Introduction to probability theory and statistical inference

Jesús Urtasun Elizari

Research Computing and Data Science

December 5, 2025

Contents

Preface

The purpose of these notes

In the following pages one will find an introductory course to the theory of probability and statistical inference, aiming to cover both foundations and basic mathematical concepts, but also practical tools to deal with real data science problems, such as bayesian probability and hypothesis testing. The text is composed by five chapters, together with some appendix sections reviewing basic mathematical notions, and a bibliographic note. The purpose of these lecture notes is to make both probability and statistical analysis an easy, engaging and exciting topic for anyone interested, without the need for prior experience.

Both, predictive probability and descriptive statistics have deep historical roots, from ancient works on chance and divination to modern scientific topics oriented towards information theory, modelling and data analysis. As one could guess, rivers of ink have been written about such topics, and endless literature sources are available. However, after following many different courses at both bachelor and postgraduate levels, and teaching such topics myself during the last three years, I have found that most resources belong, almost certainly, to one of the next three classes. Either (*i*) deeply mathematical, and hence out of reach for most experimental or clinically oriented scientists, (*ii*) laboratory oriented, focusing on inference and experimental design, and hence missing most of the mathematical background, or (*iii*) with a direct focus towards programming and computation, relying on domain specific notebooks (Python, R, Matlab, SPSS, etc), and online resources with precompiled libraries for simulation, which again miss most of the mathematical and formal intuitions.

Indeed, the misuse of statistics in experimental sciences is a critical topic in modern times, as mathematicians have extensively discussed during the last decades. The well-known article by John P. A. Ioannidis, *"Why most published research findings are false"* [1], serves as a prominent example, and it may serve as motivation for a rigorous study.

As a matter of fact, when it comes to modern statistics, data analysis or experimental design, concepts like *stochasticity, randomness, sampling, hypothesis, significance, statistic test, p-value* - just to mention some of them - are frequently used, but for most bachelor and even master's level degrees they are rarely introduced or properly defined. Indeed, for most experimental and clinically oriented degrees, they are not introduced at all, leaving the student with just a superficial knowledge relying on intuition about some particular cases. Hence, developing high-quality, simple, and accessible open source material for present and future generations, covering both probability and statistical inference from both a fundamental *and* applied level, remains an urgent task for scientists and educators.

This is intended to be a complete introductory course, and no previous mathematical background is required. By keeping the theory simple and always followed by examples, we will build the definitions and quantities from simple to more complex. All mathematical formulas will be introduced with rigorous notation, but keeping in mind that it is not the symbols or the numbers, but the intuitions and the general understanding, what we are after.

Additionally, all topics will be introduced alongside with some short historical discussion and context, as we believe that a purely technical knowledge just grasps the complexity - and beauty - of scientific topics. As one could anticipate already, a proper understanding of ideas such as uncertainty, variation, chance, probability, inference, etc, can be applied to describing a vast amount of real-world phenomena, ranging from gambling and to data analysis to modelling in physics, biology, machine learning and quantum mechanics, among many others.

As mentioned, the course is organised in five chapters. In the first two we will introduce the idea of sampling, probability and random events with simple and intuitive examples, and we will see how different approaches have been used to model information and chance in different times. The introduction here is twofold. Chapter 1 *Descriptive statistics* introduces the idea of uncertainty, sampling, and central tendency, aiming to describe and understand populations and observation sets, while Chapter 2 *Probability and random events* focuses on the mathematical definition of probability. Here we will cover the idea of random processes - also referred to as *stochastic*, probability and distribution, as a set of tools that enables mathematical predictions in uncertain cases, such as the *expected value*.

Chapter 3 *Parameter estimation* will introduce the essential difference between prediction and inference, and revisit the concept of sampling and population in more detail. We will discuss how to build *estimator* quantities out of our samples, as a way to reconstruct - or *infer* - the underlying phenomena of a population given a finite set of observations. Here we will see some general results which may sound familiar already, such as *"The Law of Large Numbers"*, the *"The Central Limit Theorem"*, or the *"Maximum Likelihood Estimation"* method.

In Chapter 4 we will discuss a group of topics commonly referred to as *hypothesis testing*. Here we will introduce the idea of hypothesis, how to quantify certainty and bias, how to model significance with the so-called *p-values*, and some common examples of statistic tests. Chapter 5 will cover with some detail conditional and Bayesian probability, revisit the idea of stochasticity and introduce the so called Markov processes.

At the end of each chapter there will be a series of exercises and coding examples to illustrate and demonstrate the concepts discussed. To avoid misconceptions, let us emphasize here that both, probability and statistics are just branches of mathematics dealing chance and information in random events, *much earlier* than computers, coding languages, Python, R or P-values were even conceived. The data-oriented, practical ways in which probability and statistics are usually taught, relying heavily on computation, is just a consequence of the fact that automatized measurements are nowadays available and trendy in modern times [...].

Example textbooks covering introduction to probability and statistical inference, for further reading:

- A simple, intuitive introduction to statistics with few mathematical concepts is provided in Spiegelhalter's *"The Art of Statistics: How to Learn from Data"* [2].

- A more foundational textbook, with more advanced mathematical approach, can be found at DeGroot and Schervish's *"Probability and Statistics"* [3].

- For a philosophical and historical perspective on probability and statistics, please find Forster and Bandyopadhyay's handbook *"Philosophy of Statistics"* [4].
- A comprehensive introduction with focus on practical applications and modern data analysis tools is can be found at Diez, Barr & Mine *"OpenIntro Statistics"* [5].
- For fundamental concepts in probability and statistics, including random variables, distributions and statistical inference, with practical examples and exercises follow Hossein Pishro-Nik's *"Probability, Statistics & Random Processes"* [6].

Introduction

Even fire obeys the laws of numbers.
— J.B. Joseph Fourier

A bit of history

As one might expect, the origins of probability and related concepts can be traced back to very ancient times. Civilizations such as the Babylonians, Egyptians, and Greeks already encountered uncertainty in various aspects of life, including commerce, games of chance, and divination. Consequently, notions of randomness and stochasticity have deep historical roots. For instance, archaeological findings suggest that the earliest known dice date back over 5,000 years, reflecting humanity's early fascination with chance and unpredictability [7]. Although these cultures had not yet developed a formal mathematical theory of probability, they recognized recurring patterns in random events and attempted to anticipate outcomes through either empirical observation or superstition. For a detailed historical overview, see Florence Nightingale's 1962 manuscript *"Games, Gods and Gambling"* [8].

While classical Greek and Roman philosophers frequently discussed the nature of chance, necessity, and determinism, their inquiries remained primarily philosophical rather than mathematical. Thinkers such as Cicero distinguished between events occurring by chance and those determined by fate, foreshadowing later developments in probability theory [9]. These early ideas, though lacking quantitative formalism, provided the intellectual foundation for later scientific inquiry into randomness and causality.

A significant shift occurred during the late medieval and early Renaissance periods, when more rigorous mathematical ideas began to shape. Italian mathematician and gambler Gerolamo Cardano (1501–1576) made substantial contributions to the mathematical analysis of chance. His work *"Liber de Ludo Aleae"* (*"Book on Games of Chance"*) [10], posthumously published in 1663, is one of the earliest known texts to explore probability through the analysis of gambling problems. However, Cardano's reasoning, while insightful, lacked the symbolic clarity and mathematical rigour of modern probability theory. Readers consulting the original manuscript will notice an ambiguous and sometimes inconsistent symbolic system, quite unlike the formal structures we use nowadays.

The formalization of probability as a mathematical discipline did not occur until the 17th century, most notably through the seminal correspondence between Blaise Pascal and Pierre de Fermat. Their work, motivated by problems such as finding a fair division of stakes in interrupted games of chance, introduced foundational concepts such as combinatorics, expected value, and variance [11]. These developments paved the way for later contributions by Christiaan Huygens, who in 1657 wrote the first published textbook on probability *"De Ratiociniis in Ludo Aleae"* [12], and Jacob Bernoulli, whose 1713 *"Ars Conjectandi"* remains among the most influential early texts in the field. Their works, along with many others, collectively laid the groundwork for the probabilistic and statistical methods that foreshadow modern scientific reasoning [13, 14].

During the 19th century, probability theory began to intertwine more deeply with statistics and the emerging mathematical analysis of physical phenomena. Florence Nightingale, best known for her pioneering role in modern nursing, also made significant contributions to statistical methodology and graphical representation of data. Her use of polar area diagrams and her advocacy for statistical reasoning in public health policy helped popularize quantitative approaches to uncertainty and variation. Around the same period, Joseph Fourier's work on heat conduction introduced Fourier series and integral transforms, tools that would later become indispensable for studying random processes, including the analysis of signals, noise, and diffusion phenomena. Although Nightingale and Fourier approached problems of uncertainty from very different perspectives—one through empirical data on human wellbeing, the other through mathematical physics—their contributions expanded the reach of probabilistic thinking and prepared the ground for future developments in stochastic analysis.

A further conceptual leap occurred in the early 20th century with the work of Andrey Markov. Motivated partly by a desire to extend the law of large numbers beyond the assumption of independent trials, Markov developed what are now known as Markov chains, thereby inaugurating the study of dependence structures in stochastic processes. His investigations demonstrated that long-run statistical regularities could emerge even when successive events were not independent, a discovery that profoundly influenced both theoretical probability and its applications in fields as diverse as statistical mechanics, linguistics, economics, and modern machine learning.

The modern axiomatic formulation of probability was introduced in the early 20th century by the Russian mathematician Andrey Kolmogorov. In his 1933 monograph *"Grundbegriffe der Wahrscheinlichkeitsrechnung"* [15], Kolmogorov synthesized classical and frequentist ideas into a rigorous mathematical framework based on measure theory. His axioms remain the standard foundation for probability theory to this day. It may seem surprising that a concept with such ancient origins was not formally axiomatized until relatively recent times, and we will return to Kolmogorov's formulation and its implications in greater detail in Chapter 5. Nevertheless, philosophical discussions about the interpretation of probability and its relation to the physical sciences—especially in the context of determinism, epistemology, and modern topics such as quantum mechanics—predate Kolmogorov's formulation and continue to evolve to this day.

Chapter 1

Descriptive statistics

Statistics is the grammar of science.
— Karl Pearson

As one might expect, the origins of probability and related concepts can be traced back to very ancient times. Civilizations such as the Babylonians, Egyptians, and Greeks already encountered uncertainty in various aspects of life, including commerce, games of chance, and divination. Consequently, notions of randomness and stochasticity have deep historical roots. For instance, archaeological findings suggest that the earliest known dice date back over 5,000 years, reflecting humanity's early fascination with chance and unpredictability [**finkel2007ancient**]. Although these cultures had not yet developed a formal mathematical theory of probability, they recognized recurring patterns in random events and attempted to anticipate outcomes through either empirical observation or superstition. For a detailed historical overview, see Florence Nightingale's 1962 manuscript *"Games, Gods and Gambling"* [**david1962games**].

While classical Greek and Roman philosophers frequently discussed the nature of chance, necessity, and determinism, their inquiries remained primarily philosophical rather than mathematical. Thinkers such as Cicero distinguished between events occurring by chance and those determined by fate, foreshadowing later developments in probability theory [**cicero45bce**]. These early ideas, though lacking quantitative formalism, provided the intellectual foundation for later scientific inquiry into randomness and causality.

A significant shift occurred during the late medieval and early Renaissance periods, when more rigorous mathematical ideas began to shape. Italian mathematician and gambler Gerolamo Cardano (1501–1576) made substantial contributions to the mathematical analysis of chance. His work *"Liber de Ludo Aleae"* (*"Book on Games of Chance"*) [**cardano1663ludo**], posthumously published in 1663, is one of the earliest known texts to explore probability through the analysis of gambling problems. However, Cardano's reasoning, while insightful, lacked the symbolic clarity and mathematical rigour of modern probability theory. Readers consulting the original manuscript will notice an ambiguous and sometimes inconsistent symbolic system, quite unlike the formal structures we use nowadays.

The formalization of probability as a mathematical discipline did not occur until the 17th century, most notably through the seminal correspondence between Blaise Pascal and Pierre de Fermat. Their work, motivated by problems such as finding a fair division of stakes in interrupted games of chance, introduced foundational concepts such as combinatorics, expected value, and variance [**devlin2008unfinished**]. These developments paved the way for later contributions by Christiaan Huygens, who in 1657 wrote the first published textbook on probability *"De Ratiociniis in Ludo Aleae"* (*"On Reasoning in Games of Chance"*), and Jacob

Bernoulli, whose 1713 *"Ars Conjectandi"* (*"The Art of Conjecturing"*) remains among the most influential early texts in the field. Their works, alongside with many others, collectively laid the groundwork for the probabilistic and statistical methods that foreshadow modern scientific reasoning [**huygens1657ratiociniis**], [**bernoulli1713ars**], [**hald1990history**].

The modern axiomatic formulation of probability was introduced in the early 20th century by the Russian mathematician Andrey Kolmogorov. In his 1933 monograph *"Grundbegriffe der Wahrscheinlichkeitsrechnung"* (*"Foundations of the Theory of Probability"*) [**kolmogorov1933grundbegriffe**], Kolmogorov synthesized classical and frequentist ideas into a rigorous mathematical framework based on measure theory. His axioms remain the standard foundation for probability theory to this day. It may seem surprising that a concept with such ancient origins was not formally axiomatized until relatively recent times, and we will return to Kolmogorov's formulation and its implications in greater detail in Chapter 5. Nevertheless, philosophical discussions about the interpretation of probability and its relation to the physical sciences - especially in the context of determinism, epistemology and modern topics such as quantum mechanics - predate Kolmogorov's formulation and continue to evolve to this day.

1.1 Population and sampling

1.2 Central tendency and variation

1.3 Data visualization

Exercises

1. Exercise [...].

2. Exercise [...].

3. Exercise [...].

Solutions

1. Solution [...].

2. Solution [...].

3. Solution [...].

Chapter 2

Foundations of Probability

It is through the calculation of probabilities that
the divine order becomes visible.
— Jacob Bernoulli

As one might expect, the origins of probability and related concepts can be traced back to very ancient times. Civilizations such as the Babylonians, Egyptians, and Greeks already encountered uncertainty in various aspects of life, including commerce, games of chance, and divination. Consequently, notions of randomness and stochasticity have deep historical roots. For instance, archaeological findings suggest that the earliest known dice date back over 5,000 years, reflecting humanity's early fascination with chance and unpredictability [**finkel2007ancient**]. Although these cultures had not yet developed a formal mathematical theory of probability, they recognized recurring patterns in random events and attempted to anticipate outcomes through either empirical observation or superstition. For a detailed historical overview, see Florence Nightingale's 1962 manuscript *"Games, Gods and Gambling"* [**david1962games**].

While classical Greek and Roman philosophers frequently discussed the nature of chance, necessity, and determinism, their inquiries remained primarily philosophical rather than mathematical. Thinkers such as Cicero distinguished between events occurring by chance and those determined by fate, foreshadowing later developments in probability theory [**cicero45bce**]. These early ideas, though lacking quantitative formalism, provided the intellectual foundation for later scientific inquiry into randomness and causality.

A significant shift occurred during the late medieval and early Renaissance periods, when more rigorous mathematical ideas began to shape. Italian mathematician and gambler Gerolamo Cardano (1501–1576) made substantial contributions to the mathematical analysis of chance. His work *"Liber de Ludo Aleae"* (*"Book on Games of Chance"*) [**cardano1663ludo**], posthumously published in 1663, is one of the earliest known texts to explore probability through the analysis of gambling problems. However, Cardano's reasoning, while insightful, lacked the symbolic clarity and mathematical rigour of modern probability theory. Readers consulting the original manuscript will notice an ambiguous and sometimes inconsistent symbolic system, quite unlike the formal structures we use nowadays.

The formalization of probability as a mathematical discipline did not occur until the 17th century, most notably through the seminal correspondence between Blaise Pascal and Pierre de Fermat. Their work, motivated by problems such as finding a fair division of stakes in interrupted games of chance, introduced foundational concepts such as combinatorics, expected value, and variance [**devlin2008unfinished**]. These developments paved the way for later contributions by Christiaan Huygens, who in 1657 wrote the first published textbook on probability *"De Ratiociniis in Ludo Aleae"* (*"On Reasoning in Games of Chance"*), and Jacob

Bernoulli, whose 1713 *"Ars Conjectandi"* (*"The Art of Conjecturing"*) remains among the most influential early texts in the field. Their works, alongside with many others, collectively laid the groundwork for the probabilistic and statistical methods that foreshadow modern scientific reasoning [**huygens1657ratiociniis**], [**bernoulli1713ars**], [**hald1990history**].

The modern axiomatic formulation of probability was introduced in the early 20th century by the Russian mathematician Andrey Kolmogorov. In his 1933 monograph *"Grundbegriffe der Wahrscheinlichkeitsrechnung"* (*"Foundations of the Theory of Probability"*) [**kolmogorov1933grundbegriffe**], Kolmogorov synthesized classical and frequentist ideas into a rigorous mathematical framework based on measure theory. His axioms remain the standard foundation for probability theory to this day. It may seem surprising that a concept with such ancient origins was not formally axiomatized until relatively recent times, and we will return to Kolmogorov's formulation and its implications in greater detail in Chapter 5. Nevertheless, philosophical discussions about the interpretation of probability and its relation to the physical sciences - especially in the context of determinism, epistemology and modern topics such as quantum mechanics - predate Kolmogorov's formulation and continue to evolve to this day.

2.1 What is probability?

2.2 Discrete random variables

2.2.1 Bernoulli events

2.2.2 Binomial distribution

2.2.3 Poisson distribution

2.2.4 Discrete uniform distribution

2.3 Continuous random variables

2.3.1 Continuous uniform distribution

2.3.2 Exponential distribution

2.3.3 Gaussian distribution

2.4 Expectation values

Exercises

1. Exercise [...].

2. Exercise [...].

3. Exercise [...].

Solutions

1. Solution [...].

2. Solution [...].

3. Solution [...].

Chapter 3

Estimation, prediction and inference

Numbers have an important story to tell, if given a voice.
— Florence Nightingale

As one might expect, the origins of probability and related concepts can be traced back to very ancient times. Civilizations such as the Babylonians, Egyptians, and Greeks already encountered uncertainty in various aspects of life, including commerce, games of chance, and divination. Consequently, notions of randomness and stochasticity have deep historical roots. For instance, archaeological findings suggest that the earliest known dice date back over 5,000 years, reflecting humanity's early fascination with chance and unpredictability [**finkel2007ancient**]. Although these cultures had not yet developed a formal mathematical theory of probability, they recognized recurring patterns in random events and attempted to anticipate outcomes through either empirical observation or superstition. For a detailed historical overview, see Florence Nightingale's 1962 manuscript *"Games, Gods and Gambling"* [**david1962games**].

While classical Greek and Roman philosophers frequently discussed the nature of chance, necessity, and determinism, their inquiries remained primarily philosophical rather than mathematical. Thinkers such as Cicero distinguished between events occurring by chance and those determined by fate, foreshadowing later developments in probability theory [**cicero45bce**]. These early ideas, though lacking quantitative formalism, provided the intellectual foundation for later scientific inquiry into randomness and causality.

A significant shift occurred during the late medieval and early Renaissance periods, when more rigorous mathematical ideas began to shape. Italian mathematician and gambler Gerolamo Cardano (1501–1576) made substantial contributions to the mathematical analysis of chance. His work *"Liber de Ludo Aleae"* (*"Book on Games of Chance"*) [**cardano1663ludo**], posthumously published in 1663, is one of the earliest known texts to explore probability through the analysis of gambling problems. However, Cardano's reasoning, while insightful, lacked the symbolic clarity and mathematical rigour of modern probability theory. Readers consulting the original manuscript will notice an ambiguous and sometimes inconsistent symbolic system, quite unlike the formal structures we use nowadays.

The formalization of probability as a mathematical discipline did not occur until the 17th century, most notably through the seminal correspondence between Blaise Pascal and Pierre de Fermat. Their work, motivated by problems such as finding a fair division of stakes in interrupted games of chance, introduced foundational concepts such as combinatorics, expected value, and variance [**devlin2008unfinished**]. These developments paved the way for later contributions by Christiaan Huygens, who in 1657 wrote the first published textbook on

probability *"De Ratiociniis in Ludo Aleae"* (*"On Reasoning in Games of Chance"*), and Jacob Bernoulli, whose 1713 *"Ars Conjectandi"* (*"The Art of Conjecturing"*) remains among the most influential early texts in the field. Their works, alongside with many others, collectively laid the groundwork for the probabilistic and statistical methods that foreshadow modern scientific reasoning [**huygens1657ratiociniis**], [**bernoulli1713ars**], [**hald1990history**].

The modern axiomatic formulation of probability was introduced in the early 20th century by the Russian mathematician Andrey Kolmogorov. In his 1933 monograph *"Grundbegriffe der Wahrscheinlichkeitsrechnung"* (*"Foundations of the Theory of Probability"*) [**kolmogorov1933grundbegriffe**], Kolmogorov synthesized classical and frequentist ideas into a rigorous mathematical framework based on measure theory. His axioms remain the standard foundation for probability theory to this day. It may seem surprising that a concept with such ancient origins was not formally axiomatized until relatively recent times, and we will return to Kolmogorov's formulation and its implications in greater detail in Chapter 5. Nevertheless, philosophical discussions about the interpretation of probability and its relation to the physical sciences - especially in the context of determinism, epistemology and modern topics such as quantum mechanics - predate Kolmogorov's formulation and continue to evolve to this day.

3.1 Prediction vs inference

3.2 The Law of Large Numbers

3.3 The Central Limit Theorem

3.4 Application to Generalized Linear Models

Exercises

1. Exercise [...].

2. Exercise [...].

3. Exercise [...].

Solutions

1. Solution [...].

2. Solution [...].

3. Solution [...].

Chapter 4

Introduction to hypothesis testing

The object of statistical science is the reduction of data to relevant information.
— Ronald A. Fisher

As one might expect, the origins of probability and related concepts can be traced back to very ancient times. Civilizations such as the Babylonians, Egyptians, and Greeks already encountered uncertainty in various aspects of life, including commerce, games of chance, and divination. Consequently, notions of randomness and stochasticity have deep historical roots. For instance, archaeological findings suggest that the earliest known dice date back over 5,000 years, reflecting humanity's early fascination with chance and unpredictability [**finkel2007ancient**]. Although these cultures had not yet developed a formal mathematical theory of probability, they recognized recurring patterns in random events and attempted to anticipate outcomes through either empirical observation or superstition. For a detailed historical overview, see Florence Nightingale's 1962 manuscript *"Games, Gods and Gambling"* [**david1962games**].

While classical Greek and Roman philosophers frequently discussed the nature of chance, necessity, and determinism, their inquiries remained primarily philosophical rather than mathematical. Thinkers such as Cicero distinguished between events occurring by chance and those determined by fate, foreshadowing later developments in probability theory [**cicero45bce**]. These early ideas, though lacking quantitative formalism, provided the intellectual foundation for later scientific inquiry into randomness and causality.

A significant shift occurred during the late medieval and early Renaissance periods, when more rigorous mathematical ideas began to shape. Italian mathematician and gambler Gerolamo Cardano (1501–1576) made substantial contributions to the mathematical analysis of chance. His work *"Liber de Ludo Aleae"* (*"Book on Games of Chance"*) [**cardano1663ludo**], posthumously published in 1663, is one of the earliest known texts to explore probability through the analysis of gambling problems. However, Cardano's reasoning, while insightful, lacked the symbolic clarity and mathematical rigour of modern probability theory. Readers consulting the original manuscript will notice an ambiguous and sometimes inconsistent symbolic system, quite unlike the formal structures we use nowadays.

The formalization of probability as a mathematical discipline did not occur until the 17th century, most notably through the seminal correspondence between Blaise Pascal and Pierre de Fermat. Their work, motivated by problems such as finding a fair division of stakes in interrupted games of chance, introduced foundational concepts such as combinatorics, expected value, and variance [**devlin2008unfinished**]. These developments paved the way for later contributions by Christiaan Huygens, who in 1657 wrote the first published textbook on

probability *"De Ratiociniis in Ludo Aleae"* (*"On Reasoning in Games of Chance"*), and Jacob Bernoulli, whose 1713 *"Ars Conjectandi"* (*"The Art of Conjecturing"*) remains among the most influential early texts in the field. Their works, alongside with many others, collectively laid the groundwork for the probabilistic and statistical methods that foreshadow modern scientific reasoning [**huygens1657ratiociniis**], [**bernoulli1713ars**], [**hald1990history**].

The modern axiomatic formulation of probability was introduced in the early 20th century by the Russian mathematician Andrey Kolmogorov. In his 1933 monograph *"Grundbegriffe der Wahrscheinlichkeitsrechnung"* (*"Foundations of the Theory of Probability"*) [**kolmogorov1933grundbegriffe**], Kolmogorov synthesized classical and frequentist ideas into a rigorous mathematical framework based on measure theory. His axioms remain the standard foundation for probability theory to this day. It may seem surprising that a concept with such ancient origins was not formally axiomatized until relatively recent times, and we will return to Kolmogorov's formulation and its implications in greater detail in Chapter 5. Nevertheless, philosophical discussions about the interpretation of probability and its relation to the physical sciences - especially in the context of determinism, epistemology and modern topics such as quantum mechanics - predate Kolmogorov's formulation and continue to evolve to this day.

4.1 Prediction vs inference revisted

4.2 General approach to hypothesis testing

4.3 Statistical tests: some examples

4.3.1 Compare sample mean with hypothesized value - One sample t-test

4.3.2 Compare sample means of two independent groups - Two sample t-test

4.3.3 Compare variation on two groups - Fisher's exact test

4.3.4 Compare variation o multiple groups - Fisher's ANOVA

4.3.5 Compare distributions and testing for normality - χ^2 test

4.4 Parametric and non-parametric tests

4.5 Comparing data and normalization

Exercises

1. Exercise [...].

2. Exercise [...].

3. Exercise [...].

Solutions

1. Solution [...].

2. Solution [...].

3. Solution [...].

Chapter 5

Modeling, dependency and correlation

The theory of probabilities is at bottom nothing but common sense reduced to calculation.
— Pierre-Simon Laplace

As one might expect, the origins of probability and related concepts can be traced back to very ancient times. Civilizations such as the Babylonians, Egyptians, and Greeks already encountered uncertainty in various aspects of life, including commerce, games of chance, and divination. Consequently, notions of randomness and stochasticity have deep historical roots. For instance, archaeological findings suggest that the earliest known dice date back over 5,000 years, reflecting humanity's early fascination with chance and unpredictability [**finkel2007ancient**]. Although these cultures had not yet developed a formal mathematical theory of probability, they recognized recurring patterns in random events and attempted to anticipate outcomes through either empirical observation or superstition. For a detailed historical overview, see Florence Nightingale's 1962 manuscript *"Games, Gods and Gambling"* [**david1962games**].

While classical Greek and Roman philosophers frequently discussed the nature of chance, necessity, and determinism, their inquiries remained primarily philosophical rather than mathematical. Thinkers such as Cicero distinguished between events occurring by chance and those determined by fate, foreshadowing later developments in probability theory [**cicero45bce**]. These early ideas, though lacking quantitative formalism, provided the intellectual foundation for later scientific inquiry into randomness and causality.

A significant shift occurred during the late medieval and early Renaissance periods, when more rigorous mathematical ideas began to shape. Italian mathematician and gambler Gerolamo Cardano (1501–1576) made substantial contributions to the mathematical analysis of chance. His work *"Liber de Ludo Aleae"* (*"Book on Games of Chance"*) [**cardano1663ludo**], posthumously published in 1663, is one of the earliest known texts to explore probability through the analysis of gambling problems. However, Cardano's reasoning, while insightful, lacked the symbolic clarity and mathematical rigour of modern probability theory. Readers consulting the original manuscript will notice an ambiguous and sometimes inconsistent symbolic system, quite unlike the formal structures we use nowadays.

The formalization of probability as a mathematical discipline did not occur until the 17th century, most notably through the seminal correspondence between Blaise Pascal and Pierre de Fermat. Their work, motivated by problems such as finding a fair division of stakes in interrupted games of chance, introduced foundational concepts such as combinatorics, expected value, and variance [**devlin2008unfinished**]. These developments paved the way for later contributions by Christiaan Huygens, who in 1657 wrote the first published textbook on

probability *"De Ratiociniis in Ludo Aleae"* (*"On Reasoning in Games of Chance"*), and Jacob Bernoulli, whose 1713 *"Ars Conjectandi"* (*"The Art of Conjecturing"*) remains among the most influential early texts in the field. Their works, alongside with many others, collectively laid the groundwork for the probabilistic and statistical methods that foreshadow modern scientific reasoning [**huygens1657ratiociniis**], [**bernoulli1713ars**], [**hald1990history**].

The modern axiomatic formulation of probability was introduced in the early 20th century by the Russian mathematician Andrey Kolmogorov. In his 1933 monograph *"Grundbegriffe der Wahrscheinlichkeitsrechnung"* (*"Foundations of the Theory of Probability"*) [**kolmogorov1933grundbegriffe**], Kolmogorov synthesized classical and frequentist ideas into a rigorous mathematical framework based on measure theory. His axioms remain the standard foundation for probability theory to this day. It may seem surprising that a concept with such ancient origins was not formally axiomatized until relatively recent times, and we will return to Kolmogorov's formulation and its implications in greater detail in Chapter 5. Nevertheless, philosophical discussions about the interpretation of probability and its relation to the physical sciences - especially in the context of determinism, epistemology and modern topics such as quantum mechanics - predate Kolmogorov's formulation and continue to evolve to this day.

5.1 Linear models

5.2 Polynomial models

5.3 GLMs and some examples

Exercises

1. Exercise [...].

2. Exercise [...].

3. Exercise [...].

Solutions

1. Solution [...].

2. Solution [...].

3. Solution [...].

Chapter 6

Introduction to conditional probability

Probability statements are just summaries of repeated observations.
— W. V. Quine

As one might expect, the origins of probability and related concepts can be traced back to very ancient times. Civilizations such as the Babylonians, Egyptians, and Greeks already encountered uncertainty in various aspects of life, including commerce, games of chance, and divination. Consequently, notions of randomness and stochasticity have deep historical roots. For instance, archaeological findings suggest that the earliest known dice date back over 5,000 years, reflecting humanity's early fascination with chance and unpredictability [**finkel2007ancient**]. Although these cultures had not yet developed a formal mathematical theory of probability, they recognized recurring patterns in random events and attempted to anticipate outcomes through either empirical observation or superstition. For a detailed historical overview, see Florence Nightingale's 1962 manuscript *"Games, Gods and Gambling"* [**david1962games**].

While classical Greek and Roman philosophers frequently discussed the nature of chance, necessity, and determinism, their inquiries remained primarily philosophical rather than mathematical. Thinkers such as Cicero distinguished between events occurring by chance and those determined by fate, foreshadowing later developments in probability theory [**cicero45bce**]. These early ideas, though lacking quantitative formalism, provided the intellectual foundation for later scientific inquiry into randomness and causality.

A significant shift occurred during the late medieval and early Renaissance periods, when more rigorous mathematical ideas began to shape. Italian mathematician and gambler Gerolamo Cardano (1501–1576) made substantial contributions to the mathematical analysis of chance. His work *"Liber de Ludo Aleae"* (*"Book on Games of Chance"*) [**cardano1663ludo**], posthumously published in 1663, is one of the earliest known texts to explore probability through the analysis of gambling problems. However, Cardano's reasoning, while insightful, lacked the symbolic clarity and mathematical rigour of modern probability theory. Readers consulting the original manuscript will notice an ambiguous and sometimes inconsistent symbolic system, quite unlike the formal structures we use nowadays.

The formalization of probability as a mathematical discipline did not occur until the 17th century, most notably through the seminal correspondence between Blaise Pascal and Pierre de Fermat. Their work, motivated by problems such as finding a fair division of stakes in interrupted games of chance, introduced foundational concepts such as combinatorics, expected value, and variance [**devlin2008unfinished**]. These developments paved the way for later contributions by Christiaan Huygens, who in 1657 wrote the first published textbook on

probability *"De Ratiociniis in Ludo Aleae"* (*"On Reasoning in Games of Chance"*), and Jacob Bernoulli, whose 1713 *"Ars Conjectandi"* (*"The Art of Conjecturing"*) remains among the most influential early texts in the field. Their works, alongside with many others, collectively laid the groundwork for the probabilistic and statistical methods that foreshadow modern scientific reasoning [**huygens1657ratiociniis**], [**bernoulli1713ars**], [**hald1990history**].

The modern axiomatic formulation of probability was introduced in the early 20th century by the Russian mathematician Andrey Kolmogorov. In his 1933 monograph *"Grundbegriffe der Wahrscheinlichkeitsrechnung"* (*"Foundations of the Theory of Probability"*) [**kolmogorov1933grundbegriffe**], Kolmogorov synthesized classical and frequentist ideas into a rigorous mathematical framework based on measure theory. His axioms remain the standard foundation for probability theory to this day. It may seem surprising that a concept with such ancient origins was not formally axiomatized until relatively recent times, and we will return to Kolmogorov's formulation and its implications in greater detail in Chapter 5. Nevertheless, philosophical discussions about the interpretation of probability and its relation to the physical sciences - especially in the context of determinism, epistemology and modern topics such as quantum mechanics - predate Kolmogorov's formulation and continue to evolve to this day.

6.1 Motivation and philosophy

6.2 Dependent and independent events

6.3 Some examples of conditional probability

Exercises

1. Exercise [...].

2. Exercise [...].

3. Exercise [...].

Solutions

1. Solution [...].

2. Solution [...].

3. Solution [...].

Chapter 7

Stochasticity and Markov Processes

The development of mathematics is a continuous process of abstraction.
— Emmy Noether

As one might expect, the origins of probability and related concepts can be traced back to very ancient times. Civilizations such as the Babylonians, Egyptians, and Greeks already encountered uncertainty in various aspects of life, including commerce, games of chance, and divination. Consequently, notions of randomness and stochasticity have deep historical roots. For instance, archaeological findings suggest that the earliest known dice date back over 5,000 years, reflecting humanity's early fascination with chance and unpredictability [**finkel2007ancient**]. Although these cultures had not yet developed a formal mathematical theory of probability, they recognized recurring patterns in random events and attempted to anticipate outcomes through either empirical observation or superstition. For a detailed historical overview, see Florence Nightingale's 1962 manuscript *"Games, Gods and Gambling"* [**david1962games**].

While classical Greek and Roman philosophers frequently discussed the nature of chance, necessity, and determinism, their inquiries remained primarily philosophical rather than mathematical. Thinkers such as Cicero distinguished between events occurring by chance and those determined by fate, foreshadowing later developments in probability theory [**cicero45bce**]. These early ideas, though lacking quantitative formalism, provided the intellectual foundation for later scientific inquiry into randomness and causality.

A significant shift occurred during the late medieval and early Renaissance periods, when more rigorous mathematical ideas began to shape. Italian mathematician and gambler Gerolamo Cardano (1501–1576) made substantial contributions to the mathematical analysis of chance. His work *"Liber de Ludo Aleae"* (*"Book on Games of Chance"*) [**cardano1663ludo**], posthumously published in 1663, is one of the earliest known texts to explore probability through the analysis of gambling problems. However, Cardano's reasoning, while insightful, lacked the symbolic clarity and mathematical rigour of modern probability theory. Readers consulting the original manuscript will notice an ambiguous and sometimes inconsistent symbolic system, quite unlike the formal structures we use nowadays.

The formalization of probability as a mathematical discipline did not occur until the 17th century, most notably through the seminal correspondence between Blaise Pascal and Pierre de Fermat. Their work, motivated by problems such as finding a fair division of stakes in interrupted games of chance, introduced foundational concepts such as combinatorics, expected value, and variance [**devlin2008unfinished**]. These developments paved the way for later contributions by Christiaan Huygens, who in 1657 wrote the first published textbook on

probability *"De Ratiociniis in Ludo Aleae"* (*"On Reasoning in Games of Chance"*), and Jacob Bernoulli, whose 1713 *"Ars Conjectandi"* (*"The Art of Conjecturing"*) remains among the most influential early texts in the field. Their works, alongside with many others, collectively laid the groundwork for the probabilistic and statistical methods that foreshadow modern scientific reasoning [**huygens1657ratiociniis**], [**bernoulli1713ars**], [**hald1990history**].

The modern axiomatic formulation of probability was introduced in the early 20th century by the Russian mathematician Andrey Kolmogorov. In his 1933 monograph *"Grundbegriffe der Wahrscheinlichkeitsrechnung"* (*"Foundations of the Theory of Probability"*) [**kolmogorov1933grundbegriffe**], Kolmogorov synthesized classical and frequentist ideas into a rigorous mathematical framework based on measure theory. His axioms remain the standard foundation for probability theory to this day. It may seem surprising that a concept with such ancient origins was not formally axiomatized until relatively recent times, and we will return to Kolmogorov's formulation and its implications in greater detail in Chapter 5. Nevertheless, philosophical discussions about the interpretation of probability and its relation to the physical sciences - especially in the context of determinism, epistemology and modern topics such as quantum mechanics - predate Kolmogorov's formulation and continue to evolve to this day.

7.1 Motivation and philosophy

7.2 Mathematical definition

7.3 Some examples of conditional probility

7.4 Stochasticity and Markov processes

Exercises

1. Exercise [...].

2. Exercise [...].

3. Exercise [...].

Solutions

1. Solution [...].

2. Solution [...].

3. Solution [...].

Chapter A

Vectors and matrices: a quick review

Chapter B

Functions and derivatives: a quick review

Chapter C

Integral calculus: a quick review

Bibliography

[1] John P. A. Ioannidis. *"Why Most Published Research Findings Are False"*. In: *PLoS Medicine* 2.8 (2005), e124.

[2] David Spiegelhalter. *"The Art of Statistics: How to Learn from Data"*. Basic Books, 2019.

[3] Morris H. DeGroot and Mark J. Schervish. *"Probability and Statistics"*. 4th ed. Pearson, 2012.

[4] P. S. Bandyopadhyay and M. R. Forster, eds. *"Philosophy of Statistics"*. Vol. 7. Handbook of the Philosophy of Science. Elsevier, 2011.

[5] M. Diez, D. Barr, and Mine Çetinkaya-Rundel. *"OpenIntro Statistics"*. OpenIntro, 2025.

[6] Hossein Pishro-Nik. *"Introduction to Probability, Statistics and Random Processes"*. Kappa Research LLC, 2014.

[7] Irving L. Finkel. *"The Ancient Origins of Dice"*. In: *Antiquity* 81.314 (2007), pp. 176–187.

[8] F. N. David. *Games, Gods and Gambling*. Griffin, 1962.

[9] Marcus Tullius Cicero. *De Divinatione*. Ancient Sources Edition, 45 BCE.

[10] Gerolamo Cardano. *Liber de Ludo Aleae*. Paris: Apud Joannem Baptistam Ferrarium, 1663.

[11] Keith Devlin. *The Unfinished Game*. Basic Books, 2008.

[12] Christiaan Huygens. *De Ratiociniis in Ludo Aleae*. Leiden: Elzevier, 1657.

[13] Jacob Bernoulli. *Ars Conjectandi*. Basel: Thurneysen Brothers, 1713.

[14] Anders Hald. *A History of Probability and Statistics and Their Applications before 1750*. Wiley, 1990.

[15] Andrey Kolmogorov. *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Berlin: Springer, 1933.