real number to each outcome of E :

$$X : E \rightarrow \mathbb{R}$$

$$\omega \rightarrow X(\omega)$$

The function rule is fixed in advance; what is uncertain is which outcome ω will actually occur, hence which numerical value $X(\omega)$ will be observed. Once the experiment is performed, the random variable takes a specific realisation, conventionally denoted by the lowercase letter x .

The range of X , written $X(E)$, is the set of all values the function can produce. Depending on whether $X(E)$ is countable (finite or countably infinite) or uncountable (typically an interval), the random variable is classified as discrete or continuous.

Example 2.1 — Two dice — the sum as a random variable

A fair six-sided die is rolled twice. The sample space has 36 equally likely outcomes: $E = \{(i, j) : 1 \leq i, j \leq 6\}$.

Define $X = \text{sum of the two dice}$. Then X is a function from E to $\mathbb{R}$:

$$X((i; j)) = i + j$$

The range is $X(E) = \{2, 3, 4, \dots, 12\}$, a finite set of 11 values. X is therefore a discrete random variable.

Example 2.2 — Waiting time as a random variable

A customer arrives at a post office. Let T denote the time (in minutes) until they are served.

The sample space E is the set of all possible arrival scenarios (queue lengths, service times). The random variable T is the function that extracts the waiting time from each scenario:

$$T : E \rightarrow [0, +\infty)$$

The range $T(E)$ is the interval $[0, +\infty)$, an uncountable set. T is therefore a continuous random variable.

2.5.2 Discrete vs continuous — the dividing line

A random variable X is **discrete** when its range $X(E)$ is finite or countably infinite. Typical examples: counts (number of defective items in a sample, number of customers arriving in an hour), binary indicators (loan repaid or not, purchase or no purchase), ordinal scores (credit ratings 1 to 5).

A random variable X is **continuous** when its range is an interval (or a union of intervals) of the real line. Typical examples: durations (time until failure, waiting times), amounts (revenue, asset returns), measurements (temperature, distance, weight).

This distinction matters because the mathematical tools used in each case differ:

- **Discrete:** probabilities are assigned to individual values; summations replace integrals.
- **Continuous:** probabilities are assigned to intervals via a density function; integrals play the role of sums.

⚠ Important — probabilities of individual values in the continuous case

- For a continuous random variable X , the probability of any specific value x is zero: $P(X = x) = 0$.
- Only probabilities of intervals, $P(a \leq X \leq b)$, are meaningful.
- This counter-intuitive property follows from the fact that the density function $f(x)$ represents probability per unit length, not probability itself.

2.5.3 The indicator variable (a special case)

Perhaps the simplest random variable is the indicator of an event A , written 1_A or I_A :

$$1_A = \text{cases}$$

defined as 1 if A occurs, and 0 if A does not occur. The indicator turns every event into a binary random variable and is the conceptual bridge between the event-based language of Chapter 1 and the variable-based language of this chapter.

A key identity connects indicators and probabilities:

$$E(1_A) = P(A)$$

That is, the expected value of the indicator of A equals the probability of A . This simple fact underlies many probabilistic arguments, including the elementary proof that the arithmetic mean of n observed indicators converges to $P(A)$: the law of large numbers in its most intuitive form.

Example 2.3 — Indicator variable for a loan default

A bank issues a loan to one borrower. Let A = "the loan defaults within one year" and $P(A) = 0.08$. Define $X = 1_A$. Then X is a discrete random variable taking two values, with:

$$P(X = 1) = 0.08, \quad P(X = 0) = 0.92$$

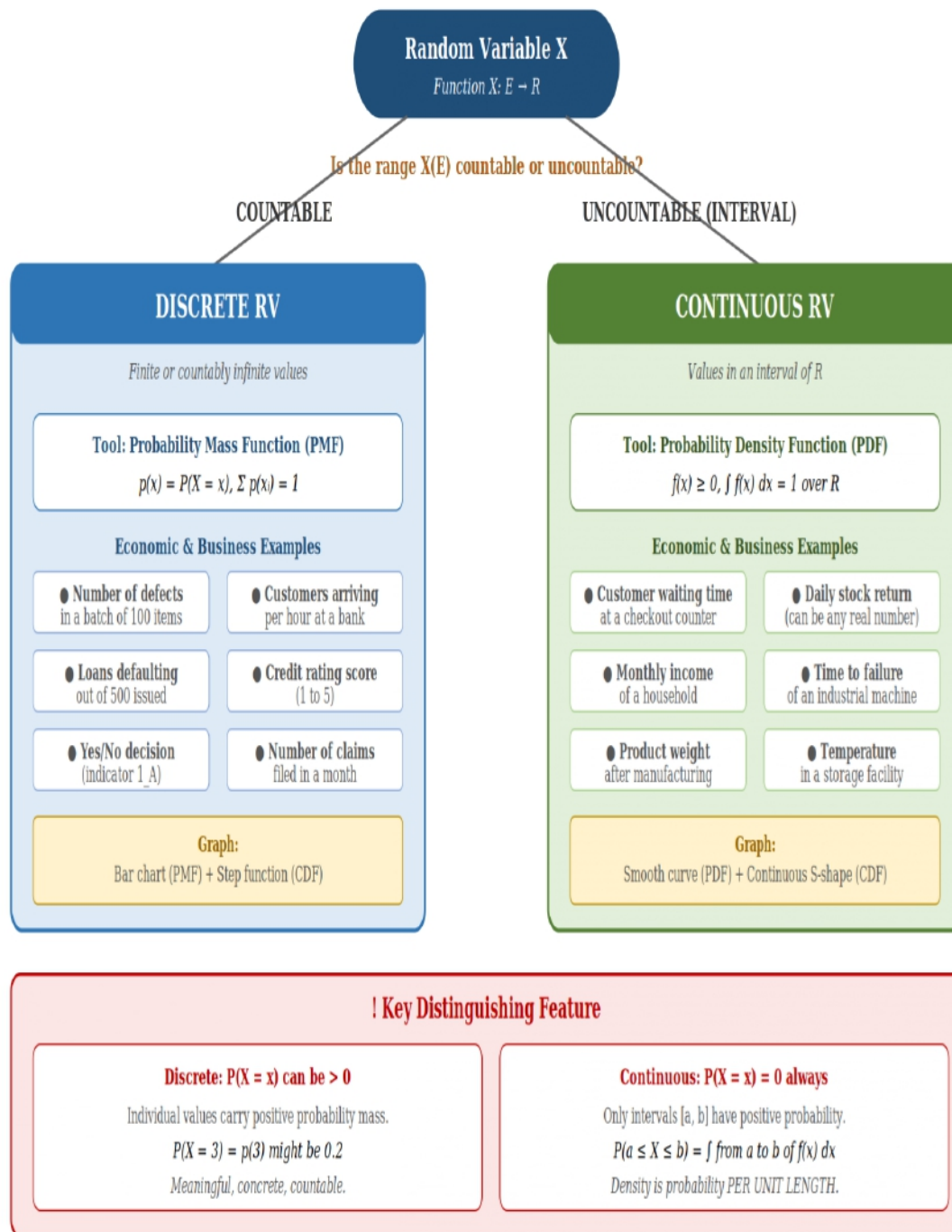
The expected value is $E(X) = 1 \times 0.08 + 0 \times 0.92 = 0.08$, which equals the probability of default, the general identity confirmed in a particular case.

Visual Summary — Mind Map of §2.5

The diagram below classifies random variables by the structure of their range and provides illustrative examples from economic life.

Figure 2.1: Random variable typology: discrete vs continuous

Mind Map 2.1 – Random Variable Typology



Source: Elaborated by the Author

2.6 Discrete Random Variables

2.6.1 The probability mass function (PMF)

Let X be a discrete random variable with range $X(E) = \{x_1, x_2, \dots\}$ (finite or countably infinite). The **probability mass function** of X , written $p(x)$ or $p_X(x)$, assigns to each possible value the probability that X takes that value:

$$p(x) = P(X = x)$$

The PMF is defined on the range $X(E)$ and satisfies two mandatory conditions:

- **Non-negativity:** $p(x_i) \geq 0$ for every x_i in $X(E)$.
- **Normalisation:** the values of p sum to one over the entire range:

$$\sum_i p(x_i) = 1$$

Conversely, any function that satisfies these two conditions can be interpreted as the PMF of some discrete random variable; the two conditions are exactly what is needed to ensure that probabilities are well defined.

Example 2.4 — PMF of the sum of two dice

Recall Example 2.1. The PMF of the sum X is obtained by counting, for each value x from 2 to 12, the number of pairs (i, j) with $i + j = x$, and dividing by 36.

x	2	3	4	5	6	7	8	9	10	11	12
$p(x)$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

The probabilities are non-negative and sum to $36/36 = 1$, confirming that we have a valid PMF. Note the triangular shape: the most likely value is $x = 7$ with probability $6/36 \approx 0.167$. This reflects the fact that more pairs (i, j) produce the sum 7 than any other value.

Example 2.5 — Constructing a PMF from a problem statement

A quality inspector tests 3 components drawn at random from a batch. Let X denote the number of defective components among the 3 tested. If the batch is very large and the defect rate is 10%, the three tests can be treated as independent.

The PMF of X is obtained by counting the ways $X = k$ defects can occur:

$$P(X = 0) = (0.90)^3 = 0.729$$

$$P(X = 1) = 3 * 0.10 * (0.90)^2 = 0.243$$

$$P(X = 2) = 3 * (0.10)^2 * 0.90 = 0.027$$

$$P(X = 3) = (0.10)^3 = 0.001$$

The four probabilities sum to $0.729 + 0.243 + 0.027 + 0.001 = 1.000$, so this is a valid PMF. We will see in Chapter 3 that this particular PMF belongs to the binomial family with parameters $n = 3$ and $p = 0.10$.

2.6.2 The cumulative distribution function (CDF)

The **cumulative distribution function** (CDF) of a random variable X , written $F(x)$ or $F_X(x)$, gives the probability that X takes a value less than or equal to x :

$$F(x) = P(X \leq x)$$

For a discrete random variable, $F(x)$ is obtained by summing the PMF over all values not exceeding x :

$$F(x) = \sum_{x_i \leq x} p(x_i)$$

The CDF is a non-decreasing step function that jumps by $p(x_i)$ at each point x_i of the range and is constant between jumps. Its key properties are:

- F is non-decreasing: $x \leq y$ implies $F(x) \leq F(y)$.
- $F(-\infty) = 0$ and $F(+\infty) = 1$.
- F is right-continuous: the value of F at a jump point equals the value after the jump.
- For any $a < b$: $P(a < X \leq b) = F(b) - F(a)$.

Example 2.6 — CDF of a discrete variable

Consider the random variable X of Example 2.5 with PMF above. The CDF is constructed by accumulating the probabilities:

x	$P(X = x)$	$F(x) = P(X \leq x)$
0	0.729	0.729
1	0.243	0.972
2	0.027	0.999
3	0.001	1.000

To compute, for example, the probability that at most 1 component is defective:

$$P(X \leq 1) = F(1) = 0.972$$

To compute the probability of at least 2 defectives:

$$P(X \geq 2) = 1 - F(1) = 1 - 0.972 = 0.028$$

The CDF is the most compact way to store the full probabilistic information about X : from F one can recover the PMF by taking differences.

2.6.3 Graphical representations

A discrete random variable is naturally represented by two distinct graphs:

- **PMF plot:** vertical bars (or points with vertical lines) of height $p(x_i)$ placed at each x_i . The heights sum to 1.
- **CDF plot:** a non-decreasing step function with jumps at each x_i of height $p(x_i)$. The function starts at 0 on the left and reaches 1 on the right.

Because the CDF is a step function, for a discrete random variable one must be careful with strict vs non-strict inequalities: $P(X < x)$ and $P(X \leq x)$ differ by exactly $p(x)$ if x is in the range.

Discrete random variable X (daily demand at a bookshop)

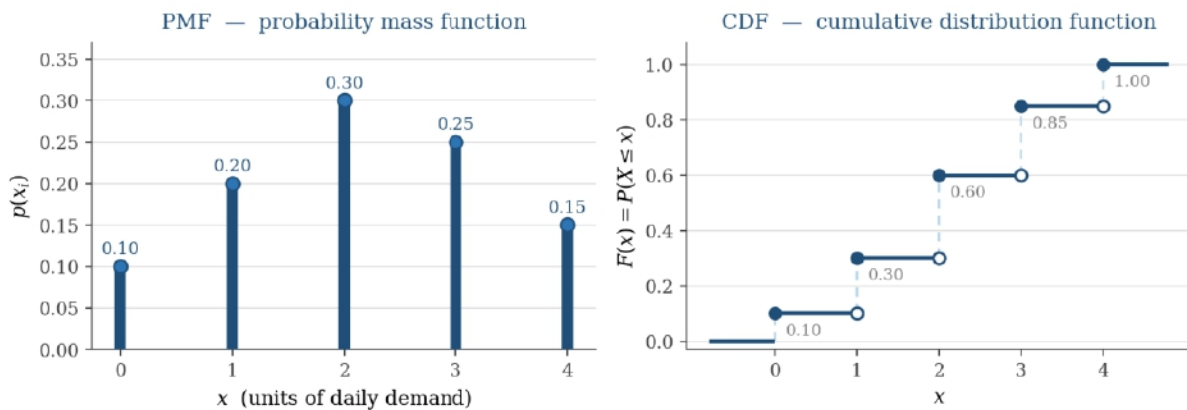


Figure 2.4 — PMF (left) and CDF (right) of a discrete random variable. The PMF shows the probability at each value x_i ; the CDF accumulates them into a non-decreasing, right-continuous step function (solid dot = endpoint included, open dot = excluded). The jumps equal the PMF heights.

2.6.4 Mode and median of a discrete distribution

The **mode** of a discrete random variable is the value x_i that maximises $p(x_i)$, the most likely individual value. A distribution may have a single mode (unimodal), two modes (bimodal), or more. Unlike the mean, the mode is always one of the possible values of X .

For the PMF of Example 2.4 (sum of two dice), the mode is $x = 7$ since $p(7) = 6/36$ is larger than any other probability. For the PMF of Example 2.5 (defective components), the mode is $x = 0$ since $p(0) = 0.729$ is the highest.

The **median** is defined, for a discrete random variable, as any value m satisfying:

$$P(X \leq m) \geq 0.5 \quad (\text{and}) \quad P(X \geq m) \geq 0.5$$

The median may not be unique for discrete distributions (e.g. if F jumps across 0.5), in which case any value in the ambiguous interval is acceptable. For the two-dice PMF, the median is 7 (the unique value with $F(6) < 0.5 < F(7)$). The median resists extreme values in a way the mean does not, which makes it the preferred measure of central tendency in distributions with long tails.

Example 2.6b — *PMF from empirical frequencies*

A call centre records the number of calls received per minute during a busy hour. Over 100 observed minutes, the frequencies are:

X (calls/min)	0	1	2	3	4	5
Frequency	12	25	30	20	10	3
Relative freq.	0.12	0.25	0.30	0.20	0.10	0.03

The relative frequencies, which sum to 1.00, give an empirical PMF for the number of calls per minute. From it, a supervisor can compute the expected call volume $E(X) = 1.95$, estimate staffing needs, and assess the probability that more than 3 calls arrive in a given minute ($= 0.13$). With more data, the empirical PMF becomes an increasingly reliable estimator of the true distribution, a consequence of the law of large numbers.

2.7 Continuous Random Variables

2.7.1 The probability density function (PDF)

Let X be a continuous random variable taking values in an interval (or the whole real line). The **probability density function** (PDF) of X , written $f(x)$ or $f_X(x)$, is a non-negative function such that, for any interval $[a, b]$:

$$P(a \leq X \leq b) = \int_a^b f(x) dx$$

That is, the probability that X falls in $[a, b]$ equals the area under the density curve between a and b . The density itself is not a probability; it has units of "probability per unit of x ". The area it encloses, however, is a probability.

A valid PDF must satisfy two conditions analogous to those of the discrete PMF:

- **Non-negativity:** $f(x) \geq 0$ for every x in $\mathbb{R}$.
- **Normalisation:** the total area under f over the entire real line equals 1:

$$\int_{-\infty}^{+\infty} f(x) dx = 1$$

As in the discrete case, any function satisfying these two conditions defines a valid continuous distribution.

⚠ Important — density is not probability

- $f(x)$ can exceed 1 at some points without violating any rule, because f is a density (per unit length).
- The probability of a single value is always zero: $P(X = x) = 0$ for any x .
- Hence $P(a \leq X \leq b) = P(a < X < b)$ for continuous variables. Strict and non-strict inequalities are interchangeable.

Example 2.7 — A simple density on $[0, 2]$

Consider the function $f(x) = x/2$ on the interval $[0, 2]$, and 0 elsewhere. Is this a valid PDF?

Non-negativity is immediate since $x/2 \geq 0$ on $[0, 2]$. For normalisation:

$$\int_0^2 \frac{x}{2} dx = \left(\frac{x^2}{4} \right)_0^2$$

Evaluating: $[x^2/4]$ from 0 to 2 = $4/4 - 0 = 1$. The function is therefore a valid PDF.

The probability that X lies between 1 and 1.5 is:

$$P(1 \leq X \leq 1.5) = \int_1^{1.5} \frac{x}{2} dx = \frac{1.5^2 - 1^2}{4} = \frac{1.25}{4} = 0.3125$$

2.7.2 The cumulative distribution function (CDF) — continuous case

The CDF is defined exactly as in the discrete case, $F(x) = P(X \leq x)$, but its computation now involves an integral rather than a sum:

$$F(x) = \int_{-\infty}^x f(t) dt$$

For a continuous random variable, F is a continuous non-decreasing function (no jumps), ranging from 0 at $-\infty$ to 1 at $+\infty$.

The key properties of the continuous CDF are:

- F is continuous and non-decreasing.
- $F(-\infty) = 0$, $F(+\infty) = 1$.
- $P(a \leq X \leq b) = F(b) - F(a)$.
- $P(X > a) = 1 - F(a)$.

Example 2.8 — CDF of the density in Example 2.7

With $f(x) = x/2$ on $[0, 2]$:

$$F(x) = \int_0^x \frac{t}{2} dt = \frac{x^2}{4}, \quad (\text{for }) 0 \leq x \leq 2$$

For $x < 0$, $F(x) = 0$; for $x > 2$, $F(x) = 1$. In one compact statement:

$$F(x) = 0 \text{ (if) } x < 0; \quad \frac{x^2}{4} \text{ (if) } 0 \leq x \leq 2; \quad 1 \text{ (if) } x > 2$$

At $x = 2$, $F(2) = 4/4 = 1$, which correctly reflects the fact that X is certain not to exceed 2. The probability $P(1 \leq X \leq 1.5)$ computed in Example 2.7 can be recovered as $F(1.5) - F(1) = 2.25/4 - 1/4 = 1.25/4 = 0.3125$, confirming the result.

2.7.3 Relation between PDF and CDF

The CDF and PDF of a continuous random variable are linked by the fundamental theorem of calculus:

$$f(x) = \frac{dF(x)}{dx}$$

That is, the density is the derivative of the cumulative function. Equivalently, the CDF is obtained from the density by integration, as we have already seen. This pair of inverse operations is the continuous analogue of the discrete relationship $F(x) = \sum p(x_i)$ over $x_i \leq x$, and the ability to pass between PDF and CDF at will is essential for practical calculations.

Continuous random variable X (PDF and its CDF)

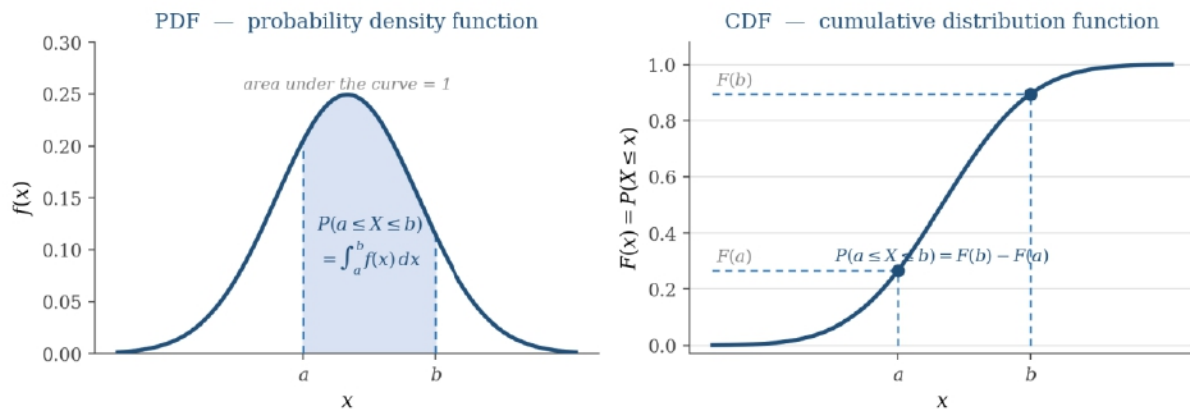


Figure 2.5 — PDF (left) and CDF (right) of a continuous random variable. Probability is the area under the density $f(x)$; the shaded region $P(a \leq X \leq b)$ equals the rise $F(b) - F(a)$ read off the CDF. The total area under $f(x)$ is 1, and F increases smoothly from 0 to 1.

i Which representation to use?

PDF is preferred when

- you need to compute probabilities of intervals via integration or find the mode (maximum of f).

CDF is preferred when

- you need $P(X \leq a)$ directly, compute medians or quantiles, or compare two distributions graphically.

2.7.4 Quantiles, percentiles, and the median

Although the mean and the variance are the two most-used summaries of a distribution, other measures of location are frequently reported. The most important are the quantiles.

For a continuous random variable X with CDF F , the **quantile of order q** (for q between 0 and 1), denoted x_q , is defined as the value such that:

$$F(x_q) = P(X \leq x_q) = q$$

That is, x_q is the threshold below which a proportion q of the probability mass lies. Three quantiles have special names:

- **Median ($x_{0.50}$):** the central value that splits the distribution into two halves of equal probability.

- **First quartile ($Q_1 = x_{0.25}$):** 25% of the distribution lies below Q_1 .
- **Third quartile ($Q_3 = x_{0.75}$):** 75% of the distribution lies below Q_3 .

Percentiles are quantiles expressed as percentages: the 90th percentile $x_{0.90}$ is the value below which 90% of realisations fall. Quantiles are particularly useful when the distribution is skewed. In that case the mean can be a misleading summary, and the median is a more reliable indicator of the "typical" value.

Example 2.8b — *Finding the median of a continuous distribution*

For the density $f(x) = x/2$ on $[0, 2]$, the CDF is $F(x) = x^2/4$. The median $x_{0.50}$ is found by solving $F(x) = 0.5$:

$$\frac{x^2}{4} = 0.5 \quad \text{rArr} \quad x^2 = 2 \quad \text{rArr} \quad x \approx 1.414$$

So the median of this distribution is about 1.414, larger than the mean of $4/3 \approx 1.333$. The gap comes from the rightward skew of the density. Half of the probability mass lies above 1.414 and half below.

The **interquartile range** $IQR = Q_3 - Q_1$ is an alternative dispersion measure: it is less sensitive to outliers than the variance and is the natural measure to report when the distribution is skewed. In this course we keep σ as the primary dispersion measure, but you should be aware of the IQR because it is standard in applied statistics.

2.8 Comparison — Discrete vs Continuous

The two settings share a common structure but use different mathematical tools. The table below sets them side by side for quick reference and doubles as a study aid.

Aspect	Discrete RV	Continuous RV
Range	Countable set (finite or infinite)	Interval of $\mathbb{R}$ (uncountable)
Probability on values	$P(X = x_i) \geq 0$	$P(X = x) = 0$
Distribution function	PMF: $p(x) = P(X = x)$	PDF: $f(x) \geq 0$
Normalisation	$\sum p(x_i) = 1$	$\int f(x) dx = 1$ over all $\mathbb{R}$
Probability of interval	$P(a \leq X \leq b) = \sum p(x_i)$	$P(a \leq X \leq b) = \int \text{from } a \text{ to } b \text{ of } f(x) dx$
CDF $F(x) = P(X \leq x)$	Step function (jumps at x_i)	Continuous non-decreasing function
Mean $E(X)$	$\sum x_i p(x_i)$	$\int x f(x) dx$
Variance $\text{Var}(X)$	$\sum (x_i - \mu)^2 p(x_i)$	$\int (x - \mu)^2 f(x) dx$
Typical examples	Count of events, counts, scores	Duration, revenue, measurement

The correspondence sum $\leftrightarrow$ integral, PMF $\leftrightarrow$ PDF, is exact: every formula derived for one case has a twin in the other. Once you recognise this symmetry, the apparent duplication of definitions becomes a unifying feature of the theory rather than a source of confusion.

★ Exam tip — how to tell them apart

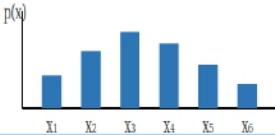
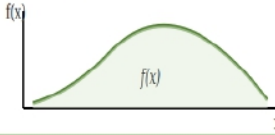
- Discrete: the problem talks about counts, numbers, categories, or integer-valued quantities.
- Continuous: the problem talks about durations, amounts, temperatures, or any quantity measured on a scale.
- Mixed cases exist (e.g. insurance claim amounts that are either zero with positive probability or any positive real), but in this course we treat the two pure cases.

Visual Summary — Mind Map of §2.8

The diagram below lays out the full comparison of discrete and continuous random variables across eight key dimensions: domain, probability function, conditions, graph, expected value, variance, CDF, and typical examples.

Figure 2.2: Discrete vs continuous: a side-by-side visual comparison

Mind Map 2.2 – Discrete vs Continuous Random Variables

DISCRETE	CONTINUOUS
Domain / Range Countable set: $X(E) = \{x_1, x_2, x_3, \dots\}$ <i>(finite or countably infinite)</i>	Domain / Range Interval of $\mathbb{R}$: $[a, b]$ or $(-\infty, +\infty)$ <i>(uncountably infinite)</i>
Probability function (PMF) $p(x) = P(X = x)$ Assigns probability to each individual value. Example: $p(2) = 0.15$, $p(5) = 0.40$, ...	Probability density (PDF) $f(x) \geq 0$, NOT a probability itself Density "per unit length" — probability is obtained by integration over an interval.
Two conditions (1) $p(x_i) \geq 0$ for all i (2) $\sum p(x_i) = 1$	Two conditions (1) $f(x) \geq 0$ for all $x \in \mathbb{R}$ (2) $\int f(x) dx = 1$ over $\mathbb{R}$
Graph: bars (sticks) 	Graph: smooth curve 
Expected value $E(X)$ $E(X) = \sum x_i \cdot p(x_i)$ Weighted average over all possible values	Expected value $E(X)$ $E(X) = \int x \cdot f(x) dx$ over $\mathbb{R}$ Integral replaces the sum
Variance $\text{Var}(X)$ $\text{Var}(X) = \sum (x_i - \mu)^2 \cdot p(x_i)$ Mean of squared deviations	Variance $\text{Var}(X)$ $\text{Var}(X) = \int (x - \mu)^2 \cdot f(x) dx$ Integral of squared deviations times f
CDF $F(x) = P(X \leq x)$ Step function: $F(x) = \sum p(x_i)$ for $x_i \leq x$ <i>(jumps at each x_i; flat in between)</i>	CDF $F(x) = P(X \leq x)$ Continuous S-curve: $F(x) = \int$ from $-\infty$ to x of $f(t) dt$ <i>(smooth; derivative = f)</i>
Typical Economic Examples ● Number of defects in a batch ● Number of customers arriving per hour ● Credit rating (1 to 5) ● Yes/No purchase decision	Typical Economic Examples ● Customer waiting time ● Daily stock return ● Monthly household income ● Time until machine failure

Source: Elaborated by the Author

2.9 Expectation of a Random Variable

The expectation, also called the mean or mathematical expectation, is the single most important numerical summary of a random variable. It captures what we would observe, on average, if the random experiment were repeated many times under identical conditions. For a manager, it is often the first quantity computed, and the one that decides whether an investment, a policy, or a strategy is worth undertaking.

2.9.1 Definition — discrete case

Let X be a discrete random variable with range $\{x_1, x_2, \dots\}$ and PMF $p(x_i)$. The **expected value** of X , written $E(X)$ or μ_X , is the weighted average of its possible values, where each value is weighted by its probability:

$$E(X) = \mu = \sum_i x_i \cdot p(x_i)$$

The expectation is a number in the same units as X . It need not be one of the possible values: for a die roll the expected value is 3.5, even though the die never shows 3.5.

Example 2.9 — Expected value — simple discrete RV

A fair six-sided die is rolled once. Let X be the face shown.

$$E(X) = 1 * \frac{1}{6} + 2 * \frac{1}{6} + 3 * \frac{1}{6} + 4 * \frac{1}{6} + 5 * \frac{1}{6} + 6 * \frac{1}{6}$$

$$E(X) = \frac{1 + 2 + 3 + 4 + 5 + 6}{6} = \frac{21}{6} = 3.5$$

Interpretation: if we rolled the die a very large number of times and averaged the results, the average would approach 3.5. No single roll gives 3.5; the expectation is a theoretical centre of gravity of the distribution.

Example 2.10 — Expected profit from a project

A company is considering a new project. The outcome depends on market conditions with the following probabilities and payoffs (in thousands of dinars):

Market condition	Probability	Payoff X (in 1 000 DZD)
Favourable	0.30	+500
Neutral	0.50	+100
Unfavourable	0.20	−200

$$E(X) = 0.30 * 500 + 0.50 * 100 + 0.20 * (-200)$$

$$E(X) = 150 + 50 - 40 = 160$$

The expected payoff is +160 (i.e. 160 000 DZD). This figure is what the company can expect, on average, if the project is repeated many times under the same probability structure. A rational manager who maximises expected profit would undertake the project, since $E(X) > 0$. This is the foundational idea of expected-utility decision-making.

2.9.2 Definition — continuous case

For a continuous random variable with density $f(x)$, the expectation is defined by an integral, the continuous analogue of the sum:

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

The integrand $x \cdot f(x)$ weights each value x by the density at that point. The operation is the same as in the discrete case: compute a weighted average over the range of X , using f instead of p and an integral instead of a sum.

Example 2.11 — Expected value of a simple density

Return to the density $f(x) = x/2$ on $[0, 2]$ from Example 2.7. The expected value is:

$$E(X) = \int_0^2 x \cdot \frac{x}{2} \, dx = \int_0^2 \frac{x^2}{2} \, dx$$
$$E(X) = \left(\frac{x^3}{6} \right)^2 = \frac{8}{6} = \frac{4}{3} \approx 1.333$$

On average, realisations of X are around 1.33, closer to the right end of $[0, 2]$ than to the middle, because the density $x/2$ is larger there. This is consistent with the shape of the density: it assigns more weight to larger values of x .

2.9.3 Properties of expectation

The expectation operator $E(\cdot)$ satisfies a number of important properties that make it easy to work with in computations. The three most used are:

- **Linearity:** $E(aX + b) = a \cdot E(X) + b$, for any constants a and b .
- **Sum:** $E(X + Y) = E(X) + E(Y)$, even if X and Y are not independent.
- **Constant:** $E(c) = c$, for any constant c .

Linearity is the workhorse: it lets us handle affine transformations (scaling, shifting) without re-computing the entire distribution. For independent random variables, a further property holds:

$$E(X \cdot Y) = E(X) \cdot E(Y) \quad , \quad (\text{if } X \text{ (and) } Y \text{ (are independent)})$$

Without independence, $E(XY)$ is generally different from $E(X) \cdot E(Y)$, and the difference is called the covariance, a concept revisited in later courses.

2.9.4 Expectation of a function of a random variable

Often we need the expectation of a transformed variable $Y = g(X)$: for instance, the expected squared return $E(X^2)$, or the expected utility $E(u(X))$ for a given utility function u . The rule, known as the law of the unconscious statistician (LOTUS), is:

Discrete:

$$E(g(X)) = \sum_i g(x_i) \cdot p(x_i)$$

Continuous:

$$E(g(X)) = \int_{-\infty}^{+\infty} g(x) * f(x) * dx$$

The key point is that we do not need to know the distribution of $g(X)$; we can compute its expectation directly from the distribution of X . For $g(x) = x^2$, this gives the expected squared value, which appears in the variance formula discussed next.

Example 2.12 — *Computing $E(X^2)$*

For the density $f(x) = x/2$ on $[0, 2]$:

$$E(X^2) = \int_0^2 x^2 * \frac{x}{2} * dx = \int_0^2 \frac{x^3}{2} * dx$$

$$E(X^2) = \left(\frac{x^4}{8}\right)^2 = \frac{16}{8} = 2$$

We will use this result in §2.10 to compute $\text{Var}(X)$.

2.10 Variance and Standard Deviation

The expectation tells us where the distribution is centred; the variance tells us how spread out it is around the centre. Two random variables can have the same mean but very different variances, and the variance often matters more for risk management than the mean itself. A portfolio with expected return 10% and variance 1 is very different from one with expected return 10% and variance 100.

2.10.1 Definition — discrete and continuous cases

The **variance** of X , written $\text{Var}(X)$ or σ^2 , is the expected squared deviation from the mean:

$$\text{Var}(X) = \sigma^2 = E((X - \mu)^2)$$

Discrete form:

$$\text{Var}(X) = \sum_i (x_i - \mu)^2 \cdot p(x_i)$$

Continuous form:

$$\text{Var}(X) = \int_{-\infty}^{+\infty} (x - \mu)^2 \cdot f(x) \cdot dx$$

The variance is measured in squared units of X . To obtain a dispersion measure in the same units as X , we take the square root:

The **standard deviation** is $\sigma = \sqrt{\text{Var}(X)}$. It has the same units as X and is typically easier to interpret than the variance.

2.10.2 Alternative formula (computational shortcut)

Expanding $(X - \mu)^2 = X^2 - 2\mu X + \mu^2$ and using the linearity of expectation gives a formula that is usually easier to compute:

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

This identity, "variance equals mean of the square minus square of the mean," is the single most useful formula in moment calculations.

Example 2.13 — Variance via the alternative formula

Continuing with $f(x) = x/2$ on $[0, 2]$:

From Example 2.11, $E(X) = 4/3$. From Example 2.12, $E(X^2) = 2$. Therefore:

$$\text{Var}(X) = 2 - \left(\frac{4}{3}\right)^2 = 2 - \frac{16}{9} = \frac{2}{9} \approx 0.222$$

$$\sigma = \sqrt{\frac{2}{9}} \approx 0.471$$

The standard deviation of X is about 0.47, roughly one-third of its mean, indicating moderate dispersion.

Example 2.14 — Variance of the project payoff

For the project of Example 2.10 with $E(X) = 160$:

$$E(X^2) = 0.30 * 500^2 + 0.50 * 100^2 + 0.20 * (-200)^2$$

$$E(X^2) = 0.30 * 250000 + 0.50 * 10000 + 0.20 * 40000 = 88000$$

$$Var(X) = 88000 - 160^2 = 88000 - 25600 = 62400$$

$$\sigma = \sqrt{62400} \approx 249.8$$

The expected payoff is 160 (thousand DZD), but the standard deviation is about 250, larger than the mean itself. This points to a high-risk project: although the expected outcome is positive, the variability is substantial, and a risk-averse manager might prefer a lower-mean, lower-variance alternative.

2.10.3 Properties of variance

The variance operator has several useful properties, some of which differ from the corresponding expectation properties:

- **Non-negativity:** $Var(X) \geq 0$, with equality if and only if X is a constant.
- **Constant:** $Var(c) = 0$, for any constant c .
- **Scaling:** $Var(aX) = a^2 \cdot Var(X)$. Note the squared factor.
- **Shift:** $Var(X + b) = Var(X)$. Adding a constant does not change the dispersion.
- **Affine:** $Var(aX + b) = a^2 \cdot Var(X)$.

For independent random variables X and Y :

$$Var(X + Y) = Var(X) + Var(Y) \quad , \quad (if \) X (and \) Y (are independent)$$

Without independence, variances do not add in general. A covariance term must be included, as you will see in later courses.

2.10.4 Standardised variables (Z-score)

Given a random variable X with mean μ and standard deviation $\sigma > 0$, the standardised variable is:

$$Z = \frac{X - \mu}{\sigma}$$

Z has mean 0 and variance 1, regardless of the original units of X:

$$E(Z) = 0 \quad , \quad Var(Z) = 1$$

Standardisation is useful for comparing distributions measured on different scales, for computing probabilities using standardised tables (as we will do with the normal distribution in Chapter 3), and for judging how unusual a particular realisation is: "2.5 standard deviations above the mean" is a more meaningful statement than "580 units above 400".

Example 2.15 — *Standardising a salary*

Monthly salaries in a firm have mean $\mu = 50\,000$ DZD and standard deviation $\sigma = 8\,000$ DZD. An employee earns 66 000 DZD. The standardised score is:

$$Z = \frac{66000 - 50000}{8000} = \frac{16000}{8000} = 2$$

The employee earns 2 standard deviations above the firm's average, a useful summary that is independent of the currency or the absolute pay scale.

2.10.5 Economic interpretation — variance as a measure of risk

In finance and managerial decision-making, variance is the standard quantitative measure of risk. Given two investments with the same expected return, the one with lower variance is said to be less risky, because its realisations are concentrated more tightly around the mean. A risk-averse investor would prefer the lower-variance option; a risk-neutral investor would be indifferent; a risk-seeking investor (rare in corporate finance) would prefer the higher-variance option.

Several refinements of the variance concept are used in practice:

- **Coefficient of variation:** $CV = \sigma / |\mu|$, a unit-free measure of relative dispersion. A CV of 0.2 means the standard deviation is 20% of the mean.
- **Sharpe-like ratio:** $E(X) / \sigma(X)$, measures expected reward per unit of risk. Used to rank investment alternatives.
- **Semi-variance:** focuses on downside deviations only ($X < \mu$), since investors typically worry more about losses than equally large gains. This is the basis of "downside risk" measures used in modern portfolio management.

Although the variance treats positive and negative deviations symmetrically, its mathematical simplicity makes it the default dispersion measure throughout this course. The more sophisticated risk measures build on it and will be encountered in later finance or econometrics modules.

i Intuition to remember

- Expectation tells you where the distribution is.

- Variance tells you how uncertain that location is.
- Two investments with identical means can have wildly different risk profiles, and a rational decision-maker must consider both.

2.10.6 Higher-order moments (optional reading)

Beyond the mean and the variance, two further summary measures are occasionally useful: skewness and kurtosis. They are included here as optional reading for students curious about the shape of distributions; the formal exam will test only $E(X)$ and $\text{Var}(X)$.

Skewness measures the asymmetry of the distribution. It is defined as:

$$Sk(X) = E\left(\left(\frac{X - \mu}{\sigma}\right)^3\right)$$

A distribution with $Sk > 0$ has a long right tail (positive skew, typical of income and revenue distributions). $Sk < 0$ indicates a long left tail (typical of age-at-death distributions). $Sk = 0$ corresponds to a symmetric distribution around the mean.

Kurtosis measures the heaviness of the tails:

$$Ku(X) = E\left(\left(\frac{X - \mu}{\sigma}\right)^4\right)$$

Distributions with high kurtosis produce extreme values more often than a normal distribution would predict. That property is central to modelling financial returns, where rare but large movements ("black swans") matter disproportionately. The normal distribution has $Ku = 3$ exactly, and distributions are often compared to it via the "excess kurtosis" $Ku - 3$.

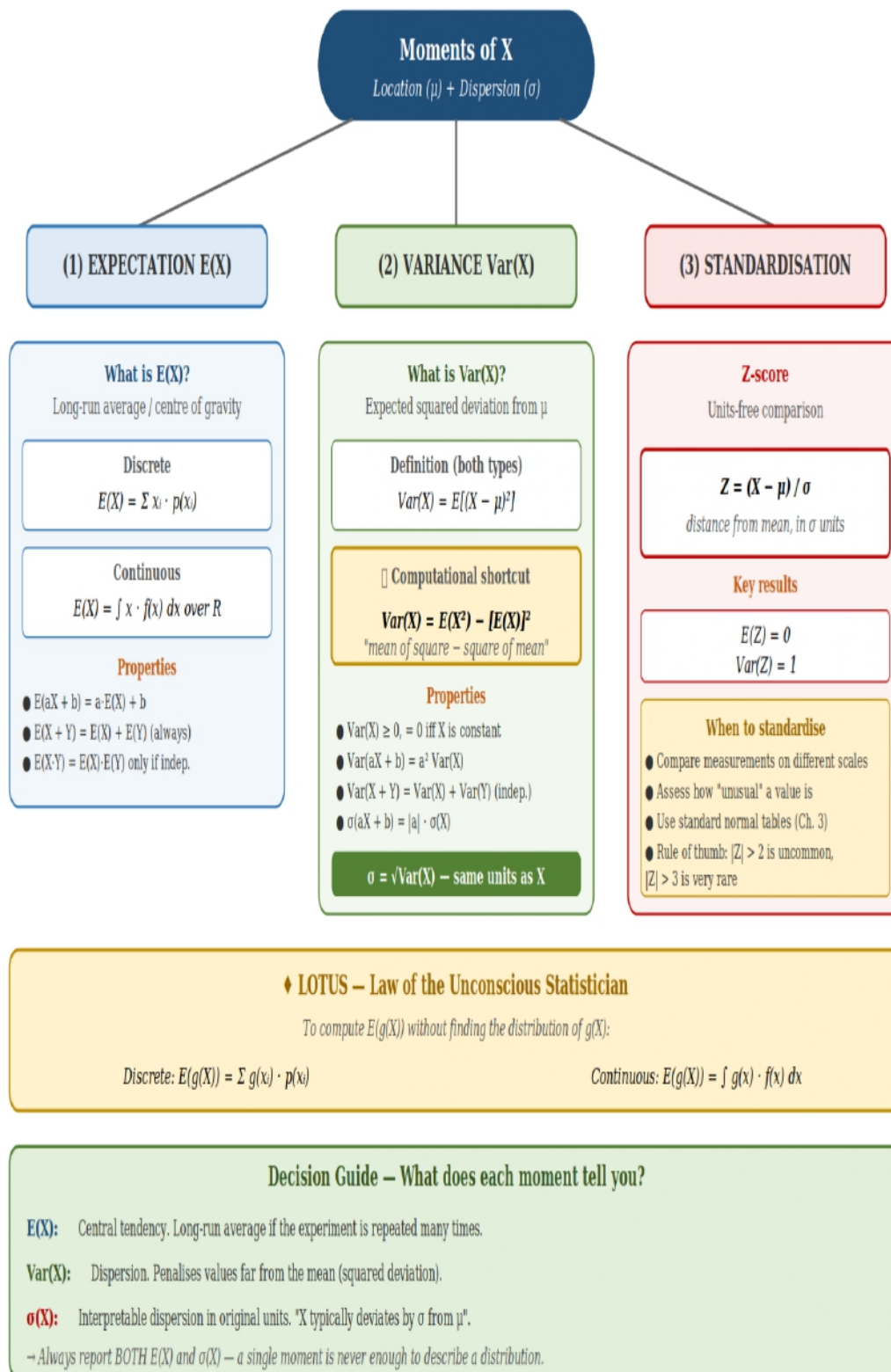
These higher moments are not required for the exam but are worth knowing: they frequently appear in applied econometrics and finance contexts.

Visual Summary — Mind Map of §2.9 – §2.10

The diagram below collects the definitions of expectation and variance, their shared properties, and the key standardisation formula into a single visual reference.

Figure 2.3: Expectation, variance, and standardisation: formulas and properties

Mind Map 2.3 – Expectation, Variance & Standardisation



Source: Elaborated by the Author

2.11 Transformations of Random Variables

Often a problem is not about X directly but about a function $Y = g(X)$. For example, X might be a raw measurement (in millimetres) and Y the same quantity in inches; or X might be a profit per unit and Y the total profit over n units. The most common case in this course is a linear (affine) transformation.

2.11.1 Linear transformation $Y = aX + b$

Let X be any random variable with mean μ and variance σ^2 , and let a, b be constants. Define $Y = aX + b$. Then:

$$E(Y) = E(aX + b) = a * \mu + b$$

$$Var(Y) = Var(aX + b) = a^2 * \sigma^2$$

$$\sigma_Y = |a| * \sigma_X$$

These three rules are direct consequences of the linearity of expectation and the scaling property of variance. They are worth memorising, as they allow many calculations to be done without re-doing integrals or sums.

2.11.2 Effects on the distribution

A linear transformation $Y = aX + b$ changes the location (mean) and the scale (standard deviation) of the distribution, but not its shape:

- The addition of b shifts the distribution by b units (left or right).
- The multiplication by $a > 0$ stretches or compresses the distribution by factor a .
- If $a < 0$, the distribution is reflected about the mean in addition to scaling.

Example 2.16 — Unit conversion

A machine produces steel rods with mean length $\mu = 100$ mm and standard deviation $\sigma = 2$ mm. In inches ($Y = X / 25.4$), the mean and standard deviation become:

$$E(Y) = \frac{100}{25.4} \approx 3.937, \quad \sigma_Y = \frac{2}{25.4} \approx 0.0787$$

The shape of the length distribution (whether bell-shaped, skewed, bimodal) is identical in the two units. Only the numerical scale changes.

Example 2.17 — Profit expressed per unit

A company's monthly total revenue (in thousand DZD) has mean $\mu = 2\,000$ and standard deviation $\sigma = 400$. Fixed costs are 500 per month, and 25% of revenue goes to variable costs. Monthly profit is:

$$P = 0.75 * X - 500$$

Using the linear-transformation rules:

$$E(P) = 0.75 * 2000 - 500 = 1000$$

$$\sigma_P = 0.75 \cdot 400 = 300$$

The expected monthly profit is 1 000 (thousand DZD), with a standard deviation of 300. The fixed cost shifts the mean but does not affect the standard deviation; the multiplicative factor 0.75 scales both.

2.11.3 Non-linear transformations — a cautionary note

The elegant rules $E(aX + b) = aE(X) + b$ and $\text{Var}(aX + b) = a^2 \text{Var}(X)$ apply only to linear (affine) transformations. For non-linear transformations $Y = g(X)$, such as $Y = X^2$, $Y = e^X$, or $Y = \log X$, the distribution of Y is generally quite different from that of X , and the moments must be computed from scratch using LOTUS.

A common misconception is to write $E(g(X)) = g(E(X))$, an identity that holds only when g is linear. For non-linear g it fails, sometimes dramatically. Jensen's inequality states that:

$$E(g(X)) \geq g(E(X)) \quad (\text{if } g \text{ is convex})$$

$$E(g(X)) \leq g(E(X)) \quad (\text{if } g \text{ is concave})$$

This observation has deep consequences in economics: if a utility function $u(w)$ is concave (as it is for risk-averse agents), then the expected utility of a random payoff is always less than the utility of the expected payoff. A risk-averse agent therefore prefers a sure payoff to a random one with the same mean; this is the formal statement of risk aversion.

Example 2.17b — Why $E(X^2) \neq [E(X)]^2$

Consider a random variable X with $P(X = 0) = P(X = 2) = 0.5$. Then:

$$E(X) = 1, \quad [E(X)]^2 = 1$$

$$E(X^2) = 0.5 * 0 + 0.5 * 4 = 2$$

The gap $E(X^2) - [E(X)]^2 = 2 - 1 = 1$ is precisely the variance of X . This gap, which must be non-negative by Jensen's inequality applied to the convex function $g(x) = x^2$, is the algebraic origin of the variance-shortcut formula.

2.12 Worked Examples in an Economic Context

Example 2.18 Inventory management — demand forecasting

A retailer studies daily demand X (in units) for a perishable product. Historical records give the following PMF:

x	0	1	2	3	4	5
$p(x)$	0.05	0.15	0.30	0.25	0.15	0.10

(a) Compute $E(X)$ and $\sigma(X)$. (b) What is the probability that daily demand is at least 3? (c) If the unit profit is 200 DZD, what is the expected daily profit?

Solution (a).

$$E(X) = 0 * 0.05 + 1 * 0.15 + 2 * 0.30 + 3 * 0.25 + 4 * 0.15 + 5 * 0.10 = 2.60$$

$$E(X^2) = 0 + 0.15 + 1.20 + 2.25 + 2.40 + 2.50 = 8.50$$

$$Var(X) = 8.50 - 2.60^2 = 8.50 - 6.76 = 1.74, \quad \sigma \approx 1.32$$

Solution (b).

$$P(X \geq 3) = 0.25 + 0.15 + 0.10 = 0.50$$

Solution (c). Expected daily profit = $200 \times E(X) = 200 \times 2.60 = 520$ DZD.

Example 2.19 Asset returns — risk measurement

Two stocks are under consideration with the following daily return distributions (in %):

Stock	$E(X)$	$\sigma(X)$
A	0.8	1.5
B	1.2	3.0

Which stock is more attractive?

Stock B has higher expected return (1.2% vs 0.8%) but also higher risk ($\sigma = 3.0\%$ vs 1.5%). The choice depends on the investor's risk tolerance. A common comparison measure is the Sharpe-like ratio $E(X)/\sigma(X)$:

$$(Stock A:) \frac{0.8}{1.5} \approx 0.53, \quad (Stock B:) \frac{1.2}{3.0} = 0.40$$

Stock A offers a higher return per unit of risk and would be preferred by a risk-averse investor. Stock B has a higher expected return in absolute terms and would be preferred by an investor willing to take additional risk for additional reward.

Example 2.20 Waiting time at a service counter

Customer waiting time T (in minutes) at a post-office counter has density $f(t) = (1/5) e^{-(t/5)}$ for $t \geq 0$, and 0 elsewhere.

- (a) Verify that f is a valid PDF.
- (b) Compute the expected waiting time $E(T)$.
- (c) Compute $P(T \leq 5)$.

Solution (a). The density is non-negative. Integrating over $[0, +\infty)$:

$$\int_0^{+\infty} \frac{1}{5} * e^{-t/5} * dt = 1$$

The integral evaluates to 1 (standard exponential integral), so f is a valid density.

Solution (b).

$$E(T) = \int_0^{+\infty} t * \frac{1}{5} * e^{-t/5} * dt = 5$$

The expected waiting time is 5 minutes, a useful planning figure for staffing decisions.

Solution (c).

$$P(T \leq 5) = \int_0^5 \frac{1}{5} * e^{-t/5} * dt = 1 - e^{-1} \approx 0.632$$

About 63.2% of customers are served within the first 5 minutes. The probability that a customer waits more than 10 minutes is $e^{(-2)} \approx 0.135$.

Example 2.21 Insurance claim size

An insurance company models the size Y (in thousand DZD) of a single claim by a uniform distribution on $[10, 50]$. The density is $f(y) = 1/40$ for $y \in [10, 50]$.

Compute the expected claim size and its standard deviation.

$$\begin{aligned} E(Y) &= \int_{10}^{50} y * \frac{1}{40} * dy = \frac{1}{40} * \left(\frac{y^2}{2} \right)_{10}^{50} \quad (\text{at bounds } 10 \rightarrow 50) \\ &= \frac{1}{80} * (2500 - 100) = 30 \\ E(Y^2) &= \int_{10}^{50} y^2 * \frac{1}{40} * dy = \frac{1}{120} * (125000 - 1000) = \frac{124000}{120} \approx 1033.3 \\ \text{Var}(Y) &= 1033.3 - 30^2 = 133.3, \quad \sigma \approx 11.55 \end{aligned}$$

The expected claim is 30 (thousand DZD) with a standard deviation of approximately 11.55. These two figures are the core inputs for setting the premium level: a company must charge enough to cover $E(Y)$, with a safety margin that reflects $\sigma(Y)$.

Example 2.22 Comparison of two investment plans

Plan A offers a sure profit of 100 000 DZD. Plan B offers 300 000 DZD with probability 0.5 and a loss of 50 000 DZD with probability 0.5. Which plan should be chosen?

Compute the expectation and variance of each:

- Plan A: $E = 100\,000$; $\text{Var} = 0$; $\sigma = 0$.
- Plan B: $E = 0.5 \times 300\,000 + 0.5 \times (-50\,000) = 125\,000$; $\text{Var} = 0.5 \times (300\,000 - 125\,000)^2 + 0.5 \times (-50\,000 - 125\,000)^2 = 30\,625\,000$; $\sigma \approx 175\,000$.

Plan B has a higher expected profit but a high dispersion. A risk-neutral investor (who cares only about E) would prefer B. A risk-averse investor (who penalises high σ) may prefer A despite its lower mean. This trade-off, formalised by utility theory, is at the heart of managerial decision-making under uncertainty.

Example 2.23 Standardising exam scores

University exam scores have mean $\mu = 12$ and standard deviation $\sigma = 3$ (on a 20-point scale). A student scores 17.

$$Z = \frac{17 - 12}{3} = \frac{5}{3} \approx 1.67$$

The student's score is approximately 1.67 standard deviations above the average, a comparison that would remain valid even if another course used a different grading scale.

Example 2.24 Mini-case: Quality-control sampling plan

A manufacturer produces circuit boards in lots of 1 000 units. Each board has an independent defect probability of 3%. A quality inspector randomly tests 20 boards from each lot. Let X be the number of defective boards found among the 20 tested.

- What are the expected value and standard deviation of X ?
- If the manufacturer accepts the lot when $X = 0$ and rejects it when $X \geq 1$, what is the probability of accepting a lot with the given defect rate?

Under the binomial assumption (anticipated in detail in Chapter 3), with $n = 20$ and $p = 0.03$:

$$E(X) = n * p = 20 * 0.03 = 0.60$$

$$\text{Var}(X) = n * p * (1 - p) = 20 * 0.03 * 0.97 = 0.582$$

$$\sigma(X) = \sqrt{0.582} \approx 0.763$$

The inspector therefore expects to find about 0.6 defective boards per sample, with a standard deviation of about 0.76.

For part (b):

$$P(X = 0) = (1 - 0.03)^{20} = 0.97^{20} \approx 0.544$$

The acceptance probability is about 54.4%. If the manufacturer desires a higher acceptance rate, they must either lower the defect rate, increase the sample size, or relax the acceptance rule (e.g. accept when $X \leq 1$). The random-variable framework makes these managerial trade-offs quantitative rather than conjectural.

Example 2.25 Mini-case: Portfolio return and risk

An investor allocates capital between two assets: 60% in Asset A and 40% in Asset B. The individual asset returns X and Y (in %) have $E(X) = 6$, $\text{Var}(X) = 4$, $E(Y) = 10$, $\text{Var}(Y) = 25$. The portfolio return is:

$$R = 0.6 X + 0.4 Y$$

Assuming X and Y are independent:

$$E(R) = 0.6 * 6 + 0.4 * 10 = 3.6 + 4.0 = 7.6 \%$$

$$\text{Var}(R) = 0.6^2 * 4 + 0.4^2 * 25 = 1.44 + 4.00 = 5.44$$

$$\sigma(R) = \sqrt{5.44} \approx 2.33 \%$$

The portfolio has expected return 7.6% and standard deviation 2.33%. Compare this to a 100% allocation in Asset B alone: $E = 10\%$, $\sigma = 5\%$. The diversified portfolio sacrifices 2.4 percentage points of expected return but cuts the standard deviation roughly in half. It is a classic illustration of the benefit of diversification, quantified entirely from the properties of expectation and variance developed in this chapter.

Note that the independence assumption is essential here: without it, a covariance term would enter $\text{Var}(R)$ and could either raise or lower the risk depending on the sign of the correlation. In real portfolios, correlations are typically positive (assets tend to move together), so the true variance is higher than the independence formula predicts. This is the subject of portfolio theory in advanced finance courses.

2.13 Common Mistakes — Watch Out

The errors below are the ones most frequently committed by first-year students. Read them carefully before attempting the exercises.

⚠ Mistake 1 — Confusing PMF and PDF

- PMF $p(x)$ is itself a probability: $0 \leq p(x) \leq 1$.
- PDF $f(x)$ is a density: it can exceed 1 without violating any rule.
- Only integrals of f over intervals give probabilities.

⚠ Mistake 2 — Treating $P(X = x) > 0$ for continuous X

- For a continuous random variable, $P(X = x) = 0$ for every x .
- Strict and non-strict inequalities give the same probability: $P(X \leq a) = P(X < a)$.
- For a discrete random variable, the two are different: $P(X \leq a) = P(X < a) + p(a)$.

⚠ Mistake 3 — Forgetting to square the coefficient in $\text{Var}(aX)$

- $\text{Var}(aX) = a^2 \text{Var}(X)$, not $a \cdot \text{Var}(X)$.
- Consequence: standard deviation scales as $|a|$, not a^2 .
- $\text{Var}(aX + b) = a^2 \text{Var}(X)$. The shift b has no effect on dispersion.

⚠ Mistake 4 — Adding variances without independence

- $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$ holds ONLY when X and Y are independent.
- Without independence, $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2 \cdot \text{Cov}(X, Y)$.

⚠ Mistake 5 — Using the wrong formula for $\text{Var}(X)$

- $\text{Var}(X) = E(X^2) - [E(X)]^2$, NOT $E(X^2) - E(X)^2$.
- The squaring is applied to the mean, not to X inside the expectation.
- Result: $\text{Var}(X)$ is always ≥ 0 (if not, you have a computational error).

⚠ Mistake 6 — Forgetting the normalisation check

- For any proposed PMF: verify $\sum p(x_i) = 1$ before proceeding.
- For any proposed PDF: verify $\int f(x) dx = 1$ over its support.
- If either condition fails, probabilities computed from that function will not be meaningful.

⚠ Mistake 7 — Mis-interpreting the expectation

- $E(X)$ is the long-run average, NOT the most likely value (that is the mode).
- $E(X)$ need not be one of the possible values of X (e.g. $E(\text{die}) = 3.5$).
- Two distributions can have the same $E(X)$ but very different shapes. Always pair $E(X)$ with $\text{Var}(X)$ or $\sigma(X)$.

⚠ Mistake 8 — Confusing σ and σ^2

- $\text{Var}(X) = \sigma^2$ is in squared units of X .
- $\sigma = \sqrt{\text{Var}(X)}$ is in the original units of X .
- Interpretations involving "X typically deviates by about σ from its mean" use σ , not σ^2 .

2.14 Key Takeaways

The most important ideas of this chapter, condensed.

- ✓ A random variable is a function $X: E \rightarrow \mathbb{R}$ that assigns a number to each outcome of the random experiment.
- ✓ Discrete RVs take values in a countable set; continuous RVs take values in an interval of $\mathbb{R}$.
- ✓ For discrete RVs: the PMF is $p(x) = P(X = x)$, with $\sum p(x_i) = 1$.
- ✓ For continuous RVs: the PDF $f(x) \geq 0$ satisfies $\int f(x) dx = 1$; $P(X = x) = 0$ for every x .
- ✓ The CDF $F(x) = P(X \leq x)$ is defined for both types. It is a step function (discrete) or a continuous non-decreasing function (continuous).
- ✓ The PDF and CDF are linked by differentiation/integration: $f(x) = F'(x)$.
- ✓ Expected value: $E(X) = \sum x_i p(x_i)$ (discrete) or $\int x f(x) dx$ (continuous).
- ✓ Expectation is linear: $E(aX + b) = a E(X) + b$, for any constants a, b .
- ✓ Variance: $\text{Var}(X) = E[(X - \mu)^2] = E(X^2) - [E(X)]^2$ in both cases.
- ✓ Standard deviation $\sigma = \sqrt{\text{Var}(X)}$ has the same units as X .
- ✓ $\text{Var}(aX + b) = a^2 \text{Var}(X)$; $\sigma(aX + b) = |a| \sigma(X)$.
- ✓ For independent X and Y : $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$ and $E(XY) = E(X)E(Y)$.
- ✓ The standardised variable $Z = (X - \mu)/\sigma$ has mean 0 and variance 1.
- ✓ $E(X)$ locates the distribution; $\text{Var}(X)$ measures its spread. Always report both.
- ✓ Linear transformations change the location and scale of the distribution but preserve its shape.

2.15 Chapter Summary Cards

The cards below condense the chapter into four visual summaries, designed for quick revision before an assessment.

Card 1 — Essential Concepts at a Glance

✓ ESSENTIAL CONCEPTS

Random variable: function $X: E \rightarrow \mathbb{R}$ assigning a number to each outcome.

Discrete RV: range countable; PMF $p(x) = P(X = x)$; $\sum p(x_i) = 1$.

Continuous RV: range uncountable; PDF $f(x) \geq 0$; $\int f(x) dx = 1$.

CDF: $F(x) = P(X \leq x)$; non-decreasing; $F(-\infty) = 0$, $F(+\infty) = 1$.

PDF/CDF link: $f(x) = F'(x)$; $F(x) = \int \text{from } -\infty \text{ to } x \text{ of } f(t) dt$.

Expectation: $E(X) = \sum x_i p(x_i)$ or $\int x f(x) dx$.

Variance: $\text{Var}(X) = E(X^2) - [E(X)]^2$; $\sigma = \sqrt{\text{Var}(X)}$.

Linear transform: $E(aX + b) = a E(X) + b$; $\text{Var}(aX + b) = a^2 \text{Var}(X)$.

Z-score: $Z = (X - \mu)/\sigma$ has mean 0 and variance 1.

LOTUS: $E(g(X)) = \sum g(x_i) p(x_i)$ or $\int g(x) f(x) dx$.

Card 2 — Terminology Table

Term	Symbol	Meaning
Random variable	X, Y, Z	Function from E to $\mathbb{R}$
Realisation	x	Observed value of X
PMF	$p(x)$	$P(X = x)$ (discrete)
PDF	$f(x)$	Density (continuous)
CDF	$F(x)$	$P(X \leq x)$ (both types)
Expected value	$E(X), \mu$	Mean / long-run average
Variance	$\text{Var}(X), \sigma^2$	Dispersion around the mean
Standard deviation	σ	$\sqrt{\text{Var}(X)}$, same units as X
Z-score	Z	$(X - \mu)/\sigma$, standardised
Support / range	$X(E)$	Set of possible values of X

Card 3 — Decision Guide: Which Formula to Use?

⚠ DECISION GUIDE

Problem talks about counts or integers → discrete RV; use PMF and sums.

Problem talks about durations, amounts, measurements → continuous RV; use PDF and integrals.

Need $P(X \leq a)$ → use CDF: $P(X \leq a) = F(a)$.

Need $P(a \leq X \leq b)$ → $F(b) - F(a)$ (both cases).

Need $E(X)$ → $\sum x_i p(x_i)$ or $\int x f(x) dx$.

Need $\text{Var}(X)$ → compute $E(X^2) - [E(X)]^2$ (usually easier than the definition).

Need $Y = aX + b$ statistics → $E(Y) = aE(X) + b$; $\sigma(Y) = |a| \sigma(X)$.

Need to compare across different scales → standardise: $Z = (X - \mu)/\sigma$.

Card 4 — Typical Exam Problems

⚠ TYPICAL EXAM PROBLEMS

Q1. Given a PMF, verify it, compute $E(X)$, $\text{Var}(X)$, and a specified probability.

Q2. Given a candidate PDF, determine the normalising constant and compute $E(X)$, $\text{Var}(X)$.

Q3. Derive the CDF from a PDF (or vice versa); use it to compute probabilities.

Q4. Apply a linear transformation $Y = aX + b$ and compute $E(Y)$, $\sigma(Y)$.

Q5. Interpret $E(X)$ and $\text{Var}(X)$ in a managerial context (risk, return, planning).

★ **EXAM TIPS**

- Always verify that the given function is a valid PMF or PDF before computing anything else.
- Use the shortcut $\text{Var}(X) = E(X^2) - [E(X)]^2$ whenever possible.
- Remember: standard deviation has the same units as X ; variance has squared units.
- Check the sign and magnitude of your final answer for reasonableness.
- Interpret numerical results in the problem's context, not just as numbers.

2.16 Practice Exercises

The exercises below form a complete set for revision and exam preparation. They are ordered by increasing difficulty. Solutions are not provided; attempt every problem with pen, paper, and the tools developed in this chapter.

2.16.1 Level 1 — Foundations

Exercise 2.1. A discrete RV X has PMF $p(0) = 0.1$, $p(1) = 0.3$, $p(2) = 0.4$, $p(3) = 0.2$. (a) Verify this is a valid PMF. (b) Compute $E(X)$, $E(X^2)$, and $\text{Var}(X)$. (c) Compute $P(X \geq 2)$ and $P(1 \leq X < 3)$.

Exercise 2.2. A continuous RV X has PDF $f(x) = 2x$ for $x \in [0, 1]$ and 0 elsewhere. (a) Verify f is a valid PDF. (b) Compute $E(X)$ and $\text{Var}(X)$. (c) Compute $P(X \leq 0.5)$ and $P(X > 0.8)$.

Exercise 2.3. Let X have mean $\mu = 10$ and standard deviation $\sigma = 2$. Compute $E(Y)$ and $\sigma(Y)$ for: (a) $Y = 3X + 5$; (b) $Y = -2X + 1$; (c) $Y = (X - 10)/2$.

Exercise 2.4. A fair coin is tossed 3 times. Let X be the number of heads. (a) Write out the PMF of X . (b) Compute $E(X)$, $\text{Var}(X)$, $\sigma(X)$. (c) Construct the CDF and use it to compute $P(X \leq 2)$.

Exercise 2.5. Let X be a continuous RV with CDF $F(x) = 0$ for $x < 0$, $F(x) = x^2/9$ for $0 \leq x \leq 3$, $F(x) = 1$ for $x > 3$. (a) Find the PDF $f(x)$. (b) Verify the normalisation condition. (c) Compute $P(1 \leq X \leq 2)$.

2.16.2 Level 2 — Applications

Exercise 2.6. Daily demand X at a bookshop (in units) has PMF $p(0) = 0.1$, $p(1) = 0.2$, $p(2) = 0.3$, $p(3) = 0.25$, $p(4) = 0.1$, $p(5) = 0.05$. Unit margin is 50 DZD. Storage cost is 20 DZD per unit left at day's end. The bookshop stocks 3 units each morning. (a) Compute expected daily profit. (b) Compute the probability that the bookshop runs out of stock.

Exercise 2.7. A continuous RV X representing customer service time (in minutes) has density $f(x) = (1/4)e^{-x/4}$ for $x \geq 0$. (a) Compute $E(X)$ and $\sigma(X)$. (b) Compute $P(X \leq 4)$, $P(X > 8)$. (c) Find the CDF $F(x)$.

Exercise 2.8. An investor's monthly return X (in %) has $E(X) = 1.5$ and $\text{Var}(X) = 9$. Let $Y = 12X - 2$ be the annualised net return after a fixed 2% fee. Compute $E(Y)$, $\text{Var}(Y)$, $\sigma(Y)$.

Exercise 2.9. A machine produces items whose weight X (in grams) follows a uniform distribution on $[95, 105]$. (a) Find the density $f(x)$. (b) Compute $E(X)$, $\text{Var}(X)$, and $\sigma(X)$. (c) Compute $P(98 \leq X \leq 102)$.

Exercise 2.10. Let X be the score (out of 20) of a randomly selected student, with $E(X) = 11$ and $\sigma(X) = 4$. Two students are compared: student A scored 15 and student B scored 7. (a)

Compute their Z-scores. (b) Which student is further from the class average and in which direction?

Exercise 2.G — Reading a distribution from its graph

Let Y be the daily number of insurance claims received by a small agency. Its probability mass function and cumulative distribution function are shown in the figure below. Using the graphs only (no formula is needed), answer the following:

- State $P(Y = 1)$ and $P(Y = 4)$.
- Read $P(Y \leq 2)$ from the CDF, then verify it equals $p(0) + p(1) + p(2)$ from the PMF.
- Compute $P(Y \geq 3)$ using the complement rule.
- Find the expected number of claims $E(Y)$ and the mode of the distribution.
- Explain why the CDF uses solid dots on the left of each step and open dots on the right.

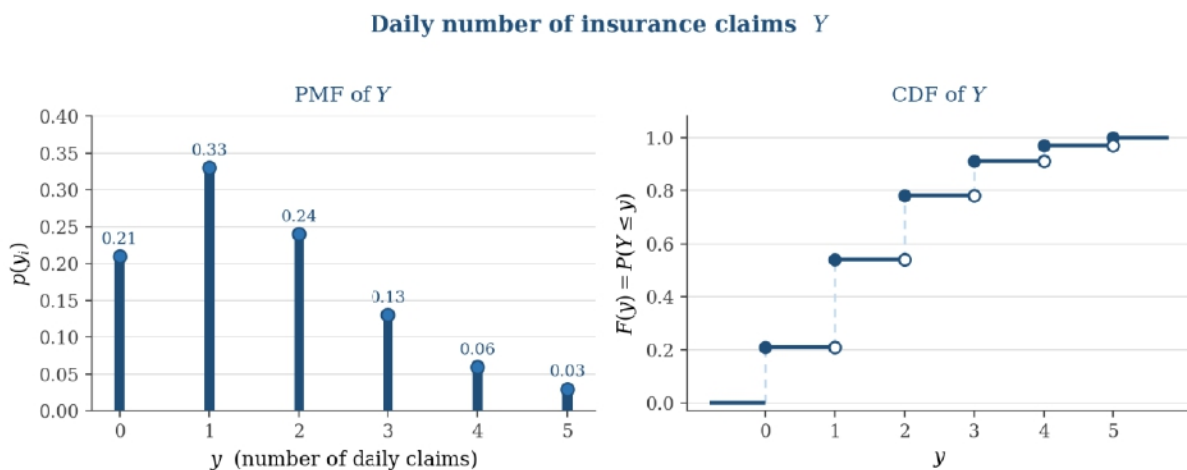


Figure 2.6 — PMF and CDF of the daily number of insurance claims Y (Exercise 2.G).

2.16.3 Level 3 — Extension

Exercise 2.11. A continuous RV X has PDF $f(x) = k \cdot x \cdot (1 - x)$ for $x \in [0, 1]$ and 0 elsewhere.

- Determine k so that f is a valid PDF.
- Compute $E(X)$, $\text{Var}(X)$, $\sigma(X)$.
- Find the CDF.
- Compute $P(0.25 \leq X \leq 0.75)$.

Exercise 2.12. Let X be a discrete RV with $P(X = -2) = 0.2$, $P(X = 0) = 0.5$, $P(X = 3) = 0.3$. Let $Y = X^2 - X$. (a) Find the PMF of Y . (b) Compute $E(Y)$ directly, and via LOTUS using the distribution of X ; check consistency.

Exercise 2.13. A firm's monthly profit (in thousand DZD) is modelled by $X \sim$ with $E(X) = 120$ and $\sigma(X) = 40$. The CEO is considering two bonus schemes: (A) 10% of profit plus 5 (thousand) fixed; (B) a fixed 15 regardless of profit. (a) Compute $E(\text{bonus})$ and $\sigma(\text{bonus})$ under each scheme. (b) Which scheme has higher expected bonus? Which has more variable bonus? Comment on the implications for the CEO's risk attitude.

Exercise 2.14. A continuous RV X has PDF $f(x) = c \cdot e^{(-2x)}$ for $x \geq 0$ and 0 for $x < 0$. (a) Determine c . (b) Compute $E(X)$ and $\text{Var}(X)$. (c) Compute $P(X > 1 \mid X > 0.5)$. Does this equal $P(X > 0.5)$? Interpret.

Exercise 2.15. Two independent RVs X and Y have $E(X) = 3$, $\text{Var}(X) = 4$, $E(Y) = 5$, $\text{Var}(Y) = 9$. Let $W = 2X - Y + 1$. (a) Compute $E(W)$ and $\text{Var}(W)$. (b) Compute $E(XY)$ and $\sigma(W)$.

2.16.4 Supplementary Exam-Style Problems

Exercise 2.16 — Mini-case: airline overbooking. An airline operates a 180-seat flight. Historical data show that each booked passenger shows up with probability 0.92, independently of others. The airline sells 195 tickets per flight. Let X be the number of passengers who show up. (a) Using $E(X) = np$ and $\text{Var}(X) = np(1-p)$, compute $E(X)$ and $\sigma(X)$. (b) What is the approximate probability that X exceeds 180 (so the airline must deny boarding to some passengers)? Assume X is approximately normal for this estimate. (c) Suggest a managerially-sensible overbooking policy based on these results.

Exercise 2.17 — Mini-case: insurance deductible. A continuous RV X represents the size (in thousand DZD) of a random insurance claim and has PDF $f(x) = 0.01 \cdot e^{(-0.01 x)}$ for $x \geq 0$ (exponential with mean 100). The insurer imposes a deductible: the insured pays the first 50 of any claim. Let Y be the amount paid by the insurer. (a) Express Y as a function of X . (b) Compute $E(Y)$. (c) Interpret the result from the insurer's pricing perspective.

Exercise 2.18 — Mini-case: quality bonus scheme. A call-centre manager offers a monthly bonus linked to the number of calls X handled per operator. The bonus formula is: Bonus = 200 if $X \geq 300$, Bonus = 50 if $200 \leq X < 300$, Bonus = 0 if $X < 200$. Suppose X follows a distribution with $P(X \geq 300) = 0.4$, $P(200 \leq X < 300) = 0.45$, $P(X < 200) = 0.15$. (a) Compute $E(\text{Bonus})$ and $\sigma(\text{Bonus})$. (b) If the manager wants to raise the expected bonus by 30% without changing the probabilities, suggest two different ways to modify the bonus amounts. Compare the variance of each modified scheme.

2.16.4 Answers to Prerequisites Diagnostic (§2.3)

(a) $\int$ from 0 to 2 of $3x^2 dx = [x^3]$ from 0 to 2 = 8.

(b) $F(x) = x^3/3 + 2x$, so $F'(x) = x^2 + 2$ and $F'(1) = 3$.

(c) Σ from $k=0$ to 3 of $(1/2)^k = 1 + 1/2 + 1/4 + 1/8 = 15/8 = 1.875$.

(d) Total probability: $P(B) = \Sigma P(A_i) P(B \mid A_i)$ over a partition $\{A_1, \dots, A_k\}$. Interpretation: the probability of B is the weighted sum of its probabilities under each cause, weighted by the prior probability of the cause.

(e) Mean = $(2+4+4+6+8+10)/6 = 34/6 \approx 5.67$. Variance ≈ 8.22 .

(f) Since the union of all events equals E and events are disjoint/exhaustive, axiom 2 gives $P(E) = 1$, and non-negativity (axiom 1) gives $P(A) \geq 0$. Since $A \subseteq E$, monotonicity gives $P(A) \leq P(E) = 1$.

2.17 References

The references below are the primary academic sources consulted for the preparation of this chapter. The booklet-wide bibliography appears at the end.

1. Blitzstein, J. K., & Hwang, J. (2019). *Introduction to probability* (2nd ed.). Chapman & Hall/CRC. <https://doi.org/10.1201/9780429428357>
2. Camm, J. D., Cochran, J. J., Fry, M. J., Ohlmann, J. W., Anderson, D. R., Sweeney, D. J., & Williams, T. A. (2024). *Statistics for business and economics* (15th ed.). Cengage Learning.
3. Devore, J. L., Berk, K. N., & Carlton, M. A. (2021). *Modern mathematical statistics with applications* (3rd ed.). Springer. <https://doi.org/10.1007/978-3-030-55156-8>
4. Hansen, B. E. (2022). *Probability and statistics for economists*. Princeton University Press.
5. Hurlin, C., & Mignon, V. (2022). *Statistique et probabilités en économie-gestion* (2e éd.). Dunod.
6. Lecoutre, J.-P. (2023). *Statistique et probabilités : Cours et exercices corrigés* (7e éd.). Dunod.
7. Ross, S. M. (2024). *A first course in probability* (10th ed., Global Edition). Pearson Education.

End of Chapter 2

You have now mastered random variables, their distributions, and moments.

Next: Chapter 3 will introduce special probability laws
(Bernoulli, Binomial, Poisson, Uniform, Exponential, Normal).

✓ **Before moving on, make sure you can:**

- Identify whether a given problem calls for a discrete or continuous model.
- Verify that a proposed PMF or PDF satisfies the two validity conditions.
- Compute $E(X)$, $E(X^2)$, $\text{Var}(X)$, and $\sigma(X)$ in both discrete and continuous settings.
- Derive the CDF from a PMF/PDF and use it to compute interval probabilities.
- Apply linear transformations and predict their effect on mean and variance.
- Interpret numerical results in an economic or managerial context.

CHAPTER 3

Special Probability Laws

Bernoulli · Binomial · Hypergeometric · Poisson · Uniform · Exponential · Normal

Chapter 2 taught us how to describe any random variable by its distribution and summarise it by its mean and variance. In practice, though, the same few distributional shapes turn up again and again across very different contexts: credit defaults, machine failures, customer arrivals, asset returns, measurement errors. Nature and society produce only a small number of recurring patterns, and once we recognise them, the tools we developed generically in Chapter 2 become concrete computational procedures.

This chapter introduces the seven most important families of distributions used in economics and management: four discrete (Bernoulli, Binomial, Hypergeometric, Poisson) and three continuous (Uniform, Exponential, Normal). For each family you will learn its definition, its parameters, its mean and variance, and the kinds of economic problems it is designed to model. By the end of the chapter, given a real situation described in words, you will be able to identify the appropriate distributional family and compute the probabilities of interest.

i In This Chapter

- §3.1 Introduction — The power of named distributions
- §3.2 Learning Objectives (Bloom's taxonomy)
- §3.3 Prerequisites Check
- §3.4 Notation and Symbols
- §3.5 Overview: The Family Tree of Distributions
- §3.6 Discrete Distributions (Bernoulli, Binomial, Hypergeometric, Poisson)
- §3.7 Continuous Distributions (Uniform, Exponential, Normal)
- §3.8 Approximations Between Distributions
- §3.9 Worked Examples in Economic Context
- §3.10 Common Mistakes — Watch Out
- §3.11 Key Takeaways
- §3.12 Chapter Summary Cards
- §3.13 Practice Exercises
- §3.14 References
- App. A** Standard Normal Table (Z-table)
- App. B** Selected Binomial Probabilities
- App. C** Selected Poisson Probabilities

3.1 Introduction

3.1.1 The power of named distributions

A distribution is the complete probabilistic description of a random variable: its PMF or PDF, its CDF, and through them its mean, variance, and higher moments. In principle, every problem could require its own distribution built from scratch. In practice, only a small number of distributional shapes recur in almost every applied context. These recurring shapes are called named distributions, and each one corresponds to a specific mechanism by which randomness is generated.

The **Bernoulli distribution** models a single yes/no trial. The **binomial** counts the successes in n independent yes/no trials. The **hypergeometric** does the same, but when sampling is done without replacement from a finite population. The **Poisson** counts rare events occurring independently in time or space. On the continuous side, the **uniform** represents equally likely outcomes on an interval; the **exponential** models waiting times; and the **normal** is the universal limit distribution of averages (Ross, 2024).

The value of knowing these seven distributions is threefold. First, they provide a vocabulary: once you recognise that customer arrivals follow a Poisson pattern, you automatically inherit every formula derived for that family. Second, they make computation tractable: closed-form expressions for means, variances, and probabilities replace numerical integration. Third, they make communication possible: a claim that returns are approximately normal with mean 8% and standard deviation 15% tells a reader everything they need to know.

3.1.2 A brief historical note

The study of specific distributions began with Jakob Bernoulli (1713), whose *Ars Conjectandi* introduced what we now call the Bernoulli and binomial distributions and stated the first version of the law of large numbers. Abraham de Moivre (1738) derived the normal distribution as an approximation to the binomial, a result that Pierre-Simon Laplace (1812) developed into the first version of the Central Limit Theorem. Siméon Denis Poisson (1837) introduced his eponymous distribution as a limit of the binomial when the number of trials is large and the probability of success is small.

The modern unified theory is largely due to twentieth-century statisticians who showed how all these distributions fit into families indexed by one or two parameters. Alongside this theoretical unification, computational tools emerged: tables of values in Karl Pearson's *Tables for Statisticians* (1914), followed by modern software that makes probabilistic computation routine. The tables you will find in the appendices of this chapter are direct descendants of Pearson's work (Blitzstein & Hwang, 2019).

3.1.3 How this chapter extends Chapter 2

Chapter 2 introduced the notion of a random variable and its numerical summaries. This chapter is the applied follow-up: rather than studying random variables in full generality, we focus on the few specific families that cover the vast majority of real-world situations. The transition is summarised as follows.

- **Chapter 2 asked:** what tools exist to describe any random variable? (Answer: PMF, PDF, CDF, E, Var.)
- **Chapter 3 asks:** which specific distributional shapes do we encounter most often? (Answer: the seven families of this chapter.)
- **Chapter 2 was abstract:** it worked with generic symbols $p(x)$, $f(x)$, $F(x)$.
- **Chapter 3 is concrete:** each named distribution has specific formulas, parameters, and applications.

At the exam, you will not be asked to derive generic properties from the axioms. You will be asked to identify which distributional family applies to a given situation, state its parameters from the problem data, and compute probabilities or expected values using the appropriate formulas. This skill is tested in every final exam of this course and in almost every quantitative course you will take afterwards.

3.2 Learning Objectives

After working through this chapter you should be able to perform the tasks listed below. Each row maps to one level of Bloom's cognitive taxonomy.

Bloom level	Action verb	Specific competency
Remember	State · Recall	The PMF/PDF, mean, and variance of each of the seven named distributions.
Understand	Explain · Distinguish	The generating mechanism behind each family; when to use Binomial vs Hypergeometric vs Poisson; when to approximate one distribution by another.
Apply	Compute · Use	Probabilities from formulas or tables; standardisation and Z-table for the Normal; Poisson approximation for rare-event problems.
Analyse	Decompose · Choose	Given a problem description in words, identify the correct distributional family and its parameters.
Evaluate	Justify · Interpret	The numerical result in the context of the original problem; comment on the adequacy of a distributional model.

These competencies map to outcomes **C3–C6** of the official course specification (fiche pédagogique) and build directly on Chapter 2.

How to study this chapter effectively

This chapter is mainly a catalogue: seven distributions, each with its own story. The study path should therefore be exhaustive rather than shortcut-seeking.

- **Step 1.** For each distribution, memorise three things: (i) the generating mechanism in words, (ii) the PMF or PDF, and (iii) the formulas for $E(X)$ and $\text{Var}(X)$.
- **Step 2.** Work through every example in §3.6 and §3.7 with pen and paper. Do not skip the small calculations.
- **Step 3.** Practise identifying the correct distribution from short verbal descriptions. This is the hardest skill and the one that distinguishes a successful exam candidate.
- **Step 4.** Learn to read the Z-table (Appendix A). Many exam questions reduce to this single skill.
- **Step 5.** Study §3.8 (Approximations) last. It makes sense only after the individual distributions are well understood.
- **Step 6.** Attempt the practice exercises in §3.13 without looking at the examples; if you cannot, return to the corresponding worked example and start again.

A realistic workload for this chapter is approximately twenty hours: four for reading, eight for worked examples, and eight for exercises. This is the heaviest chapter of the booklet, so budget accordingly.

3.3 Prerequisites Check

Before starting this chapter, you should be comfortable with the following notions. If any item seems unfamiliar, review the relevant section before continuing.

- **From Chapter 1:** independence of events; total probability; combinatorics ($C(n, k)$, $A(n, k)$, $n!$).
- **From Chapter 2:** PMF, PDF, CDF; computation of $E(X)$ and $\text{Var}(X)$ for discrete and continuous random variables; linear transformations; the Z-score.
- **From calculus:** integration of polynomials and exponentials; the number e and the exponential function; limits when $n \rightarrow \infty$.
- **From algebra:** binomial coefficient $C(n, k)$; factorials; the binomial theorem $(a + b)^n$.

The short diagnostic below lets you test yourself. Answers are given at the end of §3.13.

i Diagnostic 3.1 — Self-check before starting

- Compute $C(6, 2)$ and $C(10, 3)$.

- Expand $(a + b)^3$ using the binomial theorem.
- Compute $\int$ from 0 to ∞ of e^{-2x} dx.
- For X with $E(X) = 10$ and $\text{Var}(X) = 4$, compute the Z-score of the value 13.
- A coin is tossed 3 times. Using Chapter 2 methods, compute the probability of exactly 2 heads.
- Why is the function $f(x) = e^{-x}$, $x \geq 0$, a valid PDF? Compute $E(X)$ for it.

3.4 Notation and Symbols

The table below lists the symbols used throughout this chapter. Notation for specific distributions is introduced in the corresponding sections.

Symbol	Reading	Meaning
$X \sim B(p)$	"X follows Bernoulli(p)"	X is a Bernoulli random variable with parameter p.
$X \sim B(n, p)$	"X follows Binomial(n, p)"	X is binomial with n trials and success probability p.
$X \sim H(N, K, n)$	"X follows Hypergeometric"	Sampling n items from N, of which K are successes.
$X \sim P(\lambda)$	"X follows Poisson(λ)"	X is Poisson with rate parameter $\lambda > 0$.
$X \sim U(a, b)$	"X follows Uniform(a, b)"	X is uniformly distributed on $[a, b]$.
$X \sim \text{Exp}(\lambda)$	"X follows Exponential(λ)"	X is exponential with rate $\lambda > 0$.
$X \sim N(\mu, \sigma^2)$	"X follows Normal"	X is normal with mean μ and variance σ^2 .
$C(n, k)$	"n choose k"	Binomial coefficient: $n! / [k! (n - k)!]$.
e	Euler's number	$e \approx 2.71828$, base of the natural exponential.
λ	"lambda"	Rate parameter for Poisson or Exponential.
$\Phi(z)$	"Phi of z"	CDF of the standard normal distribution $N(0, 1)$.
Z	"Z-score"	Standardised variable: $Z = (X - \mu)/\sigma$.

3.5 The Family Tree of Distributions

The seven distributions studied in this chapter can be organised into two families according to whether the underlying random variable is discrete or continuous. Within each family, the distributions are related to one another, either by specialisation (one being a particular case of another) or by approximation (one being a limiting case of another).

3.5.1 Classification at a glance

Family	Distribution	Parameters	When to use
Discrete	Bernoulli $B(p)$	p	A single yes/no trial.
Discrete	Binomial $B(n, p)$	n, p	Count of successes in n independent trials (with replacement).
Discrete	Hypergeometric $H(N, K, n)$	N, K, n	Count of successes in n draws WITHOUT replacement.
Discrete	Poisson $P(\lambda)$	λ	Count of rare events in time or space at constant rate.
Continuous	Uniform $U(a, b)$	a, b	Equally likely outcomes on an interval $[a, b]$.
Continuous	Exponential $\text{Exp}(\lambda)$	λ	Waiting time between Poisson events.
Continuous	Normal $N(\mu, \sigma^2)$	μ, σ^2	Central-limit phenomena; measurements with symmetric errors.

3.5.2 Relationships between distributions

Several important connections link the seven families. These relationships simplify calculations and deepen understanding:

- **Bernoulli** → **Binomial**: a Binomial $B(n, p)$ is a sum of n independent Bernoulli $B(p)$ trials. The Bernoulli is the special case $n = 1$.
- **Binomial** → **Hypergeometric**: the hypergeometric describes the same situation as the binomial, but under sampling without replacement. When N is large relative to n , the two are numerically close.
- **Binomial** → **Poisson**: when n is large and p is small with $np = \lambda$ fixed, the binomial $B(n, p)$ is well approximated by Poisson(λ).
- **Poisson** → **Exponential**: if events occur according to a Poisson process at rate λ per unit time, the time between consecutive events is $\text{Exp}(\lambda)$.
- **Binomial** → **Normal**: for large n with np and $n(1-p)$ both ≥ 5 , the binomial is well approximated by a normal distribution (de Moivre–Laplace theorem).

- **Poisson** → **Normal**: for large λ , the $\text{Poisson}(\lambda)$ is well approximated by $N(\lambda, \lambda)$.

Section 3.8 formalises these approximations with the specific conditions under which each is valid.

3.6 Discrete Distributions

We begin with the four discrete families: Bernoulli, Binomial, Hypergeometric, and Poisson. Each describes a count, whether of successes, of events, or of occurrences, and each has its own characteristic mechanism.

3.6.1 The Bernoulli distribution — $B(p)$

Generating mechanism

Consider a single random experiment with exactly two possible outcomes, conventionally labelled "success" (coded 1) and "failure" (coded 0). Let $p = P(\text{success})$ be the probability of success, with $0 \leq p \leq 1$. The random variable X that records the outcome (1 or 0) follows the Bernoulli distribution with parameter p , written $X \sim B(p)$.

The Bernoulli distribution is the simplest non-trivial distribution and the building block of every other discrete distribution in this chapter.

PMF, mean, and variance

$$P(X = 1) = p, \quad P(X = 0) = 1 - p$$

Or, more compactly:

$$P(X = k) = p^k * (1 - p)^{1-k}, \quad k \in \{0, 1\}$$

$$E(X) = p$$

$$\text{Var}(X) = p * (1 - p)$$

Derivation of the mean: $E(X) = 1 \cdot p + 0 \cdot (1 - p) = p$. For the variance, $E(X^2) = 1^2 \cdot p + 0^2 \cdot (1 - p) = p$, so $\text{Var}(X) = p - p^2 = p(1 - p)$.

Notice that $\text{Var}(X)$ is maximised at $p = 0.5$ (value 0.25) and vanishes at $p = 0$ or $p = 1$. Intuitively, a 50-50 coin produces the most uncertainty, while a coin with $p = 0$ or $p = 1$ is deterministic.

Economic applications

Example 3.1 — Loan default as a Bernoulli trial

A bank classifies each loan as either "repaid on time" ($X = 0$) or "defaulted" ($X = 1$). If the default rate is $p = 0.08$:

$$E(X) = 0.08, \quad \text{Var}(X) = 0.08 * 0.92 = 0.0736, \quad \sigma = \sqrt{0.0736} \approx 0.271$$

The mean 0.08 equals the default probability, confirming the general identity $E(1_A) = P(A)$. The variance is small because p is close to zero; the outcome is fairly predictable (most loans are repaid).

Example 3.2 — Customer response to a marketing campaign

A marketing department sends an email to potential customers with a 15% probability of generating a purchase. Let $X = 1$ if the recipient purchases, 0 otherwise. Then $X \sim B(0.15)$, with:

$$E(X) = 0.15, \text{Var}(X) = 0.15 * 0.85 = 0.1275$$

For a single recipient, the expected "purchase indicator" is 0.15. When this is applied to n independent recipients, we enter the domain of the binomial distribution.

✓ Key identity for Bernoulli random variables

- $E(X) = p$, the probability of success.
- $\text{Var}(X) = p(1 - p)$, maximal at $p = 0.5$.
- The Bernoulli is the indicator variable of the event "success".

3.6.2 The Binomial distribution — $B(n, p)$

Consider n independent Bernoulli trials, each with the same success probability p . Let X be the total number of successes across the n trials. Then X follows the Binomial distribution with parameters n and p , written $X \sim B(n, p)$.

The binomial describes any "counting in independent trials" situation: the number of defective items in a production sample, the number of customers who respond to an ad, the number of heads in n coin tosses. The three defining conditions are:

- **Fixed number of trials** n , known in advance.
- **Identical trials**: each trial has the same success probability p .
- **Independent trials**: the outcome of any trial does not affect the others (equivalent to sampling with replacement).

If any of these conditions fails (for instance, if sampling is without replacement from a small population), a different distribution applies (hypergeometric, §3.6.3).

PMF, mean, and variance

The probability of observing exactly k successes in n trials is:

$$P(X = k) = \binom{n}{k} * p^k * (1 - p)^{n-k}, \quad k \in \{0, 1, \dots, n\}$$

The three factors of this formula each play a distinct role:

- $C(n, k)$ counts the number of ways to choose which k of the n trials produce the successes.
- p^k is the probability that those k specific trials succeed.
- $(1 - p)^{n-k}$ is the probability that the remaining $n - k$ trials fail.

The mean and variance of the binomial follow from its representation as a sum of n independent Bernoulli trials:

$$E(X) = n * p$$

$$Var(X) = n * p * (1 - p)$$

$$\sigma = \sqrt{n * p * (1 - p)}$$

These two formulas, np for the mean and $np(1 - p)$ for the variance, are the most important pair in this chapter and should be memorised.

Reading the binomial table

Appendix B provides tabulated values of the binomial CDF $F(k) = P(X \leq k)$ for selected combinations of n and p . To use the table:

- Locate the row corresponding to your n .

- Locate the column corresponding to your p.
- The entry gives $P(X \leq k)$; for $P(X = k)$, compute $F(k) - F(k - 1)$.
- For "at least k successes", use $1 - F(k - 1)$.

In practice, for small n (say $n \leq 20$) it is often quicker to compute $P(X = k)$ directly from the formula than to consult a table.

Economic applications

Example 3.3 — *Quality inspection*

A factory produces items with a 5% defect rate. A quality inspector randomly tests 10 items from the daily output (assuming the output is large enough that replacement is effectively immaterial). Let X be the number of defective items among the 10 tested.

$X \sim B(10, 0.05)$. The mean and variance are:

$$E(X) = 10 * 0.05 = 0.5$$

$$Var(X) = 10 * 0.05 * 0.95 = 0.475, \sigma \approx 0.689$$

On average, the inspector finds 0.5 defectives per sample, with standard deviation about 0.69. The probability of finding exactly 2 defectives is:

$$P(X = 2) = \binom{10}{2} * 0.05^2 * 0.95^8 = 45 * 0.0025 * 0.6634 \approx 0.0746$$

The probability of finding no defectives at all is:

$$P(X = 0) = 0.95^{10} \approx 0.5987$$

So about 60% of samples will find no defectives, a useful figure for calibrating the acceptance rule of the quality-control plan.

Example 3.4 — *Marketing campaign success*

A call centre contacts 20 potential customers. Each contact has a 30% probability of resulting in a sale, independently of the others. Let X be the number of sales.

$X \sim B(20, 0.30)$. The expected number of sales is:

$$E(X) = 20 * 0.30 = 6$$

$$Var(X) = 20 * 0.30 * 0.70 = 4.2, \sigma \approx 2.05$$

The probability of achieving at least 8 sales (exceeding the expected 6):

$$P(X \geq 8) = 1 - P(X \leq 7)$$

From a binomial table (Appendix B), $P(X \leq 7) \approx 0.7723$, so:

$$P(X \geq 8) \approx 1 - 0.7723 = 0.2277$$

There is roughly a 23% chance of exceeding the expected target, a figure that could inform bonus thresholds or performance benchmarks.

Example 3.5 — Portfolio of loans — a multi-Bernoulli calculation

A bank holds a portfolio of 50 small loans, each with a 4% probability of default in the next year, independently of the others. Let X be the number of defaults.

$X \sim B(50, 0.04)$.

$$E(X) = 50 * 0.04 = 2$$

$$Var(X) = 50 * 0.04 * 0.96 = 1.92, \sigma \approx 1.386$$

The expected number of defaults is 2, with standard deviation about 1.4. For risk-management purposes, the bank may ask: what is the probability that more than 5 loans default?

$$P(X > 5) = 1 - P(X \leq 5)$$

This is a small number: about 0.015 from binomial tables, roughly 1.5%. The bank thus has comfortable margin if it provisions for 5 defaults.

Note how the same binomial formula, with different parameters, serves applications as diverse as defect counting, marketing response, and credit risk. This economy of concepts is the main pedagogical reason for learning distributional families.

✓ **Binomial at a glance**

Setting: n independent, identical yes/no trials; fixed success probability p .

Range: $X \in \{0, 1, \dots, n\}$.

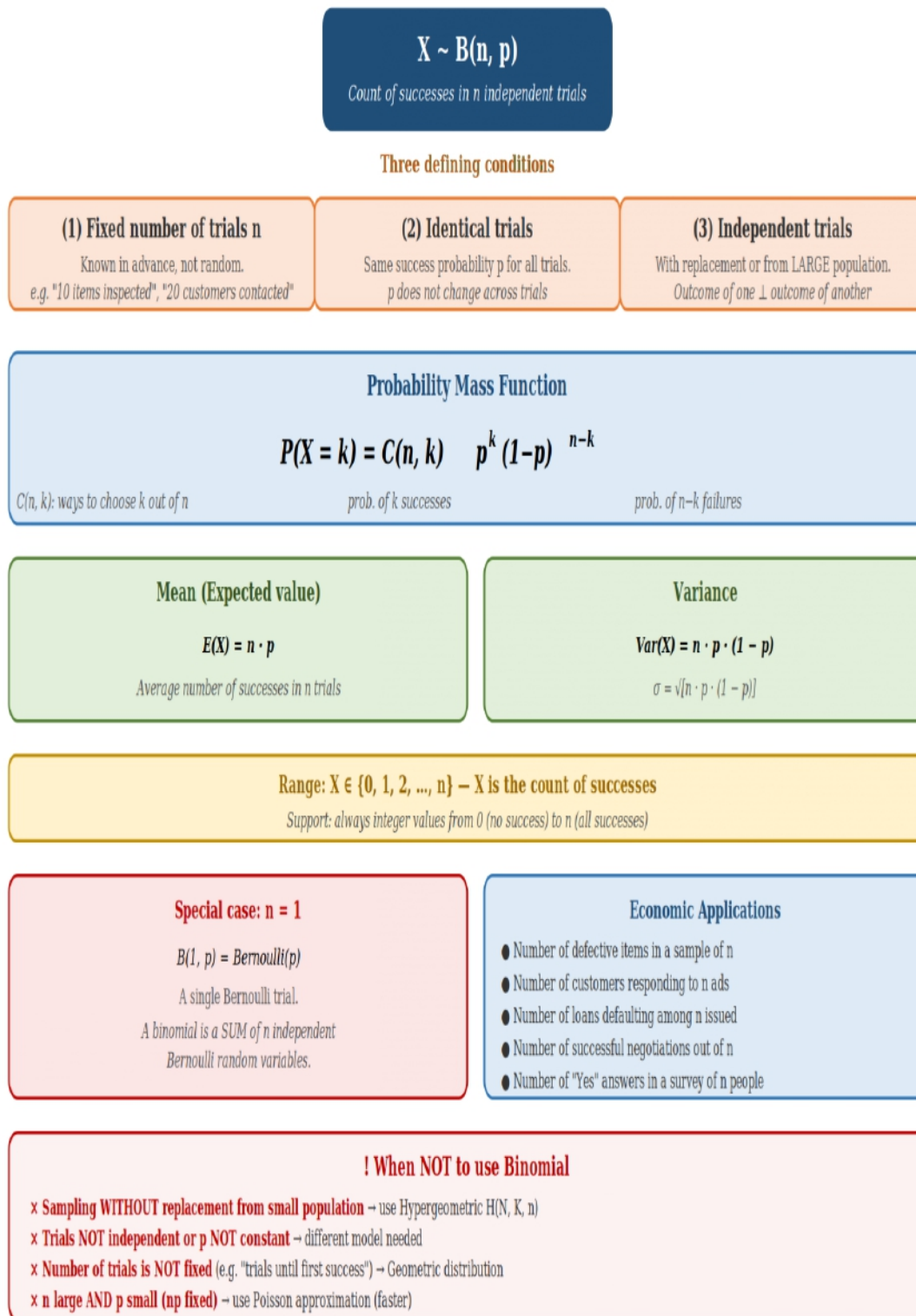
Mean: $E(X) = np$.

Variance: $Var(X) = np(1 - p)$.

When to use: counting successes in a FIXED number of trials with REPLACEMENT or from a LARGE population.

Visual Summary — Mind Map of §3.6.2

The diagram below summarises the binomial distribution and its properties, including when it applies and when a different model should be used instead.

Figure 3.1: The Binomial distribution $B(n, p)$: parameters, PMF, moments, and applications**Mind Map 3.1 – Binomial Distribution $B(n, p)$** *Source: Elaborated by the Author*

3.6.3 The Hypergeometric distribution — $H(N, K, n)$

Consider a finite population of N items, of which K are classified as "successes" and the remaining $N - K$ as "failures". A sample of n items is drawn at random, without replacement. Let X be the number of successes in the sample. Then X follows the Hypergeometric distribution with parameters N , K , and n , written $X \sim H(N, K, n)$.

The key difference from the binomial is the sampling mechanism: the binomial assumes sampling with replacement (or from a very large population where replacement is inconsequential), while the hypergeometric assumes sampling without replacement from a specific finite population. Every time an item is drawn, the composition of the remaining population changes, so the probability of a success on the next draw depends on the history.

In most practical situations involving audits, quality inspections, committee selection, or finite-population sampling, the hypergeometric is the correct model.

PMF, mean, and variance

The probability of observing exactly k successes in the sample is:

$$P(X = k) = \frac{\binom{K}{k} * \binom{N-K}{n-k}}{\binom{N}{n}}$$

The numerator counts the number of samples that contain exactly k successes and $n - k$ failures. The denominator counts all possible samples of size n from the population of N . The valid range of k is:

$$\max(0; n - (N - K)) \leq k \leq \min(n; K)$$

— you cannot draw more successes than exist in the population, nor fewer than what is forced when the failures run out. The mean and variance are:

$$E(X) = n * \frac{K}{N}$$

$$Var(X) = n * \frac{K}{N} * \left(1 - \frac{K}{N}\right) * \frac{N - n}{N - 1}$$

The mean has exactly the same form as the binomial with $p = K/N$, the proportion of successes in the population. The variance, however, includes an additional factor $(N - n)/(N - 1)$, called the finite population correction (FPC). Since the FPC is less than 1 (for $n \geq 2$), the hypergeometric variance is smaller than the binomial variance: sampling without replacement is more efficient than sampling with replacement.

★ **When Binomial and Hypergeometric give similar answers**

- When the sample size n is small compared to the population N (rule of thumb: $n/N \leq 5\%$), the FPC is close to 1 and the hypergeometric is well approximated by the binomial $B(n, K/N)$.
- When $n/N > 5\%$, the difference becomes important and the hypergeometric must be used explicitly.
- For small populations (like auditing 10 accounts out of 50), the binomial can overstate the variance significantly.

Economic applications**Example 3.6 — Quality inspection — sampling without replacement**

A shipment contains 20 electronic components, of which 5 are defective. A quality inspector randomly selects 4 components for testing (without replacement, as testing is destructive). Let X be the number of defective components in the sample.

$X \sim H(20, 5, 4)$. The probability of finding exactly 2 defectives is:

$$P(X = 2) = \frac{\binom{5}{2} * \binom{15}{2}}{\binom{20}{4}} = \frac{10 * 105}{4845} = \frac{1050}{4845} \approx 0.217$$

The expected number of defectives in the sample:

$$E(X) = 4 * \frac{5}{20} = 1$$

So on average, 1 defective appears in each sample of 4, which is what one would expect from the overall defect rate of $5/20 = 25\%$. The variance:

$$Var(X) = 4 * 0.25 * 0.75 * \frac{20 - 4}{20 - 1} = 0.75 * \frac{16}{19} \approx 0.632$$

Compare this to the binomial variance that would apply if the sampling were done with replacement: $Var = 4 \times 0.25 \times 0.75 = 0.75$. The hypergeometric variance is smaller by about 16%, which shows the information gain from not repeating draws.

Example 3.7 — Auditing — detecting errors in accounts

An audit firm receives 100 expense reports, of which 15 contain errors. A junior auditor randomly selects 10 reports for detailed review. What is the probability that at least 3 of the reviewed reports contain errors?

Let X be the number of reports with errors in the sample. $X \sim H(100, 15, 10)$.

$$P(X \geq 3) = 1 - P(X \leq 2) = 1 - \sum_{k=0}^2 \frac{\binom{15}{k} * \binom{85}{10-k}}{\binom{100}{10}}$$

Computing each term:

$$P(X = 0) = \frac{\binom{15}{0} * \binom{85}{10}}{\binom{100}{10}} \approx 0.184$$

$$P(X = 1) = \frac{\binom{15}{1} * \binom{85}{9}}{\binom{100}{10}} \approx 0.342$$

$$P(X = 2) = \frac{\binom{15}{2} * \binom{85}{8}}{\binom{100}{10}} \approx 0.291$$

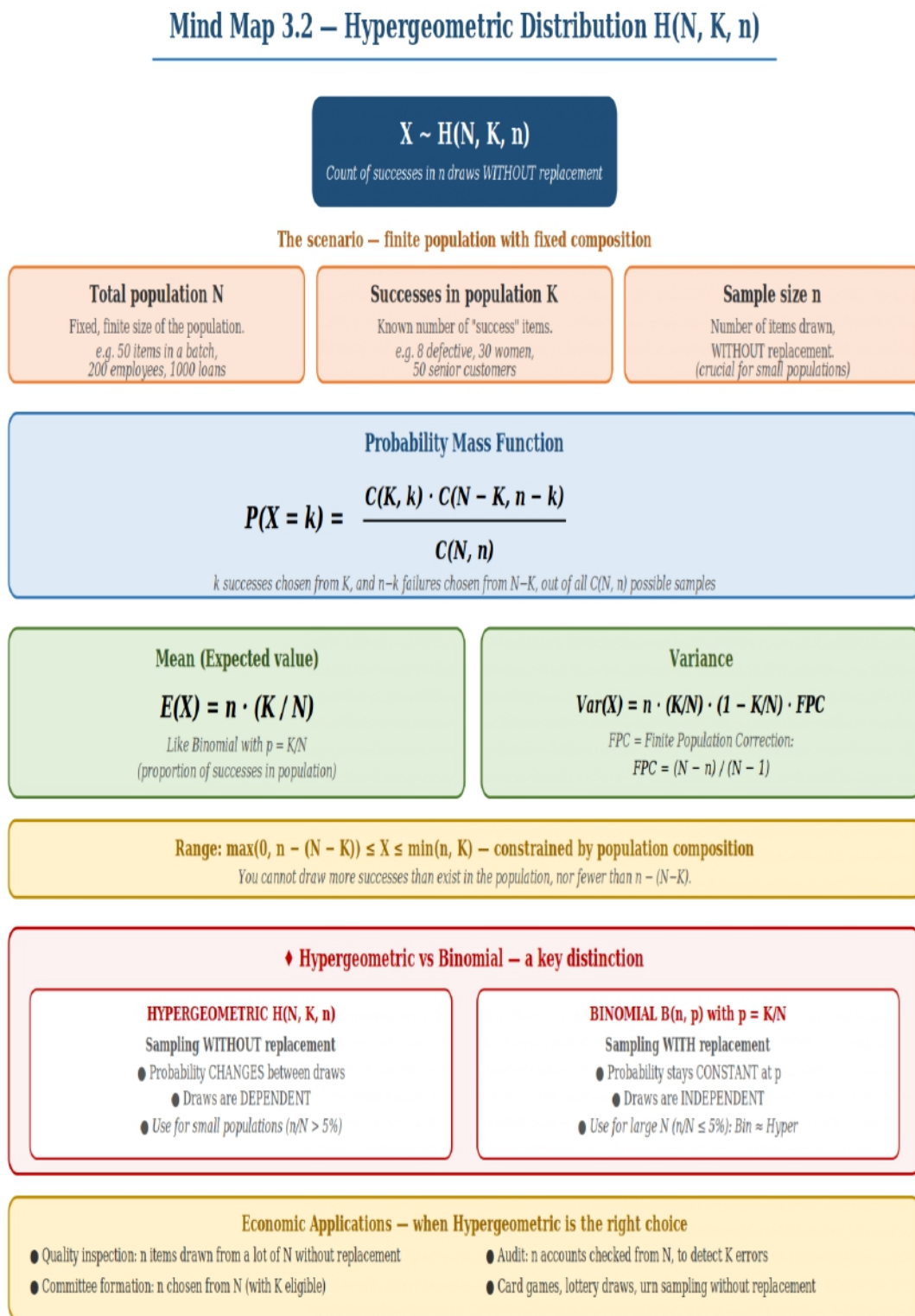
$$P(X \geq 3) = 1 - (0.184 + 0.342 + 0.291) = 1 - 0.817 = 0.183$$

There is about an 18% chance that at least 3 of the 10 reviewed reports contain errors, useful for setting the expected workload of the audit.

Visual Summary — Mind Map of §3.6.3

The diagram below summarises the hypergeometric distribution and its distinction from the binomial.

Figure 3.2: The Hypergeometric distribution $H(N, K, n)$: sampling without replacement



Source: Elaborated by the Author

3.6.4 The Poisson distribution — $P(\lambda)$

The **Poisson distribution** models the count of rare events occurring at a constant average rate over a fixed time interval or region of space. It was introduced by Siméon Denis Poisson in 1837 and is today the workhorse of any problem involving arrivals, breakdowns, claims, or other independent events that happen at a steady rate.

A Poisson process is characterised by three assumptions:

- **Constant rate:** the mean number of events per unit time, denoted λ , is the same throughout the interval.
- **Independence:** events occur independently of one another; the occurrence of an event does not change the probability of future events.
- **Rare events:** the probability of two or more events occurring simultaneously is negligible.

Under these assumptions, the number of events X observed in a fixed interval follows a Poisson distribution with parameter λ , written $X \sim P(\lambda)$.

PMF, mean, and variance

The PMF is:

$$P(X = k) = \frac{e^{-\lambda} * \lambda^k}{k!}, \quad k = 0, 1, 2, \dots$$

Unlike the binomial and hypergeometric, the Poisson has no upper limit: X can take arbitrarily large values, though the probabilities become negligible for k much larger than λ .

The Poisson has the striking property that its mean and variance are equal:

$$E(X) = \lambda$$

$$Var(X) = \lambda$$

$$\sigma = \sqrt{\lambda}$$

This identity, mean equals variance, is the signature of the Poisson distribution and provides a quick empirical test: if observed data have sample mean approximately equal to sample variance, a Poisson model may be appropriate.

Reading the Poisson table

Appendix C tabulates $P(X = k)$ for selected values of λ . For cumulative probabilities, one simply sums the relevant entries: $P(X \leq k) = \sum P(X = j)$ for j from 0 to k .

Approximation: Binomial $\rightarrow$ Poisson

When n is large and p is small (conventionally, $n \geq 20$ and $p \leq 0.05$), with $\lambda = np$ moderate, the binomial $B(n, p)$ is well approximated by the Poisson $P(\lambda)$:

$$\binom{n}{k} * p^k * (1 - p)^{n-k} \approx \frac{e^{-\lambda} * \lambda^k}{k!}, \quad (\text{where}) \lambda = np$$

This approximation is of great practical value: the binomial formula with factorials becomes cumbersome when n is large, while the Poisson formula remains simple. Historically, this was the original motivation for Poisson's derivation of his distribution.

Economic applications

Example 3.8 — Customer arrivals at a bank counter

A bank branch receives customers at an average rate of 6 per hour, and the arrivals follow a Poisson process. Let X be the number of customers arriving in a given hour. Compute:

- (a) The probability of receiving exactly 5 customers.
- (b) The probability of receiving more than 8 customers.
- (c) The expected number of customers and the standard deviation.

Solution (a). $X \sim P(6)$:

$$P(X = 5) = \frac{e^{-6} * 6^5}{5!} = \frac{0.00248 * 7776}{120} \approx 0.1606$$

Solution (b). Using complement and the Poisson table:

$$P(X > 8) = 1 - P(X \leq 8) = 1 - 0.847 = 0.153$$

Solution (c).

$$E(X) = 6, \quad \sigma = \sqrt{6} \approx 2.449$$

The branch manager can use these figures to plan staffing: six customers per hour on average, with hourly variation typically within ± 2.5 customers.

Example 3.9 — Insurance claims — using Poisson approximation

A car insurance company insures 500 similar drivers. Each driver has a 1% probability of filing a claim in a given year, independently of others. Compute the probability that the company receives exactly 6 claims.

Let X be the number of claims. Strictly, $X \sim B(500, 0.01)$. Since $n = 500$ is large and $p = 0.01$ is small, the Poisson approximation with $\lambda = np = 5$ applies:

$$P(X = 6) \approx \frac{e^{-5} * 5^6}{6!} = \frac{0.00674 * 15625}{720} \approx 0.1462$$

Computing directly with the binomial gives $P(X = 6) \approx 0.1462$, almost identical. The Poisson approximation is clearly adequate and computationally much simpler.

Example 3.10 — Defects per page in a document

A quality team finds that a printed document has on average 2 typographical errors per page, distributed according to a Poisson process. What is the probability that a randomly selected page has no errors? At least one error?

$$X \sim P(2)$$

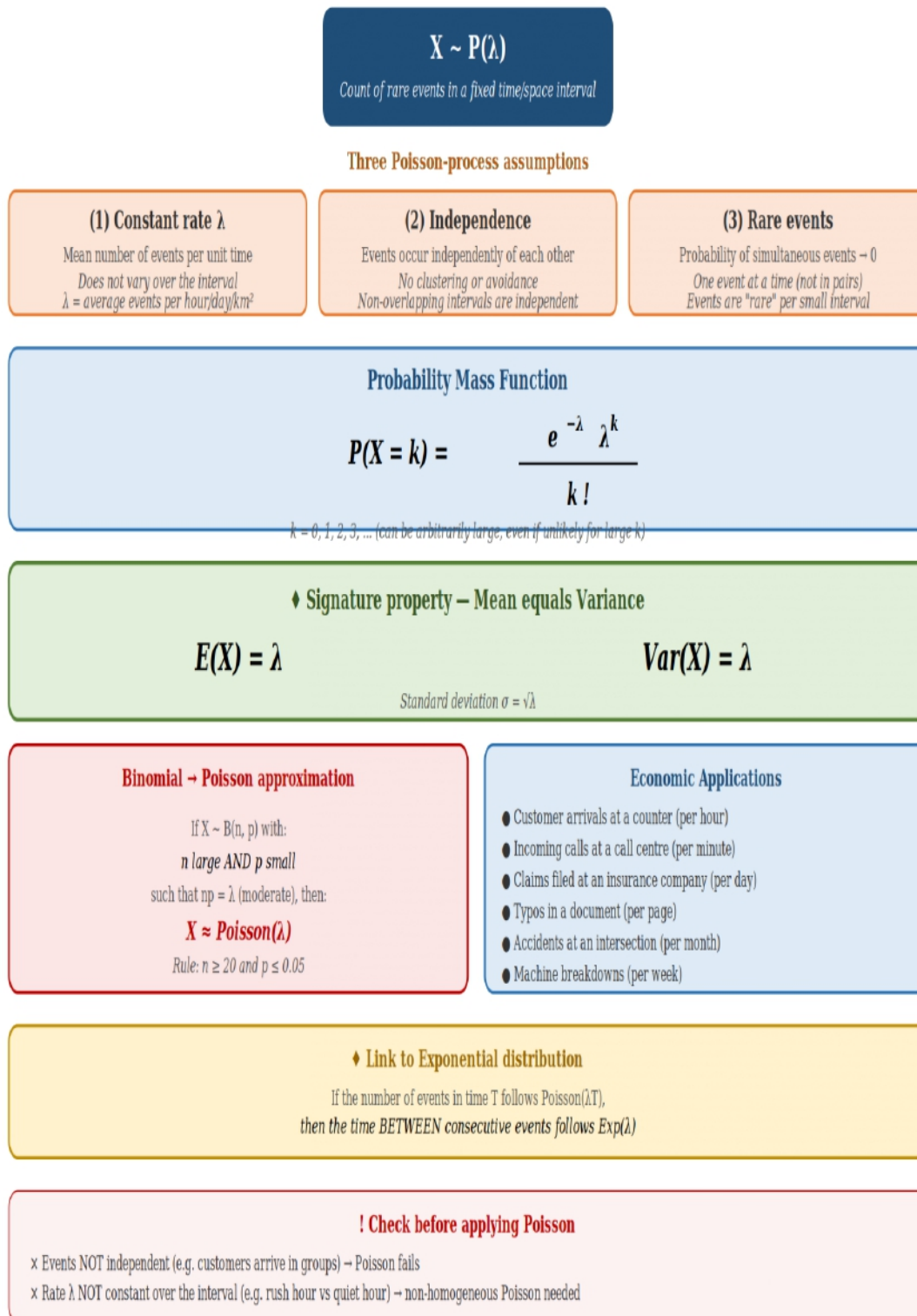
$$P(X = 0) = \frac{e^{-2} * 2^0}{0!} = e^{-2} \approx 0.1353$$

$$P(X \geq 1) = 1 - P(X = 0) \approx 1 - 0.1353 = 0.8647$$

About 86.5% of pages contain at least one error, a high rate that suggests the proof-reading procedure needs improvement.

Visual Summary — Mind Map of §3.6.4

The diagram below summarises the Poisson distribution, its assumptions, and its connection to the binomial and exponential.

Figure 3.3: The Poisson distribution $P(\lambda)$: rare events in time or spaceMind Map 3.3 – Poisson Distribution $P(\lambda)$ 

Source: Elaborated by the Author

3.7 Continuous Distributions

We now turn to the three continuous families covered in this course: Uniform, Exponential, and Normal. In contrast with the discrete distributions of §3.6, these are described by probability density functions, and probabilities are computed by integration over intervals.

3.7.1 The Uniform distribution — $U(a, b)$

Generating mechanism

A continuous random variable X is uniformly distributed on the interval $[a, b]$, written $X \sim U(a, b)$, if every sub-interval of $[a, b]$ of the same length has the same probability. The uniform distribution models complete ignorance about the value of X within the interval: there is no reason to prefer one region over another.

The uniform is the continuous analogue of "equally likely outcomes" for a discrete model. It is used as a starting point in simulation studies, as a non-informative prior in some Bayesian analyses, and to model situations like "the bus arrives at a random time within its 10-minute window".

PDF, CDF, mean, and variance

$$f(x) = \frac{1}{b - a}, \quad (\text{for } a \leq x \leq b; \quad 0 \text{ (otherwise)})$$

$$F(x) = \frac{x - a}{b - a}, \quad (\text{for } a \leq x \leq b)$$

$$E(X) = \frac{a + b}{2}$$

$$\text{Var}(X) = \frac{(b - a)^2}{12}$$

The density is constant on $[a, b]$: a rectangle of height $1/(b - a)$ and width $(b - a)$, whose area is 1. The expectation sits in the middle of the interval, as symmetry demands. The variance grows quadratically with the width of the interval.

Example 3.11 — Uniform waiting time

A public bus arrives at a stop every 10 minutes. A passenger arrives at the stop at a random time. Let T be the waiting time, measured in minutes. Since the passenger has no information about the bus schedule, $T \sim U(0, 10)$.

$$f(t) = \frac{1}{10}, \quad 0 \leq t \leq 10$$

$$E(T) = 5, \quad \text{Var}(T) = \frac{100}{12} \approx 8.33, \quad \sigma \approx 2.89$$

The expected waiting time is 5 minutes (half the inter-arrival time), with standard deviation about 2.9 minutes. The probability of waiting more than 7 minutes is:

$$P(T > 7) = 1 - F(7) = 1 - \frac{7}{10} = 0.30$$

— exactly the fraction of the interval $[0, 10]$ that lies above 7.

3.7.2 The Exponential distribution — $\text{Exp}(\lambda)$

The **Exponential distribution** models the waiting time between successive events of a Poisson process. If events occur at rate λ per unit time, the time X until the next event follows $\text{Exp}(\lambda)$. This is the continuous counterpart of the Poisson: where the Poisson counts events, the exponential measures the time between them.

Typical applications include: the time between customer arrivals at a service counter, the time until a machine fails, the duration of a phone call, the time until the next insurance claim.

PDF, CDF, mean, and variance

$$f(x) = \lambda * e^{-\lambda * x}, \quad x \geq 0; \quad 0 \text{ (otherwise)}$$

$$F(x) = 1 - e^{-\lambda * x}, \quad x \geq 0$$

$$E(X) = \frac{1}{\lambda}, \quad \text{Var}(X) = \frac{1}{\lambda^2}, \quad \sigma = \frac{1}{\lambda}$$

Notice the striking equality $\sigma = E(X)$: the standard deviation of an exponential equals its mean. This comes from the heavily skewed shape of the distribution: a few very long waits pull the mean up to match the spread.

The memoryless property

The exponential distribution has a property that no other continuous distribution possesses:

$$P(X > s + t \mid X > s) = P(X > t)$$

Verbally: given that you have already waited s units of time, the distribution of the remaining waiting time is the same as if you had just started waiting. "The exponential has no memory of the past."

This property characterises the exponential among all continuous distributions and makes it the natural choice whenever the waiting time is governed by a constant failure rate, as in reliability engineering, where a component's lifetime is not affected by how long it has already survived (provided the Poisson assumptions hold).

Connection with the Poisson distribution

If events occur according to a Poisson process at rate λ , then:

- The number of events in an interval of length T follows $P(\lambda T)$.
- The time between consecutive events follows $\text{Exp}(\lambda)$.
- The time until the n -th event follows the Gamma distribution (not covered here).

This connection explains why the same rate parameter λ appears in both distributions: λ is the fundamental rate of the underlying random process.

Example 3.12 — Time between customer arrivals

Customers arrive at a service counter at an average rate of 12 per hour (so on average one every 5 minutes). Let T be the time in minutes between two consecutive arrivals. Then $T \sim \text{Exp}(\lambda)$ with $\lambda = 1/5 = 0.2$ per minute.

$$E(T) = \frac{1}{0.2} = 5 \text{ (minutes) }, \quad \sigma = 5 \text{ (minutes) }$$

The probability that the waiting time exceeds 10 minutes:

$$P(T > 10) = e^{-0.2 * 10} = e^{-2} \approx 0.1353$$

About 13.5% of inter-arrival intervals last more than 10 minutes, useful for queue-management planning.

Example 3.13 — Machine lifetime (reliability)

A production machine has an average lifetime of 800 hours, with failures following an exponential distribution. What is the probability that the machine lasts more than 1 000 hours?

$\lambda = 1/800$ per hour. Then $T \sim \text{Exp}(1/800)$, and:

$$P(T > 1000) = e^{-1000 // 800} = e^{-1.25} \approx 0.2865$$

There is about a 29% chance of surviving beyond 1 000 hours. Using the memoryless property, if the machine has already survived 500 hours, the conditional probability it survives another 500 hours is still $e^{-(500/800)} = e^{-0.625} \approx 0.535$, the same as for a new machine.

Visual Summary — Mind Map of §3.7.2

The diagram below summarises the exponential distribution, its memoryless property, and its connection with the Poisson process.

Figure 3.4: The Exponential distribution $\text{Exp}(\lambda)$: waiting time in a Poisson process**Mind Map 3.4 – Exponential Distribution $\text{Exp}(\lambda)$** **$X \sim \text{Exp}(\lambda)$**

Continuous – Waiting time until next Poisson event

Probability Density Function

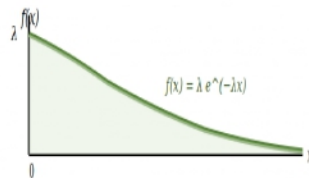
$$f(x) = \lambda \cdot e^{-\lambda x}, x \geq 0$$

Decreasing exponential; $f(0) = \lambda$ $f(x) = 0$ for $x < 0$ $\lambda > 0$ is the rate parameter**Cumulative Distribution Function**

$$F(x) = 1 - e^{-\lambda x}, x \geq 0$$

 $P(X \leq x) = 1 - e^{-\lambda x}$ $P(X > x) = e^{-\lambda x}$ (survival)

Used for "probability of waiting MORE than x"

Shape of the density – exponential decay**Mean (Expected value)**

$$E(X) = 1 / \lambda$$

Average waiting time = reciprocal of rate

Variance & Standard Deviation

$$\text{Var}(X) = 1 / \lambda^2, \sigma = 1 / \lambda$$

Standard deviation equals the mean!

♦ Memoryless property – unique to the Exponential

$$P(X > s + t \mid X > s) = P(X > t)$$

"The fact that you have already waited s minutes does not affect the distribution of the remaining waiting time."

This property characterises the Exponential – it is the ONLY continuous distribution with it.

♦ Link to Poisson distributionIf events follow $\text{Poisson}(\lambda T)$ in time T ,
then inter-arrival times follow $\text{Exp}(\lambda)$.Number of events (discrete) $\leftrightarrow$ Time between events (continuous)Poisson $\leftrightarrow$ Exponential – two sides of same coin**Economic Applications**

- Customer inter-arrival time at a counter
- Time until a machine fails (reliability)
- Duration of a phone call, service time
- Time between insurance claims
- Lifetime of an electronic component

Source: Elaborated by the Author

3.7.3 The Normal distribution — $N(\mu, \sigma^2)$

The Normal distribution, also called the Gaussian distribution or, informally, "the bell curve", is the most important distribution in all of statistics and applied probability. Its importance stems from the Central Limit Theorem (CLT): sums and averages of many independent random quantities tend to be approximately normal, regardless of the underlying distribution. As a result, the normal appears whenever a quantity is the sum of many small, independent contributions: measurement errors, aggregate demand, stock returns, physical characteristics (heights, weights), and countless others.

A continuous random variable X is Normal with mean μ and variance σ^2 (written $X \sim N(\mu, \sigma^2)$) if its density is:

$$f(x) = \frac{1}{\sigma * \sqrt{2 * \pi}} * e^{-\frac{(x - \mu)^2}{2 * \sigma^2}}$$

The density is symmetric around the mean μ , with the characteristic bell shape, and extends (with rapidly vanishing probability) to infinity in both directions.

Key properties

- **Location and scale:** μ determines the centre, σ determines the width.
- **Symmetry:** $f(\mu + x) = f(\mu - x)$; mean = median = mode = μ .
- **Moments:** $E(X) = \mu$, $\text{Var}(X) = \sigma^2$, $\sigma(X) = \sigma$.
- **Linear transformations:** if $X \sim N(\mu, \sigma^2)$, then $aX + b \sim N(a\mu + b, a^2\sigma^2)$.
- **Closure under sums:** if $X \sim N(\mu_1, \sigma_1^2)$ and $Y \sim N(\mu_2, \sigma_2^2)$ are independent, then $X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

The standard normal $N(0, 1)$

The special case $\mu = 0$, $\sigma = 1$ is called the standard normal distribution. Its variable is conventionally denoted Z :

$$Z \sim N(0, 1), \quad \varphi(z) = \frac{1}{\sqrt{2 * \pi}} * e^{-z^2 / 2}$$

The CDF of the standard normal is denoted $\Phi(z)$:

$$\Phi(z) = P(Z \leq z) = \int_{-\infty}^z \varphi(t) dt$$

This integral has no closed form, so values of $\Phi(z)$ are tabulated (see Appendix A, the standard normal table or "Z-table").

Standardisation

Any normal random variable $X \sim N(\mu, \sigma^2)$ can be converted to a standard normal by the linear transformation:

$$Z = \frac{X - \mu}{\sigma}$$

This is the Z-score introduced in Chapter 2, now applied in its most important context. Using standardisation, any probability statement about X reduces to a statement about Z :

$$P(X \leq a) = P\left(Z \leq \frac{a - \mu}{\sigma}\right) = \Phi\left(\frac{a - \mu}{\sigma}\right)$$

Values of $\Phi(z)$ are looked up in the Z-table. This reduction from "any normal" to "the standard normal" is what makes the entire normal family computationally tractable with a single table.

Reading the Z-table

The standard normal table (Appendix A) gives values of $\Phi(z)$ for z from 0.00 to 3.49 in steps of 0.01. To use the table:

- Locate the row corresponding to the first digit and first decimal of z (e.g. 1.2 for $z = 1.23$).
- Locate the column corresponding to the second decimal (0.03 for $z = 1.23$).
- The entry is $\Phi(z) = P(Z \leq z)$.
- For negative z , use $\Phi(-z) = 1 - \Phi(z)$ (symmetry of the distribution).

Common probabilities like $P(-1 \leq Z \leq 1) = 2\Phi(1) - 1 \approx 0.6827$ follow directly from the table and symmetry.

The 68-95-99.7 rule

For any normal distribution, the following empirical rule holds:

- About 68% of values lie within one standard deviation of the mean ($\mu \pm \sigma$).
- About 95% of values lie within two standard deviations ($\mu \pm 2\sigma$).
- About 99.7% of values lie within three standard deviations ($\mu \pm 3\sigma$).

This rule is useful for quick probability estimates without the Z-table and for identifying "unusual" observations: a value more than 2σ from the mean is atypical, and more than 3σ is very rare.

Central Limit Theorem (intuition)

The Central Limit Theorem states that if $X_1, X_2, \dots, X_n$ are independent, identically distributed random variables with mean μ and variance σ^2 , then for large n :

$$\frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \approx N(0, 1)$$

In other words, the sample average $\bar{X}$ is approximately normal with mean μ and variance σ^2/n , regardless of the distribution of the individual X_i . This theorem explains why the normal distribution is so ubiquitous in applied statistics: every time we aggregate many small random contributions, we obtain something approximately normal.

Example 3.14 — *Using the standard normal table*

Daily stock returns (in %) of a certain equity are modelled as $X \sim N(0.5, 2.5^2)$. Compute:

- (a) The probability of a positive return on a given day.
- (b) The probability that the return exceeds 3%.
- (c) The probability that the return is between -2% and $+2\%$.

Solution (a). We standardise:

$$\begin{aligned} P(X > 0) &= P\left(Z > \frac{0 - 0.5}{2.5}\right) = P(Z > -0.20) = 1 - \text{Phi}(-0.20) = \text{Phi}(0.20) \\ &\approx 0.5793 \end{aligned}$$

About 58% of trading days produce a positive return.

Solution (b).

$$\begin{aligned} P(X > 3) &= P\left(Z > \frac{3 - 0.5}{2.5}\right) = P(Z > 1.00) = 1 - \text{Phi}(1.00) \approx 1 - 0.8413 \\ &= 0.1587 \end{aligned}$$

About 16% of days see returns exceeding 3%.

Solution (c).

$$\begin{aligned} P(-2 \leq X \leq 2) &= \text{Phi}\left(\frac{2 - 0.5}{2.5}\right) - \text{Phi}\left(\frac{-2 - 0.5}{2.5}\right) = \text{Phi}(0.60) - \text{Phi}(-1.00) \\ &= 0.7257 - 0.1587 = 0.5670 \end{aligned}$$

About 57% of days have returns in the "typical" range $[-2, +2]$.

Example 3.15 — *Quality control — acceptable product weight*

A manufacturing process produces items with weight $X \sim N(500, 4^2)$ grams. The specification requires weight between 492 and 508 grams.

What fraction of items meets specification?

$$\begin{aligned} P(492 \leq X \leq 508) &= \text{Phi}\left(\frac{508 - 500}{4}\right) - \text{Phi}\left(\frac{492 - 500}{4}\right) = \text{Phi}(2) - \text{Phi}(-2) \\ &= 0.9772 - 0.0228 = 0.9544 \end{aligned}$$

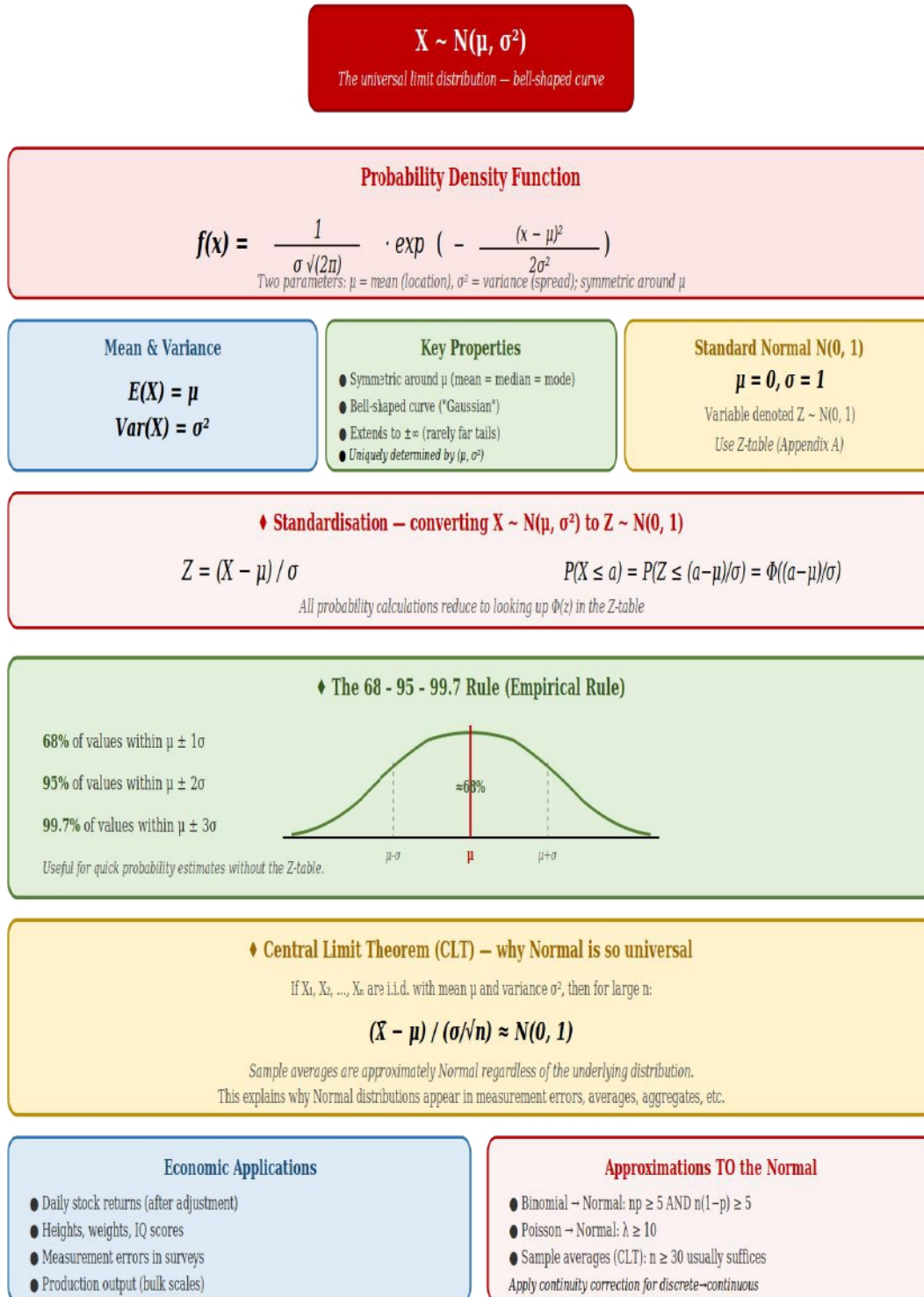
About 95.4% of items meet specification, consistent with the 68-95-99.7 rule, since the specification limits are at $\pm 2\sigma$.

Visual Summary — Mind Map of §3.7.3

The diagram below summarises the normal distribution, the standardisation formula, the empirical rule, and the Central Limit Theorem.

Figure 3.5: The Normal distribution $N(\mu, \sigma^2)$: the universal bell-shaped curve

Mind Map 3.5 – Normal Distribution $N(\mu, \sigma^2)$



Source: Elaborated by the Author

3.8 Approximations Between Distributions

Many of the distributions introduced in this chapter are related by approximations: under specific conditions, one distribution can be replaced by another, simpler one. These approximations were once indispensable, back when computation by hand was the norm, and they remain useful today for understanding the relationships between distributions and for quick estimates.

3.8.1 Binomial $\rightarrow$ Poisson

If $X \sim B(n, p)$ with n large and p small (conventionally $n \geq 20$ and $p \leq 0.05$), with $\lambda = np$ moderate (say $\lambda \leq 10$), then:

$$B(n, p) \approx P(np)$$

This approximation simplifies binomial computations when n is large: the binomial formula involves factorials that become unwieldy, while the Poisson formula uses only exponentials.

3.8.2 Binomial $\rightarrow$ Normal

If $X \sim B(n, p)$ with $np \geq 5$ and $n(1 - p) \geq 5$ (de Moivre–Laplace conditions), then:

$$B(n, p) \approx N(np, n * p * (1 - p))$$

To compute a probability $P(X \leq k)$, standardise using $\mu = np$ and $\sigma^2 = np(1 - p)$, then use the Z-table. For better accuracy, apply a continuity correction of ± 0.5 (since X is discrete and the normal is continuous).

3.8.3 Poisson $\rightarrow$ Normal

If $X \sim P(\lambda)$ with λ large (conventionally $\lambda \geq 10$), then:

$$P(\lambda) \approx N(\lambda, \lambda)$$

Again, standardise and use the Z-table. Apply continuity correction for best accuracy.

3.8.4 Summary of conditions

From	To	Condition	Key formula
Binomial $B(n, p)$	Poisson $P(np)$	$n \geq 20, p \leq 0.05$	$\lambda = np$
Binomial $B(n, p)$	Normal $N(np, np(1-p))$	$np \geq 5$ AND $n(1-p) \geq 5$	Standardise Z
Poisson $P(\lambda)$	Normal $N(\lambda, \lambda)$	$\lambda \geq 10$	$Z = (X - \lambda) / \sqrt{\lambda}$
Hypergeometric $H(N, K, n)$	Binomial $B(n, K/N)$	$n/N \leq 0.05$	$p = K/N$

★ Continuity correction (for discrete $\rightarrow$ continuous)

- When approximating a discrete distribution (Binomial or Poisson) by the Normal, apply a ± 0.5 adjustment:
- $P(X \leq k) \approx \Phi((k + 0.5 - \mu)/\sigma)$
- $P(X \geq k) \approx 1 - \Phi((k - 0.5 - \mu)/\sigma)$
- $P(X = k) \approx \Phi((k + 0.5 - \mu)/\sigma) - \Phi((k - 0.5 - \mu)/\sigma)$

3.9 Worked Examples in an Economic Context

This section brings together the seven distributions through a series of integrated economic examples. The critical skill tested here is identifying the correct distributional family from a verbal description. It is the hardest part of the chapter and the one that carries the most exam marks.

Example 3.16 Bank loan portfolio — binomial risk assessment

A regional bank holds a portfolio of 80 loans, each with an independent default probability of 3% per year. Let X be the number of defaults in the next twelve months.

(a) Identify the distribution. (b) Compute $E(X)$, $\sigma(X)$. (c) Find $P(X \leq 2)$. (d) Find $P(X \geq 5)$.

Solution (a). Fixed number of trials (80 loans), constant success probability (0.03), independence. Therefore $X \sim B(80, 0.03)$.

Solution (b).

$$E(X) = 80 * 0.03 = 2.4, \quad \sigma = \sqrt{80 * 0.03 * 0.97} \approx 1.526$$

Solution (c). Using the binomial PMF:

$$P(X \leq 2) = \sum_{k=0}^2 \binom{80}{k} * 0.03^k * 0.97^{80-k} \approx 0.570$$

Solution (d). Using the complement:

$$P(X \geq 5) = 1 - P(X \leq 4) \approx 1 - 0.907 = 0.093$$

There is about a 9% chance of 5 or more defaults, a quantity useful for capital planning.

Example 3.17 Quality audit — hypergeometric sampling

A shipment contains 40 items, of which 8 are known (from a previous test) to be defective. An auditor selects 6 items at random for inspection. Let X be the number of defectives found.

(a) Why is X hypergeometric, not binomial? (b) Compute $P(X = 2)$.

Solution (a). The sampling is without replacement from a small population ($n/N = 6/40 = 15\% > 5\%$), so each draw affects subsequent probabilities. Hence $X \sim H(40, 8, 6)$.

Solution (b).

$$P(X = 2) = \frac{\binom{8}{2} * \binom{32}{4}}{\binom{40}{6}} = \frac{28 * 35960}{3838380} \approx 0.262$$

About 26% of such audits will find exactly 2 defectives.

Example 3.18 Customer arrivals — Poisson modelling

A call centre receives an average of 4 calls per minute during peak hours. Assume a Poisson process.

- (a) Probability of receiving exactly 5 calls in a given minute.
- (b) Probability of receiving fewer than 3 calls in a given minute.
- (c) Expected number of calls in 10 minutes.

$X \sim P(4)$.

$$P(X = 5) = \frac{e^{-4} * 4^5}{5!} \approx 0.1563$$

$$P(X < 3) = P(X = 0) + P(X = 1) + P(X = 2) \approx 0.0183 + 0.0733 + 0.1465 \\ = 0.2381$$

For 10 minutes, by additivity of Poisson rates, $Y \sim P(40)$ with $E(Y) = 40$.

Example 3.19 Waiting time — exponential distribution

Continuing Example 3.18 (peak-hour arrivals at rate $\lambda = 4$ per minute), let T be the time in minutes between two consecutive calls. Then $T \sim \text{Exp}(4)$.

$$E(T) = \frac{1}{4} = 0.25 \text{ (minutes)} = 15 \text{ (seconds)}$$

$$P(T > 1) = e^{-4} \approx 0.0183$$

Only 1.8% of inter-call gaps exceed one minute, consistent with the high arrival rate.

Example 3.20 Exam scores — normal modelling

Final-exam scores in a course follow approximately $N(12, 2.5^2)$ on a 20-point scale.

- (a) Fraction of students scoring above 15.
- (b) Fraction scoring between 10 and 14.
- (c) Score exceeded by the top 10%.

Solution (a).

$$P(X > 15) = P(Z > \frac{15 - 12}{2.5}) = P(Z > 1.2) = 1 - \Phi(1.2) \approx 1 - 0.8849 \\ = 0.1151$$

Solution (b).

$$P(10 \leq X \leq 14) = \Phi(0.8) - \Phi(-0.8) = 0.7881 - 0.2119 = 0.5762$$

Solution (c). We want x_0 with $P(X > x_0) = 0.10$, i.e. $\Phi((x_0 - 12)/2.5) = 0.90$. From the Z-table, $\Phi(1.28) \approx 0.90$, so:

$$x_{0.10} = 12 + 1.28 * 2.5 = 15.2$$

The top 10% of students score above 15.2.

Example 3.21 Insurance claims — Poisson-to-Normal approximation

A company receives on average 36 insurance claims per day. Using a normal approximation ($\lambda = 36 \geq 10$), estimate the probability that a given day brings more than 45 claims.

$X \sim P(36) \approx N(36, 36)$. With continuity correction:

$$P(X > 45) \approx P(Z > \frac{45.5 - 36}{6}) = P(Z > 1.583) \approx 1 - 0.943 = 0.057$$

About 5.7% of days produce more than 45 claims.

Example 3.22 Marketing campaign — Binomial-to-Normal approximation

An email campaign is sent to 1 000 customers, each with a 20% probability of responding. Let X be the number of responses.

$X \sim B(1000, 0.20)$. Since $np = 200$ and $n(1 - p) = 800$ are both ≥ 5 , use the Normal approximation with $\mu = 200$, $\sigma^2 = 160$, $\sigma \approx 12.65$.

$$P(X \geq 220) \approx P(Z \geq \frac{219.5 - 200}{12.65}) = P(Z \geq 1.542) \approx 1 - 0.938 = 0.062$$

There is about a 6% chance that the campaign yields 220 or more responses.

Example 3.23 Uniform distribution — scheduling

A delivery truck is scheduled to arrive at a warehouse sometime between 9:00 and 11:00, with all times in that interval equally likely. Let T be the arrival time in minutes after 9:00, so $T \sim U(0, 120)$.

$$E(T) = 60 \text{ (minutes)}, \quad Var(T) = \frac{120^2}{12} = 1200, \quad \sigma \approx 34.64 \text{ (minutes)}$$

$$P(T > 90) = \frac{120 - 90}{120} = 0.25$$

A 25% probability of arriving after 10:30, useful for the warehouse to plan receiving resources.

Example 3.24 Combined use — CLT for portfolio returns

Daily returns X of a portfolio have mean $\mu = 0.08\%$ and standard deviation $\sigma = 1.5\%$, with an unknown distribution. By the CLT, the average return over $n = 60$ trading days satisfies:

$$\bar{X} \approx N(0.08, \frac{1.5^2}{60}) = N(0.08, 0.0375)$$

$$\sigma_{\bar{X}} = \sqrt{0.0375} \approx 0.194\%$$

The probability that the average daily return over 60 days exceeds 0.30%:

$$P(\bar{X} > 0.30) = P\left(Z > \frac{0.30 - 0.08}{0.194}\right) = P(Z > 1.134) \approx 1 - 0.872 = 0.128$$

About 13%. Even though individual daily returns are highly volatile, averages become much more concentrated thanks to the CLT.

Example 3.25 Identifying the right distribution — diagnostic quiz

For each situation below, identify the appropriate distribution and its parameters.

- **(a)** A surveyor contacts 200 households; each has a 40% chance of responding. Count the number of responses. $\rightarrow B(200, 0.40)$.
- **(b)** A restaurant receives customers at a constant average rate of 8 per hour. Count customers in a given hour. $\rightarrow P(8)$.
- **(c)** Waiting time for the next bus if buses pass on average every 12 minutes (Poisson process). $\rightarrow \text{Exp}(1/12)$.
- **(d)** From a box of 25 items (of which 6 are defective), 4 are drawn at random for testing. Count defectives. $\rightarrow H(25, 6, 4)$.
- **(e)** Weight of a randomly selected cereal box, with specified mean 500 g and standard deviation 5 g. $\rightarrow N(500, 25)$.
- **(f)** A single consumer either buys (with probability 0.15) or doesn't buy. $\rightarrow \text{Bernoulli}(0.15)$.
- **(g)** Random arrival time within a known 30-minute service window. $\rightarrow U(0, 30)$.

Work through each case until you can justify the choice in one sentence. This is the single most tested skill in the final exam.

3.10 Common Mistakes — Watch Out

The errors below appear in almost every cohort. Read them carefully before attempting the exercises.

⚠ Mistake 1 — Using Binomial when sampling is WITHOUT replacement

- If items are drawn from a finite population without replacement and $n/N > 5\%$, the Hypergeometric is the correct model.
- Using Binomial in this case overstates the variance.
- When in doubt, ask: "Does the probability of success change after each draw?" If yes → Hypergeometric.

⚠ Mistake 2 — Confusing Binomial and Poisson

- Binomial: fixed number of TRIALS n , each with probability p .
- Poisson: count of EVENTS in a fixed time/space interval, at a rate λ .
- The Poisson has no " n " and no upper bound: the count can in principle be arbitrarily large.

⚠ Mistake 3 — Misreading the Z-table

- The table gives $\Phi(z) = P(Z \leq z)$, the probability to the LEFT.
- For $P(Z > z)$, use $1 - \Phi(z)$.
- For $P(Z \leq -z) = 1 - \Phi(z)$ (symmetry).
- Double-check: $\Phi(0) = 0.5$, $\Phi(1) \approx 0.8413$, $\Phi(-1) \approx 0.1587$.

⚠ Mistake 4 — Forgetting to standardise before using the Z-table

- The Z-table works only for the STANDARD normal $N(0, 1)$.
- For $X \sim N(\mu, \sigma^2)$, ALWAYS compute $Z = (X - \mu)/\sigma$ first.
- Never use $P(X \leq a) = \Phi(a)$ directly; use only $P(Z \leq z) = \Phi(z)$.

⚠ Mistake 5 — Forgetting continuity correction

- When approximating a discrete distribution (Binomial or Poisson) by the Normal, apply ± 0.5 .
- $P(X \leq k) \approx \Phi((k + 0.5 - \mu)/\sigma)$, not $\Phi((k - \mu)/\sigma)$.
- Without the correction, approximation errors can be noticeable for small σ .

⚠ Mistake 6 — Mean $\neq$ Variance for Poisson (they ARE equal)

- For Poisson $P(\lambda)$: $E(X) = \text{Var}(X) = \lambda$. This is a feature, not a bug.
- If in an exercise $E(X) = \text{Var}(X)$ is given, Poisson is a natural candidate.
- For other distributions, mean and variance differ.

⚠ Mistake 7 — Memoryless property applies ONLY to Exponential

- The memoryless property $P(X > s + t \mid X > s) = P(X > t)$ is UNIQUE to the exponential distribution.
- Do not apply it to other waiting-time models (e.g. the Gamma distribution does not have it).
- Key intuition: exponential models "no aging." The distribution resets after any waited time.

⚠ Mistake 8 — Confusing λ in Poisson and Exponential

- In Poisson $P(\lambda)$, λ is the MEAN number of events per unit time.
- In Exponential $\text{Exp}(\lambda)$, λ is the RATE (events per unit time), so the mean is $1/\lambda$.
- Same symbol, different meaning: always check whether it is a count or a time.

⚠ Mistake 9 — Using approximations outside their conditions

- Binomial $\rightarrow$ Poisson: requires $n \geq 20$ and $p \leq 0.05$.
- Binomial $\rightarrow$ Normal: requires $np \geq 5$ AND $n(1-p) \geq 5$.
- Poisson $\rightarrow$ Normal: requires $\lambda \geq 10$.
- Outside these conditions, the approximation errors can be significant.

⚠ Mistake 10 — Density is not probability (Normal and other continuous)

- $f(x)$ for a continuous distribution can exceed 1 without violating any rule.
- Probabilities are areas under the density, not values of the density.
- $P(X = x) = 0$ for any continuous X . Never report $P(X = 3) = 0.15$ for a continuous case.

3.11 Key Takeaways

The most important ideas of this chapter, condensed.

- ✓ Seven distributions cover the vast majority of applied probability problems: four discrete (Bernoulli, Binomial, Hypergeometric, Poisson) and three continuous (Uniform, Exponential, Normal).
- ✓ Bernoulli $B(p)$: single yes/no trial, $E(X) = p$, $\text{Var}(X) = p(1-p)$.
- ✓ Binomial $B(n, p)$: count of successes in n independent trials with constant p ; $E(X) = np$, $\text{Var}(X) = np(1-p)$.
- ✓ Hypergeometric $H(N, K, n)$: count of successes in n draws without replacement from N ; $E(X) = nK/N$, variance includes finite population correction.
- ✓ Poisson $P(\lambda)$: count of rare events at constant rate; $E(X) = \text{Var}(X) = \lambda$.
- ✓ Uniform $U(a, b)$: equally likely values on an interval; $E(X) = (a+b)/2$, $\text{Var}(X) = (b-a)^2/12$.
- ✓ Exponential $\text{Exp}(\lambda)$: waiting time in a Poisson process; $E(X) = 1/\lambda$, $\text{Var}(X) = 1/\lambda^2$; memoryless.
- ✓ Normal $N(\mu, \sigma^2)$: bell-shaped; symmetric around μ ; uses Z-table after standardisation $Z = (X-\mu)/\sigma$.
- ✓ Central Limit Theorem: sample averages are approximately normal for large n , which explains the ubiquity of the normal distribution.
- ✓ 68-95-99.7 rule: 68% of normal values are within $\mu \pm \sigma$, 95% within $\mu \pm 2\sigma$, 99.7% within $\mu \pm 3\sigma$.
- ✓ Binomial $\rightarrow$ Poisson when n is large and p is small ($np = \lambda$, $n \geq 20$, $p \leq 0.05$).
- ✓ Binomial $\rightarrow$ Normal when $np \geq 5$ AND $n(1-p) \geq 5$ (apply continuity correction).
- ✓ Poisson $\rightarrow$ Normal when $\lambda \geq 10$ (apply continuity correction).
- ✓ Hypergeometric $\rightarrow$ Binomial when $n/N \leq 5\%$ (replacement becomes irrelevant).
- ✓ Poisson $\leftrightarrow$ Exponential: two sides of a single Poisson process, events (Poisson) vs waiting times (Exponential).
- ✓ The hardest exam skill is IDENTIFYING which distribution applies from a verbal description. Practice this systematically.

3.12 Chapter Summary Cards

The cards below condense the chapter into four visual summaries, designed for quick revision before an assessment.

Card 1 — Formula Reference Sheet

✓ ALL SEVEN DISTRIBUTIONS — KEY FORMULAS

Bernoulli $B(p)$: $P(X=1)=p$, $P(X=0)=1-p$; $E=p$, $Var=p(1-p)$.

Binomial $B(n, p)$: $P(X=k)=C(n,k) p^k (1-p)^{(n-k)}$; $E=np$, $Var=np(1-p)$.

Hypergeometric $H(N,K,n)$: $P(X=k)=C(K,k)C(N-K,n-k)/C(N,n)$; $E=nK/N$.

Poisson $P(\lambda)$: $P(X=k)=e^{(-\lambda)} \lambda^k / k!$; $E=Var=\lambda$.

Uniform $U(a, b)$: $f(x)=1/(b-a)$; $E=(a+b)/2$, $Var=(b-a)^2/12$.

Exponential $Exp(\lambda)$: $f(x)=\lambda e^{(-\lambda x)}$, $F(x)=1-e^{(-\lambda x)}$; $E=1/\lambda$, $Var=1/\lambda^2$.

Normal $N(\mu, \sigma^2)$: Standardise $Z=(X-\mu)/\sigma$; use Z-table for $\Phi(z)$ values.

Approximations: Bin $\rightarrow$ Poisson ($n \geq 20$, $p \leq 0.05$); Bin $\rightarrow$ Normal ($np \geq 5$, $n(1-p) \geq 5$);
Poisson $\rightarrow$ Normal ($\lambda \geq 10$).

Card 2 — Decision Guide: Which Distribution to Use?

△ STEP-BY-STEP IDENTIFICATION

Q1. Is the variable discrete or continuous?

Discrete ↓ → proceed to Q2

Continuous ↓ → proceed to Q5

Q2. A single trial with yes/no outcome? → *Bernoulli* $B(p)$

Q3. n independent trials, constant p (with replacement)? → *Binomial* $B(n, p)$

Q4a. Sampling without replacement from finite N? → *Hypergeometric* $H(N, K, n)$

Q4b. Count of rare events at constant rate? → *Poisson* $P(\lambda)$

Q5. Equally likely values on [a, b]? → *Uniform* $U(a, b)$

Q6. Waiting time or time between events? → *Exponential* $Exp(\lambda)$

Q7. Bell-shaped, symmetric, result of many aggregated causes? → *Normal* $N(\mu, \sigma^2)$

Card 3 — Mean and Variance Quick Reference

Distribution	Mean $E(X)$	Variance $Var(X)$	Range of X
Bernoulli $B(p)$	p	$p(1-p)$	$\{0, 1\}$
Binomial $B(n, p)$	np	$np(1-p)$	$\{0, 1, \dots, n\}$
Hypergeometric $H(N, K, n)$	$n \cdot K/N$	$n(K/N)(1-K/N)(N-n)/(N-1)$	constrained
Poisson $P(\lambda)$	λ	λ	$\{0, 1, 2, \dots\}$
Uniform $U(a, b)$	$(a+b)/2$	$(b-a)^2/12$	$[a, b]$
Exponential $Exp(\lambda)$	$1/\lambda$	$1/\lambda^2$	$[0, +\infty)$
Normal $N(\mu, \sigma^2)$	μ	σ^2	$(-\infty, +\infty)$

Memorise this table cold. It saves time on every exam question.

Card 4 — Typical Exam Problems

△ TYPICAL EXAM PROBLEMS

- Q1.** Given a verbal description, identify the distribution and its parameters.
- Q2.** Compute $P(X = k)$ or $P(X \leq k)$ from a Binomial or Poisson formula.
- Q3.** Use the Z-table to find $P(X \leq a)$ for $X \sim N(\mu, \sigma^2)$.
- Q4.** Apply a Binomial $\rightarrow$ Poisson or Poisson $\rightarrow$ Normal approximation.
- Q5.** Use Hypergeometric for a sampling-without-replacement problem.
- Q6.** Find a quantile: given $P(X \leq x_0) = q$, solve for x_0 using the Z-table.
- Q7.** Interpret a numerical result in the context of the original problem.

★ EXAM TIPS

- Identify the distribution FIRST before writing any formula. Half the battle is conceptual.
- State the parameters explicitly: for Binomial, write n and p ; for Normal, μ and σ .
- Use the complement rule whenever the problem asks "at least" or "more than".
- When approximating, clearly state the conditions you have verified.
- Always apply continuity correction when using Normal for discrete variables.
- Check your answer: probabilities in $[0, 1]$; posteriors sum to 1 over a partition.

3.13 Practice Exercises

The exercises below form a complete set for revision and exam preparation. They are ordered by increasing difficulty. Solutions are not provided; attempt every problem with the tools developed in this chapter.

3.13.1 Level 1 — Foundations

Exercise 3.1. A fair die is rolled 10 times. Let X be the number of sixes obtained. (a) Identify the distribution of X and its parameters. (b) Compute $E(X)$ and $\sigma(X)$. (c) Compute $P(X = 2)$ and $P(X \geq 3)$.

Exercise 3.2. A fair coin is tossed 100 times. Let Y be the number of heads. Using the normal approximation to the binomial (with continuity correction), estimate $P(Y \geq 55)$.

Exercise 3.3. A call centre receives an average of 12 calls per hour, following a Poisson process. (a) Probability of receiving exactly 10 calls in an hour. (b) Probability of receiving fewer than 8. (c) Expected number of calls in 8 hours.

Exercise 3.4. A continuous random variable $X \sim N(100, 225)$. Compute: (a) $P(X \leq 110)$; (b) $P(85 \leq X \leq 115)$; (c) the value x_0 such that $P(X > x_0) = 0.05$.

Exercise 3.5. A box contains 20 items, 6 of which are defective. Four items are drawn at random without replacement. (a) Identify the distribution. (b) Compute $P(\text{exactly 1 defective in the sample})$.

3.13.2 Level 2 — Applications

Exercise 3.6. A small bank issues 150 short-term loans. Each has a 4% default probability, independently. Let X be the number of defaults. (a) Compute the exact $P(X \leq 10)$ using the binomial. (b) Using Poisson approximation, estimate the same probability. (c) Compare the two results.

Exercise 3.7. Customers arrive at a café at a rate of 15 per hour. Let T be the time until the next customer (in minutes). (a) Identify the distribution of T . (b) Compute $P(T > 5)$. (c) Verify the memoryless property: $P(T > 10 \mid T > 5) = P(T > 5)$.

Exercise 3.8. Monthly incomes in a region follow approximately $N(50\,000, 10\,000^2)$ DZD. (a) What proportion of households earn above 65 000 DZD? (b) Find the income level exceeded by the top 20% of households.

Exercise 3.9. A lottery sells 1 000 tickets. Each ticket has a 0.2% chance of winning a prize, independently. Using the Poisson approximation, estimate the probability that the lottery awards exactly 2 prizes.

Exercise 3.10. A machine has a lifetime exponentially distributed with mean 2 000 hours. (a) What is the probability it survives more than 3 000 hours? (b) Given it has survived 2 000 hours, what is the probability it survives another 1 000?

3.13.3 Level 3 — Extension

Exercise 3.11. An insurance company receives an average of 50 claims per day. (a) Using the normal approximation, estimate the probability of receiving more than 65 claims on a given day. (b) Discuss whether the normal approximation is adequate here.

Exercise 3.12. An audit selects 15 accounts from a population of 80, of which 12 are known to contain errors. (a) Identify the distribution of X , the number of errors found. (b) Compute $E(X)$ and $\text{Var}(X)$. (c) Compute $P(X \geq 3)$.

Exercise 3.13. Let X and Y be independent with $X \sim N(10, 4)$ and $Y \sim N(15, 9)$. Let $W = 2X + Y$. (a) Find the distribution of W . (b) Compute $P(W > 40)$.

Exercise 3.14. A professor grades exams that have mean 11 and standard deviation 3 (on 20). In a class of 40 students, what is the approximate probability that the class average is above 12? (Hint: use the CLT.)

Exercise 3.15. Defects in a 10 m^2 panel appear at a Poisson rate of 0.3 per m^2 . (a) What is the probability that a 2 m^2 panel has no defects? (b) What is the probability that a 5 m^2 panel has exactly 2 defects?

3.13.4 Supplementary Exam-Style Problems

Exercise 3.16 — Mini-case: airline no-shows. An airline flight has 200 seats; 210 tickets are sold. Each passenger has a 95% probability of showing up, independently. Let X be the number of passengers who show up. (a) Identify the distribution. (b) Using the normal approximation, estimate $P(X > 200)$, the probability of overbooking.

Exercise 3.17 — Mini-case: insurance overflow. A company receives on average 25 claims per week (Poisson). The budget covers up to 35 claims per week. Estimate the probability of exceeding the budget in a given week.

Exercise 3.18 — Mini-case: production scheduling. Customer arrival times at a store are uniformly distributed over a 4-hour window. Three customers are expected in the window. Compute the probability that all three arrive in the first hour. (Hint: arrivals are independent uniformly distributed.)

Exercise 3.19 — Quality control. A lot of 50 items contains 5 defectives. An inspector samples 8 items without replacement. (a) Compute $P(\text{exactly 1 defective in the sample})$. (b) Compare with the binomial approximation using $p = 0.10$; comment on the accuracy.

Exercise 3.20 — Portfolio duration. An investment has an exponentially distributed duration with mean 5 years until redemption. (a) Probability it is redeemed within the first 3 years. (b) Probability it survives at least 7 years.

3.13.5 Answers to Prerequisites Diagnostic (§3.3)

(a) $C(6, 2) = 15$; $C(10, 3) = 120$.

(b) $(a + b)^3 = a^3 + 3a^2b + 3ab^2 + b^3$.

(c) $\int \text{from } 0 \text{ to } \infty \text{ of } e^{(-2x)} dx = 1/2$.

(d) $Z = (13 - 10) / \sqrt{4} = 1.5$.

(e) $P(X = 2) = C(3, 2) \times 0.5^2 \times 0.5 = 3 \times 0.125 = 0.375$.

(f) $f(x) = e^{(-x)} \geq 0$ and $\int \text{from } 0 \text{ to } \infty \text{ of } e^{(-x)} dx = 1$, so it is a valid PDF. $E(X) = \int \text{from } 0 \text{ to } \infty \text{ of } x \cdot e^{(-x)} dx = 1$.

3.14 References

The references below are the primary academic sources consulted for the preparation of this chapter.

1. Blitzstein, J. K., & Hwang, J. (2019). *Introduction to probability* (2nd ed.). Chapman & Hall/CRC. <https://doi.org/10.1201/9780429428357>
2. Camm, J. D., Cochran, J. J., Fry, M. J., Ohlmann, J. W., Anderson, D. R., Sweeney, D. J., & Williams, T. A. (2024). *Statistics for business and economics* (15th ed.). Cengage Learning.
3. Chan, S. H. (2021). *Introduction to probability for data science*. Michigan Publishing. <https://doi.org/10.3998/mpub.12387745>
4. Devore, J. L., Berk, K. N., & Carlton, M. A. (2021). *Modern mathematical statistics with applications* (3rd ed.). Springer. <https://doi.org/10.1007/978-3-030-55156-8>
5. Hansen, B. E. (2022). *Probability and statistics for economists*. Princeton University Press.
6. Hurlin, C., & Mignon, V. (2022). *Statistique et probabilités en économie-gestion* (2e éd.). Dunod.
7. Lecoutre, J.-P. (2023). *Statistique et probabilités : Cours et exercices corrigés* (7e éd.). Dunod.
8. Linde, W. (2024). *Probability theory: A first course in probability theory and statistics* (2nd ed.). De Gruyter. <https://doi.org/10.1515/9783111325064>
9. Ross, S. M. (2024). *A first course in probability* (10th ed., Global Edition). Pearson Education.

Appendix A — Standard Normal Distribution Table (Z-table)

Values of $\Phi(z) = P(Z \leq z)$ for the standard normal distribution $N(0, 1)$. The row gives the first decimal of z ; the column gives the second decimal. For negative z , use $\Phi(-z) = 1 - \Phi(z)$.

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633

Appendix A (continued) — Z-table for $z = 1.8$ to 3.4

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

★ How to use the Z-table

- For $z \geq 0$, read $\Phi(z)$ directly from the table.
- For $z < 0$, use $\Phi(-z) = 1 - \Phi(z)$. Example: $\Phi(-1.23) = 1 - \Phi(1.23) = 1 - 0.8907 = 0.1093$.
- For $P(a \leq Z \leq b)$, compute $\Phi(b) - \Phi(a)$.
- For $P(Z > z)$, compute $1 - \Phi(z)$.
- For $|z| > 3.49$, use $\Phi(z) \approx 1$ (upper tail negligible) or $\Phi(z) \approx 0$ (lower tail negligible).

Appendix B — Selected Binomial Probabilities

Values of $P(X = k)$ for $X \sim B(n, p)$, for selected values of n and p . For other values, use the formula $P(X = k) = C(n, k) p^k (1-p)^{(n-k)}$ directly.

$n = 10$

k	p=0.05	p=0.10	p=0.20	p=0.30	p=0.50	p=0.70	p=0.90
0	0.5987	0.3487	0.1074	0.0282	0.0010	<.0001	<.0001
1	0.3151	0.3874	0.2684	0.1211	0.0098	0.0001	<.0001
2	0.0746	0.1937	0.3020	0.2335	0.0439	0.0014	<.0001
3	0.0105	0.0574	0.2013	0.2668	0.1172	0.0090	<.0001
4	0.0010	0.0112	0.0881	0.2001	0.2051	0.0368	0.0001
5	<.0001	0.0015	0.0264	0.1029	0.2461	0.1029	0.0015
6	<.0001	0.0001	0.0055	0.0368	0.2051	0.2001	0.0112
7	<.0001	<.0001	0.0008	0.0090	0.1172	0.2668	0.0574
8	<.0001	<.0001	<.0001	0.0014	0.0439	0.2335	0.1937
9	<.0001	<.0001	<.0001	0.0001	0.0098	0.1211	0.3874
10	<.0001	<.0001	<.0001	<.0001	0.0010	0.0282	0.3487

$n = 20$

k	p=0.05	p=0.10	p=0.20	p=0.30	p=0.50
0	0.3585	0.1216	0.0115	0.0008	<.0001
1	0.3774	0.2702	0.0576	0.0068	<.0001
2	0.1887	0.2852	0.1369	0.0278	0.0002
3	0.0596	0.1901	0.2054	0.0716	0.0011
4	0.0133	0.0898	0.2182	0.1304	0.0046
5	0.0022	0.0319	0.1746	0.1789	0.0148
6	0.0003	0.0089	0.1091	0.1916	0.0370
7	<.0001	0.0020	0.0545	0.1643	0.0739
8	<.0001	0.0004	0.0222	0.1144	0.1201
9	<.0001	<.0001	0.0074	0.0654	0.1602
10	<.0001	<.0001	0.0020	0.0308	0.1762
11	<.0001	<.0001	0.0005	0.0120	0.1602
12	<.0001	<.0001	<.0001	0.0039	0.1201
13	<.0001	<.0001	<.0001	0.0010	0.0739
14	<.0001	<.0001	<.0001	0.0002	0.0370
15	<.0001	<.0001	<.0001	<.0001	0.0148
16	<.0001	<.0001	<.0001	<.0001	0.0046
17	<.0001	<.0001	<.0001	<.0001	0.0011
18	<.0001	<.0001	<.0001	<.0001	0.0002
19	<.0001	<.0001	<.0001	<.0001	<.0001
20	<.0001	<.0001	<.0001	<.0001	<.0001

For cumulative probabilities $P(X \leq k)$, sum the individual $P(X = j)$ for $j = 0$ to k .

Appendix C — Selected Poisson Probabilities

Values of $P(X = k)$ for $X \sim P(\lambda)$, for selected values of λ and k .

k	$\lambda=0.5$	$\lambda=1.0$	$\lambda=2.0$	$\lambda=3.0$	$\lambda=4.0$	$\lambda=5.0$	$\lambda=7.0$	$\lambda=10.0$
0	0.6065	0.3679	0.1353	0.0498	0.0183	0.0067	0.0009	<.0001
1	0.3033	0.3679	0.2707	0.1494	0.0733	0.0337	0.0064	0.0005
2	0.0758	0.1839	0.2707	0.2240	0.1465	0.0842	0.0223	0.0023
3	0.0126	0.0613	0.1804	0.2240	0.1954	0.1404	0.0521	0.0076
4	0.0016	0.0153	0.0902	0.1680	0.1954	0.1755	0.0912	0.0189
5	0.0002	0.0031	0.0361	0.1008	0.1563	0.1755	0.1277	0.0378
6	<.0001	0.0005	0.0120	0.0504	0.1042	0.1462	0.1490	0.0631
7	<.0001	<.0001	0.0034	0.0216	0.0595	0.1044	0.1490	0.0901
8	<.0001	<.0001	0.0009	0.0081	0.0298	0.0653	0.1304	0.1126
9	<.0001	<.0001	0.0002	0.0027	0.0132	0.0363	0.1014	0.1251
10	<.0001	<.0001	<.0001	0.0008	0.0053	0.0181	0.0710	0.1251
11	<.0001	<.0001	<.0001	0.0002	0.0019	0.0082	0.0452	0.1137
12	<.0001	<.0001	<.0001	<.0001	0.0006	0.0034	0.0263	0.0948
13	<.0001	<.0001	<.0001	<.0001	0.0002	0.0013	0.0142	0.0729
14	<.0001	<.0001	<.0001	<.0001	<.0001	0.0005	0.0071	0.0521
15	<.0001	<.0001	<.0001	<.0001	<.0001	0.0002	0.0033	0.0347
16	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.0014	0.0217
17	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.0006	0.0128
18	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.0002	0.0071
19	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.0037
20	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.0019

For cumulative probabilities $P(X \leq k)$, sum the values up to and including k . For values of λ not shown, either interpolate or compute directly using the formula $P(X = k) = e^{(-\lambda)} \lambda^k / k!$.

★ Using the Poisson table

- The table gives $P(X = k)$ directly, not the CDF.
- For $P(X \leq k)$, sum entries from 0 to k .
- For $P(X \geq k)$, compute $1 - P(X \leq k - 1)$.
- For $\lambda \geq 10$, the normal approximation $N(\lambda, \lambda)$ (with continuity correction) is usually faster and accurate enough.

End of Chapter 3

You have now mastered the seven most important probability distributions.

These distributions, together with the techniques of Chapters 1 and 2, cover the vast majority of quantitative problems you will encounter in economics, management, finance, and applied statistics.

✓ **Before closing this booklet, make sure you can:**

- Identify the appropriate distributional family from a verbal problem description.
- State the parameters, mean, and variance of each of the seven named distributions.
- Compute Binomial, Hypergeometric, and Poisson probabilities from their formulas or tables.
- Standardise a Normal variable and read probabilities from the Z-table.
- Apply the main approximations (Binomial→Poisson, Binomial→Normal, Poisson→Normal) with their conditions.
- Interpret numerical results in an economic or managerial context.

Unified Bibliography

The bibliography below consolidates the references cited throughout the three chapters of this booklet. All entries follow the APA 7 style and are ordered alphabetically by the first author's surname.

Readers who wish to pursue a topic in greater depth are encouraged to consult the original sources listed here. Many are available in the library of the University of Sétif 1 or through the Algerian Virtual Library (SNDL).

I. Textbooks — Probability and Statistics

1. Blitzstein, J. K., & Hwang, J. (2019). *Introduction to probability* (2nd ed.). Chapman & Hall / CRC. <https://doi.org/10.1201/9780429428357>
2. Camm, J. D., Cochran, J. J., Fry, M. J., Ohlmann, J. W., Anderson, D. R., Sweeney, D. J., & Williams, T. A. (2024). *Statistics for business and economics* (15th ed.). Cengage Learning.
3. Devore, J. L., Berk, K. N., & Carlton, M. A. (2021). *Modern mathematical statistics with applications* (3rd ed.). Springer. <https://doi.org/10.1007/978-3-030-55156-8>
4. Grinstead, C. M., & Snell, J. L. (2019). *Introduction to probability* (2nd rev. ed.). American Mathematical Society. <https://math.dartmouth.edu/~prob/prob/prob.pdf>
5. Hansen, B. E. (2022). *Probability and statistics for economists*. Princeton University Press.
6. Newbold, P., Carlson, W. L., & Thorne, B. M. (2023). *Statistics for business and economics* (10th Global ed.). Pearson Education.
7. Pishro-Nik, H. (2021). *Introduction to probability, statistics, and random processes*. Kappa Research. <https://www.probabilitycourse.com>
8. Ross, S. M. (2024). *A first course in probability* (10th ed., Global Edition). Pearson Education.
9. Sahu, S. K. (2024). *Introduction to probability, statistics & R: Foundations for data-based sciences*. Springer. <https://doi.org/10.1007/978-3-031-37865-2>
10. Wasserman, L. (2020). *All of statistics: A concise course in statistical inference* (Corrected reprint). Springer. <https://doi.org/10.1007/978-0-387-21736-9>

II. Références en langue française

1. Foata, D., & Fuchs, A. (2020). *Calcul des probabilités : Cours, exercices et problèmes corrigés* (4e éd.). Dunod.
2. Hurlin, C., & Mignon, V. (2022). *Statistique et probabilités en économie-gestion* (2e éd.). Dunod.
3. Lecoutre, J.-P. (2023). *Statistique et probabilités : Cours et exercices corrigés* (7e éd.). Dunod.
4. Tassi, P., & Legait, S. (2021). *Théorie des probabilités en vue des applications statistiques* (3e éd.). Éditions Technip.

III. Applied probability in economics and management

1. Anderson, D. R., Sweeney, D. J., Williams, T. A., Camm, J. D., Cochran, J. J., Fry, M. J., & Ohlmann, J. W. (2020). *Essentials of modern business statistics with Microsoft Excel* (8th ed.). Cengage Learning.

2. Bluman, A. G. (2019). *Elementary statistics: A step-by-step approach* (10th ed.). McGraw-Hill Education.
3. Keller, G. (2023). *Statistics for management and economics* (12th ed.). Cengage Learning.
4. Levine, D. M., Stephan, D. F., & Szabat, K. A. (2021). *Statistics for managers using Microsoft Excel* (9th Global ed.). Pearson Education.
5. Sharpe, N. R., De Veaux, R. D., & Velleman, P. F. (2021). *Business statistics* (4th ed.). Pearson Education.

Further Reading & Interactive Resources for Self-Study

The works above are the academic backbone of the course. The two resources below serve a different purpose: both are free, openly available, and built for self-study. Use them the way you would use a tutor. When a definition in these notes feels abstract, read the matching section in one of these sources first, then come back to the formal statement here. Read actively: reproduce each derivation by hand, build a small numerical example of your own, and check it against the worked solutions. Concepts such as conditional probability, the law of large numbers, and the central limit theorem become much clearer once you have seen them move as their parameters change.

1. Diez, D., Çetinkaya-Rundel, M., & Barr, C. D. (2019). *OpenIntro statistics* (4th ed.). OpenIntro. <https://www.openintro.org/book/os/>

How to use it: a free PDF with real-data examples. Its chapters on random variables and the normal distribution parallel Chapters 2 and 3 here. Read them when you want the same idea explained a second way, then attempt its end-of-chapter problems for extra practice.

2. Devlin, T., Guo, J., Kunin, D., & Xiang, D. (2018). *Seeing theory: A visual introduction to probability and statistics*. Brown University. <https://seeing-theory.brown.edu>

How to use it: interactive visualisations. Drag the parameters and watch a distribution, the law of large numbers, or the central limit theorem respond in real time. This is the fastest way to build intuition before you meet the formulas in Chapters 2 and 3.

List of References

1. Anderson, D. R., Sweeney, D. J., Williams, T. A., Camm, J. D., Cochran, J. J., Fry, M. J., & Ohlmann, J. W. (2020). *Essentials of modern business statistics with Microsoft Excel* (8th ed.). Cengage Learning.
2. Blitzstein, J. K., & Hwang, J. (2019). *Introduction to probability* (2nd ed.). Chapman & Hall / CRC. <https://doi.org/10.1201/9780429428357>
3. Bluman, A. G. (2019). *Elementary statistics: A step-by-step approach* (10th ed.). McGraw-Hill Education.
4. Camm, J. D., Cochran, J. J., Fry, M. J., Ohlmann, J. W., Anderson, D. R., Sweeney, D. J., & Williams, T. A. (2024). *Statistics for business and economics* (15th ed.). Cengage Learning.
5. Chan, S. H. (2021). *Introduction to probability for data science*. Michigan Publishing. <https://doi.org/10.3998/mpub.12387745>
6. Devlin, T., Guo, J., Kunin, D., & Xiang, D. (2018). *Seeing theory: A visual introduction to probability and statistics*. Brown University. <https://seeing-theory.brown.edu>
7. Devore, J. L., Berk, K. N., & Carlton, M. A. (2021). *Modern mathematical statistics with applications* (3rd ed.). Springer. <https://doi.org/10.1007/978-3-030-55156-8>
8. Diez, D., Çetinkaya-Rundel, M., & Barr, C. D. (2019). *OpenIntro statistics* (4th ed.). OpenIntro. <https://www.openintro.org/book/os/>
9. Foata, D., & Fuchs, A. (2020). *Calcul des probabilités : Cours, exercices et problèmes corrigés* (4e éd.). Dunod.
10. Grinstead, C. M., & Snell, J. L. (2019). *Introduction to probability* (2nd rev. ed.). American Mathematical Society. <https://math.dartmouth.edu/~prob/prob/prob.pdf>
11. Hansen, B. E. (2022). *Probability and statistics for economists*. Princeton University Press.
12. Hurlin, C., & Mignon, V. (2022). *Statistique et probabilités en économie-gestion* (2e éd.). Dunod.
13. Keller, G. (2023). *Statistics for management and economics* (12th ed.). Cengage Learning.
14. Lecoutre, J.-P. (2023). *Statistique et probabilités : Cours et exercices corrigés* (7e éd.). Dunod.
15. Levine, D. M., Stephan, D. F., & Szabat, K. A. (2021). *Statistics for managers using Microsoft Excel* (9th Global ed.). Pearson Education.
16. Linde, W. (2024). *Probability theory: A first course in probability theory and statistics* (2nd ed.). De Gruyter. <https://doi.org/10.1515/9783111325064>
17. Newbold, P., Carlson, W. L., & Thorne, B. M. (2023). *Statistics for business and economics* (10th Global ed.). Pearson Education.
18. Pishro-Nik, H. (2021). *Introduction to probability, statistics, and random processes*. Kappa Research. <https://www.probabilitycourse.com>

19. Ross, S. M. (2024). *A first course in probability* (10th ed., Global Edition). Pearson Education.
20. Sahu, S. K. (2024). *Introduction to probability, statistics & R: Foundations for data-based sciences*. Springer. <https://doi.org/10.1007/978-3-031-37865-2>
21. Sharpe, N. R., De Veaux, R. D., & Velleman, P. F. (2021). *Business statistics* (4th ed.). Pearson Education.
22. Tassi, P., & Legait, S. (2021). *Théorie des probabilités en vue des applications statistiques* (3e éd.). Éditions Technip.
23. Wasserman, L. (2020). *All statistics: A concise course in statistical inference* (Corrected reprint). Springer. <https://doi.org/10.1007/978-0-387-21736-9>

End of the Handout

Statistics 2 — Probability Theory

A complete pedagogical booklet for first-year LMD students

✓ A word of thanks

- To my students, whose questions and feedback shape every chapter of this handout.
- To the Faculty of Economic, Commercial and Management Sciences for continuing to support open, high-quality teaching materials.
- To the colleagues who read drafts of this work and offered constructive comments.
- May this handout serve you well throughout your studies and beyond.

Dr. Manel REBBOUH

Associate Professor, SETIF 1 University— Ferhat Abbas

Academic Year 2025 / 2026