Springer Series in Synergetics

Springer COMPLEXITY

Peter Schuster

Stochasticity in Processes

Fundamentals and Applications to Chemistry and Biology

Springer

Springer Complexity

Springer Complexity is an interdisciplinary program publishing the best research and academic-level teaching on both fundamental and applied aspects of complex systems – cutting across all traditional disciplines of the natural and life sciences, engineering, economics, medicine, neuroscience, social and computer science.

Complex Systems are systems that comprise many interacting parts with the ability to generate a new quality of macroscopic collective behavior the manifestations of which are the spontaneous formation of distinctive temporal, spatial or functional structures. Models of such systems can be successfully mapped onto quite diverse "real-life" situations like the climate, the coherent emission of light from lasers, chemical reaction-diffusion systems, biological cellular networks, the dynamics of stock markets and of the internet, earthquake statistics and prediction, freeway traffic, the human brain, or the formation of opinions in social systems, to name just some of the popular applications.

Although their scope and methodologies overlap somewhat, one can distinguish the following main concepts and tools: self-organization, nonlinear dynamics, synergetics, turbulence, dynamical systems, catastrophes, instabilities, stochastic processes, chaos, graphs and networks, cellular automata, adaptive systems, genetic algorithms and computational intelligence.

The three major book publication platforms of the Springer Complexity program are the monograph series "Understanding Complex Systems" focusing on the various applications of complexity, the "Springer Series in Synergetics", which is devoted to the quantitative theoretical and methodological foundations, and the "SpringerBriefs in Complexity" which are concise and topical working reports, case-studies, surveys, essays and lecture notes of relevance to the field. In addition to the books in these two core series, the program also incorporates individual titles ranging from textbooks to major reference works.

Springer Series in Synergetics

Founding Editor: H. Haken

The Springer Series in Synergetics was founded by Herman Haken in 1977. Since then, the series has evolved into a substantial reference library for the quantitative, theoretical and methodological foundations of the science of complex systems.

Through many enduring classic texts, such as Haken's *Synergetics and Information and Self-Organization*, Gardiner's *Handbook of Stochastic Methods*, Risken's *The Fokker Planck-Equation* or Haake's *Quantum Signatures of Chaos*, the series has made, and continues to make, important contributions to shaping the foundations of the field.

The series publishes monographs and graduate-level textbooks of broad and general interest, with a pronounced emphasis on the physico-mathematical approach.

More information about this series at http://www.springer.com/series/712

Peter Schuster

Stochasticity in Processes

Fundamentals and Applications to Chemistry and Biology

Springer

Peter Schuster
Institut für Theoretische Chemie
Universität Wien
Wien, Austria

ISSN 0172-7389 ISSN 2198-333X (electronic)
Springer Series in Synergetics
ISBN 978-3-319-39500-5 ISBN 978-3-319-39502-9 (eBook)
DOI 10.1007/978-3-319-39502-9

Library of Congress Control Number: 2016940829

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG Switzerland

Dedicated to my wife Inge

Preface

The theory of probability and stochastic processes is often neglected in the education of chemists and biologists, although modern experimental techniques allow for investigations of small sample sizes down to single molecules and provide experimental data that are sufficiently accurate for direct detection of fluctuations. Progress in the development of new techniques and improvement in the resolution of conventional experiments have been enormous over the last 50 years. Indeed, molecular spectroscopy has provided hitherto unimaginable insights into processes at atomic resolution down to time ranges of a hundred attoseconds, whence observations of single particles have become routine, and as a consequence current theory in physics, chemistry, and the life sciences cannot be successful without a deeper understanding of fluctuations and their origins. Sampling of data and reproduction of processes are doomed to produce interpretation artifacts unless the observer has a solid background in the mathematics of probabilities. As a matter of fact, stochastic processes are much closer to observation than deterministic descriptions in modern science, as indeed they are in everyday life, and presently available computer facilities provide new tools that can bring us closer to applications by supplementing analytical work on stochastic phenomena with simulations.

The relevance of fluctuations in the description of real-world phenomena ranges, of course, from unimportant to dominant. The motions of planets and moons as described by celestial mechanics marked the beginning of modeling by means of differential equations. Fluctuations in these cases are so small that they cannot be detected, not even by the most accurate measurements: sunrise, sunset, and solar eclipses are predictable with almost no scatter. Processes in the life sciences are entirely different. A famous and typical historical example is Mendel's laws of inheritance: regularities are detectable only in sufficiently large samples of individual observations, and the influence of stochasticity is ubiquitous. Processes in chemistry lie between the two extremes: the deterministic approach in conventional chemical reaction kinetics has not become less applicable, nor have the results become less reliable in the light of modern experiments. What has increased dramatically are the accessible resolutions in amounts of materials, space, and

time. Deeper insights into mechanisms provide new access to information regarding molecular properties for theory and practice.

Biology is currently in a state of transition: the molecular connections with chemistry have revolutionized the sources of biological data, and this sets the stage for a new theoretical biology. Historically, biology was based almost exclusively on observation and theory in biology engaged only in the interpretation of observed regularities. The development of biochemistry at the end of the nineteenth and the first half of the twentieth century introduced quantitative thinking concerning chemical kinetics into some biological subdisciplines. Biochemistry also brought a new dimension to experiments in biology in the form of *in vitro* studies on isolated and purified biomolecules. A second influx of mathematics into biology came from population genetics, first developed in the 1920s as a new theoretical discipline uniting Darwin's natural selection and Mendelian genetics. This became part of the theoretical approach more than 20 years before evolutionary biologists completed the so-called *synthetic theory*, achieving the same goal.

Then, in the second half of the twentieth century, molecular biology started to build a solid bridge from chemistry to biology, and the enormous progress in experimental techniques created a previously unknown situation in biology. Indeed, the volume of information soon went well beyond the capacities of the human mind, and new procedures were required for data handling, analysis, and interpretation. Today, biological cells and whole organisms have become accessible to complete description at the molecular level. The overwhelming amount of information required for a deeper understanding of biological objects is a consequence of two factors: (i) the complexity of biological entities and (ii) the lack of a universal theoretical biology.

Primarily, apart from elaborate computer techniques, the current flood of results from molecular genetics and genomics to systems biology and synthetic biology requires suitable statistical methods and tools for verification and evaluation of data. However, analysis, interpretation, and understanding of experimental results are impossible without proper modeling tools. In the past, these tools were primarily based on differential equations, but it has been realized within the last two decades that an extension of the available methodological repertoire by stochastic methods and techniques from other mathematical disciplines is inevitable. Moreover, the enormous complexity of the genetic and metabolic networks in the cell calls for radically new methods of modeling that resemble the mesoscopic level of description in solid state physics. In mesoscopic models, the overwhelming and for many purposes dispensable wealth of detailed molecular information is cast into a partially probabilistic description in the spirit of *dissipative particle dynamics* [358, 401], for example, and such a description cannot be successful without a solid mathematical background.

The field of stochastic processes has not been bypassed by the digital revolution. Numerical calculation and computer simulation play a decisive role in present-day stochastic modeling in physics, chemistry, and biology. Speed of computation and digital storage capacities have been growing exponentially since the 1960s, with a doubling time of about 18 months, a fact commonly referred to as Moore's law

[409]. It is not so well known, however, that the spectacular exponential growth in computer power has been overshadowed by progress in numerical methods, as attested by an enormous increase in the efficiency of algorithms. To give just one example, reported by Martin Grötschel from the Konrad Zuse-Zentrum in Berlin [260, p. 71]:

> The solution of a benchmark production planning model by linear programming would have taken – extrapolated – 82 years CPU time in 1988, using the computers and the linear programming algorithms of the day. In 2003 – fifteen years later – the same model could be solved in one minute and this means an improvement by a factor of about 43 million. Out of this, a factor of roughly 1 000 resulted from the increase in processor speed whereas a factor of 43 000 was due to improvement in the algorithms.

There are many other examples of similar progress in the design of algorithms. However, the analysis and design of high-performance numerical methods require a firm background in mathematics. The availability of cheap computing power has also changed the attitude toward exact results in terms of complicated functions: it does not take much more computer time to compute a sophisticated hypergeometric function than to evaluate an ordinary trigonometric expression for an arbitrary argument, and operations on confusingly complicated equations are enormously facilitated by symbolic computation. In this way, present-day computational facilities can have a significant impact on analytical work, too.

In the past, biologists often had mixed feelings about mathematics and reservations about using too much theory. The new developments, however, have changed this situation, if only because the enormous amount of data collected using the new techniques can neither be inspected by human eyes nor comprehended by human brains. Sophisticated software is required for handling and analysis, and modern biologists have come to rely on it [483]. The biologist Sydney Brenner, an early pioneer of molecular life sciences, makes the following point [64]:

> But of course we see the most clear-cut dichotomy between hunters and gatherers in the practice of modern biological research. I was taught in the pregenomic era to be a hunter. I learnt how to identify the wild beasts and how to go out, hunt them down and kill them. We are now, however, being urged to be gatherers, to collect everything lying about and put it into storehouses. Someday, it is assumed, someone will come and sort through the storehouses, discard all the junk and keep the rare finds. The only difficulty is how to recognize them.

The recent developments in molecular biology, genomics, and organismic biology, however, seem to initiate this change in biological thinking, since there is practically no way of shaping modern life sciences without mathematics, computer science, and theory. Brenner advocates the development of a comprehensive theory that would provide a proper framework for modern biology [63]. He and others are calling for a *new theoretical biology* capable of handling the enormous biological complexity. Manfred Eigen stated very clearly what can be expected from such a theory [112, p. xii]:

> Theory cannot remove complexity but it can show what kind of 'regular' behavior can be expected and what experiments have to be done to get a grasp on the irregularities.

Among other things, the new theoretical biology will have to find an appropriate way to combine randomness and deterministic behavior in modeling, and it is safe to predict that it will need a strong anchor in mathematics in order to be successful.

In this monograph, an attempt is made to bring together the mathematical background material that would be needed to understand stochastic processes and their applications in chemistry and biology. In the sense of the version of Occam's razor attributed to Albert Einstein [70, pp. 384–385; p. 475], viz., "everything should be made as simple as possible, but not simpler," dispensable refinements of higher mathematics have been avoided. In particular, an attempt has been made to keep mathematical requirements at the level of an undergraduate mathematics course for scientists, and the monograph is designed to be as self-contained as possible. A reader with sufficient background should be able to find most of the desired explanations in the book itself. Nevertheless, a substantial set of references is given for further reading. Derivations of key equations are given wherever this can be done without unreasonable mathematical effort. The derivations of analytical solutions for selected examples are given in full detail, because readers interested in applying the theory of stochastic processes in a practical context should be in a position to derive new solutions on their own. Some sections that are not required if one is primarily interested in applications are marked by a star ($\star$) for skipping by readers who are willing to accept the basic results without explanations.

The book is divided into five chapters. The first provides an introduction to probability theory and follows in part the introduction to probability theory by Kai Lai Chung [84], while Chap. 2 deals with the link between abstract probabilities and measurable quantities through statistics. Chapter 3 describes stochastic processes and their analysis and has been partly inspired by Crispin Gardiner's handbook [194]. Chapters 4 and 5 present selected applications of stochastic processes to problem-solving in chemistry and biology. Throughout the book, the focus is on stochastic methods, and the scientific origin of the various equations is never discussed, apart from one exception: chemical kinetics. In this case, we present two sections on the theory and empirical determination of reaction rate parameters, because for this example it is possible to show how Ariadne's red thread can guide us from first principles in theoretical physics to the equations of stochastic chemical kinetics. We have refrained from preparing a separate section with exercises, but case studies which may serve as good examples of calculations done by the reader himself are indicated throughout the book. Among others, useful textbooks would be [84, 140, 160, 161, 194, 201, 214, 222, 258, 290, 364, 437, 536, 573]. For a brief and concise introduction, we recommend [277]. Standard textbooks in mathematics used for our courses were [21, 57, 383, 467]. For dynamical systems theory, the monographs [225, 253, 496, 513] are recommended.

This book is derived from the manuscript of a course in stochastic chemical kinetics for graduate students of chemistry and biology given in the years 1999, 2006, 2011, and 2013. Comments by the students of all four courses were very helpful in the preparation of this text and are gratefully acknowledged. All figures in this monograph were drawn with the *COREL* software and numerical computations were done with *Mathematica 9*. *Wikipedia, the free encyclopedia*, has been used

extensively by the author in the preparation of the text, and the indirect help by the numerous contributors submitting entries to Wiki is thankfully acknowledged.

Several colleagues gave important advice and made critical readings of the manuscript, among them Edem Arslan, Reinhard Bürger, Christoph Flamm, Thomas Hoffmann-Ostenhof, Christian Höner zu Siederissen, Ian Laurenzi, Stephen Lyle, Eric Mjolsness, Eberhard Neumann, Paul E. Phillipson, Christian Reidys, Bruce E. Shapiro, Karl Sigmund, and Peter F. Stadler. Many thanks go to all of them.

Wien, Austria Peter Schuster
April 2016

Contents

Chapter 1
Probability

The man that's over-cautious will achieve little.
Wer gar zu viel bedenkt, wird wenig leisten.
Friedrich Schiller, Wilhelm Tell, III

Abstract Probabilistic thinking originated historically when people began to analyze the chances of success in gambling, and its mathematical foundations were laid down together with the development of statistics in the seventeenth century. Since the beginning of the twentieth century statistics has been an indispensable tool for bridging the gap between molecular motions and macroscopic observations. The classical notion of probability is based on counting and dealing with finite numbers of observations. Extrapolation to limiting values for hypothetical infinite numbers of observations is the basis of the frequentist interpretation, while more recently a *subjective* approach derived from the early works of Bayes has become useful for modeling and analyzing complex biological systems. The Bayesian interpretation of probability accounts explicitly for the incomplete but improvable knowledge of the experimenter. In the twentieth century, set theory became the ultimate basis of mathematics, thus constituting also the foundation of current probability theory, based on Kolmogorov's axiomatization of 1933. The modern approach allows one to handle and compare finite, countably infinite, and also uncountable sets, the most important class, which underlie the proper consideration of continuous variables in set theory. In order to define probabilities for uncountable sets such as subsets of real numbers, we define Borel fields, families of subsets of sample space. The notion of random variables is central to the analysis of probabilities and applications to problem solving. Random variables are elements of discrete and countable or continuous and uncountable probability spaces. They are conventionally characterized by their distributions.

Classical probability theory, in essence, can handle all cases that are modeled by discrete quantities. It is based on counting and accordingly runs into problems when it is applied to uncountable sets. Uncountable sets occur with continuous variables and are therefore indispensable for modeling processes in space as well as for handling large particle numbers, which are described as continuous concentrations in chemical kinetics. Current probability theory is based on set theory and can handle variables on discrete—hence countable—as well as continuous—hence

P. Schuster, *Stochasticity in Processes*, Springer Series in Synergetics,
DOI 10.1007/978-3-319-39502-9_1

uncountable—sets. After a general introduction, we present a history of probability theory through examples. Different notions of probability are compared, and we then provide a short account of probabilities which are derived axiomatically from set theoretical operations. Separate sections deal with countable and uncountable sample spaces. Random variables are characterized in terms of probability distributions and those properties required for applications to stochastic processes are introduced and analyzed.

1.1 Fluctuations and Precision Limits

When a scientist reproduces an experiment, what does he expect to observe? If he were a physicist of the early nineteenth century, he would expect the same results within the precision limits of the apparatus he is using for the measurement. Uncertainty in observations was considered to be merely a consequence of technical imperfection. Celestial mechanics comes close to this ideal and many of us, for example, were witness to the outstanding accuracy of astronomical predictions in the precise timing of the eclipse of the sun in Europe on August 11, 1999. Terrestrial reality, however, tells that there are limits to reproducibility that have nothing to do with lack of experimental perfection. Uncontrollable variations in initial and environmental conditions on the one hand and the broad intrinsic diversity of individuals in a population on the other hand are daily problems in biology. Predictive limitations are commonplace in complex systems: we witness them every day when we observe the failures of various forecasts for the weather or the stock market. Another no less important source of randomness comes from the irregular thermal motions of atoms and molecules that are commonly characterized as thermal fluctuations. The importance of fluctuations in the description of ensembles depends on population size: they are—apart from exceptions—of moderate importance in chemical reaction kinetics, but highly relevant for the evolution of populations in biology.

Conventional chemical kinetics handles molecular ensembles involving large numbers of particles,[1] $N \approx 10^{20}$ and more. Under the majority of common conditions, for example, at or near chemical equilibrium or stable stationary states, and in the absence of autocatalytic self-enhancement, random fluctuations in particle numbers are proportional to $\sqrt{N}$. This so-called $\sqrt{N}$ *law* is introduced here as a kind of heuristic, but we shall derive it rigorously for the Poisson distribution in Sect. 2.3.1 and we shall see many specific examples where it holds to a good approximation. Typical experiments in chemical laboratories deal with amounts of

[1] In this monograph we shall use the notion of particle number as a generic term for discrete population variables. Particle numbers may be numbers of molecules or atoms in a chemical system, numbers of individuals in a population, numbers of heads in sequences of coin tosses, or numbers of dice throws yielding the same number of pips.

substance of about 10^{-4} mol—of the order of $N = 10^{20}$ particles—so these give rise to natural fluctuations which typically involve $\sqrt{N} = 10^{10}$ particles, i.e., in the range of $\pm 10^{-10}N$. Under such conditions the detection of fluctuations would require an accuracy of the order of $1{:}10^{10}$, which is (almost always) impossible to achieve in direct measurements, since most techniques in analytical chemistry encounter serious difficulties when concentration accuracies of $1{:}10^{-6}$ or higher are required.

Exceptions are new techniques for observing single molecules (Sect. 4.4). In general, the chemist uses concentrations rather than particle numbers, i.e., $c = N/(N_L V)$, where $N_L = 6.022 \times 10^{23}$ mol^{-1} and V are Avogadro's constant[2] and the volume in dm^3 or liters. Conventional chemical kinetics considers concentrations as continuous variables and applies deterministic methods, in essence differential equations, for analysis and modeling. It is thereby implicitly assumed that particle numbers are sufficiently large to ensure that the limit of infinite particle numbers is essentially correct and fluctuations can be neglected. This scenario is commonly not justified in biology, where particle numbers are much smaller than in chemistry and uncontrollable environmental effects introduce additional uncertainties.

Nonlinearities in chemical kinetics may amplify fluctuations through autocatalysis in such a way that the random component becomes much more important than the $\sqrt{N}$ law suggests. This is already the case with simple autocatalytic reactions, as discussed in Sects. 4.3.5, 4.6.4, and 5.1, and becomes a dominant effect, for example, with processes exhibiting oscillations or deterministic chaos. Some processes in physics, chemistry, and biology have no deterministic component at all. The most famous is *Brownian motion*, which can be understood as a visualized form of microscopic diffusion. In biology, other forms of entirely random processes are encountered, in which fluctuations are the only or the major driving force of change. An important example is random drift of populations in the space of genotypes, leading to fixation of mutants in the absence of any differences in fitness. In evolution, after all, particle numbers are sometimes very small: every new molecular species starts out from a single variant.

In 1827, the British botanist Robert Brown detected and analyzed irregular motions of particles in aqueous suspensions. These motions turned out to be independent of the nature of the suspended materials—pollen grains or fine particles of glass or minerals served equally well [69]. Although Brown himself had already

[2]The amount of a chemical compound A is commonly specified by the number N_A of molecules in the reaction volume V, via the number density $C_\mathsf{A} = N_\mathsf{A}/V$, or by the concentration $c_\mathsf{A} = N_\mathsf{A}/N_L V$, which is the number of moles in one liter of solution, where N_L is Avogadro's constant $N_L = 6.02214179 \times 10^{23}$ mol^{-1}, i.e., the number of atoms or molecules in one mole of substance. Loschmidt's constant $n_0 = 2.6867774 \times 10^{25}$ m^{-3} is closely related to Avogadro's constant and counts the number of particles in one liter of ideal gas at standard temperature and pressure, which are 0° and 1 atm = 101.325 kPa. Both quantities have physical dimensions and are not *numbers*, a point often ignored in the literature. In order to avoid ambiguity errors we shall refer to Avogadro's constant as N_L, because N_A is needed for the number of particles A (for units used in this monograph see appendix Notation).

demonstrated that the motion was not caused by any (mysterious) biological effect, its origin remained something of a riddle until Albert Einstein [133], and independently Marian von Smoluchowski [559], published satisfactory explanations in 1905 and 1906, respectively.[3] These revealed two main points:

(i) The motion is caused by highly frequent collisions between the pollen grain and the steadily moving molecules in the liquid in which the particles are suspended, and
(ii) the motion of the molecules in the liquid is so complicated and irregular that its effect on the pollen grain can only be described probabilistically in terms of frequent, statistically independent impacts.

In order to model Brownian motion, Einstein considered the number of particles per unit volume as a function of space[4] and time, viz., $f(x,t) = N(x,t)/V$, and derived the equation

$$\frac{\partial f}{\partial t} = D\frac{\partial^2 f}{\partial x^2}\,, \quad \text{with solution } f(x,t) = \frac{C}{\sqrt{4\pi D}}\frac{\exp(-x^2/4Dt)}{\sqrt{t}}\,,$$

where $C = N/V = \int f(x,t)\,\mathrm{d}x$ is the number density, the total number of particles per unit volume, and D is a parameter called the *diffusion coefficient*. Einstein showed that his equation for $f(x,t)$ was identical to the differential equation of diffusion already known as Fick's second law [165], which had been derived 50 years earlier by the German physiologist Adolf Fick. Einstein's original treatment was based on small discrete time steps $\Delta t = \tau$ and thus contains a—well justified—approximation that can be avoided by application of the modern theory of stochastic processes (Sect. 3.2.2.2). Nevertheless, Einstein's publication [133] represents the first analysis based on a probabilistic concept that is actually comparable to current theories, and Einstein's paper is correctly considered as the beginning of stochastic modeling. Later Einstein wrote four more papers on diffusion with different derivations of the diffusion equation [134]. It is worth mentioning that 3 years after the publication of Einstein's first paper, Paul Langevin presented an alternative mathematical treatment of random motion [325] that we shall discuss at length in the form of the Langevin equation in Sect. 3.4. Since the days of Brown's discovery, interest in Brownian motion has never ceased and publications on recent theoretical and experimental advances document this fact nicely—two interesting recent examples are [344, 491].

[3] The first mathematical model of Brownian motion was conceived as early as 1880, by Thorvald Thiele [330, 528]. Later, in 1900, a process involving random fluctuations of the Brownian motion type was used by Louis Bachelier [31] to describe the stock market at the Paris stock exchange. He gets the credit for having been the first to write down an equation that was later named after Paul Langevin (Sect. 3.4). For a recent and detailed monograph on Brownian motion and the mathematics of normal diffusion, we recommend [214].

[4] For the sake of simplicity we consider only motion in one spatial direction x.

From the solution of the diffusion equation, Einstein computed the diffusion parameter D and showed that it is linked to the mean square displacement $\langle \Delta x^2 \rangle$ of the particle in the x-direction:

$$D = \frac{\langle \Delta x^2 \rangle}{2t} , \quad \text{or} \quad \lambda_x = \sqrt{\langle \Delta x^2 \rangle} = \sqrt{2Dt} .$$

Here Δx is the net distance the particle travels during the time interval t. Extension to three-dimensional space is straightforward and results only in a different numerical factor: $D = \langle \Delta x^2 \rangle / 6t$. Both quantities, the diffusion parameter D and the mean displacement λ_x, are measurable, and Einstein concluded correctly that a comparison of the two quantities should allow for an experimental determination of Avogadro's constant [450].

Brownian motion was indeed the first completely random process that became accessible to a description within the frame of classical physics. Although James Clerk Maxwell and Ludwig Boltzmann had identified thermal motion as the driving force causing irregular collisions of molecules in gases, physicists in the second half of the nineteenth century were not interested in the details of molecular motion unless they were required in order to describe systems in the thermodynamic limit. In statistical mechanics the measurable macroscopic functions were, and still are, derived by means of global averaging techniques. By the first half of the twentieth century, thermal motion was no longer the only uncontrollable source of random natural fluctuations, having been supplemented by quantum mechanical uncertainty as another limitation to achievable precision.

The occurrence of complex dynamics in physics and chemistry has been known since the beginning of the twentieth century through the groundbreaking theoretical work of the French mathematician Henri Poincaré and the experiments of the German chemist Wilhelm Ostwald, who explored chemical systems with periodicities in space and time. Systematic studies of dynamical complexity, however, required the help of electronic computers and the new field of research on complex dynamical systems was not initiated until the 1960s. The first pioneer of this discipline was Edward Lorenz [354] who used numerical integration of differential equations to demonstrate what is nowadays called *deterministic chaos*. What was new in the second half of the twentieth century were not so much the concepts of complex dynamics but the tools to study it. Easy access to previously unimagined computer power and the development of highly efficient algorithms made numerical computation an indispensable technique for scientific investigation, to the extent that it is now almost on a par with theory and experiment.

Computer simulations have shown that a large class of dynamical systems modeled by nonlinear differential equations exhibit irregular, i.e., nonperiodic, behavior for certain ranges of parameter values. Hand in hand with complex dynamics go limitations on predictability, a point of great practical importance: although the differential equations used to describe and analyze chaos are still deterministic, initial conditions of an accuracy that could never be achieved in reality would be required for correct long-time predictions. Sensitivity to small

changes makes a stochastic treatment indispensable, and solutions were indeed found to be extremely sensitive to small changes in initial and boundary conditions in these *chaotic regimes*. Solution curves that are almost identical at the beginning can deviate exponentially from each other and appear completely different after sufficiently long times. Deterministic chaos gives rise to a third kind of uncertainty, because initial conditions cannot be controlled with greater precision than the experimental setup allows. It is no accident that Lorenz first discovered chaotic dynamics in the equations for atmospheric motions, which are indeed so complex that forecasts are limited to the short or mid-term at best.

In this monograph we shall focus on the mathematical handling of processes that are irregular and often simultaneously sensitive to small changes in initial and environmental conditions, but we shall not be concerned with the physical origin of these irregularities.

1.2 A History of Probabilistic Thinking

The concept of probability originated much earlier than its applications in physics and resulted from the desire to analyze by rigorous mathematical methods the chances of winning when gambling. An early study that has remained largely unnoticed, due to the sixteenth century Italian mathematician Gerolamo Cardano, already contained the basic ideas of probability. However, the beginning of classical probability theory is commonly associated with the encounter between the French mathematician Blaise Pascal and a professional gambler, the Chevalier de Méré, which took place in France a 100 years after Cardano. This tale provides such a nice illustration of a pitfall in probabilistic thinking that we repeat it here as our first example of conventional probability theory, despite the fact that it can be found in almost every textbook on statistics or probability.

On July 29, 1654, Blaise Pascal addressed a letter to the French mathematician Pierre de Fermat, reporting a careful observation by the professional gambler Chevalier de Méré. The latter had noted that obtaining at least one *six* with one die in 4 throws is successful in more than 50 % of cases, whereas obtaining at least one *double six* with two dice in 24 throws comes out in fewer than 50 % of cases. He considered this paradoxical, because he had calculated naïvely and erroneously that the chances should be the same:

$$\text{4 throws with one die yields} \quad 4 \times \frac{1}{6} = \frac{2}{3}\,,$$

$$\text{24 throws with two dice yields} \quad 24 \times \frac{1}{36} = \frac{2}{3}\,.$$

Blaise Pascal became interested in the problem and correctly calculated the probability as we would do it now in classical probability theory, by careful counting of events:

$$\text{probability} = P = \frac{\text{number of favorable events}}{\text{total number of events}} \,. \tag{1.1}$$

According to (1.1), the probability is always a positive quantity between zero and one, i.e., $0 \leq P \leq 1$. The sum of the probabilities that a given event has either occurred or not occurred is always one. Sometimes, as in Pascal's example, it is easier to calculate the probability q of the unfavorable case and to obtain the desired probability by computing $p = 1 - q$. In the one-die example, the probability of not throwing a *six* is $5/6$, while in the two-die case, the probability of not obtaining a *double six* is $35/36$. Provided the events are independent, their probabilities are multiplied[5] and we finally obtain for 4 and 24 trials, respectively:

$$q(1) = \left(\frac{5}{6}\right)^4 \quad \text{and} \quad p(1) = 1 - \left(\frac{5}{6}\right)^4 = 0.51775 \,,$$

$$q(2) = \left(\frac{35}{36}\right)^{24} \quad \text{and} \quad p(2) = 1 - \left(\frac{35}{36}\right)^{24} = 0.49140 \,.$$

It is remarkable that Chevalier de Méré was able to observe this rather small difference in the probability of success—indeed, he must have watched the game very often!

In order to see where the Chevalier made a mistake, and as an exercise in deriving correct probabilities, we calculate the first case—the probability of obtaining at least one *six* in four throws—by a more direct route than the one used above. We are throwing the die four times and the favorable events are: 1 time *six*, 2 times *six*, 3 times *six*, and 4 times *six*. There are four possibilities for 1 *six*—the *six* appearing in the first, the second, the third, or the fourth throw, six possibilities for 2 *sixes*, four possibilities for 3 *sixes*, and one possibility for 4 *sixes*. With the probabilities $1/6$ for obtaining a *six* and $5/6$ for any other number of pips, we get finally

$$\binom{4}{1} \times \frac{1}{6} \left(\frac{5}{6}\right)^3 + \binom{4}{2} \times \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^2 + \binom{4}{3} \times \left(\frac{1}{6}\right)^3 \frac{5}{6} + \binom{4}{4} \times \left(\frac{1}{6}\right)^4 = \frac{671}{1296} \,. \qquad \square$$

For those who want to become champion probability calculators, we suggest calculating $p(2)$ directly as well.

[5] We shall come back to a precise definition of independent events later, when we introduce modern probability theory in Sect. 1.6.4.

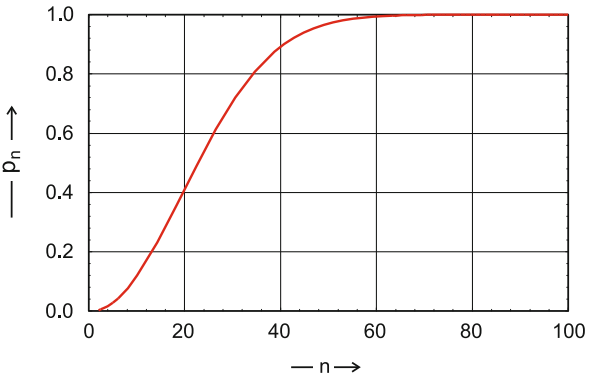

Fig. 1.1 The birthday problem. The *curve* shows the probability p_n that two people in a group of n people celebrate their birthday on the same day of the year

The second example presented here is the *birthday problem*.[6] It can be used to demonstrate the common human inability to estimate probabilities:

> Let your friends guess – without calculating – how many people you need in a group so that there is a fifty percent chance that at least two of them celebrate their birthday on the same day. You will be surprised by some of the answers!

With our knowledge of the gambling problem, this probability is easy to calculate. First we compute the negative event, that is, when everyone celebrates their birthday on a different day of the year, assuming that it is not a leap year, so that there are 365 days. For n people in the group, we find[7]

$$q = \frac{365}{365} \cdot \frac{364}{365} \cdot \frac{363}{365} \cdot \ldots \cdot \frac{365-(n-1)}{365} \quad \text{and} \quad p = 1 - q \,.$$

The function $p(n)$ is shown in Fig. 1.1. For the above-mentioned 50 % chance, we need only 27 people. With 41 people, we already have more than 90 % chance that two of them will celebrate their birthday on the same day, while 57 would yield a probability above 99 %, and 70 a probability above 99.9 %. An implicit assumption in this calculation has been that births are uniformly distributed over the year, i.e., the probability that somebody has their birthday on some particular day does not depend on that particular day. In mathematical statistics, such an assumption may be subjected to test and then it is called a *null hypothesis* (see [177] and Sect. 2.6.2).

Laws in classical physics are considered to be deterministic, in the sense that a single measurement is expected to yield a precise result. Deviations from this result

[6] The birthday problem was invented in 1939 by Richard von Mises [557] and it has fascinated mathematicians ever since. It has been discussed and extended in many papers, such as [3, 89, 255, 430], and even found its way into textbooks on probability theory [160, pp. 31–33].

[7] The expression is obtained by the following argument. The first person's birthday can be chosen freely. The second person's must not be chosen on the same day, so there are 364 possible choices. For the third, there remain 363 choices, and so on until finally, for the n th person, there are $365 - (n-1)$ possibilities.

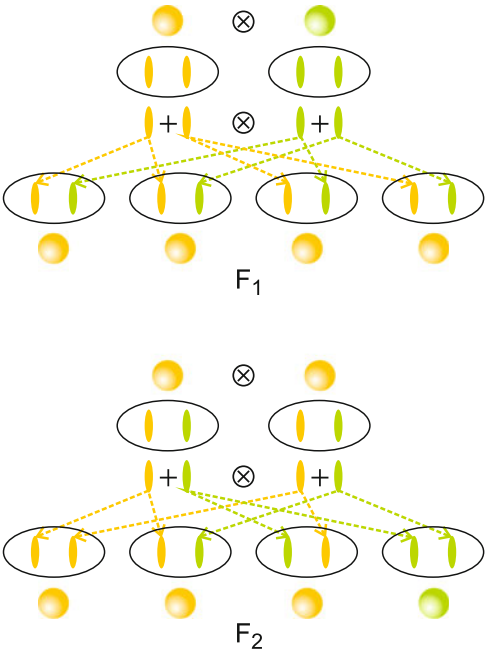

Fig. 1.2 Mendel's laws of inheritance. The sketch illustrates Mendel's laws of inheritance: (i) the law of segregation and (ii) the law of independent assortment. Every (diploid) organism carries two copies of each gene, which are separated during the process of reproduction. Every offspring receives one randomly chosen copy of the gene from each parent. *Encircled* are the genotypes formed from two alleles, yellow or green, and above or below the genotypes are the phenotypes expressed as the colors of seeds of the garden pea (*pisum sativum*). The *upper part* of the figure shows the first generation (F_1) of progeny of two homozygous parents—parents who carry two identical alleles. All genotypes are heterozygous and carry one copy of each allele. The yellow allele is dominant and hence the phenotype expresses yellow color. Crossing two F_1 individuals (*lower part* of the figure) leads to two homozygous and two heterozygous offspring. Dominance causes the two heterozygous genotypes and one homozygote to develop the dominant phenotype and accordingly the observable ratio of the two phenotypes in the F_2 generation is 3:1 on the average, as observed by Gregor Mendel in his statistics of fertilization experiments (see Table 1.1)

are then interpreted as due to a lack of precision in the equipment used. When it is observed, random scatter is thought to be caused by variations in experimental conditions that are not sufficiently well controlled. Apart from deterministic laws, other regularities are observed in nature, which become evident only when sample sizes are made sufficiently large through repetition of experiments. It is appropriate to call such regularities *statistical laws*. Statistical results regarding the biology of plant inheritance were pioneered by the Augustinian monk Gregor Mendel, who discovered regularities in the progeny of the *garden pea* in controlled fertilization experiments [392] (Fig. 1.2).

As a third and final example, we consider some of Mendel's data in order to exemplify a statistical law. Table 1.1 shows the results of two typical experiments

Table 1.1 Statistics of Gregor Mendel's experiments with the garden pea (*pisum sativum*)

	Form of seed			Color of seed		
Plant	Round	Wrinkled	Ratio	Yellow	Green	Ratio
1	45	12	3.75	25	11	2.27
2	27	8	3.38	32	7	4.57
3	24	7	3.43	14	5	2.80
4	19	10	1.90	70	27	2.59
5	32	11	2.91	24	13	1.85
6	26	6	4.33	20	6	3.33
7	88	24	3.67	32	13	2.46
8	22	10	2.20	44	9	4.89
9	28	6	4.67	50	14	3.57
10	25	7	3.57	44	18	2.44
Total	336	101	**3.33**	355	123	**2.89**

In total, Mendel analyzed 7324 seeds from 253 hybrid plants in the second trial year. Of these, 5474 were round or roundish and 1850 angular and wrinkled, yielding a ratio 2.96:1. The color was recorded for 8023 seeds from 258 plants, out of which 6022 were yellow and 2001 were green, with a ratio of 3.01:1. The results of two typical experiments with ten plants, which deviate more strongly because of the smaller sample size, are shown in the table

distinguishing roundish or wrinkled seeds with yellow or green color. The ratios observed with single plants exhibit a broad scatter. The mean values for ten plants presented in the table show that some averaging has occurred in the sample, but the deviations from the ideal values are still substantial. Mendel carefully investigated several hundred plants, whence the statistical law of inheritance demanding a ratio of 3:1 subsequently became evident [392].[8] In a somewhat controversial publication [176], Ronald Fisher reanalyzed Mendel's experiments, questioning his statistics and accusing him of intentionally manipulating his data, because the results were too close to the ideal ratio. Fisher's publication initiated a long-lasting debate in which many scientists spoke up in favor of Mendel [427, 428], but there were also critical voices saying that most likely Mendel had unconsciously or consciously eliminated outliers [127]. In 2008, one book declared *the end of the Mendel–Fisher controversy* [186]. In Sect. 2.6.2, we shall discuss statistical laws and Mendel's experiments in the light of present day mathematical statistics, applying the so-called χ^2 test.

Probability theory in its classical form is more than 300 years old. It is no accident that the concept arose in the context of gambling, originally considered to be a domain of chance in stark opposition to the rigours of science. Indeed it was rather a long time before the concept of probability finally entered the realms

[8] According to modern genetics this ratio, like other ratios between distinct inherited phenotypes, are idealized values that are found only for completely independent genes [221], i.e., lying either on different chromosomes or sufficiently far apart on the same chromosome.

of scientific thought in the nineteenth century. The main obstacle to the acceptance of probabilities in physics was the strong belief in determinism that held sway until the advent of quantum theory. Probabilistic concepts in nineteenth century physics were still based on deterministic thinking, although the details of individual events at the microscopic level were considered to be too numerous to be accessible to calculation. It is worth mentioning that probabilistic thinking entered physics and biology almost at the same time, in the second half of the nineteenth century. In physics, James Clerk Maxwell pioneered statistical mechanics with his dynamical theory of gases in 1860 [375–377]. In biology, we may mention the considerations of pedigree in 1875 by Sir Francis Galton and Reverend Henry William Watson [191, 562] (see Sect. 5.2.5), or indeed Gregor Mendel's work on the genetics of inheritance in 1866, as discussed above. The reason for the early considerations of statistics in the life sciences lies in the very nature of biology: sample sizes are typically small, while most of the regularities are probabilistic and become observable only through the application of probability theory. Ironically, Mendel's investigations and papers did not attract a broad scientific audience until they were rediscovered at the beginning of the twentieth century. In the second half of the nineteenth century, the scientific community was simply unprepared for quantitative and indeed probabilistic concepts in biology.

Classical probability theory can successfully handle a number of concepts like conditional probabilities, probability distributions, moments, and so on. These will be presented in the next section using set theoretic concepts that can provide a much deeper insight into the structure of probability theory than mere counting. In addition, the more elaborate notion of probability derived from set theory is absolutely necessary for extrapolation to countably infinite and uncountable sample sizes. Uncountability is an unavoidable attribute of sets derived from continuous variables, and the set theoretic approach provides a way to define probability measures on certain sets of real numbers $\mathbf{x} \in \mathbb{R}^n$. From now on we shall use only the set theoretic concept, because it can be introduced straightforwardly for countable sets and discrete variables and, in addition, it can be straightforwardly extended to probability measures for continuous variables.

1.3 Interpretations of Probability

Before introducing the current standard theory of probability we make a brief digression into the dominant philosophical interpretations:

(i) the classical interpretations that we have adopted in Sect. 1.2,
(ii) the frequency-based interpretation that stand in the background for the rest of the book, and
(iii) the Bayesian or *subjective* interpretation.

The *classical interpretation of probability* goes back to the concepts laid out in the works of the Swiss mathematician Jakob Bernoulli and the French mathematician

and physicist Pierre-Simon Laplace. The latter was the first to present a clear definition of probability [328, pp. 6–7]:

> The theory of chance consists in reducing all the events of the same kind to a certain number of equally possible cases, that is to say, to such as we may be equally undecided about in regard of their existence, and in determining the number of cases favorable to the event whose probability is sought. The ratio of this number to that of all possible cases is the measure of this probability, which is thus simply a fraction whose numerator is the number of favorable cases and whose denominator is the number of all possible cases.

Clearly, this definition is tantamount to (1.1) and the explicitly stated assumption of equal probabilities is now called the *principle of indifference*. This classical definition of probability was questioned during the nineteenth century by the two British logicians and philosophers George Boole [58] and John Venn [549], among others, initiating a paradigm shift from the classical view to the modern frequency interpretations of probabilities.

Modern interpretations of the concept of probability fall essentially into two categories that can be characterized as *physical probabilities* and *evidential probabilities* [228]. Physical probabilities are often called *objective* or frequency-based probabilities, and their advocates are referred to as *frequentists*. Besides the pioneer John Venn, influential proponents of the frequency-based probability theory were the Polish–American mathematician Jerzy Neyman, the British statistician Egon Pearson, the British statistician and theoretical biologist Ronald Fisher, the Austro-Hungarian–American mathematician and scientist Richard von Mises, and the German–American philosopher of science Hans Reichenbach. Physical probabilities are derived from some real process like radioactive decay, a chemical reaction, the turn of a roulette wheel, or rolling dice. In all such systems the notion of probability makes sense only when it refers to some well defined experiment with a random component.

Frequentism comes in two versions: (i) *finite frequentism* and (ii) *hypothetical frequentism*. Finite frequentism replaces the notion of the *total number of events* in (1.1) by the *actually recorded number of events*, and is thus congenial to philosophers with empiricist scruples. Philosophers have a number of problems with finite frequentism. For example, we may mention problems arising due to small samples: one can never speak about probability for a single experiment and there are cases of unrepeated or unrepeatable experiments. A coin that is tossed exactly once yields a relative frequency of heads being either zero or one, no matter what its bias really is. Another famous example is the spontaneous radioactive decay of an atom, where the probabilities of decaying follow a continuous exponential law, but according to finite frequentism it decays with probability one only once, namely at its actual decay time. The evolution of the universe or the origin of life can serve as cases of unrepeatable experiments, but people like to speak about the probability that the development has been such or such. Personally, I think it would do no harm to replace *probability* by *plausibility* in such estimates dealing with unrepeatable single events.

Hypothetical frequentism complements the empiricism of finite frequentism by the admission of infinite sequences of trials. Let N be the total number of repetitions

of an experiment and n_A the number of trials when the event A has been observed. Then the relative frequency of recording the event A is an *approximation of the probability* for the occurrence of A :

$$\text{probability}\,(A) = P(A) \approx \frac{n_A}{N}\,.$$

This equation is essentially the same as (1.1), but the claim of the hypothetical frequentists' interpretation is that there exists a *true frequency* or *true probability* to which the relative frequency would converge if we could repeat the experiment an infinite number of times[9]:

$$P(A) = \lim_{N\to\infty} \frac{n_A}{N} = \frac{|A|}{|\Omega|}\,, \quad \text{with } A \in \Omega\,. \tag{1.2}$$

The probability of an event A relative to a sample space Ω is then defined as the limiting frequency of A in Ω. As N goes to infinity, $|\Omega|$ becomes infinitely large and, depending on whether $|A|$ is finite or infinite, $P(A)$ is either zero or may be a nonzero limiting value. This is based on two a priori assumptions that have the character of axioms:

(i) *Convergence*. For any event A, there exists a limiting relative frequency, the probability $P(A)$, satisfying $0 \le P(A) \le 1$.
(ii) *Randomness*. The limiting relative frequency of each event in a set Ω is the same for any *typical* infinite subsequence of Ω.

A typical sequence is *sufficiently random*[10] in order to avoid results biased by predetermined order. As a negative example of an acceptable sequence, consider *heads, heads, heads, heads, ...* recorded by tossing a coin. If it was obtained with a fair coin—not a coin with two heads—$|A|$ is 1 and $P(A) = 1/|\Omega| = 0$, and we may say that this particular event has measure zero and the sequence is not typical. The sequence *heads, tails, heads, tails, ...* is not typical either, despite the fact that it yields the same probabilities for the average number of *heads* and *tails* as a fair coin. We should be aware that the extension to infinite series of experiments leaves the realm of empiricism, leading purist philosophers to reject the claim that the interpretation of probabilities by hypothetical frequentism is more *objective* than others.

Nevertheless, the frequentist probability theory is not in conflict with the mathematical axiomatization of probability theory and it provides straightforward

[9]The absolute value symbol $|A|$ means here the size or cardinality of A, i.e., the number of elements in A (Sect. 1.4).

[10]Sequences are sufficiently random when they are obtained through recordings of random events. *Random sequences* are approximated by the sequential outputs of *pseudorandom number generators*. 'Pseudorandom' implies here that the approximately random sequence is created by some deterministic, i.e., nonrandom, algorithm.

guidance in applications to real-world problems. The pragmatic view that prefigures the dominant concept in current probability theory has been nicely put by William Feller, the Croatian–American mathematician and author of the two-volume classic introduction to probability theory [160, 161, Vol. I, pp. 4–5]:

> The success of the modern mathematical theory of probability is bought at a price: the theory is limited to one particular aspect of 'chance'. (...) we are not concerned with modes of inductive reasoning but with something that might be called physical or statistical probability.

He also expresses clearly his attitude towards pedantic scruples of philosophic purists:

> (...) in analyzing the coin tossing game we are not concerned with the accidental circumstances of an actual experiment, the object of our theory is sequences or arrangements of symbols such as 'head, head, tail, head, ...'. There is no place in our system for speculations concerning the probability that the sun will rise tomorrow. Before speaking of it we should have to agree on an idealized model which would presumably run along the lines 'out of infinitely many worlds one is selected at random ...'. Little imagination is required to construct such a model, but it appears both uninteresting and meaningless.

We shall adopt the frequentist interpretation throughout this monograph, but give brief mention here briefly to two more interpretations of probability in order to show that it is not the only reasonable probability theory.

The *propensity interpretation* of probability was proposed by the American philosopher Charles Peirce in 1910 [448] and reinvented by Karl Popper [455, pp. 65–70] (see also [456]) more than 40 years later [228, 398]. *Propensity* is a tendency to do or achieve something. In relation to probability, the propensity interpretation means that it makes sense to talk about the probabilities of single events. As an example, we can talk about the probability—or propensity—of a radioactive atom to decay within the next 1000 years, and thereby conclude from the behavior of an ensemble to that of a single member of the ensemble. Likewise, we might say that there is a probability of 1/2 of getting 'heads' when a fair coin is tossed, and precisely expressed, we should say that the coin has a propensity to yield a sequence of outcomes in which the limiting frequency of scoring 'heads' is 1/2. The single case propensity is accompanied by, but distinguished from, the *long-run* propensity [215]:

> A long-run propensity theory is one in which propensities are associated with repeatable conditions, and are regarded as propensities to produce in a long series of repetitions of these conditions frequencies, which are approximately equal to the probabilities.

In these theories, a long run is still distinct from an infinitely long run, in order to avoid basic philosophical problems. Clearly, the use of *propensities* rather than *frequencies* provides a somewhat more careful language than the frequentist interpretation, making it more acceptable in philosophy.

Finally, we sketch the most popular example of a theory based on *evidential probabilities*: Bayesian statistics, named after the eighteenth century British mathematician and Presbyterian minister Thomas Bayes. In contrast to the frequentist view, probabilities are *subjective* and exist only in the human mind. From a

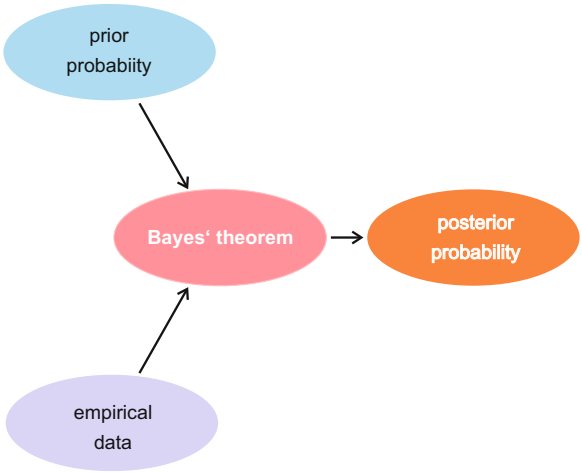

Fig. 1.3 A sketch of the Bayesian method. Prior information on probabilities is confronted with empirical data and converted by means of Bayes' theorem into a new distribution of probabilities called *posterior* probability [120, 507]

practitioner's point of view, one major advantage of the Bayesian approach is that it gives a direct insight into the way we improve our knowledge of a given subject of investigation. In order to understand Bayes' theorem, we need the notion of conditional probability, presented in Sect. 1.6.4. We thus postpone a precise formulation of the Bayesian approach to Sect. 2.6.5. Here we sketch only the basic principle of the method in a narrative manner.[11]

In physics and chemistry, we common deal with well established theories and models that are assumed to be essentially correct. Experimental data have to be fitted to the model and this is done by adjusting unknown model parameters using fitting techniques like the maximum-likelihood method (Sect. 2.6.4). This popular statistical technique is commonly attributed to Ronald Fisher, although it has been known for much longer [8, 509]. Researchers in biology, economics, social sciences, and other disciplines, however, are often confronted with situations where no commonly accepted models exist, so they cannot be content with parameter estimates. The model must then be tested and the basic formalisms improved.

Figure 1.3 shows schematically how Bayes' theorem works: the inputs of the method are (i) a preliminary or *prior* probability distribution derived from the initial model and (ii) a set of empirical data. Bayes theorem converts the inputs into a *posterior* probability distribution, which encapsulates the improvement of the model in the light of the data sample.[12] What is missing here is a precise probabilistic formulation of the process shown in Fig. 1.3, but this will be added in Sect. 2.6.5.

[11] In this context it is worth mentioning the contribution of the great French mathematician and astronomer the Marquis de Laplace, who gave an interpretation of statistical inference that can be considered equivalent to Bayes' theorem [508].

[12] It is worth comparing the Bayesian approach with conventional data fitting: the inputs are the same, a model and data, but the nature of the probability distribution is kept constant in data fitting methods, whereas it is conceived as flexible in the Bayes method.