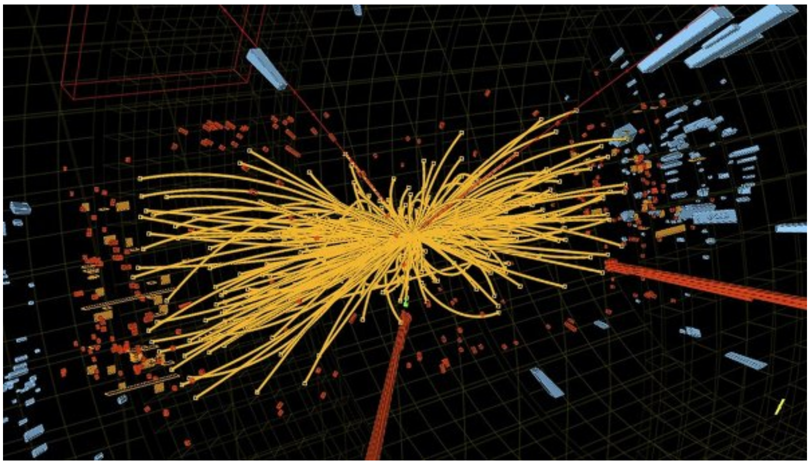
Alessandro Candolini

Statistical Data Analysis

Theory, methods, algorithms and modern techniques
from high-energy and computational physics to information technology



March 14, 2021

E-MAIL: alessandro.candolini@gmail.com

First printing, September 2014

> “There is nothing more practical than a good theory.”
>
> — Kurt Lewin, 1945

CONTENTS

v Advanced material and applications

vi Appendix

PREFACE

This preface and these notes are written from a physicist' perspective. This will ultimately drive the approach towards the subject and it will bias the selection of topics presented here, the examples discussed throughout this work, and the level of mathematical rigour (higher than what is customary in several books of practical statistics) , leading to prioritize definitely what is likely to be of major relevance or more common for an audiance of physicists, nevertheless trying to emphasize at the same time the impacts, implications and applications in contemporary information technology and big data community.

A through understanding of traditional and modern statistical inference (theory, methodologies, techniques, and algorithms) has always been playing quite a prominent role in the education and training of a scientist (in general) and of a physicist (in particular), providing the framework and the tools to analyse in a reliable and rigorous way the outcomes of either actual experiments or virtual computer simulations, in order to get insights, estimate parameters and their accuracy and precision, make comparisons between theory and experiments, validate hypothesis tests, assess the validity of a theoretical model and its underlying assumptions against experimental data, and make predictions.

The need of a powerful background in statistics has probably become even more demanding in recent years for practitioners in data analysis. The exponential growth of computing resources during the past few decades has made the last decades increasingly data-intensive. Beyond "traditional methods (e. g., probabilitic modelling, linear and non-linear regression theory, confidence intervals, theory of estimators, statistical significance, hypothesis test, time series analysis, etc), new computer-intensive powerful techniques has emerged in the last decades, including new progress in Bayesian statistics, maximum entropy methods, computational statistics (montecarlo methods, expectation-maximization algorithm, hidden Markov models, resampling techniques such as the statistical bootstrap, etc), statistical learning (artificial neural networks, support vector machines, k-means clustering, and related topics in pattern recognition, etc), just to mention few examples.

In this respect, physics has always been a rich source of probabilistic models. (statistical mechanics, etc). statistical mechanics (lattice spin models such as Potts models or spin glasses, ergodic theory), group theoretical methods and concepts borrower from relativistic quantum field theory and high energy particle physics provide valuable insights to better understand the mechanism behind these tools.

From a theoretical viewpoint, is a beautiful and fashinating field of research, sharing important connections with information theory, mathematics, artificial intelligence and robotics, physics (e. g., statistical mechanics), etc. But the relevance of these methods is far from being just of theoretical interest.

From a practical viewpoint, statistical methods are undergoing a tremendus development in recent years. This is not only relevant for physics however. playing an increasing widespread role in modern information technology and big data community, Warning about the misleading analysis and abuse of some statistical technique, nevertheless helping in this way to popularize some cornerstone machine learning techniques. while the size of data is posing challanging questions on how to efficiently pro and they have proved successful to approach data analysis in many different areas, not only including physics, impacting

Pervasiveness of distributed cloud-based clusters, crucially providing both the computational and the storage resources necessary for processing massive datasets;

The data-driven paradigm is emerging, the internet of things (IoT) is promising an even more increasing volume of data to be digested in the upcoming years, pressing for successful strategies capable of processing at high rate (or even in real-time) massive distributed datasets.

Mastering these topics requires a mix of knowledge, skills and experience to succeed, together with high proficiency with related algorithms and their implementation. Familiarity with algorithms and ability to efficiently implement them is an essential part, and it can become even less trivial when the data requires distributed or high-performance computing.

Two schools: frequentist and Bayesian. Unfortunately, there is no general consensus among statisticians (and even among physicists)

The aim of this notes is slightly ambitious. First, thoroughly discuss at graduate level the theoretical underpinnings of traditional and modern statistical tools and their mathematical foundations, establishing the main results and developing theoretical tools. I firmly believe that it is not possible to perform a reliable data analysis without a solid knowledge of the theory behind those tools: theory is essential to understand and control the range of validity of a given method, to understand how to inspect the results, and to highlight artifacts (which are always present), and in case to develop custum solutions for the problem at hand. In the second part, we will discuss concrete examples, together with actual implementation of algorithms to handle data analysis in real examples. This will help to illustrate pitfalls, best practises, implementation strategies, libraries and platforms, etc

Trieste, March 14, 2021 Alessandro Candolini

Part I

PROBABILITY

This part deals with the theory of probability. The rigorous definition of probability is given, according to Kolmogorov. Then, its properties are derived, a number of recurrent common probability distributions are introduced together with tools to handle them (among them are: probability distributions, characteristic functions, moments, etc) and finally few important limit theorems are established. The theory of probability is important by its own and it is at the same time the natural prerequisite for statistics.

1

KOLMOGOROV AXIOMS

Rigorous foundation of mathematical probability is provided by Kolmogorov's measure-theoretical approach[*]. In this framework, the mathematical notion of probability is defined in an abstract way by all and only the rules that any probability must fulfill. Among the strenghts of this approach is the fact that it completely *decouples* the definition of probability from empirical notions and conceptual interpretations. Kolmogorov's axioms provides the rules of the game, allowing to establish properties of probability and ways to manipulate mathematical probabilities on a general ground, without relying on a specific way to assign values to a probability. In order to apply the theory to study a particular situation, one should supply the framework with a concrete "probabilistic model" (i. e., she/he should assign a probability) suitable for the example at hand; the choice and the construction of the probabilistic model is an empirical "input" which should be provided externally to the Kolmogorov framework.

It is worth mentioning that Kolmogorov's approach to mathematical probability is not free from criticism. An introductory review of approaches to probability and their conceptual interpretation are briefly sketched in section 1.9.

1.1 CRASH COURSE IN ABSTRACT MEASURE THEORY

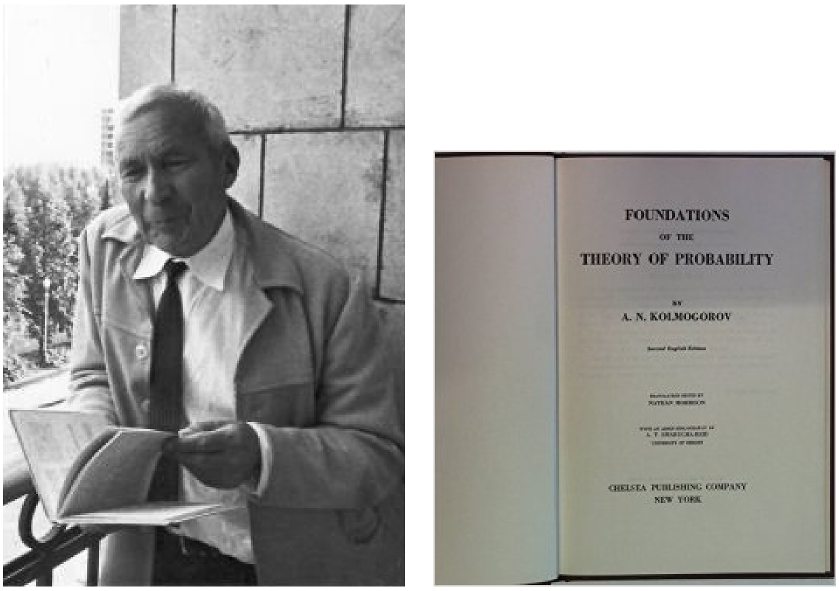
Kolmogorov axioms are rooted in abstract measure theory. Since this is a fairly advanced topic in mathematics, we give here a brief account of the very basic of this subject in order the Reader to be accustomed to it at least to some extend[†].

Roughly speaking, the ultimate goal is that of appending some reasonable notion of "measure" to the subsets of a given non-empty set (the precise mathematical definition of measure coming soon). Troubles eventually arises however when dealing with sets having uncountable many items if one tries to apply the notion of measure naively to all subsets of such sets. The famous Banach-Tarsky paradox is the prototype of the kind of difficulties one encounters.

> For those who are not familiar, a discussion of the Banach-Tarsky paradox can be found, e. g., in K. Stromberg, "The

[*] A. N. Kolmogorov, *Foundations of the theory of probability*, 2nd ed., Chelsea Pub. Co., New York 1956.

[†] For a more comprehensive treatment and its application to rigorous probability, refer to, e. g., Kolmogorov, *op. cit.*; P. Billingsley, *Probability and Measure*, 3rd ed., Wiley Series in Probability and Statistics, Wiley, Hoboken, NJ 1995; J. S. Rosenthal, *A first look at rigorous probability theory*, World Scientific, Singapore 2006.



FOUNDATIONS
OF THE
THEORY OF PROBABILITY

BY
A. N. KOLMOGOROV

Second English Edition

[illegible]
NATHAN MORRISON

[illegible]
A. T. BHARUCHA-REID
[illegible]

CHELSEA PUBLISHING COMPANY
NEW YORK

(A) (B)

Figure 1.1. (a) Russian mathematician Andrey Nikolaevich Kolmogorov (1903–1987). Source: Wikipedia. (b) Title page of english edition of Kolmogorov's *Foundations of the theory of probability*.

Banach-Tarski Paradox", *The American Mathematical Monthly*, 86, 3 (1979), pp. 151-161 and references therein. Proof of Banach-Tarsky paradox is highly non-trivial and it relieas technically on the axiom of choice. Nevertheless, the statement of the theorem itself is paradigmatic on the kind of difficulties one meets when dealing with uncountable sets.

To overcome the aforementioned difficulties when trying to "measure" a subset of a set, at a techical level one resorts to constrain the definition of "measure" over suitable collections of subsets called σ-algebras.

Let Ω be any set. Hereafter, 2^{Ω} will denote the "power set" of Ω, i. e., the set of all and only the subsets of Ω. In axiomatic set theory, the existence of 2^{Ω} for every set Ω is a postulate.[*]

Definition 1.1 (σ-algebra): Let Ω be any *non-empty* set and $\Sigma \subseteq 2^{\Omega}$ any arbitrary subset of the power set. Σ is called a σ-algebra over Ω if

(a) $\Omega \in \Sigma$;

(b) for every $A \subseteq \Omega$, if $A \in \Sigma$ then $\complement A\Omega \in \Sigma$;

(c) for every sequence $(A_n)_{n \in \mathbb{N}} \colon \mathbb{N} \to \Sigma$ of subsets of Ω belonging to Σ, $\bigcup_{n \in \mathbb{N}} A_n \in \Sigma$.

[*] The axiom of power set is one of the Zermelo-Fraenkel axioms. It allows a simple definition of the Cartesian product of two sets, which in turn allows to define the notion of application between sets.

In words: (a) implies that any σ-algebra is *non-empty*. (b) means a σ-algebra is closed under the complementation: if an arbitrary subset A belongs to Σ then also its complementary $\complement A\Omega$ belogs to it. (c) means that the σ-algebra is closed under *countable* unions. This latter assumption might look rather technical at this level. Why not just restrict ourselves to the union of a *finite* (instead of countably infinite) number of subsets? As we shall see later, (c) allows to handle convergence theorems, to take limits of sequences of subsets, etc.

Remark. The empty set $\varnothing$ also belongs to any Σ-algebra. Proof: by (a) Ω belongs to Σ; (b), its complement $\complement\Omega\Omega = \varnothing$ must belong to Σ.

Example 1.1: Given any set Ω, the subset $\{\,\Omega, \varnothing\,\} \subseteq 2^{\Omega}$ consisting only of the empty set $\varnothing$ and the set Ω itself is a σ-algebra over Ω; it is called the "minimal" or "trivial" σ-algebra over Ω.

Example 1.2: Let Ω be a non-empty set having a non-trivial subset $A \subseteq \Omega$. The subset $\left\{\, \Omega, A, \complement A\Omega, \varnothing \,\right\} \subseteq 2^{\Omega}$ is a σ-algebra over Ω.

Example 1.3: The power set 2^{Ω} of a set Ω is a σ-algebra over Ω, called the "discrete" σ-algebra.

A set endowed with a sigma algebra is called a "measurable space"; the reason for this name is that it is always possible to define a "measure" on it. More formally, we have the following definition.

Definition 1.2 (Measurable space): An ordered pair (Ω, Σ) where Ω is an arbitrary set and Σ is any σ-algebra over Ω is called a "measurable space".

Lemma 1.1: Let (Ω, Σ) be any measurable space. Then, for every sequence $(A_n)_{n\in\mathbb{N}} \colon \mathbb{N} \to \Sigma$ of elements in the σ-algebra Σ, we have

$$\bigcap_{n\,\in\mathbb{N}} A_n \in \Sigma\,,$$

i. e., any σ-algebra is also closed with respect to countable intersection.

Proof. DeMorgan rules + induction. ■

1.2 SET THEORY DEFINITION OF PROBABILITY

Dice: ⚀

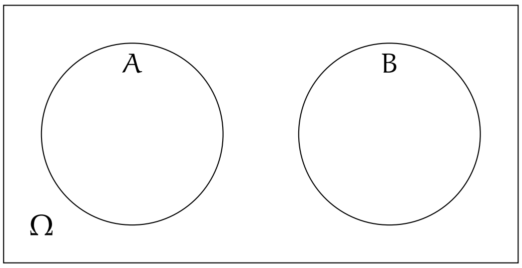


Figure 1.2. Union

1.3 FIRST PROPERTIES OF PROBABILITY

1.4 BOREL-CANTELLI LEMMAS

1.5 DISCRETE PROBABILITY SPACE

Before continuing with the discussion of probability measures in their full generality, it is helpful to consider the simpler case where the sample space Ω is finite or countable.

1.5.1 *Dices*

1.5.2 *Urns*

1.5.3 *Coupon collector problem*

1.6 BERTRAND PARADOX

1.7 CONDITIONAL PROBABILITIES AND INDEPENDENT EVENTS

Definition 1.3 (Conditional probability): Let $(\Omega, \Sigma, P(\cdot))$ be a probability space and $A, B \in \Sigma$ such that $P(B) > 0$. The "conditional probability of A given B is defined as

$$P(A \mid B) := \frac{P(A \cap B)}{P(B)} . \tag{1.1}$$

Remark. The definition makes sense since $A, B \in \Sigma$ implies $A \cap B \in \Sigma$ (being Σ a σ-algebra). The condition $P(B) > 0$ is required otherwise the denominator on the right-hand side would vanish.

Set interpretation of definition 1.3: $P(A \mid B)$ can be visualized as if we were restricting the sample space to the subset B alone. The logic behind this equation is that if the outcomes are restricted to B, this set serves as the new sample space.

1.8 bayes' theorem

1.9 a panorama of approaches to probability

No attempt will be make at discussing the deep philosophical implications of applying the Kolmogorov's framework.

A review is done by W. Feller, *An introduction to probability theory and its applications*, 3rd ed., Wiley mathematical statistics series, Wiley, New York 1966, vol. 1, § 1. See also R. T. Cox, "Probability, frequency and reasonable expectation", *Am. J. Phys.* 17 (1946), pp. 1-13. For a review of Cox, see K. S. Van Horn, "Constructing a logic of plausible inference: A guide to Cox's theorem", *International Journal of Approximate Reasoning*, 34 (2003), pp. 3-24. See also E. Jaynes and G. Bretthorst, *Probability Theory. The Logic of Science*, Cambridge University Press, Cambridge 2003.

1.10 what's the use of all this?

The usefulness of abstract measure theory extends far beyond the Kolmogorov's foundation of mathematical probability, and find broad applications in other areas of applied mathematics and physics as well[*].

From a practical viewpoint, it would have been possible to develop a more intuitive theory of probability relying on the more familiar concepts of probability distributions and densities, to be developed in the next chapter[†]; these tools are very useful to study the simplest cases. Nevertheless, such approach would have been less general and inadeguate to formulate rigorously some limit theorems. Using sets rather than distributions to define the probability allows for a more general and powerful treatment, even though it leads to some mathematically-intensive formulation at the beginning.

Kolmogorov's axioms provide a theoretical framework to define probability in a abstract way, emphasizing its mathematical "structure" and properties, regardless on the specific empiric applications. In order to to apply this framework to real-life applications, one must first build a suitable probabilistic model (i. e., provide a specific choice of the probability measure) to this framework. Statistics (to be discussed in later chapters) is concerned with inferring the "most appropriate" (a least to some extend) probabilistic model (out of some range of models) from inspection of the available empirical data. The basic properties

[*] Just to make one example, refer to, e. g., A. Fasano and S. Marmi, *Analytical Mechanics: An Introduction*, Oxford graduate texts in mathematics, Oxford University Press, Oxford 2006, § 13, for an application of measure-theoretical concepts to classical ergodic theory, chaotic dynamical systems and classical equilibrium statistical mechanics.

[†] An example of this approch is offered by the first chapter of N. G. Van Kampen, *Stochastic processes in physics and chemistry*, 3rd ed., Elsevier, Amsterdam 2007.

of probability established on this chapter will be use throughout all the following chapters.

The importance of Bayes theorem cannot be understimated: beyond being of major importance by its own, it plays a decisive role in the foundation of Bayesian inference (more on this in later chapters).

REFERENCES FOR CHAPTER 1

Billingsley, P., *Probability and Measure*, 3rd ed., Wiley Series in Probability and Statistics, Wiley, Hoboken, NJ 1995. (Cit. on p. 3.)

Cox, R. T., "Probability, frequency and reasonable expectation", *Am. J. Phys.* 17 (1946), pp. 1-13. (Cit. on p. 7.)

Fasano, A. and S. Marmi, *Analytical Mechanics: An Introduction*, Oxford graduate texts in mathematics, Oxford University Press, Oxford 2006. (Cit. on p. 7.)

Feller, W., *An introduction to probability theory and its applications*, 3rd ed., Wiley mathematical statistics series, Wiley, New York 1966, vol. 1. (Cit. on p. 7.)

Jaynes, E. and G. Bretthorst, *Probability Theory. The Logic of Science*, Cambridge University Press, Cambridge 2003. (Cit. on p. 7.)

Kolmogorov, A. N., *Foundations of the theory of probability*, 2nd ed., Chelsea Pub. Co., New York 1956. (Cit. on p. 3.)

Rosenthal, J. S., *A first look at rigorous probability theory*, World Scientific, Singapore 2006. (Cit. on p. 3.)

Stromberg, K., "The Banach-Tarski Paradox", *The American Mathematical Monthly*, 86, 3 (1979), pp. 151-161. (Cit. on p. 3.)

Van Horn, K. S., "Constructing a logic of plausible inference: A guide to Cox's theorem", *International Journal of Approximate Reasoning*, 34 (2003), pp. 3-24. (Cit. on p. 7.)

Van Kampen, N. G., *Stochastic processes in physics and chemistry*, 3rd ed., Elsevier, Amsterdam 2007. (Cit. on p. 7.)

2 RANDOM VARIABLES

Discrete and continuous, univariate and multivariate real random variables are discussed, together with a bunch of practical tools helping to handle with them, specifically: probability density function, cumulative distribution function, moments (i. e., mean, variance, skewness, kurtosis, etc). All of this provides a more concrete, visualizable approach to probability built on top of the construction of the previous chapter by means of probability measures; it should be stressed however that the probability measures provides a more general framework. This chapter also covers functions of random variables, change of variables (i. e., what physicists are accustomed to call "propagation of errors"), etc. Generating functions — another useful tool to deal with probability distributions — are postponed to chapter 4: an entire chapter is devoted to them since their theory and their application deserve special attention. More general random variables (e. g., matrix-valued random variables) although useful are not presented here (see however chapter 11, which is devoted to a brief account of random matrix theory).

2.1 DEFINITIONS AND BASIC PROPERTIES

Let (a) $(\Omega, \Sigma, P(\cdot))$ be any probability space; (b) (E, ϵ) be any measure space; (c) $X\colon \Sigma \to E$ be any (measurable) function from Σ to E. In this context: Ω is called the "sample space", E is called "state space", X is called a "random variable", "aleatory variable" or "stochastic variable".

Whenever the range $X(\Sigma) \subseteq E$ of X has finite or countably infinite cardinality, the random variable X is called a "discrete" random variable. When the range of X is uncountably infinite instead, X is called a "continous" random variable.

In this chapter, we will be mainly interested in the case of *real*-valued (discrete or continuous) univariate (i. e., $E \subset \mathbb{R}$) or multivariate (i. e., $E \subseteq \mathbb{R}^N$, $N \in \mathbb{N}$, $N \geqslant 2$) random variables (the Borel algebra being the underlying σ-algebra in the state space). For such case, it makes sense and it proves useful to introduce the important concepts of moments, distribution functions, etc. (Those definitions are slightly different whether the random variable is discrete or continuous and univariate or multivariate, as we will see shortly.) However, the definition of random variables applies to more general settings; for example, in chapter 11 the case where E is a suitable space of matrices will be taken into account. Other scenarios are also possible (e. g., random graphs, etc) some being relevant for statistical mechanics, machine learning, etc (see later chapters).

It is worth notice that random variables take value over the set E, nevertheless E is not the probability space itself; the random variable instead is defined as a mapping defined over a suitable σ-algebra of an underlying probability space Ω and targetting the (measure) state space E.

What is the benefit of having defined a random variable as a mapping from a probability space to the state space E instead of having E itself equipped with a suitable probability measure?

2.2 PROBABILITY DISTRIBUTION

In most applications, the underlying probability space defining a random variable remains hidden. Instead, a "distribution function" is introduced which

2.2.1 *Univariate discrete random variable*

2.2.2 *Univariate continuous random variable*

2.2.3 *Multivariate discrete random variable*

2.2.4 *Multivariate continuous random variable*

2.3 MOMENTS OF A DISTRIBUTION

2.4 TRANSFORMATIONS OF RANDOM VARIABLES

2.5 BUFFON'S NEEDLE

2.6 WHAT'S THE USE OF ALL THIS?

3

CATALOG OF PROBABILITY DISTRIBUTIONS

The tools developed in chapter 2 are applied to a list of important discrete and continuous, univariate and multivariate probability distributions. These distributions are the building blocks of several probabilistic models and are relevant for later applications in statistics, so it is worth browsing through all the details and calculations.

3.1 DISCRETE DISTRIBUTIONS

3.1.1 *Bernoulli distribution*

3.1.2 *Binomial distribution*

3.1.3 *Poisson distribution*

3.1.4 *Hypergeometric distribution*

3.1.5 *Multinomial distribution*

3.2 CONTINUOUS DISTRIBUTION

3.2.1 *Uniform distribution*

3.2.2 *Exponential distribution*

3.2.3 *Cauchy distribution*

3.2.4 *Univariate normal distribution*

3.2.5 *Multivariate normal distribution*

3.2.6 *Chi-square distribution*

3.2.7 t*-student distribution*

3.2.8 *Gamma distribution*

3.2.9 *Beta distribution*

3.2.10 *Weibull distribution*

3.3 WHAT'S THE USE OF ALL THIS?

4 GENERATING FUNCTIONS

Generating functions are a overwealming weapon to attack irresistable problems which fail a more straightforward approach. In this section we will introduce the notion of generating function, we will establish the relevant properties and we apply the technique of generating functions to few prototypical examples, in particular involving the famous Fibonacci numbers and the integer partitions. In later chapters, the technique of generating functions will prove useful for example to deal with Galton-Walton problem. In statistical mechanics, the generating function approach is related to the grancanonical partition function. The main source for this chapter is H. S. Wilf, *Generatingfunctionology*, 2nd ed., Academic Press, London 1994.

4.1 FIBONACCI NUMBERS AS A PROTOTYPE

The sequence $(F_n)_{n\in\mathbb{N}}$ of Fibonacci numbers can be defined in a number of equivalent ways and historically made its first appearance in connection with a combinatorial model (the rabbit's population model described below). One definition is the following: the sequence $(F_n)_{n\in\mathbb{N}}$ of Fibonacci numbers is defined as the *unique* solution of

$$F_{n+2} = F_n + F_{n+1}, \quad n \in \mathbb{N} \tag{4.1}$$

satisfying the initial conditions

$$F_0 = 1\,, \quad F_1 = 1\,. \tag{4.2}$$

Equation (4.1) is a second-order homogeneous linear recursive equation with constant coefficients, the general theory of such equations ensure that there exists one and only one solution of it satisfying the initial valued problem.

4.2 DEFINITION AND PROPERTIES OF GENERATING FUNCTIONS

4.3 SOLUTION OF FIBONACCI RECURRENCE USING GENERATING FUNCTIONS

REFERENCES FOR CHAPTER 4

Wilf, H. S., *Generatingfunctionology*, 2nd ed., Academic Press, London 1994. (Cit. on p. 13.)

5 LIMIT THEOREMS

5.1 INEQUALITIES

Part II

ADDITIONAL TOPICS IN PROBABILITY

The material in this part might appear to be slightly departing from the core topics of these notes, namely statistics, however much of the material is relevant for stochastic modelling.

6 STOCHASTIC PROCESSES

7 MARKOV PROCESSES

An important class of stochastic processes is provided by the so-called Markov processes. Markov processes are ubiquitous. They are a relevant ingredient for stochastic modelling and they provide the theoretical basis of important statistical and coomputational techniques, ranging from Markov-chain Montecarlo to hidden Markov models, discussed in later parts. This chapter serves as an introduction to their definitions and first properties, and illustrates few beginning examples. The material is indispensable also for later chapters.

The following stochastic processes *need not* to be Markovian. They can be Markovian in some special cases but they do not need to be Markovian in general and allow for special treatment. For example, the simple random walk is Markovian; the self-avoiding random walk is a prototypical example of a *non*-Markovian process. For this reason, these processes will be discuss in later chapters:

- random walks;
- branching processes.

As prototypes of Markov processes we will discuss:

- Erenfest urn;
- Google™ PageRank™ ranking algorithm;
- Applications of Markov processes to shopping;
- Solution of the coupon problem via Markov chains.

In later chapters, we will discuss modern topics like

- Markov chain Montecarlo and Metropolis-Hasting algorithm in computational statistical mechanics and computational Bayesian inference;
- Hidden Markov models.

Both heavily relies on Markov processes. The stochastic calculus and the chapter on stochastic differential equation makes use of Wigner integrals which are based on Markov processes.

7.1 MARKOV PROPERTY

Intuitively speaking, Markov processes are *memoryless*: the probability of evolving from the current state to the next one does not depend on

Figure 7.1. Russian mathematician Andrei Andreyevich Markov (1856–1922), the father of Markov processes.

the previous history of the process. The exact mathematical implementation of this concept can look a bit technical for Markov processes defined on continuous state spaces and continuous index space, but for finite settings it should be easier to grasp. For this reason, we first define the general notion for arbitrary (continuous) processes but later we will mainly focus on the finite setting, where Markov processes are more often called "Markov chains". Markov chains will be enough for most of our purposes. General Markov processes occur however in many situations (Brownian motion, Markovian evolution of open quantum systems, etc); for this reason, we give the general definitions.

7.2 ERGODICITY

7.3 STATIONARY DISTRIBUTION

7.4 REMARKS ON NON-MARKOVIAN PROCESSES

[*]

7.5 THE COUPON PROBLEM REVISITED

7.6 ERENFEST URN

7.7 SHOPPING MODEL

7.8 GOOGLE$^{\text{TM}}$ PAGERANK$^{\text{TM}}$ ALGORITHM

In the early days of web search engine,

[*] N. G. Van Kampen, "Remarks on Non-Markov Processes", *Brazilian Journal of Physics*, 28, 2 (June 1998).

Web search engines usually work in this way: (a) Web Crawling (b) Indexing (c) Ranking First of all, an automated Web crawler retrieves and stores information from the HTML markup of several many web pages. Data is analyzed and stored in a index database for use in later queries. Finally, when a query is performed by the user, the list of matching results needs to be sorted by some criteria (ranking).

The PageRank™ algorithm was has a lot of interesting mathematics behind it. In this note we will use the PageRank™ algorithm a a prototypical model and as a pathfinder to explore and to illustrate some interesting mathematics behind it. We will give also a toy implementation of the algorithm to start playing with it so to discover some of its features.

REFERENCES FOR CHAPTER 7

Van Kampen, N. G., "Remarks on Non-Markov Processes", *Brazilian Journal of Physics*, 28, 2 (June 1998). (Cit. on p. 22.)

8 POISSON PROCESSES

Poisson processes and their characterization is presented.

*8.1 LORENTZ CLASSICAL MODEL OF ELETTRICAL CONDUCTIVITY IN METALS

8.2 WHAT'S THE USE OF ALL THIS?

9 RANDOM WALKS

10

BRANCHING PROCESSES

For a superb account of branching processes, see Harris[*].

REFERENCES FOR CHAPTER 10

Harris, T. E., *The Theory of Branching Processes*, Dover phoenix editions, Dover Publications, New York 2002. (Cit. on p. 29.)

[*] T. E. Harris, *The Theory of Branching Processes*, Dover phoenix editions, Dover Publications, New York 2002.

11 GLIMPSE AT RANDOM MATRIX THEORY

12

ITÔ STOCHASTIC DIFFERENTIAL EQUATIONS

Part III

STATISTICAL INFERENCE

After introducing the basis of descriptive statistics, the core of frequentist and Bayesian inference are throughly studied.

Part IV

GLIMPSE AT STATISTICAL LEARNING

Machine learning techniques have been providing state-of-art performance in the analysis of massive datasets (even though some analysis at a rather heuristic level). This part deals with a rather introdoctory overview of some key algorithms rooted in machine learning and their statistical meaning.

13 A PANORAMA OF MACHINE LEARNING TECHNIQUES

The main source of material for this part is provided by:

- Hastie *et al.*[*]
- Bishop[†]

13.1 what's the use of all this?

references for chapter 13

Bishop, C., *Pattern Recognition and Machine Learning*, Information Science and Statistics, Springer, New York 2006. (Cit. on p. 39.)

Hastie, T., R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning. Data Mining, Inference, and Prediction*, 2nd ed., Springer Series in Statistics, Springer, New York 2009. (Cit. on p. 39.)

[*] T. Hastie *et al.*, *The Elements of Statistical Learning. Data Mining, Inference, and Prediction*, 2nd ed., Springer Series in Statistics, Springer, New York 2009.

[†] C. Bishop, *Pattern Recognition and Machine Learning*, Information Science and Statistics, Springer, New York 2006.

14

ARTIFICIAL NEURAL NETWORKS

14.1 models of electrical conductivity in neurons

14.2 perceptron

14.3 feed-forward neural networks and backpropagation algorithm

14.4 kohonen neural networks

14.5 associative memories

14.6 what's the use of all this?

15

SUPPORT VECTOR MACHINES

15.1 what's the use of all this?

16

GLIMPSE AT DEEP LEARNING

Part V

ADVANCED MATERIAL AND APPLICATIONS

A number of modern additional topics in statistics is covered, in particular modern computational tools (expectation-maximization algorithm, hidden-markov models).

17 KALMAN FILTERING

17.1 what's the use of all this?

18

MAXIMUM ENTROPY PRINCIPLE

18.1 what's the use of all this?

19 EXPECTATION-MAXIMIZATION ALGORITHM

19.1 what's the use of all this?

20 HIDDEN MARKOV MODELS

20.1 what's the use of all this?

21

POTTS MODELS AND SPIN GLASSES FOR IMAGE RESTORATION

21.1 what's the use of all this?

22 UNFOLDING

22.1 what's the use of all this?

Part VI

APPENDIX

A

IMPLEMENTATION NOTES

It is worth summarizing few notes about implementation.

A.1 GLIMPSE AT FUNCTIONAL PROGRAMMING LANGUAGES

Functional programming languages has gained attraction over imperative ones in recent years due to their ability to handle effortlessly and in robust way concurrency and multithreading/multicore computing, making them particularly suitable to distributed frameworks.

Historically, concurrent programming was not the original motivation for functional programming. After the early functional-flavored language Lisp and the development of APL in the early 1960s by Iverson, functional programming was eventually formalized by John Bakus (the father of Fortran) in his cornerstone 1977 Turing Award Lecture *Can Programming Be Liberated From the von Neumann Style? A Functional Style and its Algebra of Programs*, where he presented his FP programming language and emphasizes how liberating a programming language from the traditional imperative style would have lead to a programming language closer to mathematical abstractions.

For several years, functional programming was relegated mainly as an academic tool. Nevertheless, several powerful functional programming languages have been developed in the last decades. It is worth mentioning at this point: Haskell, Erlang, OCam, Closure. Scala was build

A.2 REACTIVE PARADIGM

A.3 PYTHON

A.4 IPYTHON AND OTHER NOTEBOOKS

A.5 SCALA

A.6 CERN ROOT FOR HEP

In high-energy physics

A.7 DISTRIBUTED PROGRAMMING: APACHE SPARK

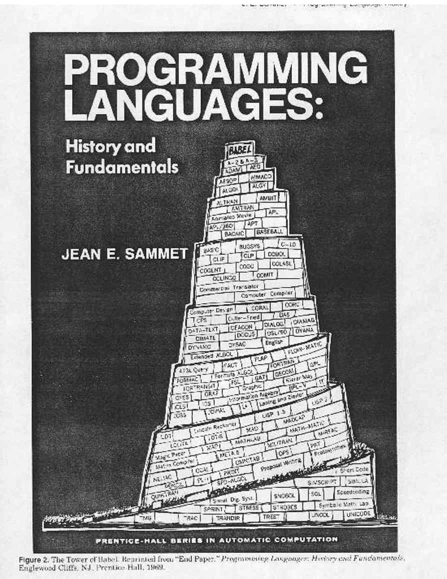

PROGRAMMING LANGUAGES:
History and Fundamentals
JEAN E. SAMMET
BABEL
PRENTICE-HALL SERIES IN AUTOMATIC COMPUTATION
Figure 2. The Tower of Babel. Reprinted from "End Paper." Programming Languages: History and Fundamentals, Englewood Cliffs, NJ, Prentice Hall, 1969.

Figure A.1. The Babel's Tower of programming languages, taken from the cover of J. Sammet's *Programming Languages: History and Fundamentals* (1969). Being a 1969 book, the picture misses several important additions, nevertheless it is still representative of the difficulty in choosing a suitable programming language.

REFERENCES FOR APPENDIX A

COMTET, L., *Advanced Combinatorics. The Art of Finite and Infinite Expansions*, Springer Netherlands, Dordrecht 1974.

FELLER, W., *An introduction to probability theory and its applications*, 3rd ed., Wiley mathematical statistics series, Wiley, New York 1966, vol. 1.

KRONENBURG, M. J., *The Binomial Coefficient for Negative Arguments*, version 2, 2015, arXiv: 1105.3689 [math].

ROSS, S. M., *A First Course in Probability*, 8th ed., Pearson Prentice Hall, Upper Saddle River, NJ 2010.

BIBLIOGRAPHY

REFERENCES FOR CHAPTER 1

Billingsley, P., *Probability and Measure*, 3rd ed., Wiley Series in Probability and Statistics, Wiley, Hoboken, NJ 1995. (Cit. on p. 3.)

Cox, R. T., "Probability, frequency and reasonable expectation", *Am. J. Phys.* 17 (1946), pp. 1-13. (Cit. on p. 7.)

Fasano, A. and S. Marmi, *Analytical Mechanics: An Introduction*, Oxford graduate texts in mathematics, Oxford University Press, Oxford 2006. (Cit. on p. 7.)

Feller, W., *An introduction to probability theory and its applications*, 3rd ed., Wiley mathematical statistics series, Wiley, New York 1966, vol. 1. (Cit. on p. 7.)

Jaynes, E. and G. Bretthorst, *Probability Theory. The Logic of Science*, Cambridge University Press, Cambridge 2003. (Cit. on p. 7.)

Kolmogorov, A. N., *Foundations of the theory of probability*, 2nd ed., Chelsea Pub. Co., New York 1956. (Cit. on p. 3.)

Rosenthal, J. S., *A first look at rigorous probability theory*, World Scientific, Singapore 2006. (Cit. on p. 3.)

Stromberg, K., "The Banach-Tarski Paradox", *The American Mathematical Monthly*, 86, 3 (1979), pp. 151-161. (Cit. on p. 3.)

Van Horn, K. S., "Constructing a logic of plausible inference: A guide to Cox's theorem", *International Journal of Approximate Reasoning*, 34 (2003), pp. 3-24. (Cit. on p. 7.)

Van Kampen, N. G., *Stochastic processes in physics and chemistry*, 3rd ed., Elsevier, Amsterdam 2007. (Cit. on p. 7.)

REFERENCES FOR CHAPTER 4

Wilf, H. S., *Generatingfunctionology*, 2nd ed., Academic Press, London 1994. (Cit. on p. 13.)

REFERENCES FOR CHAPTER 7

Van Kampen, N. G., "Remarks on Non-Markov Processes", *Brazilian Journal of Physics*, 28, 2 (June 1998). (Cit. on p. 22.)

REFERENCES FOR CHAPTER 10

Harris, T. E., *The Theory of Branching Processes*, Dover phoenix editions, Dover Publications, New York 2002. (Cit. on p. 29.)

REFERENCES FOR CHAPTER 13

BISHOP, C., *Pattern Recognition and Machine Learning*, Information Science and Statistics, Springer, New York 2006. (Cit. on p. 39.)

HASTIE, T., R. TIBSHIRANI, and J. FRIEDMAN, *The Elements of Statistical Learning. Data Mining, Inference, and Prediction*, 2nd ed., Springer Series in Statistics, Springer, New York 2009. (Cit. on p. 39.)

REFERENCES FOR APPENDIX A

COMTET, L., *Advanced Combinatorics. The Art of Finite and Infinite Expansions*, Springer Netherlands, Dordrecht 1974.

FELLER, W., *An introduction to probability theory and its applications*, 3rd ed., Wiley mathematical statistics series, Wiley, New York 1966, vol. 1.

KRONENBURG, M. J., *The Binomial Coefficient for Negative Arguments*, version 2, 2015, arXiv: 1105.3689 [math].

ROSS, S. M., *A First Course in Probability*, 8th ed., Pearson Prentice Hall, Upper Saddle River, NJ 2010.