

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.

Editorial and Programme Advisory Board

Henry Abarbanel, Institute for Nonlinear Science, University of California, San Diego, USA

Dan Braha, New England Complex Systems Institute and University of Massachusetts Dartmouth, USA

Péter Érdi, Center for Complex Systems Studies, Kalamazoo College, USA and Hungarian Academy of Sciences, Budapest, Hungary

Karl Friston, Institute of Cognitive Neuroscience, University College London, London, UK

Hermann Haken, Center of Synergetics, University of Stuttgart, Stuttgart, Germany

Viktor Jirsa, Centre National de la Recherche Scientifique (CNRS), Université de la Méditerranée, Marseille, France

Janusz Kacprzyk, System Research, Polish Academy of Sciences, Warsaw, Poland

Kunihiko Kaneko, Research Center for Complex Systems Biology, The University of Tokyo, Tokyo, Japan

Scott Kelso, Center for Complex Systems and Brain Sciences, Florida Atlantic University, Boca Raton, USA

Markus Kirkilionis, Mathematics Institute and Centre for Complex Systems, University of Warwick, Coventry, UK

Jürgen Kurths, Nonlinear Dynamics Group, University of Potsdam, Potsdam, Germany

Andrzej Nowak, Department of Psychology, Warsaw University, Poland

Hassan Qudrat-Ullah, York University, Toronto, Ontario, Canada

Linda Reichl, Center for Complex Quantum Systems, University of Texas, Austin, USA

Peter Schuster, Theoretical Chemistry and Structural Biology, University of Vienna, Vienna, Austria

Frank Schweitzer, System Design, ETH Zurich, Zurich, Switzerland

Didier Sornette, Entrepreneurial Risk, ETH Zurich, Zurich, Switzerland

Stefan Thurner, Section for Science of Complex Systems, Medical University of Vienna, Vienna, Austria

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

Springer Series in Synergetics

ISBN 978-3-319-39500-5

DOI 10.1007/978-3-319-39502-9

ISSN 2198-333X (electronic)

ISBN 978-3-319-39502-9 (eBook)

Library of Congress Control Number: 2016940829

© Springer International Publishing Switzerland 2016

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

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 (★) 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
April 2016

Peter Schuster

Contents

1	Probability	1
1.1	Fluctuations and Precision Limits	2
1.2	A History of Probabilistic Thinking.....	6
1.3	Interpretations of Probability	11
1.4	Sets and Sample Spaces	16
1.5	Probability Measure on Countable Sample Spaces.....	20
1.5.1	Probability Measure	21
1.5.2	Probability Weights	24
1.6	Discrete Random Variables and Distributions	27
1.6.1	Distributions and Expectation Values	27
1.6.2	Random Variables and Continuity	29
1.6.3	Discrete Probability Distributions	34
1.6.4	Conditional Probabilities and Independence	38
1.7	* Probability Measure on Uncountable Sample Spaces	44
1.7.1	* Existence of Non-measurable Sets.....	46
1.7.2	* Borel σ -Algebra and Lebesgue Measure	49
1.8	Limits and Integrals	55
1.8.1	Limits of Series of Random Variables.....	55
1.8.2	Riemann and Stieltjes Integration	59
1.8.3	Lebesgue Integration	63
1.9	Continuous Random Variables and Distributions	70
1.9.1	Densities and Distributions	71
1.9.2	Expectation Values and Variances	76
1.9.3	Continuous Variables and Independence	77
1.9.4	Probabilities of Discrete and Continuous Variables	78
2	Distributions, Moments, and Statistics	83
2.1	Expectation Values and Higher Moments.....	83
2.1.1	First and Second Moments.....	84
2.1.2	Higher Moments.....	91
2.1.3	* Information Entropy	95

2.2	Generating Functions	101
2.2.1	Probability Generating Functions.....	101
2.2.2	Moment Generating Functions	103
2.2.3	Characteristic Functions	105
2.3	Common Probability Distributions	107
2.3.1	The Poisson Distribution	109
2.3.2	The Binomial Distribution	111
2.3.3	The Normal Distribution	115
2.3.4	Multivariate Normal Distributions.....	120
2.4	Regularities for Large Numbers	124
2.4.1	Binomial and Normal Distributions	125
2.4.2	Central Limit Theorem.....	130
2.4.3	Law of Large Numbers.....	133
2.4.4	Law of the Iterated Logarithm	135
2.5	Further Probability Distributions	137
2.5.1	The Log-Normal Distribution.....	137
2.5.2	The χ^2 -Distribution	140
2.5.3	Student's t-Distribution	143
2.5.4	The Exponential and the Geometric Distribution	147
2.5.5	The Pareto Distribution	151
2.5.6	The Logistic Distribution	154
2.5.7	The Cauchy–Lorentz Distribution	156
2.5.8	The Lévy Distribution.....	159
2.5.9	The Stable Distribution	161
2.5.10	Bimodal Distributions.....	166
2.6	Mathematical Statistics	168
2.6.1	Sample Moments	169
2.6.2	Pearson's Chi-Squared Test	173
2.6.3	Fisher's Exact Test	180
2.6.4	The Maximum Likelihood Method	182
2.6.5	Bayesian Inference	190
3	Stochastic Processes	199
3.1	Modeling Stochastic Processes	203
3.1.1	Trajectories and Processes	203
3.1.2	Notation for Probabilistic Processes.....	208
3.1.3	Memory in Stochastic Processes.....	209
3.1.4	Stationarity	214
3.1.5	Continuity in Stochastic Processes	216
3.1.6	Autocorrelation Functions and Spectra.....	220
3.2	Chapman–Kolmogorov Forward Equations	224
3.2.1	Differential Chapman–Kolmogorov Forward Equation	225
3.2.2	Examples of Stochastic Processes	235
3.2.3	Master Equations	260

3.2.4	Continuous Time Random Walks	273
3.2.5	Lévy Processes and Anomalous Diffusion	284
3.3	Chapman–Kolmogorov Backward Equations	303
3.3.1	Differential Chapman–Kolmogorov Backward Equation ...	305
3.3.2	Backward Master Equations	307
3.3.3	Backward Poisson Process	310
3.3.4	Boundaries and Mean First Passage Times	313
3.4	Stochastic Differential Equations	319
3.4.1	Mathematics of Stochastic Differential Equations	321
3.4.2	Stochastic Integrals	323
3.4.3	Integration of Stochastic Differential Equations	337
4	Applications in Chemistry	347
4.1	A Glance at Chemical Reaction Kinetics	350
4.1.1	Elementary Steps of Chemical Reactions	351
4.1.2	Michaelis–Menten Kinetics	358
4.1.3	Reaction Network Theory	372
4.1.4	Theory of Reaction Rate Parameters	388
4.1.5	Empirical Rate Parameters	407
4.2	Stochasticity in Chemical Reactions	415
4.2.1	Sampling of Trajectories	416
4.2.2	The Chemical Master Equation	418
4.2.3	Stochastic Chemical Reaction Networks	425
4.2.4	The Chemical Langevin Equation	432
4.3	Examples of Chemical Reactions	435
4.3.1	The Flow Reactor	436
4.3.2	Monomolecular Chemical Reactions	441
4.3.3	Bimolecular Chemical Reactions	450
4.3.4	Laplace Transform of Master Equations	459
4.3.5	Autocatalytic Reaction	477
4.3.6	Stochastic Enzyme Kinetics	485
4.4	Fluctuations and Single Molecule Investigations	490
4.4.1	Single Molecule Enzymology	491
4.4.2	Fluorescence Correlation Spectroscopy	500
4.5	Scaling and Size Expansions	509
4.5.1	Kramers–Moyal Expansion	509
4.5.2	Small Noise Expansion	512
4.5.3	Size Expansion of the Master Equation	514
4.5.4	From Master to Fokker–Planck Equations	521
4.6	Numerical Simulation of Chemical Master Equations	526
4.6.1	Basic Assumptions	527
4.6.2	Tau-Leaping and Higher-Level Approaches	531
4.6.3	The Simulation Algorithm	533
4.6.4	Examples of Simulations	542

- 5 Applications in Biology** 569
 - 5.1 Autocatalysis and Growth 572
 - 5.1.1 Autocatalysis in Closed Systems 572
 - 5.1.2 Autocatalysis in Open Systems 575
 - 5.1.3 Unlimited Growth 580
 - 5.1.4 Logistic Equation and Selection 583
 - 5.2 Stochastic Models in Biology 585
 - 5.2.1 Master Equations and Growth Processes 585
 - 5.2.2 Birth-and-Death Processes 589
 - 5.2.3 Fokker–Planck Equation and Neutral Evolution 605
 - 5.2.4 Logistic Birth-and-Death and Epidemiology 611
 - 5.2.5 Branching Processes 631
 - 5.3 Stochastic Models of Evolution 649
 - 5.3.1 The Wright–Fisher and the Moran Process 651
 - 5.3.2 Master Equation of the Moran Process 658
 - 5.3.3 Models of Mutation 665
 - 5.4 Coalescent Theory and Phylogenetic Reconstruction 673

- Notation** 679

- References** 683

- Author Index** 707

- Index** 711